University of Reading

Labour markets, technological change, and natural disasters

With special reference to the race between education and technology, the task content of jobs and, the demand for ICT labour after disasters

by

Jorge Abner Campos González

Thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Department of Applied Economics and Marketing
School of Agriculture, Policy, and Development

July 2022

Declaration of original authorship

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged

Jorge Abner Campos González
11th July 2022

Abstract

This thesis examines the interactions between labour markets, technological change, and natural disasters in Chile, which has been considered as one of most successful Latin American countries according to economic growth and structural economic reforms over the last decades. One of these major reforms was trade liberalization implying an important absorption of foreign technologies biased towards skilled labour which increased from 10.1 per cent in 1980 to 30.4 per cent in 2018. At the same time, the Chilean higher education experienced substantial growth between the 1980s and 2010s, showing that people enrolled in tertiary education sextupled. In 1984, 11% of the 18 – 24 enrolled in tertiary education while in in 2018, it was 67%. In this regard, Chile supplies an environment particularly well suited to study the technological change driving the skill premium evolution. in Chile, like most Latin American countries, the skill premium is suggested as the main force driving the observed rise and fall of income inequality in recent decades (Acosta et al., 2019; Guerra-Salas, 2018; Parro & Reyes, 2017). The decline in income inequality is an important step in improving the population's assessment of their well-being. Furthermore, Chile supplies a unique location for studying the potential for technological upgrading in the aftermath of catastrophes since it is characterized by recurring severe earthquakes. Earthquakes supply the opportunity to analyse technological upgrading due to their unexpected occurrence and destructive ability. Therefore, examining the interactions between technology, labour markets and natural disasters has important implications for our understanding of the interrelations between these economic forces. The thesis is divided into three essays.

The first essay aims to test the Race between Education and Technology, RBET, model empirically for Chile using recurrent bi-annual labour survey data from 1980 to 2018. The main aspects that motivate this research are the lack of evidence in the post-2000 period and "estimation difficulties" reported by past studies. These difficulties imply mainly the computation of positive coefficients standing for the expected negative relationship between the skill premium, i.e., the gap between skilled and unskilled wages, and the relative supply of skilled labour as posited by the RBET theory. Besides, a positive coefficient would imply the computation of a negative elasticity of substitution between skilled and unskilled labour. We also find "estimation difficulties" using cointegration techniques. Alternatively, we apply an Unobserved Component Model, UCM, estimated by Bayesian inference, UCM-Bayesian, whose results are more consistent with the RBET model. We find that both demand and supply factors explain the evolution of the skill premium. In the context of the race between technology and education, in the pre-2000 period, the relative demand attributable to *skill-biased technological change*, SBTC, with its rapid acceleration contributing to a high skill premium, is suggested as the dominant factor. However, in the post-2000 span, the demand factor was surpassed by strong increases in the relative supply, suggesting education as the dominant factor inducing a declining trend in the skill premium. Furthermore, our estimate for the elasticity of substitution is 6.5. The value

greater than one implies that skilled and unskilled workers are imperfect substitutes but more substitutable than commonly thought, given the past estimates for this parameter.

The second essay evaluates the influence on the skill premium for the task-content of jobs and specific workers' abilities. We exploit the text data from job posting ads covering 2009-2018 (approx. 189,000 ads) to capture demand for tasks and skills. Our task-related analysis tests the expected complementarity between skilled labour and non-routine cognitive (analytical and interactive) and routine cognitive tasks. In our skills-related analysis, we evaluate whether cognitive and social abilities influence the skill premium. Our results show weak evidence for non-routine cognitive tasks as drivers of the skill premium progress, while routine cognitive tasks do not explain this wages differential. Also, we do not find evidence that cognitive or social abilities, separately or in combination, explain the evolution of the skill premium. The apparently inferior importance of cognitive tasks and abilities might imply an inefficient educational investment or unwanted changes in the occupational ladder for higher educated workers.

The potential impact natural disasters have in improving demand for labour in the Information and Communication Technologies, ICT, sector is explored in the third essay. We explore whether disasters can accelerate the current technical progress featured by ICT, assuming that updated and ICT compatible equipment replaces the destroyed equipment. In turn, this faster rate of technological adoption would lead to increases in demand for ICT labour. We use a severe earthquake (8.8 M_w) that struck Chile's Central Region on February 27, 2010, as a natural experiment and a subsample from online job posting data used in the second essay. We implement a structural topic model, STM, to estimate the difference in the prevalence of topics related to ICT labour (as proxies for upgrading new technologies) and the Construction sector labour by comparing two years before to two years after the earthquake. Our results show that the prevalence of the ICT labour topic does not substantially change, suggesting that there was no technological replacement following the earthquake. In contrast, the prevalence of the Construction labour topic was significantly different after the disaster, suggesting that reconstruction activities lead to employment differences in the Construction sector.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my primary supervisor Professor Kelvin Balcombe, not only for his continuous support, patience, motivation, and knowledge but also for the hard questions which encouraged me to widen my research from various perspectives. Besides my main supervisor, I would like to thank the co-supervisors of my research, Dr Ariane Kehlbacher and Dr Giacomo Zanello, for their insightful comments and stimulating discussions.

I am fortunate to have been a part of the School of Agriculture, Policy, and Development, and the Department of Applied Economics and Marketing: in particular, to have been part of Office 304, where I met wonderful people.

I wish to acknowledge the financial support from the National Agency for Research and Development (ANID) / Scholarship Program / DOCTORADO BECAS CHILE/2017 – 72180253. I also with to express my gratitude to Trabajando.com for granting me access to online job postings databases and Carolyn Lyle for her proofreading help.

Lastly, my family deserves endless gratitude: my beloved wife Katherine for her continuous support and patience, my son Ariel for teaching me that writing in English can be fun, and my daughter Catalina for keeping me on track by asking me every day when I would finish my thesis. I am indebted also to several members of our extended families in Chile who have been our support these years as immigrants in the United Kingdom.

Contents

List of Tables

List of Figures

Abbreviations and acronyms

ADF	Augmented Dickey-Fueller
ALM	The Autor, Levy and Murnane (2003) model
BIC	Schwarz Bayesian Criterion
CES	Constant Elasticity of Substitution
CIUO08-CL	Chilean Standard Classification of Occupations
DTM	Document-Term Matrix
ECM	Error Correction Model
EOD	*Encuesta de Ocupación y Desempleo del Gran Santiago*
HQC	Hannan-Quinn Criterion
ICT	Information and Communication Technologies
IRF	Impulse-Response Function
ISCO-08	International Standard Classification of Occupations
KPSS	Kwiatkowski, Phillips, Schmidt, and Shin
MCMC	Monte Carlo Markov Chains
M_w	Moment Magnitude Scale
OLS	Ordinary Least Squares
RBET	The Race between Education and Technology Model
SBTC	Skill Biased Technological Change
STM	Structural Topic Model
SVM	Support Vector Machines
TM	Topic Model
UCM	Unobserved Component Model
VAR	Vector Autoregressive
VECM	Vector Error Correction Model

1. Introduction

The relationship between technological progress and labour markets has been the subject of a long-standing debate, as technology is a significant force driving employment and earnings (Autor et al., 1998). In recent decades, Information and Communications Technologies (ICT from now on) and other computer-based emergent advancements, such as automation and robotics, have further fuelled this controversy as drivers for much of the technological change in production (Acemoglu & Autor, 2011; Almeida et al., 2020). This debate has motivated a vast literature examining the interactions between technology and labour outputs, such as the skill premium, i.e., the gap between skilled and unskilled wages[1]. The skill premium is particularly important as a measure of income inequality between workers with different skills, showing how the relative prices of skills evolve (Acemoglu & Autor, 2011). The evolution of the skill premium also provides opportunities to understand the characteristics of the process of economic development, especially how economic forces like technological change may influence the demand for, and supply of, highly qualified workers.

The literature has developed several economic models to understand how technological advancements drive the demand for labour based on their skills. One is the Race between Education and Technology, RBET, model (Acemoglu, 2002; Acemoglu & Autor, 2011; Autor et al., 2008; Katz & Murphy, 1992), also known as the Skill-Biased Technological Change, SBTC, model since it assumes that technological change is *biased* towards skilled workers, i.e., there is an increasing demand for skilled workers coming from technology. On the supply side, the educational system supplies these skills or qualifications, affecting the educational attainment of the workforce. These simultaneous shifts in supply and demand implicitly refer to a *race between education and technology* as posited by the RBET model, where the effect of technological change on labour is based solely on the skills endowment of workers.

The RBET model, like the ALM model proposed by Autor, Levy and Murnane (2003), is among the economic models most frequently applied to the study of these interactions (Goos, 2018). The ALM model posits that technical progress depends on skills and the task content of jobs. One of its main predictions is that technological progress complements non-routine cognitive tasks (e.g., researching, decision making, persuasion) typically performed by skilled workers (e.g., managers, professionals) (see Table 3.1). Thus, the ALM model can account for the interactions among skills, tasks and technologies, becoming a revised version of the SBTC effect established by the RBET model (Acemoglu & Autor, 2011). However, beyond the expected complementarity between cognitive tasks and skilled labour, some have documented a reversal in demand for cognitive abilities (Beaudry et al.,

[1] In this thesis we define skilled labour as suitable for college or post-secondary graduates and unskilled labour as suitable for graduates of high-school or secondary education or those whose education has not reached these levels. To consider different workers' educational attainments (e.g., high-school dropouts) we define additional variables to adjust for these and other compositional labour structure. See section 2.5 for methodological details.

2016; Deming & Noray, 2020). At the same time, some have noted a rising complementarity between cognitive and social skills (e.g., communication, cooperation with others) with a positive impact on wages (Deming, 2017; Edin et al., 2017). Consequently, the joint demand for both cognitive and social skills might result in better labour outcomes such as the skill premium, for skilled labour.

Recapitulating, a critical aspect of the RBET and the ALM models is that they assume that technological progress steadily evolves (Acemoglu, 2002; Acemoglu & Autor, 2011; Autor et al., 2008). Alternatively, it has been suggested that the pace of technological change might be affected by episodes or substantial events like natural disasters (Crespo Cuaresma et al., 2008; Okuyama, 2003; Okuyama et al., 2004; Skidmore & Toya, 2002). Such events might lead to technological upgrading because updated capital replaces the old equipment destroyed by natural catastrophes. In turn, the technological replacement might lead to changes in demand for labour based on the complementarity or substitutability between the new equipment and labour. If, as supported by the literature, ICT technologies currently drive much of the technological change in production (see e.g., Acemoglu & Autor, 2011; Almeida et al., 2020), equipment compatible with ICT will replace the equipment damaged by recent natural disasters. In turn, this technological upgrading might positively affect the demand for ICT employment. Conceptually, extensions of growth models like the Solow-Swan approach (Solow, 1956; Swan, 1956) and a more literal assumption of the creative-destruction hypothesis (Aghion & Howitt, 1990; Schumpeter, 1976) allow us to examine these interactions. Thus, we can examine how labour markets respond to both regular economic forces i.e., the role of demand factors like the technological change in the skill premium evolution, and unexpected shocks like natural disasters considering the potential technological upgrading proxied by changes in demand for specialized labour i.e., ICT labour in the aftermath of catastrophes. This conceptual link between the expected technological upgrading leading to improvements in demand for ICT labour is detailed in section 4.2.2.

1.1. Labour, and educational aspects of Chile and relevance of the country to investigate the interactions between labour, technology, and disasters

Chile has been considered as one of most successful Latin American countries (LAC) according to economic growth and far-reaching economic reforms over the last 50 years (Gallego, 2012; Murakami & Nomura, 2020). Most major economic reforms in Chile occurred between 1975 and 1995, with trade liberalization being the most relevant (Beyer et al., 1999). Researchers suggest that one of the implications of this openness was the absorption of foreign technologies biased towards skilled labour (Beyer et al., 1999; Gallego, 2012; Robbins, 1994a). In the 1980s and 1990s, Chile imported about 85% of non-transportation machinery and equipment from the US and OECD countries (Gallego, 2012). In this regard, it has been suggested that the relative demand for skilled labour increased significantly in most sector of economy during the 1980s and 1990s (Gallego, 2012) and continued increasing over time. The employed workforce in the Chilean economy in 1980 was around 3 million workers and in 2018, around 8 million workers while the share of skilled labour rises from 10.1 per cent in 1980 to 30.4 per cent in 2018 (INE, 2017; University of Chile, 2020). Over this period, researchers have linked this higher share to the expansion of tertiary education and the exit of the older and less educated cohorts (Murakami & Nomura, 2020; Parro & Reyes, 2017). The observed data shows that Chilean higher education experienced substantial growth in recent decades. According to educational and census data for 1984-2018, people enrolled in tertiary education sextupled (INE, 2017; MINEDUC, 2020). The 18–24 age group enrolled in tertiary education grew from 189,151 (11% of this age group) in 1984 to 521,882 (31%) in 2002. In 2018, it exceeded 1.2 million (approximately 67% of the 18–24 age group). Apart from the endogenous response of agents to the increase in returns to education, these changes in educational attainments were also fuelled by educational reforms, starting in the 1980s, that expanded and diversified the Chilean tertiary educational system (Murakami & Nomura, 2020).

Regarding Chile as a country where we can investigate the interactions between labour, technology, and disasters Chile supplies an environment particularly well suited to study the technological change driving the skill premium evolution and technological upgrading due to the occurrence of natural catastrophes. On the one side, in Chile, like most Latin American countries, the skill premium is suggested as the main force driving the observed rise and fall of income inequality in recent decades (Acosta et al., 2019; Guerra-Salas, 2018; Parro & Reyes, 2017). The decline in income inequality is an important step in improving the population's assessment of their well-being. This is especially relevant to Chile: as the country has grown economically, even reaching high-income status[2], its high and persistent income inequality has come to the fore. In this regard, the study of the skill premium evolution for Chile may provide lessons for other economies since the country is often considered a *model* for other middle-income *emergent* economies, particularly in Latin America (Ramos et al., 2013; Sánchez-Páramo & Schady, 2003). Countries transitioning to higher income levels have sometimes seen Chile as a successful case of escaping the *middle-income trap* where factors such as political and economic stability, trade and financial liberalizations, and investment in education played a role (Galeano & Gallego, 2018). Besides, the transition from middle to high-income status may be even more difficult as globalisation increases (Eeckhout & Jovanovic, 2007). Of 101 middle-income countries in 1960, only 16 became high-income by 2012; among them were only two Latin American countries: Chile and Uruguay[3] (World Bank, 2013, 2020). As noted above, although Chile has reached high levels of income, it lags behind in terms of dimensions of well-being, including income inequality, compared to high-income economies in other regions (OECD et al., 2019). Therefore, examining technological change as

[2] The World Bank classifies countries into four income groups—low, lower-middle, upper-middle, and high-income countries using thresholds based on Gross National Income (GNI) per capita in current USD Income. Chile in 2012 was assigned to the high-income category since its GNI per capita was higher than USD$12,615 in that year (World Bank, 2020).
[3] Other examples of countries in the 16's group by region: Europe (Greece, Poland, Portugal) and Asia (Republic of Korea, Oman).

a contributor to the skill premium evolution has important implications for well-being and our understanding of the economic development process.

Furthermore, Chile supplies a unique location for studying the potential for technological upgrading in the aftermath of catastrophes since it is characterized by recurring severe earthquakes. Ten of the most destructive earthquakes (8 M_w[4] and above) hit Chile in the past century (Barrientos & CSN Team, 2018). In the last decade, three earthquakes over this magnitude affected different Chilean regions in 2010, 2014 and 2015. Earthquakes supply the opportunity to analyse technological upgrading due to their unexpected occurrence and destructive ability. We can measure differences in demand for specific kinds of labour, like ICT labour, as a proxy of technological upgrading, given the technological change in production. We can also study the expected increases in demand for labour devoted to reconstruction activities, e.g., Construction labour.

Assessment of the implications of the RBET and ALM models reveals that understanding the role of cognitive and social skills, as well as the potential for technological upgrading in the aftermath of disasters, is essential for efficient policy. In the context of a race between supply and demand driving the skill premium, policymakers need to consider the possible effects as the dominant factor changes over time since interactions between both factors can lead to increases and decreases in the skill premium. Therefore, coordination between intersectoral policymakers (e.g., labour and education) is required to balance the supply of skills provided by the education system with the needs of labour markets. Similarly, in the case of the ALM model and the study of cognitive and social skills, some policy implications arise from the relevance of tasks and abilities that are expected to increase the demand for better-educated workers. The examination of technological replacements in the aftermath of a disaster, however, can contribute to the recovery process by encouraging technical adoption in planning and labour policies focusing on the employment of lesser favoured labour in reconstruction activities.

The rest of this chapter is as follows. We first discuss the RBET and ALM models' implications and the importance of their analysis for Chile, given the inconclusive nature and small quantity of previous research on this country. Similar motivation underlies our analysis of cognitive and social skills driving the skill premium and our inquiry on how an earthquake can be considered as a substantial event affecting the pace of technological change and, consequently, demand for specialized employment like ICT labour. From these discussions, we proceed to introduce the research questions that motivate this work. Subsequently, we state the objectives of this research, followed by the relevance and contribution of this thesis. We conclude with an overview of the thesis structure.

[4] M_w refers to the Moment Magnitude scale, which is usually used for measuring earthquakes' "size". The M_w values are proportional to an earthquake's total energy release (NOAA, 2019).

1.2. Problem statement

In the light of the overview above, this section provides specific backgrounds of the issues that will be discussed in this work, as well as its implications for the Chilean context. For the sake of clarity, we describe these issues in three subsections with their respective key questions.

1.2.1. On the RBET model

As introduced, the RBET model allows us to examine how supply and demand factors influence the skill premium at the same time. This concurrency represents a *race* between education as a supplier of skills, and technology, as a force demanding more skilled labour. Numerous empirical studies support the evidence for the RBET model, particularly in high-income countries (see, e.g., Acemoglu & Autor, 2011; Autor et al., 2020, 1998; Goldin & Katz, 2008; Katz & Autor, 1999). Conceptually, the RBET model employs an aggregate production function with Constant Elasticity of Substitution, CES (Acemoglu, 2002; Goldin & Katz, 2008, 2009). The CES primary inputs are wages and quantities of skilled and unskilled workers. The value of elasticity of substitution between skilled and unskilled labour is a critical parameter in the model because it shows how changes in either technology or supplies affect demand and wages. The model predicts that the skill premium rises if increases in the relative supply of skilled workers do not compensate for their relative demand growth. Alternatively, if the supply rises faster than the pace of demand, the skill premium will decrease.

Bringing the data to the RBET model allows us to estimate the model parameters empirically. However, the estimation is problematic because insufficiently general models or inappropriate estimation approaches have either led to estimates incompatible with the RBET model or imposed incorrect assumptions (Varella, 2008b). For example, unexpected signs and magnitudes for parameters used to derive the elasticity of substitution between skilled and unskilled labour have been obtained. According to the RBET model, a negative sign for the supply factor coefficient is expected to establish the negative relationship between this factor and the skill premium (see section 2.2.1). However, some estimation methods report positive signs generating an inconsistent estimate which can also lead to the computation of a negative elasticity of substitution between skilled and unskilled labour (see section 2.2.2). In this regard, not only the unexpected sign but also the magnitude of the supply factor coefficient has attracted interest. Commonly, the elasticity of substitution between skilled and unskilled labour is computed as the reciprocal of the supply factor coefficient, where small estimates of this coefficient translate to *implausibly* large elasticities. Most of the prominent RBET literature developed until the 2000s decade reports estimates in the interval [1, 3] (see, e.g., Acemoglu, 2002; Ciccone & Peri, 2005; Gallego, 2012; Goldin & Katz, 2009; Robbins, 1996), and this range has been taken as a *consensus* (see footnote 5). However, researchers have recently reported higher values, in some cases, after extending the analysis period using the same data (Acemoglu & Autor, 2011; Blankenau & Cassou, 2011; Varella, 2008a, 2011). Additionally, the elasticity estimation can also be difficult because of modelling and data structure, among other issues (Acemoglu & Autor, 2011; Acosta et al., 2019; Blankenau & Cassou, 2011; Borjas et al., 2012). Therefore, bringing the data to the RBET model can be problematic (Acosta et al., 2019).

In the case of Chile, the RBET evidence is inconclusive, and most of it focussed on a period before the 2000s. Although similar estimation methods have been applied, they have yielded both expected and contradictory estimates for different periods of interest. For example, Gallego (2012) found support for the RBET model using cointegration techniques on data between 1960-2000. In that period, both the skill premium and the relative supply of skilled labour increased. The study reported a negative coefficient for the relative supply of skilled workers, as expected, and the elasticity of substitution was between one and two. Conversely, Murakami (2014) used the same Chilean data source for 1974-2007 using similar estimation methods. However, the Murakami (2014) results did not support the RBET predictions since the coefficient representing the supply factor was non-significant, and it has sometimes been positive, which is an unexpected result according to the theory. As discussed above, a positive sign leads to the computation of a negative elasticity. Some reported that the obtention of negative values might result from imprecision in data and methods (Blankenau & Cassou, 2011; Havranek et al., 2020) or “improbable estimation results”.

Recapitulating, some documented changes in the path of the data might explain the Murakami (2014) results for the Chilean case. The 2000s and 2010s witnessed a rapid decline in the skill premium, while the relative supply of skilled labour showed an increasing trend (Fernández & Messina, 2018; Murakami, 2014). Similar changing patterns in the skill premium have been reported for other Latin American countries, and some authors have warned about the estimation challenges to the evaluation of these changing patterns (Acosta et al., 2019; Fernández & Messina, 2018). To try to fill this gap in the context of the RBET model for the Chilean case, we propose to consider the following two key questions:

- ❖ Does the evidence in Chile over 1980-2018 support the RBET model, its predictions, and implications?
- ❖ To what extent can we address the methodological difficulties that arise in estimating and testing the RBET model?

1.2.2. On the ALM model and the complementarity between cognitive and social skills

The ALM model has arisen as a candidate to explain some trends and changes in labour demand not explained by the RBET model. For example, the RBET model fails to elucidate the faster growth of high and low skilled occupations along with the drop in middle-skilled jobs observed in countries such as the US and some European countries, known as job polarisation (Autor et al., 2006; Goos &

Manning, 2007; Van Ark & O'Mahony, 2016). In contrast, the ALM model can account for the interactions among skills, tasks and technologies (Acemoglu & Autor, 2011). Acemoglu and Autor (2011) define a *task* as "a unit of work activity that produces outputs" and *skill* as "a worker's endowments of capabilities for performing various tasks". Some empirical studies document the interactions predicted by the ALM model (Autor et al., 2003; Goos et al., 2014; Goos & Manning, 2007; Sebastian, 2018; Spitz-Oener, 2006).

In Chile, studies examining some implications from the ALM model are recent, and the evidence contradicts its main predictions. For example, Almeida et al. (2020) analysed the impact of complex software as a proxy for ICT-related technologies, finding a displacement of skilled workers since cognitive tasks started to be performed by the analysed software. These findings contradict the ALM model prediction of technology as a complement for tasks performed by high-skilled workers. The results of Almeida et al. (2020) are in line with recent studies showing a broader class of jobs at risk due to the potential displacement role of frontier technologies (Arntz et al., 2016; Frey & Osborne, 2017). Another Chilean study reported movements of workers from low-skilled occupations towards high-skilled and middle-skilled occupations (Zapata-Román, 2021). Hence, most educated workers face competition from advanced innovations and labour with a lower endowment of skills despite high-skill analytical abilities. In this regard, it is still an open question whether cognitive tasks might be less important in jobs typically employing skilled labour leading to changes in the skill premium.

Regarding the suggested reversal in demand for cognitive skills and the increasing importance of social or people skills (e.g., communication, cooperation with others) as complements for the former, to the writer's knowledge, the influence of the complementarity between cognitive and social abilities on the skill premium has not been studied in the case of Chile. Consequently, we are investigating the importance of tasks and the endowment of abilities of skilled labour under the ALM model and recent documented trends, respectively, in that country. Specifically, in our setting, two questions arise:

- To what extent does the evidence support the ALM model, its predictions, and implications, particularly those related to skilled labour, for Chile in 2009 -2018?
- To what extent do cognitive and/or social skills drive the skill premium for Chile in 2009 - 2018?

1.2.3. On natural disasters and demand for ICT labour

ICT and related technologies have been considered one of the major driving forces for modern technical change (see, e.g., Acemoglu & Autor, 2011; Almeida et al., 2020; Hwang & Shin, 2017). For this reason, it can be reasonably supposed that equipment endowed with ICT will replace physical capital damaged by catastrophes occurring in recent decades. For Chile, indicators covering assets like hardware, telecommunications and software show that the share of ICT in total investment has been growing, reaching values around 10% in 2010, in line with high-income countries like Spain and Italy, resulting in important ICT capital formation (ECLAC, 2013). Therefore, we might expect an increase in ICT technological adoption would imply that much of the technology replacement after a disaster in recent decades will be based on ICT-compatible capital goods.

The rapid move towards ICT-based equipment might result in greater demand for workers involved in the ICT sector. We assume that these workers are in high demand throughout the adoption phase of ICT-compatible equipment (O'Mahony et al., 2008), but researchers have ignored this issue in the context of the impacts of natural disasters on labour. Overall, researchers have studied the impact on labour as a substitute for damaged capital, leading to improvements in labour demand and wages (Belasen & Polachek, 2009; Benson & Clay, 2004; Cavallo et al., 2013; Skidmore & Toya, 2002). With greater relevance to a technological replacement, Leiter et al. (2009) reported higher physical capital accumulation and employment growth in regions affected by disasters. However, if a catastrophe promotes a more significant capital stock, it does not necessarily imply positive impacts on labour participation. Tanaka (2015) found a negative impact on employment, despite over-investment in physical capital. The author speculates that the decrease of population in the affected area may be a possible reason. This assumption is in line with the view that labour markets promote technical improvements since technology upgrading requires workers' skills to be profitable (Acemoglu & Autor, 2012).

The demand for ICT employment has gone largely unnoticed in the literature that examines the interactions between natural disasters and labour markets compared to other shocks to the workforce, like recessions and pandemics. For instance, recessions can affect ICT employment negatively (Holm & Østergaard, 2015). Conversely, the recent COVID-19 pandemic has affected the ICT workforce less than other occupations, given that their jobs mainly require teleworking and involve less exposure to social or face-to-face interactions (Pouliakas & Branka, 2020; Redmond & Mcguinness, 2020). Our work tries to address this gap in understanding the effect of natural disasters on specific types of labour, such as ICT employment, proposing the consideration of the following key question:

- ❖ To what extent do recent disasters accelerate the pace of technological change using the demand for ICT labour as a proxy for technological replacement?

1.3.Research objectives

As noted above, the literature has gaps in empirical research related to the RBET and ALM models and on the impacts of natural catastrophes on labour. Specifically, little research has involved applying the RBET and ALM model to Chile as well as the analysis of the ability's endowment of skilled labour in order to understand how technological change drives the skill premium. This section outlines the specific research aims, which will be pursued by three essays trying to fill the identified gaps.

We first explore the race between the demand for skilled labour coming from technology and the supply of skills coming from the education system using the RBET model's conceptualization. Both supply and demand factors can explain the skill premium evolution, but from a time perspective, one or the other can be the dominant factor or the *winner* in the context of a race. Besides, the difficulties in bringing the model to the data give us our motivation to explore and propose alternative estimation strategies, and given the importance of the elasticity of substitution between skilled and unskilled labour, we also investigate the implications of estimates beyond the *consensus*.

We then aim to investigate how the task-content of jobs and specific abilities drive the skill premium. We use same approach as in the RBET model study to configure the skill premium although with some technical adjustments due to data availability (see section 3.5.1.1 for details). Using the ALM model's conceptualization, we examine three kinds of tasks that are relevant to skilled labour: i) non-routine analytical (e.g., researching, analysing); ii) non-routine interactive (e.g., negotiating, coordinating); iii) routine cognitive (e.g., calculating, bookkeeping) (See Table 3.1 and Table 3.2 for details of these task classifications). Also, we investigate the extent to which the demand for cognitive and social abilities, separately and simultaneously, and software skills (as a proxy for recent technological advancements) influence the skill premium (see Table 3.6 for details of these skill categories).

A third goal is to study how a severe earthquake might affect the pace of technological change, using changes in demand for ICT labour as a proxy for technological replacements. We assume that equipment compatible with ICT replaces the destroyed machinery since it has been suggested that ICT and related technologies drive the technical change in production. Additionally, we examine the Construction labour, considering some expected reconstruction activities taking place after disasters.

We also aim to make policy recommendations based on our results. Mainly, we inform labour and educational policymakers as well those designing post-disaster recovery and labour strategies.

1.4.Relevance and contributions of the thesis

Most of the evidence for the RBET model for Chile is focussed on the pre-2000 period (see, e.g., Beyer et al., 1999; Gallego, 2012), while attempts to extend the period under analysis by using post-2000 data fail due to some estimation difficulties (Murakami, 2014). The skill premium grew considerably in the pre-2000 period, but it has been declining since the 2000s (Gallego, 2012; Murakami, 2014; Murakami & Nomura, 2020; Parro & Reyes, 2017). This changing pattern gives insights on the dominance, in the pre-2000 period, of the demand or SBTC factor, and, in the post-2000 period, of the labour supply. We empirically analyse this evidence for Chile, considering the difficulties of applying the model to the data. We apply an Unobserved Component Model, UCM, with Bayesian inference, UCM-Bayesian, to overcome the difficulties seen in other estimation methods, mainly cointegration techniques. To our best knowledge, this is the first study that applies UCM-Bayesian to test the RBET model's implications. Thus, we contribute to the literature by extending the race between education and technology to cover the last four decades. Besides, we contribute to the evidence on how labour market conditions such as unemployment rates and minimum wages affect the evolution of the skill premium. In the case of our UCM-Bayesian implementation, it might help other applied researchers using skill premium data featured by changing patterns (e.g., in some other Latin American countries).

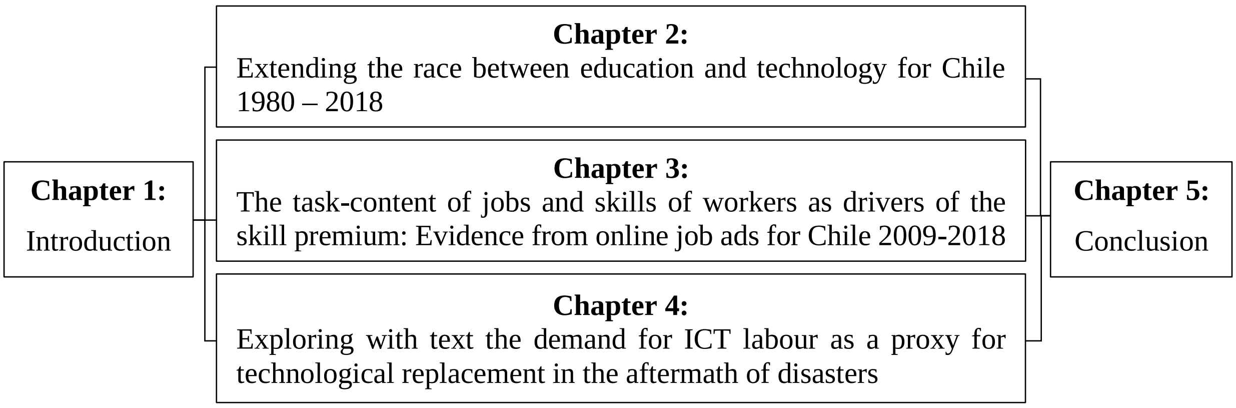
Furthermore, our examination of some of the economic relationships suggested by the ALM model and the complementarity between cognitive and social skills contribute to this fairly recent discussion for Chile (Almeida et al., 2020; Zapata-Román, 2021). Some marginal contributions beyond the focus of this research are exploiting online job postings as data and constructing dictionaries of tasks and skills to give more tools for labour analysis.

In the literature on natural disasters and labour markets, there is no study that examines the role of natural disasters as promotors of technological upgrading by observing changes in ICT labour. We contribute to fill this gap. Our study also adds to a growing literature on the analysis of text in online job postings, specifically in the context of disasters' impacts on labour markets.

1.5.Thesis Structure

The current chapter has supplied the relevant background and outlined the purposes of this thesis with reference to the ways in which each essay will fill the identified literature gaps. The structure of the thesis and the subjects of its three component essays are outlined in Figure 1.1..

Figure 1.1. Thesis structure



Each essay is self-contained, i.e., it is composed of an introduction and literature review, followed by a conceptual and empirical framework, data, estimation methods, empirical results, discussion, and concluding section. We have introduced the use of three separate and unrelated theoretical backgrounds, i.e., the RBET and ALM models, and a conceptual framework to understand how natural disasters can affect technological replacement, in Chapter 2, Chapter 3 and Chapter 4, respectively. However, in this thesis these conceptual frameworks are interrelated in the following two aspects. First, the RBET and ALM models assumes that technological change is *biased* towards skilled workers, but the first only analyse the impact on the skill premium of the relative demand for skilled labour, while ALM model also considers the task-content of jobs. Then, both models can help us to understand the interactions between labour and technology from alternative points of view but using the same labour output i.e. the skill premium. Second, both RBET and ALM models assume that technological progress steadily evolves. Conversely, in our modelling to understand the impact on technology replacement of natural disasters we assume that substantial events like natural disasters might affect the pace of technological change. Thus, we can conceptually examine how labour markets respond to regular forces like technological change and unexpected shocks like catastrophes considering the potential technological upgrading.

Chapter 2 examines and tests the RBET model for Chile. We review past studies examining the implications of the model for Chile. Given the methodological difficulties in some of these studies testing the theoretical implications of the model, we also review the literature focusing on modelling and methods of estimation. Using biannual labour survey data for the period 1980 – 2018, our estimates for the skill premium in Chile show an inverted U-shaped pattern growing up to the early 2000s and reducing since that decade. The relative supply of skilled workers increased over the period under analysis. To estimate the RBET parameters, we apply cointegration techniques (e.g., the Vector Error Correction Model) as in past studies on Chile (e.g. Gallego, 2012; Murakami, 2014). Like Murakami (2014), our estimation procedures yield coefficients that are inconsistent with the model's conceptualization. To overcome this problem, we apply a UCM-Bayesian model, which yields parameter estimates that are consistent with the underlying theory. Our results in the context of the race between technology and education show that, in the pre-2000 period, the increasing relative demand for skilled labour attributable to SBTC contributing to the increasing skill premium appears to be the dominant factor. In contrast, after 2000, the relative supply occupies a dominant position, inducing a declining trend in the skill premium. Our estimate for the elasticity of substitution between skilled and unskilled is 6.5, implying that these kinds of labour are more substitutable than common thought. From these findings we draw conclusions and show some of the implications of our results from both methodological and policy perspectives. Limitations of the study and suggestions for further research are considered.

Testing for implications based on the ALM model (particularly how cognitive tasks and skills drives the skill premium) is examined in Chapter 3. We evaluate how analytical and interactive non-routine work activities drive skilled relative to unskilled wages. We provide the conceptual and empirical framework by reviewing current literature on the ALM model. Furthermore, we provide a review of specific abilities of workers, such as cognitive and social skills, individually and in combination. Using samples of online job ads, we model the interactions between the skill premium and variables representing tasks and skills using the Vector Autoregressive framework, VAR (Sims, 1980). Our results show that non-routine cognitive tasks, both analytical and interactive, weakly drive the skill premium. We do not, however, find evidence of the impact on the skill premium arising from simultaneous demand for cognitive and social skills. Conclusions and implications are drawn from these results.

The impacts of natural disasters on technological upgrading proxied by changes in ICT labour are explored in Chapter 4. Our conceptual and empirical framework is built by reviewing the literature on extensions of growth models and more literal assumptions of the creative-destruction hypothesis focusing on impacts on labour. We explore the demand for ICT labour as a proxy for technological replacements to analyse the impact of the 27^{th} of February 2010 earthquake using a subsample of online job ads used in Chapter 3. It implements the Structural Topic Model, STM, developed by Roberts et al. (2016, 2013) to discover the *hidden topics* in our collection of job ads. The estimation of the model yields 53 topics. The prevalence of the ICT labour topic does not change after the earthquake. Conversely, the Construction labour topic's prevalence is significantly different after the disaster, suggesting that reconstruction activities lead to differences in this sector. Our results suggest that there was no substantial technological replacement in the aftermath of the 2010 Biobío earthquake.

Chapter 5 summarises the research findings, draws general conclusions from those findings and discusses some policy implications of the results. The limitations of each analysis are also discussed, along with suggestions for further research.

2. Essay I: Extending the race between education and technology for Chile 1980 – 2018

Using recurrent bi-annual labour survey data from 1980 to 2018, we extend the race between education and technology, the RBET model, for Chile. This model allows us to evaluate how both demand and supply factors explain the evolution of the skill premium. Our motivation is the lack of evidence in the post-2000 period and "estimation difficulties" reported by past studies. These difficulties mainly imply the computation of parameters inconsistent with RBET theory. For example, the RBET model establishes an inverse relationship between the skill premium and the relative supply of skilled labour: this is, a negative coefficient on the relative supply of skilled labour is expected. However, some estimated positives coefficients. Besides, a positive coefficient leads to the computation of negative estimates for the elasticity of substitution between skilled and unskilled labour, which is also against the RBET conceptualization. We also find "estimation difficulties" using cointegration techniques. Alternatively, we apply an Unobserved Component Model estimated by Bayesian inference, UCM-Bayesian, whose results are more consistent with the RBET model. We find that both demand and supply factors drive the evolution of the skill premium. In the context of the race between technology and education, in the pre-2000 period, the relative demand attributable to skill-biased technological change, SBTC, with its rapid acceleration contributing to a high skill premium, is suggested as the dominant factor. However, in the post-2000 span, the demand factor has been surpassed by strong increases in the relative supply, suggesting education as the dominant factor inducing a declining trend in the skill premium. Furthermore, our estimate for the elasticity of substitution is 6.5. The value greater than one implies that skilled and unskilled workers are imperfect substitutes but more substitutable than commonly thought, given the past estimates for this parameter.

Keywords: RBET, SBTC, technological change, skill premium, elasticity, substitution, education
JEL Classification: D24, I26, J23, J24, J31, N36, O15, O33

2.1.Introduction

In Chile, like most Latin American countries, the skill premium, i.e. the skilled labour wages relative to unskilled labour or the wages gap between tertiary-educated workers and those with less education, has been suggested as the main force driving the observed rise and fall of income inequality in recent decades (Acosta et al., 2019; Guerra-Salas, 2018; Parro & Reyes, 2017). Since Latin America has been the most unequal region globally since the last century (Williamson, 2010), more attention needs to be paid to the evolution of its primary contributor. In this regard, there is a joint agreement about the inverted U-shaped pattern shown by the skill premium evolution during the last five decades in Chile. It grew considerably since the mid-1970s, peaked in 1987, then held steady over the 1990s, and it has been declining since the 2000s (Gallego, 2012; Murakami, 2014; Murakami & Nomura, 2020; Parro & Reyes, 2017). In this case, distinguishing between two periods, i.e., pre-2000 and post-2000, provides insights on the rise and fall of the skill premium. Most of the major economic reforms that feature the Chilean economy occurred in the pre-2000 period, with trade liberalization as the most relevant (Beyer et al., 1999). This openness allowed the absorption of foreign technologies, most of them biased toward skilled labour, leading to higher demand for skilled labour and increasing the skill premium (Gallego, 2012). At the same time, economic development from physically intensive sectors, i.e. agriculture and manufacturing, moving to less physically demanding and more knowledge intensive sectors such as services, also led to higher demand for better-educated workers before the 2000s (Buera & Kaboski, 2012).

In the post-2000 period, the skill premium decline has been linked to the increasing availability of skilled workers due to the expansion of tertiary education (Murakami & Nomura, 2020; Parro & Reyes, 2017). This expansion, which was fuelled by critical educational reforms in the 1980s and 1990s (Schneider, 2013; Valiente et al., 2020), has resulted in fewer returns to a lower educational level (Murakami & Nomura, 2020). Therefore, we can evaluate the skill premium evolution by following the interaction between forces encouraging the demand for skilled workers and affecting their supply. In particular, examining the role of technologies biased towards better-educated workers and education as a supplier of skills as significant forces driving the skill premium evolution has motivated a vast literature (see, e.g., Acemoglu & Autor, 2011; Autor et al., 2020, 2008; Katz & Murphy, 1992).

On the demand side, it has been suggested that technological change is *biased* towards skilled workers, enhancing their productivity and wages (Acemoglu, 2002; Acemoglu & Autor, 2011; Autor et al., 2008; Katz & Murphy, 1992). Consequently, an increasing demand for skilled workers arises from technology. On the supply side, the educational system supplies these skills or qualifications, affecting workers' educational attainment. Therefore, in a context of increased relative demand for skilled workers arising from skill-biased technological change (SBTC henceforth), the skill premium also responds to changes in the relative supply of better-educated workers. These simultaneous shifts in demand and supply factors implicitly refer to a *race between education and technology* (RBET henceforth) model (Acemoglu & Autor, 2011; Autor et al., 2020; Goldin & Katz, 2008; Katz & Murphy, 1992). The idea behind the RBET model was initially noted by Tinbergen (1972, 1974). There is ample empirical evidence supporting its main predictions with pioneer applications such as Katz & Murphy (1992), Levy & Murnane (1996), Johnson (1997), and Acemoglu (2002), among others. The RBET model is also known as the SBTC model or framework and supply-demand framework: the literature generally uses these different names to refer to the same model. In this research, we use the RBET acronym to refer to this conceptual idea.

The fundamental notion behind the RBET model is that there is steady growth in the demand for skilled labour from SBTC, and there is varying growth in the relative supply of skilled workers. Since both forces co-occur, the impact on the skill premium of one of them depends on how the other variable responds. Thus, it rises if the supply of skilled workers does not compensate for technology's demand for skilled labour growth. Alternatively, if the supply rises faster than the demand, the skill premium will decrease.

Conceptually, this framework relies on a production function with Constant Elasticity of Substitution, CES, where skilled and unskilled labour are imperfect substitutes. The elasticity of substitution between both kinds of labour plays a pivotal role since its value approaching to zero, one or positive infinity rules the RBET framework's predictions (see section 2.2.1.1). For example, an elasticity of a value equal to one would imply that changes in relative quantities of both kinds of labour are precisely proportional to their relative changes in wages. Alternatively, elasticities below one would imply complementarity between them while elasticities greater than one, skilled and unskilled are substitutes (details in section 2.2.1.1). In addition, the elasticity will show the strength of the influence of both the SBTC and the relative supply of skilled labour on the skill premium. For example, an elasticity higher than one might imply a more substantial SBTC effect on the skill premium, assuming that skilled workers become relatively more productive due to technology improvements. Therefore, several interpretations and assumptions rely on the empirical estimation of the elasticity of substitution. Some of these interpretations might lead to important policy issues. For example, a higher substitution level between skilled and unskilled labour might imply that skilled workers are moving to less skilled positions resulting in worker’s skills underutilisation and inefficient investment in education or training.

The implementation required to test the RBET model empirically relies on specifications proposed in the most prominent studies in this literature (see, e.g., Acemoglu, 2002; Acemoglu & Autor, 2011; Katz & Murphy, 1992). Overall, we need variables representing the skill premium, the SBTC term and the relative supply of skilled labour. Typically, we can obtain the skill premium and the relative supply by using observed wages and skilled and unskilled labour quantities, respectively. In the case of the demand coming from technology or the SBTC term, a standard procedure is using linear trends to capture its dynamics (since we do not directly observe this component) as in most of the influential studies (see . e.g., Acemoglu, 2002; Acemoglu & Autor, 2011; Katz & Murphy, 1992).

More generally, estimation strategies depend on the nature of the data, the chosen estimation method and the specification of both the model and the variables. However, researchers have warned that the RBET estimation is difficult, beset by numerous methodological and data problems (Acosta et al., 2019; Borjas et al., 2012; Fernández & Messina, 2018; Varella, 2008b). In the case of Chile, some produced theoretically unfeasible results due to the appearance of a wrong sign for the coefficient representing the relative supply of skilled labour, i.e., a positive sign (Murakami, 2014; Robbins, 1994b). This output is theoretically unfeasible because a positive sign is against the expected negative relationship between the skill premium and the relative supply of skilled workers (see section 2.2.1.3). Besides, a positive coefficient leads to negative estimates for the elasticity of substitution between skilled and unskilled labour (see Eq. (2.12) related statements). These "improbable estimation results" have received less attention in this literature. One reason for this lack of interest might be that most research focuses on high-income countries like the US. In these countries, the skill premium has continued to show a long-run increasing pattern (Autor et al., 2020). In contrast, as noted above, in Latin American countries like Chile, the skill premium shows an inverted U-shaped pattern in recent decades. In this context, researchers warned that the evaluation of skill premium drivers in a context of changing patterns is problematic, and it might impose incorrect interpretations or assumptions (Acosta et al., 2019; Havranek et al., 2020; Varella, 2008b). In the light of the issues discussed, it is proper to emphasise alternative estimation methods of implementing and empirically testing the RBET framework predictions.

The empirical testing of the RBET framework for Chile is sparse and inconclusive. Some studies analysing data in the pre-2000 period support the RBET model by documenting an SBTC effect leading to the increasing skill premium and elasticities of substitution between skilled and unskilled labour between one and two (Beyer et al., 1999; Gallego, 2012; Robbins, 1994a). In this period, the relative supply of skilled labour plays a minor role. In contrast, studies that do not support this evidence cite the appearance of theoretically unfeasible results, i.e., "improbable estimation results", or elasticities beyond the *consensus*[5], i.e., the range [1, 3], as a reason to reject the RBET model for Chile (Murakami, 2014; Robbins, 1994b; Sánchez-Páramo & Schady, 2003).

On one side, Robbins (1994b) and Murakami (2014) reported that the relative supply changes in some of their models could not explain the skill premium for 1975-1992 and 1974-2007, respectively. In both cases, the rejection of the RBET predictions was due to the appearance of a wrong sign for the coefficient representing the relative supply of skilled labour, i.e., a positive sign, which is theoretically unfeasible (see section 2.2.1.3). Robbins (1994b) and Murakami (2014) suggested that the differences in quality education between traditional and private universities, whose creation and development were

[5] The notion of a so-called *consensus* related to the values of the elasticity of substitution between skilled and unskilled labour in the range [1, 3] was proposed by some researchers such as Cantore et al. (2017) and Johnson (1997) based on the estimates observed in some of the most prominent papers of this literature (see e.g., Acemoglu, 2002; Goldin & Katz, 2009; Katz & Murphy, 1992).

fuelled by major educational reforms in the 1980s and 1990s (Valiente et al., 2020), as a possible reason. However, studies applying cohort analyses reported that the quality between these higher education institutions did not influence the skill premium (Gallego, 2012; Gindling & Robbins, 2001). Besides, a positive coefficient on the supply factor can lead to the computation of negative elasticities of substitution between skilled and unskilled labour (see section 2.2.2). Some suggested that the obtention of negative elasticities might arise from imprecision in data and methods (Blankenau & Cassou, 2011; Havranek et al., 2020) or "improbable estimation results", as discussed above.

On the other side, researchers have declared "implausible" some elasticity of substitution estimates beyond the *consensus*. For example, Sánchez-Páramo & Schady (2003) estimated elasticities around 10 for 1970-1999, arguing that such values were imprecise and improbable without questioning the *consensus* since the RBET conceptualization does not consider an upper threshold for elasticities. However, elasticities around four are frequent in the RBET literature, while elasticities around five or six are less frequent (Havranek et al., 2020). In Latin American countries, empirical estimates suggested elasticities around three and four (Acosta et al., 2019; Manacorda et al., 2010) and around 11 for the important maquiladora industry in Mexico (Varella & Ibarra-Salazar, 2013). These examples suggest that there should be no upper threshold for reporting positive estimates. Also, publication biases have been suggested since most published estimates adhere to the *consensus* (Havranek et al., 2020). Thus, it seems that the evidence favouring the rejection of the RBET framework for Chile has relied on theoretically unfeasible results or the appearance of *non-consensual* elasticity values. More generally, studies using Chilean data that reported wrong results or larger elasticities of substitution do not refer to any imprecision in data and methods (see, e.g., Murakami, 2014; Sánchez-Páramo & Schady, 2003). Some wrong results were due to the application of cointegration techniques (e.g., Murakami, 2014), which have some limitations in their ability to test causal relationships, leading to a rejection of expected theoretical relationships (Guisan, 2001; Moosa, 2017).

To recapitulate, most studies using data over the pre-2000 period report that the SBTC and the relative supply of skilled labour drive the skill premium. Given the increasing pattern in the skill premium, this evidence shows that increases in the relative supply of skilled workers did not compensate for the growth in technology's demand for skilled labour. Therefore, within the RBET model, the SBTC appears to be the *winner* or the dominant factor. On the other hand, research examining data beyond 2000 rejected the RBET predictions due to "improbable estimation results". Therefore, we cannot declare a winner within the RBET model in the post-2000 period. However, the skill premium decrease in the post-2000 has been linked to the higher availability of skilled workers fuelled by the significant expansion of higher education institutions, but without strong empirical evidence (see, e.g., Murakami & Nomura, 2020; Parro & Reyes, 2017). In this regard, a still open question is what will happen if the demand for skilled labour due to SBTC does not grow fast enough to meet the increased relative supply of skilled labour. In this case, under the RBET model, we expect that the dominant factor in the post-

2000 period will be education, even though the SBTC is operating, resulting in the skill premium decline.

This study aims to test the RBET model empirically for Chile using data from 1980 to 2018. This conceptual idea has successfully explained the influence on the skill premium of changes in demand and supply factors. However, its empirical testing has been a much-debated topic in recent decades. In this regard, the main aspects that motivate this research are the lack of evidence, mainly in the post-2000 period, and "estimation difficulties" reported by past studies. We apply (among other methods) cointegration techniques within a Vector Error Correction Model, VECM. Researchers have previously reported that both the skill premium and the relative supply for Chile are not trend-stationary variables, i.e., there are unit roots in the data (Beyer et al., 1999; Gallego, 2012; Murakami, 2014). A VECM allows us to analyse non-stationary variables, which might lead to spurious relationships using standard regression estimation. Our VECM yielded the wrong sign for the coefficient representing the relative supply of skilled labour; this was similar to the experience of Murakami (2014) and Robbins (1994b). Given these "estimation difficulties", we apply an Unobserved Component Model, UCM, estimated by Bayesian inference as an alternative strategy, UCM-Bayesian. Some advantages of our UCM-Bayesian strategy are its flexibility, allowing the model components to vary over time, and the direct estimation of elasticity (with VECM, it is obtained as a reciprocal, the usual procedure in this literature, as presented in Eq. (2.12) related statements). Bayesian estimation also allows us to include the expected value for the elasticity of substitution according to the *consensus* and past studies for Chile (Beyer et al., 1999; Gallego, 2012) and for other countries in the region (Manacorda et al., 2010) as *priors*.

Our UCM-Bayesian results support the empirical evidence for the RBET model. We found that both forces, demand, and supply factors, play a role in explaining the evolution of the skill premium in Chile between 1980 and 2018. In the context of the race between technology and education, in the pre-2000 period, the relative demand attributable to SBTC with its rapid acceleration contributing to a high skill premium is suggested as the dominant factor. However, in the post-2000 span, the demand factor started to be surpassed by strong increases in the relative supply, suggesting education as the new *winner*, inducing a declining trend in the skill premium. Furthermore, our estimate for the elasticity of substitution is greater than one: this is 6.5, which would imply that both kinds of workers are imperfect substitutes but more substitutable than commonly thought, given the past estimates for this parameter. These results contribute to the understanding of the relationship between demand and supply as forces driving the skill premium by revisiting the evidence for Chile and the implications underlying the race between technology and education during the last four decades. It is also important to emphasise alternative estimation methods like UCM-Bayesian, which can handle the assumptions imposed by the RBET model since, as noted above, the technical implementation of the RBET can be problematic.

The Essay is structured as follows. We begin by presenting the concepts and theory behind the RBET model and how is its typical empirical implementation. We then review the literature in three subsections, encompassing the evidence of the RBET model for Chile, the elasticity of substitution issues and a review of the estimation techniques applied by researchers to the empirical implementation and testing of the RBET model. After this review, we develop our empirical models, followed by the description of the data and the methods. Next, we present and discuss our results. Conclusions will be presented in the final section.

2.2. The RBET conceptualization and estimation

This section presents the conceptualization behind the RBET model and how its main parameters are estimated and interpreted. Also, we discuss some limitations.

2.2.1. The RBET model conceptualization

Conceptually, and as introduced above, the RBET model relies on an aggregate production function CES (Solow, 1957), representing the relationship between different inputs generating an output. Solow (1957) introduced a "technical term" into this CES function to allow for technical change, i.e., any variation such as speedups or slowdowns, among others. However, this technical term is a form of factor-neutral technological change. Thus, by definition, technical change is neutral on changes in relative prices: in other words, technical improvements do not cause changes in the relative prices or wages for a particular kind of labour. Alternatively, the technical change cannot be viewed as factor-neutral given the increases in the skill premium in the last decades in high-income countries (mainly in the US) in conjunction with increases in their relative supply (Violante, 2016). Therefore, to evaluate the interactions between the skill premium as the output or the dependent variable and skilled and unskilled labour quantities as inputs or explanatory variables, researchers extended the Solow insights by adding a *factor-biased technical change* or SBTC (Acemoglu, 2002; Goldin & Katz, 2008, 2009; Katz & Murphy, 1992). The form of the CES function with skilled and unskilled quantities modelled with factor-specific productivities is:

$$Q = [(A_S S)^{\rho} + (A_U U)^{\rho}]^{1/\rho}, \tag{2.1}$$

where Q is aggregated output, S and U are quantities of skilled and unskilled workers, respectively, A_S is the factor augmenting technology for the skilled and A_U is the factor augmenting technology for the unskilled. The term ρ, with $\rho \leq 1$, is the substitution parameter.

The economic interpretation of Eq. (2.1) can be applied to several situations. For example, it can refer to an economy with more than one good and consumers' utility functions involved or to the existence of two different economic sectors producing imperfect substitute goods, among others (Acemoglu & Autor, 2011). For convenience, in this conceptual approach, skilled and unskilled workers are imperfect substitutes, producing only one good. Besides, technology is exogenous, and its central assumption is that the technological change is *skill-biased* because of the complementarity between technology and most educated workers. As a result, this complementarity might lead to higher demand for skilled labour.

The following sections show how this conceptualization configures the elasticity of substitution and the skill premium from Eq. (2.1). We also summarize the main predictions and implications of the model.

2.2.1.1.The elasticity of substitution

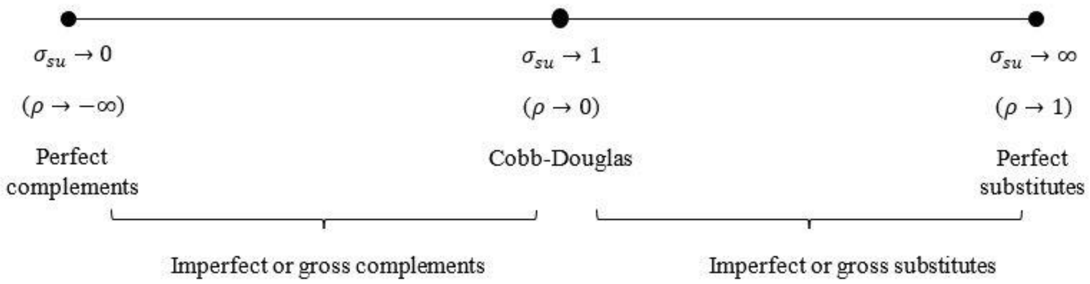
This section discusses how the RBET model configures the elasticity of substitution between skilled and unskilled labour. We note the elasticity as σ_{SU} (the subscript SU refers to skilled and unskilled). Formally, from Eq. (2.1) σ_{SU} is given by:

$$\sigma_{SU} \equiv 1/(1-\rho), \rho \in (-\infty, 1), \tag{2.2}$$

where ρ is as in Eq. (2.1). The value of elasticity of substitution σ_{SU} establishes the level of substitution or complementarity between skilled and unskilled labour. Additionally, it shows how changes in either technology (given by A_S and A_U) or supplies (S and U) affect demand and wages.

There are three special cases for σ_{SU} given that $\rho \in (-\infty, 1)$ (Acemoglu, 2002). First, when $\sigma_{SU} \to 0$ (or $\rho \to -\infty$), skilled and unskilled workers will be perfect complements and they are used in fixed proportions (output function is Leontief). Secondly, when $\sigma_{SU} \to \infty$ (or $\rho \to 1$), both kinds of workers are perfect substitutes. One implication of this case is that the relative availabilities of each kind of labour are not related to changes in wages; therefore, skilled workers can be placed in unskilled job positions and *vice versa*. Third, when $\sigma_{SU} \to 1$ (or $\rho \to 0$), the output function tends to be Cobb Douglas, which implies that, as production grows, skilled wages grow in the same proportion as unskilled wages. Figure 2.1 summarizes the three cases.

Figure 2.1. The three special cases for the elasticity of substitution



2.2.1.2. The skill premium setting and main implications of the model

The skill premium configuration assumes competitive labour markets with many firms and factors paid at the marginal product value. From Eq. (2.1), the wage for skilled labour is

$$w_S = \frac{\partial Q}{\partial S} = A_S^{\rho}\left[A_U^{\rho}(S\,/\,U)^{-\rho} + \ A_S^{\rho}\right]^{(1-\rho)/\rho}, \tag{2.3}$$

and, for unskilled,

$$w_U = \frac{\partial Q}{\partial U} = A_U^{\rho}\left[A_U^{\rho} + A_S^{\rho}(S/U)^{\rho}\right]^{(1-\rho)/\rho}, \tag{2.4}$$

where w_S and w_U are skilled wages and unskilled wages, respectively. Two relevant implications result from Equations (2.3) and (2.4). First, since $\frac{\partial w_S}{\partial S/U} < 0$ in Eq. (2.3), this result implies that the skilled labour demand curve is downward sloping. Therefore, *ceteris paribus*, as the skilled workforce increases, its wages should fall. Secondly, since $\frac{\partial w_U}{\partial S/U} > \ 0$ in Eq. (2.4), it implies that, *ceteris paribus*, as the number of skilled workers rises in the workforce, the wages of unskilled workers should rise. In this case, both groups of workers are q-complements, where higher quantities of the one improves the marginal product of the other. In other words, intensifications in the use of skilled labour lead to improvements in the marginal productivity of unskilled labour and *vice versa.*

To set the skill premium, ω, as the ratio between the skilled and unskilled wages, the Eq. (2.3) and (2.4) are combined, and the elasticity re-ordered, as follows:

$$\omega = \frac{w_S}{w_U} = \left(\frac{A_S}{A_U}\right)^{\rho}\left(\frac{S}{U}\right)^{-(1-\rho)} = \left(\frac{A_S}{A_U}\right)^{(\sigma_{SU}-1)/\sigma_{SU}}\left(\frac{S}{U}\right)^{-1/\sigma_{SU}}. \tag{2.5}$$

Rewriting Eq. (2.5) by taking logs of both sides yields

$$\ln\omega = \left(\frac{\sigma_{SU}-1}{\sigma_{SU}}\right)\ln\left(\frac{A_S}{A_U}\right) - \frac{1}{\sigma_{SU}}\ln\left(\frac{S}{U}\right). \tag{2.6}$$

The Eq. (2.6) links the skill premium defined as log wage differentials between skilled and unskilled wages, $ln\ \omega$, to the SBTC term represented by $ln\left(\frac{A_S}{A_U}\right)$ and to the relative supply of skills, $ln\left(\frac{S}{U}\right)$. These relationships refer to the conceptual foundations of the RBET model with the skill premium as the output and the SBTC and the educational attainment of the workforce as inputs.

2.2.1.3.Summarizing the RBET model predictions

The representation in Eq. (2.6) shows that the association between the skill premium, SBTC and the relative supply of skilled labour can be expressed as a simple log-linear relationship. Therefore, we can summarize the expected primary outcomes in terms of the interactions between these variables (Acemoglu & Autor, 2011).

Formally, to evaluate how the skill premium responds to SBTC, we differentiate the Eq. (2.6) as follows:

$$\frac{\partial \ln\omega}{\partial \ln(A_S/A_U)} = \frac{\sigma_{SU}-1}{\sigma_{SU}}. \tag{2.7}$$

Given the values of elasticity of substitution σ_{SU} presented above (see section 2.2.1.1 and Figure 2.1), a value of $\sigma_{SU} > 1$, i.e., skilled and unskilled labour are imperfect substitutes, in the Eq. (2.7) implies that relative improvements in the SBTC term increase the skill premium. Hence, we expect skilled workers to become relatively more productive due to technological improvements. Conversely, if $\sigma_{SU} < 1$, i.e., skilled and unskilled groups are gross complements, then we expect an increase in the SBTC term to shift the relative demand curve inward and reduce the skill premium.

Regarding the effect of the provision of skills on the skill premium, the differentiation of the Eq. (2.6) to the relative supply factor $ln\left(\frac{S}{U}\right)$ yields

$$\frac{\partial \ln \omega}{\partial \ln(\frac{S}{U})} = -\frac{1}{\sigma_{SU}} < 0. \tag{2.8}$$

Eq. (2.8) implies that, for a given *skill bias of technology* captured here by the SBTC term, an increase in the relative supply of skills $ln\left(\frac{S}{U}\right)$ reduces the skill premium. Therefore, there is an inverse relationship between both variables. In other words, the higher availability of skilled workers might lead to relative lower wages for this kind of labour. Therefore, the elasticity of substitution rules these interactions and establishes the level of substitution or complementarity between skilled and unskilled labour.

If the RBET model is viewed as a race, both forces compete to be a more significant influence on the skill premium; therefore, the impact of one of them depends on how the other variable changes. In practical terms, the skill premium rises if increases in the SBTC term are not counteracted by increases in the supply of skilled workers. Conversely, the skill premium will decrease if the supply of skills increases more rapidly than the SBTC term.

2.2.2. The estimation of the RBET model parameters and some limitations

The last section reviewed the conceptual foundations of the RBET model, the specification of its main variables and parameters and the expected primary outcomes. In this section, we show how the RBET model represented by Eq. (2.6) can be applied to the data. Firstly, we need to obtain variables standing for the skill premium, the SBTC term and the relative supply. While the skill premium and the relative supply can be quantified by using observed wages and quantities of skilled and unskilled labour, the SBTC term $ln\left(\frac{A_S}{A_U}\right)$ is not directly observed. However, it has been assumed that the SBTC dynamics can be captured by a linear trend in the most prominent studies in this literature (see, e.g., Acemoglu, 2002; Acemoglu & Autor, 2011; Katz & Murphy, 1992) as follows:

$$ln\left(\frac{A_S}{A_U}\right) = \beta_0 + \beta_1 t, \tag{2.9}$$

where t is calendar time, β_0 is the intercept, and the β_1 coefficient measures the rate of change of the SBTC term over time. Recalling the assumption about the technical change biased to better-educated workers, Eq. (2.9) represents a linear trend increase in the demand for skilled workers coming from technology. Thus, substituting the SBTC dynamics formalized in Eq. (2.9) into Eq. (2.6) and adding time subscripts to the components, except for σ_{SU} which is assumed fixed, yields

$$\ln \omega_t = \left(\frac{\sigma_{SU}-1}{\sigma_{SU}}\right)\beta_0 + \left(\frac{\sigma_{SU}-1}{\sigma_{SU}}\right)\beta_1 t \; - \frac{1}{\sigma_{SU}} \ln\left(\frac{S}{U}\right)_t. \tag{2.10}$$

By reordering and adding the parameters to be estimated, the specification is

$$\ln \omega_t = \beta_0 + \beta_1 t \; - \beta_2 \ln\left(\frac{S}{U}\right)_t + e_t, \tag{2.11}$$

where β_0 and β_1 are as in Eq. (2.9). There is a SBTC effect if $\beta_1 > 0$ (see Eq. (2.7) related statements). As discussed earlier, the coefficient on $\left(\frac{S}{U}\right)_t$, $-\beta_2$, is expected to be negative (see Eq. (2.8) related statements). Then, for a given *skill bias of technology* captured by the SBTC term, an increase in the relative supply of skills reduces the skill premium. Overall, the relationships displayed in Eq. (2.11) imply that the pace of technological progress grows steadily, and the relative supply of skilled workers can vary over time. If both forces do not cancel each other out, the skill premium will change according to the expected outcomes above (see section 2.2.1.3). Also, $-\beta_2$ supplies an estimate of $-\left(\frac{1}{\sigma_{SU}}\right)$, therefore, it allows us to estimate the elasticity of substitution σ_{SU}, which is usually obtained as a reciprocal. From Eq. (2.10) and Eq. (2.11) we see that

$$-\beta_2 = -\left(\frac{1}{\sigma_{SU}}\right), \tag{2.12}$$

then, the estimated $-\beta_2$ must be inverted and multiplied by -1 to compute the elasticity of substitution. In this sense, we require negative β_2 estimates to generate theoretically plausible elasticity estimates (Blankenau & Cassou, 2011). If not, our estimation could not yield reasonable elasticity estimates, i.e., a negative elasticity, in the context of the RBET conceptualization (see Figure 2.1).

The RBET model has been demonstrated to be workable, theoretically attractive and empirically successful in estimating these interactions (Acemoglu & Autor, 2011). However, as introduced above, some limitations have emerged such as the "improbable estimation results" due to the wrong sign results for the relative supply of skilled labour coefficient. In addition, data and method choices might lead to a bias in reported elasticities (Havranek et al., 2020). Overall, these difficulties in bringing the model to the data might impose incorrect interpretations or assumptions (Acosta et al., 2019; Havranek et al., 2020; Varella, 2008b). Also, some have warned that using linear time trends and unit roots in the data might imply analysing using non-stationary variables. However, the testing of stationarity or presence of unit roots is rarely reported in this literature (Varella, 2008b). Besides, some suggest that the usual computation of the elasticity of substitution, σ_{SU}, as a reciprocal might be inaccurate since small differences in the relative supply coefficients can lead to large variations in elasticity estimations (Behar, 2009; Havranek et al., 2020). In this sense, the computation of direct estimates would be appropriate. We review these methodological issues in one of our Literature Review segments (see section 2.4.3). In the next section, we present our empirical models.

2.3.Empirical models

This section formulates the empirical models for estimation based on the conceptual RBET model. We estimate a base and an extended model. Recapitulating from the RBET model specified in Eq. (2.11), we specify our empirical base model as

$$\ln \omega_t = \beta_0 + \beta_1 t - \beta_2 \ln \left(\frac{S}{U}\right)_t + \beta_3 Ch98_t + e_t, \tag{2.13}$$

where ω_t is the skill premium at time t, β_0 and β_1 represent the trend component that acts as a proxy for the SBTC, and $\left(\frac{S}{U}\right)_t$ is the skilled labour supply relative to unskilled at time t. $Ch98$ is a dummy variable for methodological change in the data categorization of educational attainment (1 = March 1998 and onwards, 0 = before March 1998). This change in data categorization consisted of splitting secondary education into regular secondary education and vocational secondary education from March 1998 and onwards. Our extended model includes variables related to institutional controls as follows:

$$\ln \omega_t = \beta_0 + \beta_1 t - \beta_2 \ln \left(\frac{S}{U}\right)_t + \beta_3 Ch98_t + \beta_4 Unem_t - \beta_5 MinW_t + e_t, \tag{2.14}$$

where $Unem_t$ is the unemployment rate in time t and $MinW_t$ is the minimum wage in time t. In the case of Chile, unemployment rates and minimum wages are considered labour market conditions that might also affect the evolution of the skill premium as reported in previous studies (Gallego, 2012; Gindling & Robbins, 2001; Murakami, 2014). A positive relationship between the skill premium and unemployment, i.e., a higher unemployment rate leading to an increase in wage differential, might suggest that a disproportionately high number of unskilled workers are represented among the unemployed. Consequently, their wages would probably fall more rapidly than the wages of the skilled labourers, leading to a greater skill premium (Gindling & Robbins, 2001). On the other hand, if unemployment affects predominantly skilled labour, a negative relationship between unemployment and the skill premium might occur (Larrañaga, 2001). Thus, our results will provide additional insights on the influence of this variable on the skill premium.

Regarding labour policies to establish minimum wages, it is assumed that these interventions affect the wages of unskilled labour. Therefore, without changes in skilled labour wages, the increases in minimum wages might lead to a decline in the skill premium. Previous studies show evidence of this inverse relationship like Murakami (2014), although from models reporting “estimation difficulties”. Others reported no statistically significant estimates (Gallego, 2012; Gindling & Robbins, 2001). Therefore, in our extended model, we expect a negative and significant coefficient standing for the expected inverse relationship between the skill premium and minimum wages.

2.4.Literature review

This literature review encompasses three strands. First, we describe the past attempts to test empirically the RBET model for Chile, establishing the presence of self-contradictory evidence.

Overall, the evidence for the Chilean case is inconclusive, and researchers have highlighted unexpected results or "improbable estimation results" related to estimating the coefficient on the supply factor and, consequently, the computation of the elasticity of substitution between skilled and unskilled labour. These reported difficulties motivate us to review the reporting of the elasticity of substitution and how researchers have empirically tested the RBET model. These two aspects constitute our second and third strands, respectively.

2.4.1. The Skill Premium and the RBET evidence for Chile

The skill premium in Chile has shown an inverted U-shaped pattern during the last six decades. After continuous growth since the 1960s (Gallego, 2012), the 2000s witnessed a fall in the skill premium (Murakami, 2014; Parro & Reyes, 2017). In this context, the distinction of two periods, i.e., pre-2000 and post-2000, to identify the rise and fall of the skill premium, respectively, might not be trivial. Most major economic reforms in Chile occurred between 1975 and 1995, with trade liberalization being the most relevant (Beyer et al., 1999). Also, the process of economic development arising from physically intensive sectors (e.g., agriculture, manufacturing) moving to less physically and more knowledge intense sectors (e.g., services) led to higher demand for skilled labour, increasing the skill premium (Buera & Kaboski, 2012).

The increasing pattern of the skill premium in Chile between 1960 and 2000 can be evaluated in terms of inter-decade growth. On average, it increased from 0.4 in the 1960s to 1.3 and 1.5 in the 1980s, and 1990s, respectively (Gallego, 2012), suggesting that the relative demand for skilled workers increased in most economic sectors. The trade liberalization that took place in the 1980s and the 1990s was one of Chile's most crucial major economic reforms in the pre-2000 period. Researchers suggest that one of the implications of this openness was the absorption of foreign technologies biased towards skilled labour (Beyer et al., 1999; Gallego, 2012; Robbins, 1994a). Before 2000, Chile imported about 85% of non-transportation machinery and equipment from the US and OECD countries (Gallego, 2012). Therefore, the SBTC, fuelled by the trade liberalization that began in the 1980s, might explain the increased relative demand for skilled labour in the pre-2000 period.

Regarding the skill premium decline observed in the post-2000 period, the increasing availability of skilled labour in the labour market has been suggested as the leading cause. Researchers have linked this higher relative supply to the expansion of tertiary education and the exit of the older and less educated cohorts (Murakami & Nomura, 2020; Parro & Reyes, 2017). The observed data shows that Chilean higher education experienced substantial growth in recent decades. According to educational and census data for 1984-2018, people enrolled in tertiary education sextupled (INE, 2017; MINEDUC, 2020). The 18–24 age group enrolled in tertiary education grew from 189,151 (11% of this age group) in 1984 to 521,882 (31%) in 2002. In 2018, it exceeded 1.2 million (approximately 67% of the 18–24 age group). Apart from the endogenous response of agents to the increase in returns to education, these changes in educational attainments were also fuelled by educational reforms, starting in the 1980s, that expanded and diversified the Chilean tertiary educational system (Murakami & Nomura, 2020). However, the skill premium reversal can also be explained by demand-side factors, such as the increases in unskilled workers' wages due to the boom in commodity prices (in particular, copper) observed in the 2000s (Pellandra, 2015). Therefore, in the pre-2000 period, the skill premium increased, fuelled mainly by economic reforms that drove the demand for skilled labour. As the skill premium rose, the relative supply of workers responded to this higher premium faster than the relative demand for skilled labour, pushing the skill premium down in the post-2000 period. Despite the observed educational expansion in the post-2000 period, however, "improbable estimation results" have prevented successful testing of its role under the RBET model.

The empirical testing of the RBET model for Chile is sparse and inconclusive. Studies covering the 1960-2000 period documented an SBTC leading to an increased skill premium with relative supply playing a minor role and elasticities in the range [1, 2] (Beyer et al., 1999; Gallego, 2012; Robbins, 1994a). In contrast, studies that do not support this evidence suggested, on the one side, the production of results that were inconsistent with theoretical expectations and, on the other side, higher elasticities, i.e. out of the range of the *consensus*, as causes leading to rejection of some RBET predictions (Murakami, 2014; Robbins, 1994b; Sánchez-Páramo & Schady, 2003). For example, Robbins (1994b) and Murakami (2014) reported that, in some of their models, the relative supply changes could not explain the skill premium for 1975-1992 and 1974-2007, respectively. In both cases, the rejection was due to results that were counter to expectations given by the RBET conceptualization, i.e., a positive sign for the coefficient representing the relative supply, which is theoretically unfeasible (see Eq. (2.11) and (2.12) related statements). Murakami (2014) and Robbins (1994b) suggested that the differences in the quality education provided by traditional and private universities as a possible reason. However, studies applying cohort analyses reported that the difference in quality between these higher education institutions did not influence the skill premium (Gallego, 2012; Gindling & Robbins, 2001). As in Murakami (2014) and Robbins (1994b), researchers have generally not attempted to interpret these theoretically unfeasible results, which might result from imprecision in data and methods. Measurement error, noise in wages and labour supply data are also usually noted in this literature (Havranek et al., 2020). Besides, researchers have warned of several estimation and data difficulties when the skill premium evolution is changing (Acosta et al., 2019). As reviewed above, the skill premium shows an inverted U-shaped pattern, which might explain the unexpected results that Murakami (2014) reported.

Regarding elasticities beyond the *consensus*, Sánchez-Páramo & Schady (2003) questioned their estimates of around 10 for Chile analysing the span 1970-1999, arguing that such values were imprecise and improbable given the so-called *consensus*. Sánchez-Páramo & Schady (2003) suggested the increasing (almost monotonic) pattern of the skill premium as a potential reason for these larger elasticities, without questioning the *consensus*. Researchers have suggested a bias in the publication of elasticity estimates, with most published estimates belonging to the *consensus*, i.e., values in the range

[1, 3], while larger elasticities (four and over) are rare in this literature (Havranek et al., 2020). Therefore, it seems that the evidence cited as grounds for rejecting the RBET model for Chile has been derived results countering the RBET conceptualisation and the production of non-consensual elasticities values. Furthermore, these studies do not refer to any imprecision in data and methods. For instance, Murakami (2014) applied cointegration techniques that rely on assumptions that may not hold. In this regard, researchers have warned that the cointegration approach can lead to results inconsistent with the underlying theory (Guisan, 2001; Moosa, 2017).

This section has outlined the evidence regarding the skill premium and the RBET model for Chile. The changes in the skill premium seen in Chile during the last five decades show different forces behind its pre-2000 rise and post-2000 fall, without conclusive evidence for or against the RBET predictions. The steadily rising relative supply of skilled labour over the period might suggest a dominant role for this variable, given the skill premium decline after 2000. However, there is no evidence of the expected negative relationship between the skill premium and the relative supply of skilled labour under the RBET model. Studies using post-2000 data reported "improbable estimation results" due to wrong results that might lead to negative elasticities. Also, elasticities beyond the *consensus* have been pronounced implausible for Chile. These elasticity and estimation issues motivate us to review, in the following sections, some issues related to this parameter and the most frequently applied methods focusing on both estimation and modelling concerns in the RBET literature.

2.4.2. The elasticity of substitution

The elasticity of substitution between skilled and unskilled labour is one of the most frequently estimated parameters in labour economics (Havranek et al., 2020). As presented in sections 2.2.1.1 and 2.2.1.3, it governs the subsequent predictions of the RBET model, showing how changes in either SBTC or the educational attainment of the workforce affect the skill premium. The elasticity of substitution determines if skilled and unskilled workers are complements or substitutes. Also, some researchers have computed changes in the relative demand as a residual using elasticity as a critical parameter (see, e.g., Gallego, 2012). Conceptually, we would expect zero or positive values for the elasticity of substitution (see section 2.2.1.1 and Figure 2.1). The reporting of negative elasticities in previous studies highlights that such findings violate standard economic theory (see, e.g., Blankenau & Cassou, 2011; Kearney 1997). However, despite being positive and theoretically plausible, larger elasticities, i.e., beyond the *consensus*, have sometimes been described as implausible (see, e.g., Sánchez-Páramo & Schady, 2003; Varella, 2008). Overall, elasticities around four are common in the RBET literature, while elasticities around five or six are less frequent (Havranek et al., 2020). Some estimated elasticities are above three and four for several Latin American countries (Acosta et al., 2019; Manacorda et al., 2010) and around 11 for the crucial maquiladora industry in Mexico (Varella & Ibarra-Salazar, 2013). These values show that there should be no upper threshold for elasticities reporting.

The computation of larger elasticities might partly be explained by data granularity, the features of the analysed period, and the use of different definitions for skilled labour, among other factors. For instance, it has been suggested that higher than annual data granularity might expand the elasticity, which can be due to measurement error associated with higher frequency data (Havranek et al., 2020). A measurement error in the relative supply variable results in attenuation bias in the estimated regression parameter. Then, smaller coefficients translate to larger elasticities. Also, where particular periods are concerned, higher elasticities have been linked to a more rapid SBTC in the US, but the evidence is inconclusive since values between one and two also have featured periods of SBTC acceleration (Acemoglu, 1998; Katz & Murphy, 1992). Also, some larger elasticities can be found in studies using different measurements for skills or analysing substitution between age groups. For instance, Varella (2011) estimated a global measure of elasticity using different definitions for skilled labour. When skilled labour was defined as secondary graduates, the values varied from three to four, while values for the definition as primary graduates were between six and 12. For a pool of Latin American countries, including Chile, suggested elasticities were around three when the relative supply was based on worked hours and above five when the relative supply was established on the workforce or population estimations (Manacorda et al., 2010). For Germany, the elasticity between secondary and primary graduates was four (Glitz & Wissmann, 2017). Regarding the use of workers' groups based on their age, Glitz & Wissmann (2017) reported an elasticity of substitution between young and old workers of 8.2 in Germany, which is somewhat higher than the comparable estimates by Card & Lemieux (2001) of around five for the US and six for Canada.

Overall, large elasticities might have some implications: firstly, the possibility of switching between skilled and unskilled workers is higher, and the impact on the skill premium for an observed relative supply of skilled labour will be less important than the demand factor (Katz & Murphy, 1992; Varella & Ibarra-Salazar, 2013). Also, Acemoglu (2002) suggests that if the elasticity is greater than two, the skill premium will be an increasing function of the relative supply of skilled labour: in other words, the SBTC is endogenous. This endogeneity implies that technological change is biased by profit incentives where the market availability of skilled labour drives the creation and adoption of technologies. In other words, a rise in the relative supply encourages so much SBTC that the demand for skills increases more than enough to counterbalance the potential increase in the supply of skilled workers. Thus, both the declines and increases in the skill premium might be related to increases in the supply of skilled labour.

In terms of estimation procedures, the computation of the elasticity of substitution as a reciprocal of the relative supply estimated coefficient may also contribute to larger elasticities. As introduced earlier, this procedure dominates most of the prominent RBET literature (see, e.g., Acemoglu, 2002; Goldin & Katz, 2009; Katz & Murphy, 1992). Besides, it has been argued that the magnitude of the estimates applying least-squares methods (e.g., Ordinary Least Squares, OLS) in a regression analysis is usually smaller than expected since they are downward biased towards zero because of error measurement (Hausman, 2001). Thus, small coefficients translate to large elasticities of substitution. Consequently, the inverted coefficient estimated using least-squares methods or its variations may be inconsistent. We review the OLS applications in this literature and other methods in the next section 2.4.3.

The calculation of the elasticity of substitution as a reciprocal as the standard procedure within the RBET literature might potentially be one of the principal reasons why it is replete with elasticity estimates above one. Havranek et al. (2020) reported that inverted estimates tend to be 1.5 larger than direct estimates. In this context, researchers have suggested that alternative modelling strategies to estimate the elasticity directly are needed (Ciccone & Peri, 2005). For example, some have modelled the relative supply as the dependent variable (Li, 2010). Others applied different specifications (e.g., translog cost function) instead of the CES production function to estimate the elasticity directly (Askilden & Nilsen, 2005; Bergström & Panas, 1992). Therefore, it is possible that direct estimation methods might yield more accurate elasticity of substitution estimates, though they are employed less frequently in the RBET literature.

The issues outlined above show that bringing the data to the RBET framework can be problematic. Therefore, we should proceed with care when choosing empirical strategies to implement and test the RBET model empirically. In the next section, we review some of the strategies most frequently employed to empirically implement and test the RBET model, focusing on methods, data and modelling issues that can lead to model misspecification and estimation imprecision. Where larger elasticities that fall outside the consensus are concerned, it should be recognised that these values do not represent a rejection of the RBET model since there is no upper threshold for positive elasticities. While implausibly large elasticities may prompt debate about the applicability of the RBET model, it is the estimation of unfeasible theoretically results that are, therefore, incompatible with the underlying model that perhaps warrants greater attention.

2.4.3. Methods and modelling issues in the RBET literature

This section reviews some of the estimation strategies most commonly applied to implement and empirically test the RBET model. Although the RBET literature is vast, studies that give extensive detail about their econometric strategies are scarce. We also pay particular attention to the use of linear time trends to proxy the SBTC component and how this approach can influence the elasticity of substitution estimates.

One of the simplest ways to estimate the RBET model parameters is using OLS in a standard linear regression. Pioneering studies applied this approach using the US data, obtaining elasticities between one and three (Acemoglu, 2002; Goldin & Katz, 2008, 2009; Katz & Murphy, 1992). Similarly, Acosta et al. (2019) applied OLS to a pool of 16 Latin American countries, but an insignificant relative supply parameter was estimated (it would have implied an elasticity above four). However, while OLS might yield expected outcomes using aggregated data for a particular country or group, it may perform inadequately with less aggregated data (e.g. at industry or economic sector level). The estimate of industry-specific regressions performed by Blankenau & Cassou (2011) resulted in values that were substantially distant from each other, i.e. in the range [-436, 500], compared to the aggregate estimate of around two. The authors posit that, in the case of negative elasticities, these findings are evidence of a methodological failing when using data at the industry level compared to the aggregated level. Besides, Blankenau & Cassou (2011) suggested an additional difficulty in the regressions at the industry level related to the potential endogeneity between the skill premium and the relative supply as the main reason. This concern arises from the assumption that relative supply is an exogenous variable that may be positively correlated with the skill premium in the long run, which is an idea used in conjunction with OLS. However, these adjustments can be more immediate at the industry level (Ciccone & Peri, 2005; Varella, 2011). Researchers have applied alternative estimation strategies, such as models including omitted fixed effects (e.g. country, region or time) or instrumental variables to overcome potential endogeneity issues. The work on the US conducted by Ciccone & Peri (2005), using OLS without taking into account fixed effects, resulted in an unexpected sign for the relative supply parameter and the computation of negative elasticities. Once they specified models with the state and time as fixed effects, the OLS yielded elasticities of around three. Similarly, Acosta et al. (2019) estimated specifications with country and year as fixed effects, finding elasticities between 3.1 and 3.5 for selected Latin American countries. Other applications using fixed-effects specifications with OLS have been reported for the US (Mallick & Sousa, 2017), Portugal (Nogueira et al., 2017), etc.

Regarding using instrumental variables, IVs, as a modelling approach to control the potential endogeneity described above, this approach has usually been combined with estimation strategies other than OLS. For example, Varella Mollick (2011) aims to estimate the world elasticity of substitution from aggregated data of 52 countries using three IVs related to the relative supply: the educational enrolment ratio, the GDP share of national expenditure on education and the proportion of government spending on education concerning global government expenditure. Using the Generalized Method of Moments, GMM, the world elasticity of substitution values varied between two and three. A similar approach was employed by Blankenau & Cassou (2011) at the industry level in the US, but modelling instruments related to the industry share and wage ratios.

More sophisticated techniques combining IVs and fixed effects specifications are utilised in Ciccone & Peri (2005). Using instruments based on laws regarding child labour and child compulsory school attendance at the state-level in the US, the estimates were obtained using Two-Stage Least Squares, 2SLS, Limited-Information Maximum Likelihood, LIML, and Fuller-LIML, which yielded elasticities of substitution around 1.5. For low and middle-income countries, estimation by GMM has also provided robust evidence regarding technology measures driving the demand for skilled labour (Conte & Vivarelli, 2011). In addition to the estimation using GMM, Razzak & Timmins (2008) employed the Estimated Generalised Least Squares, EGLS, methodology with data from New Zealand.

As a main result, they reported a GMM-EGLS estimation with constant, trend term and four lags as instruments. The estimated coefficient had an implied elasticity of 2.9 for the period under analysis. Similar Maximum Likelihood, ML, approaches using the Seemingly Unrelated Regression, SUR, a multiple-equation system approach, were applied to the US with elasticities above two (McAdam & Willman, 2018). Also, Glitz & Wissmann (2017) employed a SUR approach for Germany, with elasticity estimates as low as 1.6 between skilled and unskilled workers and as high as eight between older and younger workers.

Here, we reviewed the main methods used to overcome the potential endogeneity between the variables evaluated under the RBET model. However, this estimation can be more problematic because of other issues, such as how the trend term that proxies the SBTC component is specified (Acemoglu & Autor, 2011; Borjas et al., 2012) or how the data structure evolves (Acosta et al., 2019). According to the conceptual framework, the SBTC component is modelled as a linear time trend to capture its assumed steady evolution over time (see section 2.2.1.3). Some suggest the inclusion of richer specifications (e.g. spline trend, cubic and quadratic time trends) to capture some deviations in the skill premium (Acemoglu & Autor, 2011; Autor et al., 2008). It has also been suggested that the time trends modelling leads to estimating reasonable elasticities (Blankenau & Cassou, 2011; Ciccone & Peri, 2005). In contrast, the ability of time terms to proxy the SBTC term has been questioned in non-high-income countries since other forces (e.g. external shocks and trade openness) might also affect the relative demand for skilled labour (Acosta et al., 2019). In this regard, adding these variables might make the time trend effect more negligible. Besides, the sensitivity of the elasticity estimates depends on trend specification (Borjas et al., 2012; Fernández & Messina, 2018). For instance, Borjas et al., (2012) studied the elasticity of substitution between native and immigrant workers. They found that adopting a quadratic trend yielded a negative elasticity, and using a cubic trend led to additional volatility. Therefore, the modelling of the time trend is an issue that requires attention, given its implications for the elasticity of substitution estimation. To overcome some of these issues, researchers have prefered the use of time dummies or time fixed-effects models like those reviewed above (Glitz & Wissmann, 2017).

Other difficulties linked to the evolution of the data structure arise. For example, the potential presence of trends in both the skill premium and the relative supply present a challenge to modelling and estimation strategies because there is the possibility of unit roots in the data. Consequently, inferences can be derived from spurious regressions (Granger & Newbold, 1974). However, testing for stationarity or the presence of unit roots is rarely reported in the RBET literature (Razzak & Timmins, 2008; Varella Mollick, 2008). As exceptions, Razzak & Timmins (2008) and Dupuy & Marey (2008) found unit roots in their data for New Zealand and the US, respectively. The data from Colombia, Brazil and Mexico also shows the presence of unit roots (Medina & Posso, 2010; Varella Mollick, 2008). For Chile, Gallego (2012) and Murakami (2014) reported the existence of unit roots in the skill premium and the relative supply of skilled labour, but Beyer et al. (1999) did not find support. Most of these studies applied cointegration techniques (in section 2.6.1.3 we give a detailed description of cointegration). These techniques posit a relationship between non-stationary variables such that some linear combination of these variables is stationary.

Recurrent approaches to estimating cointegrating relationships include the Granger-Engle estimation (Engle & Granger, 1987) and VECM (Johansen, 1995), among others. These estimation strategies rely mainly on OLS and or ML techniques. For instance, Dupuy & Marey (2008) used cointegration techniques based on OLS, while Medina & Posso (2010) and Varella Mollick (2008) estimated their coefficients of interest using ML on VECMs. For the Chilean case, Gallego (2012) applied ML to a VECM, reporting that the skill premium and the relative supply are cointegrated, and the estimation yielded plausible results. Conversely, Murakami (2014) reported inconclusive evidence about cointegration using Engel-Granger cointegration tests. However, some models yielded the "wrong" sign for the relative supply of skilled labour resulting in the computation of a negative elasticity, our so-called "improbable estimation results". According to Murakami (2014), this discrepancy was not caused by "improbable estimation results" but potential bias in the data (discussed above in section 2.4.1).

Overall, the use of cointegration techniques is frequent in the RBET literature. On the one hand, some have applied these techniques using data for Chile (Beyer et al., 1999; Gallego, 2012), Mexico (Varella, 2008b), Colombia (Medina & Posso, 2010), the US (Balleer & Rens, 2013; Dupuy & Marey, 2008), Norway (Von Brasch, 2016), Germany (Hutter & Weber, 2017), and New Zealand (Razzak & Timmins, 2008), among others, with most of them yielding theoretically expected results. On the other hand, some researchers have warned of difficulties with cointegration methods. For example, Razzak & Timmins (2008) reported that unit roots tests are sensitive to lag order, making the estimation of the RBET model difficult. Also, Von Brasch (2016) shows that some caution must be applied to the consideration of VECM assumptions (e.g., weak exogeneity between variables, significant lags). Additionally, concerns have emerged regarding the estimation of the elasticity of substitution when using cointegration. Firstly, it has been suggested that the elasticity is sensitive to the configuration of the cointegration strategy in terms of the presence or absence of time trends and lag order (Varella, 2008b). Secondly, using a VECM implies the computation of elasticity as a reciprocal, which can also contribute to less precise elasticities. Havranek et al. (2020) report that inverted estimates tend to be 1.5 larger than direct estimates. Thirdly, cointegration might yield smaller elasticity estimates than time fixed effects or instrumental variables strategies (Havranek et al., 2020). Therefore, these estimation difficulties might lead to incorrect assumptions or the reporting of biased estimates (Von Brasch, 2016). More generally, researchers have warned about some limitations of cointegration techniques that might make them unsuitable for testing causal relationships in Economics and Econometrics (see. e.g., Guisan, 2001; Moosa, 2017).

Most of the estimation strategies reviewed above rely on classical frequentist methods. Rarer are methods using Bayesian estimation. For example, Balleer & Rens (2013) used a Bayesian Vector

Autorregresion framework to explore the implications of SBTC for business cycle fluctuations in the US, where the skill premium is one of the several variables analysed. Also, for the US, Guvenen & Kuruscu (2009) applied Bayesian methods to model and estimate individuals' expectations regarding the advent of SBTC. Since we applied a UCM-Bayesian strategy to this research, we give additional coverage of Bayesian estimation in section 2.6.2.2.

2.5.Data and estimation of variables

This section introduces the data used in this research and the estimation of the skill premium and the relative supply of skilled labour. We have accumulated these variables from a series of biannual cross-section labour surveys for 1980-2018. We follow closely the strategies from previous studies for Chile and RBET literature.

We use data from the Employment and Unemployment Survey for Greater Santiago (in Spanish, *Encuesta de Ocupación y Desempleo del Gran Santiago* or EOD) carried out by the University of Chile since 1956 (University of Chile, 2020). We use biannual data (March and June), which has been available since 1980. Each biannual survey covers about 3,000 households and interviews all household members (about 10,000 individuals). The survey is considered a good representation of the Chilean labour market (Gallego, 2012; Robbins, 1994c). The EOD has applied practically the same questionnaire from its creation, which reinforces the comparability of its data. This feature has been helpful for the design and evaluation of labour policies. Also, the EOD has been widely used in studies examining wage differentials and their drivers in Chile (e.g., Beyer et al., 1999; Gallego, 2012; Gindling & Robbins, 2001; Murakami, 2014; Robbins, 1994b). As noted above, we use biannual survey data over 1980 and 2018, i.e., 76 time periods.

Regarding the method of constructing estimates of the skill premium and the relative supply of skilled workers, we closely follow the strategies of Autor et al. (2008), Card & Lemieux (2001), and Ciccone & Peri (2005), among others. Other researchers also have applied these strategies to Chile (see, e.g., Gallego, 2012; Murakami, 2014; Beyer et al., 1999). In particular, we have adopted the definitions and thresholds of Murakami (2014) for our computation of skilled and unskilled labour variables. To compute the skill premium, we define skilled labour as suitable for college or post-secondary graduates and unskilled labour as suitable for graduates of high-school or secondary education or those whose education has not reached these levels. Similarly, in the computation of the relative supply, we consider skilled labourers as equivalent to college graduates and unskilled labourers as equivalent to high-school graduates, as in previous studies (Card & Lemieux, 2001; Ciccone & Peri, 2005; Gallego, 2012; Murakami, 2014).

To estimate the skill premium, we regress the monthly log earnings for each of the 76 time periods on the usual determinants of wages (e.g., education level, experience). Then, we compute the predicted log wages difference between the college graduates (our skilled group) and high-school graduates (our unskilled group) as our proxy for the skill premium. We focus on monthly earnings according to the EOD, and our group of interest is restricted to salaried and full time (more than 30 hours a week) male workers aged 14-65 years[6]. To adjust for compositional changes, we use weighted averages from the construction of education by experience subgroups. The skill premium estimation consists of the following three steps:

Step 1) Construction of education by experience subgroups to adjust for compositional labour changes (e.g., different skills levels) within each sub-group: we define five education categories as our measure of schooling in order to classify different workers' school attainments (primary, high school dropouts, high school graduates, some college and college graduates), and four potential work experience[7] groups (0-9, 10-19, 20-29 and 30 or more years). Combining the education and experience categories, we construct 20 education by experience sub-groups. We use the average share of total hours worked monthly for each sub-group as weights.

Step 2) Estimation of the wage equation for skilled and unskilled workers regressing a Mincer type equation: we regress (log) earnings for each one of the time periods in our study, estimating the following standard wage equation (for conciseness ignoring the error term) [8]:

$$\ln W_{i,t} = \gamma_0 + \gamma_1 educ\_cat'_{i,t} + \gamma_2 exp_{i,t} + \gamma_3 exp^2_{i,t} + X'_i\delta \tag{2.15}$$

where, $W_{i,t}$ are hourly wages for individual i in time t. W is computed by dividing monthly wages by monthly working hours, and W is expressed in December 2018 Chilean pesos using the *Unidad de Fomento* as a deflator[9]; $educ\_cat$ are educational categories defined in Step 1) with primary education as the base category; exp is work experience; X is a vector containing dummies for workers classified as the heads of households and employed in the public sector. Also, it includes *industry* which is a variable factor including eight industries such as agriculture, mining, and construction, among others, with manufacturing as the base category. We use the results from Eq. (3.3) to compute the predicted wages for skilled and unskilled workers as detailed in the next step.

Step 3) Prediction of the average wage for skilled and unskilled groups and computation of the skill premium. We estimate the predicted log wages using regression results from Step 2) evaluated according to the corresponding experience level (5, 15, 25, or 35 years based on experience categories) and the base categories included in vector X. We compute the predicted log wages difference between the college graduates and high-school graduates as our proxy for the skill premium. We use the average share of monthly hours worked for each education x experience sub-group (formed in Step 1) as

[6] Following Beyer et al. (1999), Card & Lemieux (2001), Gallego (2012), Murakami (2014) and Rothwell (2012), we exclude women because of potential sample selection biases generated by changes in their labour participation.

[7] Potential work experience is computed as follows: potential work experience = age - years of schooling – 6 (age for starting compulsory education): cat 1. 0-9, cat 2. 10-19, cat 3. 20-19 and cat 4.30+.

[8] This method allows control by other demographic characteristics of the labour supply which are not related to the education premium.

[9] The Unidad de Fomento (UF) is a Chilean unit of account. The exchange rate between the UF and the Chilean peso is constantly adjusted for inflation.

weights. Thus, we quantify the difference between skilled and unskilled wages, which is our skill premium measure.

Regarding calculating the relative supply of skilled labour, we refer to skilled and unskilled labour as college graduates and high-school graduate equivalents, respectively. Therefore, we define our relative supply measure as the ratio of monthly hours worked by the former to the latter. This estimation is based on Card & Lemieux (2001), Ciccone & Peri (2005) and Murakami (2014). Since the RBET model assumes only two production factors, skilled and unskilled labour, we classify all workers into categories of college graduates and high-school graduate equivalents. First, for each observation and, based on total monthly worked hours, we calculate the amounts of unskilled labour or High School equivalents, Lu, in each observation t as follows:

$$Lu_t \equiv L_t^{PS}\left(\frac{\overline{w}_t^{PS}}{\overline{w}_t^{HSG}}\right) + L_t^{HSD}\left(\frac{\overline{w}_t^{HSD}}{\overline{w}_t^{HSG}}\right) + L_t^{HSG} + L_t^{SC}\left(\frac{\left(\frac{\overline{w}_t^{UG}}{\overline{w}_t^{SC}} - 1\right)}{\left(\frac{\overline{w}_t^{SC}}{\overline{w}_t^{HSG}} - 1\right) + \left(\frac{\overline{w}_t^{UG}}{\overline{w}_t^{SC}} - 1\right)}\right) \tag{2.16}$$

where $L_t^{PS}, L_t^{HSD}, L_t^{HSG}$ and L_t^{SC} are total monthly worked hours by workers who have completed school up to the primary level, dropped out of high school, graduated from high school, and attended some college[10], respectively. $\overline{w}_t^{PS}$, $\overline{w}_t^{HSD}$, $\overline{w}_t^{HSG}$, $\overline{w}_t^{SC}$ and $\overline{w}_t^{UG}$ are average wages for each t for workers who have completed school up to the primary level, dropped out of high school, graduated from high school, have attended some college, and are university graduates, respectively. Similarly, the amounts of skilled or college equivalents, Ls, in t are:

$$Ls_t \equiv L_t^{UG} + L_t^{SC}\left(\frac{\left(\frac{\overline{w}_t^{SC}}{\overline{w}_t^{HSG}} - 1\right)}{\left(\frac{\overline{w}_t^{SC}}{\overline{w}_t^{HSG}} - 1\right) + \left(\frac{\overline{w}_t^{UG}}{\overline{w}_t^{SC}} - 1\right)}\right) \tag{2.17}$$

where L_t^{UG} is the total monthly worked hours by college graduates' workers, and rest of the variables are as in Eq. (2.16).

Regarding our institutional control variables stated in our extended empirical model in Eq. (2.14), the unemployment rate and minimum wages are obtained from Banco Central de Chile (2020) and Biblioteca del Congreso (2020), respectively. As in wages above, minimum wages are expressed in December 2018 Chilean pesos using the *Unidad de Fomento* as a deflator (see footnote 9).

2.6.Methods

This section introduces our methods, VECM and UCM-Bayesian, to test empirically the models represented in Eq. (2.13) and (2.14) as described in section 2.3.

[10] Following Murakami (2014), workers with some college education are "split" between college equivalents and high-school equivalents based on relative wages.

2.6.1. VECM

The VECM is a reparameterization of a Vector Autoregressive VAR, used in time series analysis to model stationary and non-stationary variables and their interrelationships. Stationarity implies that stochastic properties of a time series are time-invariant, i.e., mean, variance or covariance with other variables do not change over time[11]. Stationarity or stationarity around a deterministic trend (henceforth trend stationarity) is a desirable property in data under analysis since regression analysis using stationary variables does not lead to spurious regressions (Granger & Newbold, 1974).

VECMs allow the modelling of non-stationary variables. This framework can both test for cointegration and estimate cointegrating relationships (i.e., where linear combinations of non-stationary variables are stationary or trend-stationary). Therefore, we need to test stationarity or trend stationarity in the data-generating process in order to specify the VECM specification correctly (Cryer & Chan, 2008; Lutkepohl, 2005).

We follow the next steps for our VECM estimation. Firstly, we test stationarity or trend stationarity in the data. Secondly, we select and estimate the best unrestricted VAR model for our dependent variable in terms of lag order. Thirdly, if the variables are non-stationary, we assess if both variables are cointegrated, and fourthly, we estimate our VECM using the selected parameters in previous steps.

2.6.1.1. Stationarity testing

The stationarity testing is based on the presence of unit roots or unit root processes that are non-stationary. The stationarity can be determined through integration order estimation and stationarity property itself. For example, a series is said to be integrated of order 1, $I(1)$, if it has one unit root, and an $I(d)$ series has d unit-roots. The number of unit roots (at the zero frequency) equals the number of times a series needs to be differenced to make it stationary or trend-stationary (Lutkepohl, 2005). This testing is required since estimation and inference in VARs and VECMs become non-standard in the presence of unit roots in the data.

In practice, however, practitioners often refer to a series as being $I(0)$ or "stationary" if it is "mean reverting" (i.e., has a mean) and tests are generally not directed at determining if the variance or the covariance properties of the series are stable. Additionally, it is commonly said that a variable is "stationary around trend" if it reverts (i.e., if it has a long-run tendency to return) towards some stable function of time. Here we will also adopt this convention of referring to a series as being stationary if it is mean-reverting or stationary-around-trend if it reverts to some stable function of time. The

[11] In other words, stationarity implies that the stochastic process which generates the data is time unvarying. Specifically, the stochastic process $\{Y_t\}$ is strictly stationarity if the joint distribution of $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ is the same as the joint distribution of $Y_{t_1-k}, Y_{t_2-k}, \dots, Y_{t_n-k}$ for all choices of time points $t_1, t_2, \dots, t_n$ and all choices of time lag k. Then, when $n = 1$ the univariate distribution of Y_t is the same as that of Y_{t-k} for all t and k (the Y's are marginally identically distributed) and the mean and variance are constant over time (Neusser, 2016).

stationarity property of a series is, therefore, confirmed by applying tests for unit-roots presence and stationarity property itself (Neusser, 2016).

In this research, we conduct testing on each variable individually, applying the Augmented Dickey-Fuller (ADF) and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) (Kwiatkowski et al., 1992). While the null hypothesis of ADF is the existence of a unit root, which implies a non-stationary series, in KPSS, the null is the presence of stationarity. Some noted that this hypothesis swap allows complementarity between both tests (Neusser, 2016). However, researchers have warned of contradictory results when KPSS is used as *confirmatory analysis* for ADF results since both tests are similar in terms of power or size properties (Maddala & Kim, 1998). The implementation of unit root tests involves regressions that contain lags or contain kernel density estimates to control for autocorrelation in the errors. The choice of the optimal lag order has been selected using information criteria as described below.

2.6.1.2. Optimal lag order in VAR

Optimal lag order is required to reduce residual correlation in obtaining the proper VAR specification. In the context of VECM, the VAR lag order can influence the estimation of cointegrating relationships (Gonzalo, 1994). The typical approach estimates VAR models with different lag orders beginning with higher-order lags and then decreasing the lag order. The selected lag order can be obtained by evaluating the minimum values of statistical information criteria, which reward goodness of fit but penalize additional parameters. In this research, we apply the Schwarz Bayesian criterion, BIC, and Hannan-Quinn criterion, HQC, which lead to consistent estimates of the optimal lag order (Neusser, 2016). An alternative approach is to select the minimum number of lags where a test for no-serial correlation cannot be rejected. This approach has not been employed here.

2.6.1.3. Cointegration and VECM specification and estimation

This section defines cointegration and shows how VECM is specified as a representation of a cointegrated system. Some noted that cointegration and error correction might be seen as “twin concepts” and that VECM is able to model economic equilibrium relationships using a relatively rich time-series specification (Cottrell & Lucchetti, 2021). Next, we describe the steps to test the cointegration relationship following the Johansen (1995) approach and the estimation of the VECM parameters.

2.6.1.3.1. Definition of cointegration

Formally, the cointegration theorem of Engle & Granger (1987) posits that for a given set of variables $X_1, X_2, \dots, X_k$ integrated of order d, $I(d)$, exists a vector $\beta = [\beta_1, \beta_2, \dots, \beta_m]$ such that the linear combination $\varepsilon = \sum_{i=1}^{k} \beta_i X_{\mathrm{i}}$ is integrated of order $I(d-b)$, where $d \geq b > 0$, the variables $X_1, X_2, \dots, X_k$ are said cointegrated of order d and b or $CI(d, b)$. The vector β is called the *cointegration vector*, and the linear combination between the variables (also referred to as an r-dimensional vector of trend-stationary variables) describes a 'stable' economic relationship which is more stationary, $I(0)$, than the variables under analysis (Johansen, 1995). In this regard, we are testing a structural equation formulated by the RBET model between the levels of the variables of interest (see the empirical models specified in section 2.3). If the variables are non-stationary, we can reformulate them, for convenience, in terms of levels and differences. Thus, if a stationary relationship models a structural association, these may lead us to consider stationary relationships between levels, that is, cointegration relationships (Johansen, 1995). For the sake of clarity, if we represent the cointegration vector as coefficients in a regression model, i.e., $\beta = [1, -\beta_1 \dots, \beta_{k-1m}]$, we can represent the cointegration residual as $\varepsilon = Y - \beta_1 X_1 - \dots - \beta_{k-1} X_{k-1}$. This last representation corresponds to the standard regression model, $Y = \beta_1 X_1 + \dots + \beta_{k-1} X_{k-1} + \varepsilon$, assuming that ε is $I(d-b)$. Therefore, if the series under study are non-stationary, for example they are $I(1)$, then their cointegration order or the linear combination will be $CI(0)$ or trend-stationary. In general, the use of auto-regressive processes generates processes of class $CI(1,1)$ (Johansen, 1995).

2.6.1.3.2. VECM as a representation of a system of cointegrated variables

As introduced above, VECM is a particular VAR model when variables are cointegrated. Engle & Granger (1987) show that cointegrated variables can be represented by error correction models, ECM, as follows.

Let us consider an autocorrelation relationship or VAR of order *p*, with a deterministic part (e.g., trend) given by μ_t which is usually polynomial in time, for an n-variate process y_t:

$$y_t = \mu_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t. \tag{2.18}$$

Since $y_{t-i} \equiv y_{t-1} - (\Delta y_{t-2} + \Delta y_{t-3} + \ldots + y_{t-i+1})$, the Eq. (2.18) can re-write as:

$$\Delta y_t = \mu_t + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \tag{2.19}$$

$$\Pi = \sum_{i-1}^{p} \phi_i - I \tag{2.20}$$

$$\Gamma_i = -\sum_{j=i+1}^{p} \phi_i, \tag{2.21}$$

where Δ is the first-difference operator, and p is the optimal lag order. Π and Γ_i are coefficients matrices. The Eq. (2.19) is the VECM representation of Eq. (2.18), and its interpretation depends on the rank, r, of the matrix Π since r will indicate the number of independent cointegration vectors (Engle & Granger, 1987; Johansen, 1995). In this regard, there are three cases. Firstly, if $r = 0$, it means that the matrix

Π is compound by zeros, and the processes are all $I(1)$ and not cointegrated, which implies that the representation is just the VAR model for the process in differences, i.e., the model can be estimated using the first-difference approach. Secondly, if $r = n$, the matrix Π is invertible or of full rank. It means that the series are stationary; therefore, any combination of them will be stationary. Hence, the model can be estimated directly. Thirdly, if $0 < r < n$ there are r independent and linear relationships in the Π matrix which implies cointegration between the model variables, and the Π matrix can be expressed as $\alpha\beta'$ where α and β' are arrays of $n \times r$, and the r columns of β' are linearly independent cointegration vectors. In other words, if y_t has cointegration relationship, then $\Pi y_{t-1} \sim I(0)$ and, in this case, the Eq. (2.19) can be written as

$$\Delta y_t = \mu_t + \alpha\beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t. \tag{2.22}$$

As noted above, we can represent cointegrated variables as ECM. Then, from Eq. (2.22), we can express $\beta' y_{t-1} = ect_{t-1}$, where ect_{t-1} is the lagged error correction term. Theoretically, a cointegrated process can be represented as an ECM solution (Engle & Granger, 1987; Johansen, 1995). In this regard, the ect_{t-1} term reflects the disequilibrium generated by the long-term equilibrium relationships between variables since the deviations in an observed variable depend on the changes from the equilibrium between variables in a given relationship (Lutkepohl, 2005). We rewrite Eq. (2.22) with the ect_{t-1} term as follows:

$$\Delta y_t = \mu_t + \alpha(ect_{t-1}) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t. \tag{2.23}$$

The Eq. (2.23) is the VECM, in which each n-variate equation is an ECM.

Also, some additional constraints can be placed on the trend term μ_t. For example, there is no trend, or the deterministic trend belongs exclusively to the long-run relationship or the cointegration relationship. Therefore, there is no deterministic trend in the first difference of the variables. In this regard, five cases can represent these assumptions about the deterministic trend (Johansen, 1995). In this research, we specified case 4, “unrestricted constant + restricted trend”, which considers that the cointegration equation includes a trend, but the first difference of the series does not. The assumption of the trend being restricted to the cointegrated system comes from our empirical model specification (see section 2.3), where we include a trend parameter as a proxy for the SBTC effect according to the RBET model estimation strategy.

2.6.1.3.3. Cointegration test using the Johansen approach

The Johansen method enables us to determine the cointegration of n variables $I(1)$ (there can be more than one cointegration relationship). The approach investigates the cointegration rank r of β or the number of cointegrating vectors through maximum likelihood algorithms. There are two log-likelihood ratio tests determining r. First, the maximum eigenvalue or “λ-max” test examines the following hypotheses:

$$H_0: r = r^*$$
$$H_1: r = r^* + 1.$$

This test is applied sequentially for $r^* = 0,1,2,\dots,n-1$ and the first no rejection of the null hypothesis is considered the estimation of r. Second, the "trace" test, which examines the following hypotheses:

$$H_0: r = r^*$$
$$H_1: r = r > r^*.$$

The trace test is applied sequentially for $r^* = 0,1,2,\dots,n-1$ and the first no rejection of the null hypothesis is considered the estimation of r. If there is no evidence to reject $H_0: r = 1$ under the described tests, then a cointegration relationship has been found. Consequently, we will estimate the VECM with $r = 1$.

2.6.1.3.4. VECM estimation of our empirical models

Rewriting our base empirical model from Eq. (2.13) under the specification of the VECM system from Eq. (2.23), we specify the ECM equation on the skill premium, ω, as follows:

$$\Delta \ln \omega_t = \beta_0 + \alpha(ect_{t-1}) + \sum_{i=1}^{p-1} \rho_i \Delta \ln \omega_{t-i} + \sum_{i=1}^{p-1} \gamma_i \Delta \ln \left(\frac{S}{U}\right)_{t-i} + \varepsilon_t \tag{2.24}$$

where $\Delta \ln \omega_{t-i}$ and $\Delta \ln \left(\frac{S}{U}\right)_{t-i}$ are the differences that capture short-run variations in the skill premium, $\ln \omega$, and the relative supply of skilled labour, $\ln \left(\frac{S}{U}\right)$, respectively. ρ, γ, α and ect_{t-1} are coefficients to be estimated and ε_t is white noise. In this specification, we normalize α and ect_{t-1} on the target variable, the skill premium, where ect_{t-1} define the long-run relationship if the skill premium and the relative supply of skilled labour are cointegrated. Therefore, our base empirical model from Eq. (2.13), expressed in terms of the ect_{t-1} , considering that we have implemented case 4 of the Johansen modelling (a trend term within the cointegration equation as discussed above in section 2.6.1.3.2), yields the following cointegration equation to be estimated:

$$ect_{t-1} = \ln \omega_{t-1} - \beta_1 t + \beta_2 \ln \left(\frac{S}{U}\right)_{t-1} - \beta_3 Ch98_{t-1}. \tag{2.25}$$

Reordering Eq. (2.25) on the normalized skill premium yields

$$\ln \omega_{t-1} = \beta_1 t - \beta_2 \ln \left(\frac{S}{U}\right)_{t-1} + \beta_3 Ch98_{t-1} + ect_{t-1}. \tag{2.26}$$

Eq. (2.26) refers to our base empirical model in Eq. (2.13). We expect that β_2 coefficient in Eq. (2.26) to be negative and significant as evidence for an inverse relationship between the skill premium and the relative supply of skilled labour, as posited by the RBET empirical modelling (see section 2.2.1.3). Besides, a negative β_2 coefficient allow us to compute a positive elasticity (see Eq. (2.12) related statements). The same procedures apply to our extended empirical model from Eq. (2.14).

The estimation of the VECM proceeds in two stages: firstly, the application of the Johansen approach (Johansen, 1995) to find the cointegration rank r (if any) and, secondly, for a given r, the estimation of the VECM parameters. All procedures detailed in this section are estimated using the statistical software `Gretl` (Baiocchi & Distaso, 2003; Cottrell & Lucchetti, 2021).

2.6.2. The UCM-Bayesian

This section outlines the UCM formalization following the notation and descriptions given mainly by Pelagatti (2016) and Durbin & Koopman (2012). Then, we shall present the main Bayesian estimation features. Next, we shall show our empirical models under UCM-Bayesian specifications and detail the procedure and techniques involved in estimating them following Koop (2003) and Gelman et al. (2020), among others.

2.6.2.1. UCM formalization

The UCM, also known as structural time series models, STSM, or local level models, LLM, is the basic structure used to represent a time series. It is specified directly in terms of its components of interest (e.g., trend, seasonal and error components) plus additional relevant terms (e.g., a regressor). The main UCM feature is that the model components are modelled as stochastic processes. Thus, we can write a general representation of a UCM in its additive form as:

$$Y_t = \underset{trend}{\mu_t} + \underset{seasonal}{\gamma_t} + \underset{cycle}{\psi_t} + \underset{regressor}{\beta_t} + \underset{noise}{\varepsilon_t}\ , \qquad t = 1, \dots . n \tag{2.27}$$

where Y_t is the observed data, μ_t a slowly changing component, typically called the "trend", the seasonal component γ_t is a periodic term, ψ_t is a cyclical component, β_t is a regressor of interest, and ε_t, is the irregular model component.

Overall, this approach allows us to infer relevant properties of the trend from the observed data (Durbin & Koopman, 2012). We specified the trend as Local Linear Trend, LLT, which can be interpreted as a linear trend with intercept and slope evolve synchronized over time as random walks[12] (Pelagatti, 2016). In this regard, the LLT specification is defined by two state equations modelling the level and the slope, as follow:

$$\mu_t = \mu_{t-1} + \nu_{t-1} + \eta_t \tag{2.28}$$

$$\nu_t = \nu_{t-1} + \zeta_t, \tag{2.29}$$

where μ_t represents the stochastic level of the trend at t, and ν_t *is* the stochastic slope of the trend (or the increment of level between t and $t+1$). The terms η_t and ζ_t are independent white noise

[12] A random walk time series model such as $\{S_t,\ t = 0,1,2, \dots .\}$ is obtained by cumulatively summing (or "integrating") i.i.d. random variables. Then, a random walk with zero mean is obtained by defining $S_0 = 0$ and $S_t = Y_1 + Y_2 + \cdots + Y_t$ for $t = 1,2, \dots$. where $\{Y_t\}$ is iid noise. In other words, in the RW process the contemporaneous value of the variable is composed of its past value plus an error, $y_t = y_{t-1} + \varepsilon_t$. Thus, the change in Y is strictly random.

sequences[13]. The initial conditions for level and slope, μ_0 and ν_0, respectively, are usually unknown. The LLT specification above nests different special cases of interest by fixing the values of the level and slope variances, σ_η^2 and σ_ζ^2, respectively, or the initial slope ν_0 to zero. For example, setting $\sigma_\eta^2 = \sigma_\zeta^2 = 0$ would reduce μ_t to be constant, i.e., we obtain a deterministic linear trend or $\mu_t = \mu_0 + \nu_0 t$. Other special cases include μ_t being a random walk or $\mu_t = \mu_{t-1} + \eta_t$ ($\sigma_\zeta^2 = \nu_0 = 0$, i.e. the slope is constant and equals zero), a random walk with drift or $\mu_t = \mu_{t-1} + \nu_0 + \eta_t$ ($\sigma_\zeta^2 = 0$, i.e. the slope becomes constant, and we obtain a random walk with drift ν_0) or an integrated random walk which is a very smooth trend where $\sigma_\eta^2 = 0$.

To illustrate the genesis of the LLT specification, let us consider the following linear function:

$$\mu_t = \mu_0 + \nu_0 t \tag{2.30}$$

where u_0 and ν_0 are the intercept and the slope, respectively. The linear function in Eq. (2.30) can be represented in incremental form or as a difference equation such that

$$\mu_t = \mu_{t-1} + \nu_0 \tag{2.31}$$

Eq. (2.30) and (2.31) define the same linear function since the iteration for $t' = 1,2,\ldots,t$ can be represented as

$$\begin{gathered} \mu_1 = \mu_0 + \nu_0 \\ \mu_2 = \mu_1 + \nu_0 = \mu_0 + 2\nu_0 \\ \ldots\ldots\ldots\ldots\ldots.. \\ \mu_t = \mu_{t-1} + \nu_0 = \mu_0 + t\nu_0. \end{gathered} \tag{2.32}$$

Once we add the white noise η_t to Eq. (2.31), we obtain

$$\mu_t = \mu_{t-1} + \nu_0 + \eta_t \tag{2.33}$$

where μ_t (the level) progresses as a random walk with drift[14], which can also be represented as

$$\mu_t = \mu_0 + \nu_0 t + \textstyle\sum_{s=1}^{t} \eta_s. \tag{2.34}$$

In the representation given by Eq. (2.31), μ_t is a linear trend with a random walk intercept, but the slope is constant. However, the slope can also be added as a random walk to be time-varying. The result is the pair of equations defining the LLT, Eq. (2.28) and (2.29), which refer to a linear trend where both level or intercept and slope evolve as random walks over time. Estimating these trend components according to this approach is a proper way of measuring the relevance of the trend behaviour and its role in a time series model (Koop, 2003).

---

[13] The assumption of white noise residuals implies formally they are a sequence of uncorrelated random variables and they come from a Normal distribution with mean zero and variance $\sigma^2 < \infty$ (Brockwell & Davis, 2016)

[14] A *random walk with drift* has a nonzero constant term which is a deterministic linear trend in the mean. If the process starts at $t = 0$, it generates a time series with upward trend (Lutkepohl, 2005).

2.6.2.2. Bayesian estimation

The Bayesian estimation relies on probability models whose conditional probability distributions characterise variables and unknown parameters. Consequently, this approach allows us to evaluate the probability of our empirical modelling specified in section 2.3, in terms of parameters and their relationships, given observed data.

The core of Bayesian estimation is *Baye's Rule* or Baye's Theorem, which is stated as:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} \tag{2.35}$$

where $p(B|A)$ is the conditional probability of an event B given A and, $p(A|B)$ is A given B. $p(A)$ and $p(B)$ are the marginal probabilities of A and B, respectively. The Bayesian approach replaces events A and B above with vectors or matrices of data and parameters in the econometric estimation. To illustrate, a given vector or matrix y, which contains time-series data, can be linked to the vector or matrix of parameters, θ, which would be able to explain the data, to obtain:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \tag{2.36}$$

where the term $p(\theta|y)$ allows us to estimate the unknown parameters given the known data, which can also be referred to as the *posterior* density. The term $p(y|\theta)$ is the *likelihood function*, a probability density function for the y given the θ, and it often is described as the data generating process. The term $p(\theta)$ is the *prior* density, which summarizes the prior knowledge about the parameters before observing the data. Since the term $p(y)$ refers to the probability of producing the data and it does not refer to θ, the usual Baye's formula representation can be written as a proportionality relationship, where $p(y)$ is a normalizing constant, which makes the *posterior* a true probability distribution that integrates to 1, as follows:

$$p(\theta|y) \propto p(y|\theta)p(\theta). \tag{2.37}$$

The relationship posits in Eq. (2.37) refers to as "*posterior is proportional to likelihood times prior*" (Koop, 2003). It implies Bayes' theorem is an "update" of $p(\theta)$ since the data is updating our prior beliefs about θ; therefore, the *posterior* probability distribution will be a combination of data and non-data information. In this regard, the *posterior* distribution will be the basis for all inference of the unknown model parameters. To illustrate, assuming that we are interested in the *posterior* mean as a point estimate and suppose θ is a vector of k elements, $\theta = (\theta_1, \ldots, \theta_k)'$, we obtain the *posterior* mean of any element of θ as:

$$E[(\theta)|y] = \int \theta\, p(\theta|y)d\theta. \tag{2.38}$$

The form with an involved integral in Eq. (2.38) applies to all parameters such as the posterior standard deviation and other measures, including transforms and functions of parameters. Thus, all these probabilistic statements or functions calculations from the *posterior* have the form:

$$E[g(\theta)|y] = \int g(\theta)\, p(\theta|y) d\theta \tag{2.39}$$

where $g(\theta)$ is a *function of interest* (e.g., $g(\theta) = \theta$ and $g(\theta) = \theta^2$, for the mean and variance, respectively) (Koop, 2003). The quantities in Eq. (2.39) can often not be estimated analytically. In these cases, these measures are obtained by simulating the *posterior*.

2.6.2.3. The empirical models under UCM-Bayesian specification

This section shows how we specify our base empirical (and extended) model from Eq. (2.13) under UCM, i.e., a system compounds by an observation equation and two additional state equations modelling the level and the slope. Remarkably, the Bayesian estimation allows us to specify the parameter standing for the elasticity of substitution between skilled and unskilled labour, σ_{SU}, directly[15] in our UCM specification. Recalling our empirical models in section 2.3 and considering that the $-\beta_2$ parameter in these models supplies an estimate of $-\left(\frac{1}{\sigma_{SU}}\right)$ (see Eq. (2.12) related statements), our UCM specification for our base empirical model represented by Eq. (2.13) is:

$$\ln \omega_t = \mu_t - \frac{\ln\left(\frac{S}{U}\right)_t}{\sigma_{SU}} + \alpha Ch98_t + \gamma S_t + \varepsilon_t \tag{2.40}$$

$$\mu_t = \mu_{t-1} + \nu_{t-1} + \eta_t \tag{2.41}$$

$$\nu_t = \nu_{t-1} + \zeta_t \tag{2.42}$$

and the extended model represented by Eq. (2.14),

$$\ln \omega_t = \mu_t - \frac{\ln\left(\frac{S}{U}\right)_t}{\sigma_{SU}} + \alpha Ch98_t + \gamma S_t + \epsilon Unem_t - \delta MinW_t + \varepsilon_t \tag{2.43}$$

$$\mu_t = \mu_{t-1} + \nu_{t-1} + \eta_t \tag{2.44}$$

$$\nu_t = \nu_{t-1} + \zeta_t \tag{2.45}$$

where ω_t is the skill premium at time t, μ_t is the trend component or the level of the series at time t, and $\left(\frac{S}{U}\right)_t$ is the skilled labour supply relative to unskilled at time t. σ_{SU} is the parameter to be estimated standing for the elasticity of substitution between skilled and unskilled labour. $Ch98_t$, $Unem_t$ and $MinW_t$ are as in Eq. (2.13) and (2.14). Unlike the VECM, there is no lag controlling seasonality; therefore, we include a seasonal dummy, S, which controls for seasonality given the biannual data (1 = March; 0= June). ν_t is the slope and ε_t, η_t and ζ_t are independent white noise sequences[16]. As σ_{SU}, the parameters α, γ, ϵ, and δ also be estimated.

[15] The estimation is "direct" since traditional estimation in most of the literature arrives at this estimate from the inverse of the coefficient of the relative supply of skilled labour in the skill premium equation (see section 2.2.2). See section 2.4.2 and 2.4.3 for a review and implications.

[16] The assumption of white noise residuals implies formally they are a sequence of uncorrelated random variables and they come from a Normal distribution with mean zero and variance $\sigma^2 < \infty$ (Brockwell & Davis, 2016)

2.6.2.4. UCM-Bayesian estimation

As noted above, Bayesian estimation evaluates probability models where conditional probability distributions characterize all variables and unknown parameters. Therefore, to express our empirical UCM base model specified by Eq. (2.40), (2.41) and (2.42) as a probability model, they must be expressed in terms of observations and unknown parameters regarding the proper probability distributions. Since the model is a linear regression where the residuals are assumed to follow a Normal distribution[17], then the UCM base model can be written as the next group of equations (same for the extended model):

$$ln\ \omega_t \sim N\left(\mu_t - \frac{\ln\left(\frac{S}{U}\right)_t}{\sigma_{SU}} + \alpha Ch98_t + \gamma S_t, \sigma_\varepsilon^2\right) \tag{2.46}$$

$$\mu_t \sim N(\mu_{t-1} + \upsilon_{t-1}, \sigma_\eta^2) \tag{2.47}$$

$$\upsilon_t \sim N(\upsilon_{t-1}, \sigma_\zeta^2) \tag{2.48}$$

where σ_{SU}, α and γ are component parameters and σ_ε^2, σ_η^2 and σ_ζ^2 are the variance parameters for white noise innovations.

Recalling that from the Bayesian point of view, our parameters are seen as random variables which have associated prior probability distributions, we will update these distributions as we observe data. In this regard, we condition all the parameters specified by Eq. (2.46), (2.47) and, (2.48) belonging to $\mathbb{R}$ with some precision. For our elasticity of substitution parameter, σ_{su}, the prior distribution represents our beliefs about the possible values that the parameter can take. We incorporate current beliefs about the elasticity of substitution using a Normal distribution with a parameter sampling space in the range [0.01, 10]. For Chile, past studies reported values in the *consensus* range [1, 3] (see, e.g., Gallego, 2012) and values around 10 (see, e.g., Sánchez-Páramo & Schady, 2003) as reviewed in section 2.4.2. Also, since we use `Stan` (Stan Development Team, 2019) as software to estimate our UCM-Bayesian models (additional details below), this tool defines the Normal on the standard deviation, σ, instead the variance, σ^2. Therefore, the prior probability distribution for our elasticity of substitution parameter, σ_{su}, is defined on the standard deviation (see Eq. (2.49) below).

In the case of the white noise parameters, σ_ε^2, σ_η^2 and σ_ζ^2, the prior distributions are Cauchy and conditioned with a lower threshold of 0.01 without an upper threshold. The use of a Cauchy with centre zero (mean) and scale (standard deviation) equalling ten implies also the use of relatively noninformative prior distribution (Gelman, 2006; Gelman et al., 2008; Stan Development Team, 2019). We use the same specification for α and γ. For the state equations of μ_t, Eq. (2.47), and υ_t, Eq. (2.48),

[17] In typical linear regression model with y, X and e as dependant variable, vector of covariates and residuals, respectively, the assumption of Normal distributed residuals implies that $y = X\beta + e$ and $e \sim N(0, \sigma^2) \Longrightarrow y \sim N(X\beta, \sigma^2)$. It is given since the addition of $X\beta$ to the e mean yields a distribution with mean $X\beta$ and variance σ^2 (or standard deviation, σ). Regarding "Normal" behaviour of residuals, it implies that most error values would be around zero and fewer of them around tails.

it is suggested the use of hierarchical priors (Koop, 2003). Moreover, we specify noninformative uniform priors for μ_0 and υ_0, i.e., the prior distributions are Uniform, U, to give same probability to all the possible values since we cannot properly specify our prior knowledge of these parameters. The *priors'* distributions are (where [,] denotes the prior range of the distribution):

$$\begin{gathered}\sigma_{su} \sim N(0.1, 3)\ [0.01,10] \\ \sigma_{\varepsilon}^2 \sim Cauchy(0, 10)[0.01, \infty] \\ \sigma_{\eta}^2 \sim Cauchy(0, 10)\ [0.01, \infty] \\ \sigma_{\zeta}^2 \sim Cauchy(0, 10)\ [0.01, \infty] \\ \alpha \sim Cauchy(0, 10) \\ \gamma \sim Cauchy(0, 10) \\ \mu_0 \sim U(0,1) \\ \nu_0 \sim U(0,1)\end{gathered} \tag{2.49}$$

The Bayesian formulation of our UCM probability model specified in Eq. (2.46), (2.47) and (2.48) plus the specified priors in Eq. (2.49), consists of the *Likelihood* function $p(y|\mu, \sigma_{su}, \alpha, \gamma, \sigma_{\varepsilon}^2)$, the *Prior* distributions given to μ, σ_{su}, α, γ, and σ_{ε}^2, and the *Posterior* distribution $p(\mu, \sigma_{su}, \alpha, \gamma, \sigma_{\varepsilon}^2|y)$. As discussed above in Eq. (2.39) related statements, the *posterior* is estimated by simulation using sampling algorithms. These approaches imply that a particular parameter from the *posterior* is approximated numerically by simulating draws to evaluate the *function of interest* at the random sample (e.g., the mean, the variance).

The sampling method used in our procedure relies on Monte Carlo Markov Chains, MCMC, techniques. The MCMC sampling follows the Monte Carlo Integration theorem. Following Koop's (2003) notation, this theorem posits the subsequent statements. Let $\theta^{(s)}$ for $s = 1, \dots, S$ be a random sample from $p(\theta|y)$, and define $\hat{g}s = \frac{1}{S}\sum_{s=1}^{S} g\left(\theta^{(s)}\right)$ then $\hat{g}s$ converges to $E[g(\theta)|y]$ as S goes to ∞. Then, the sampling from the *posterior* will be the *posterior* simulation and $\theta^{(s)}$ is a draw or replication. In practical terms, an extensive sampling from any probability distribution can be used to explain the main features of the distribution of interest.

The MCMC simulation gives an opportunity for a statistical relationship between each draw and the next to build the Markov Chain. The method requires chains of great length, i.e., many iterations since initial arbitrary values must be discarded along with the performance of some additional adjustments (e.g., techniques to minimize autocorrelation)[18]. In addition, MCMC estimates several chains simultaneously to explore convergency diagnostics related to how each chain converges from initial arbitrary values towards the *posterior* target distribution.

The MCMC estimation of parameters in this study uses the Hamiltonian Monte Carlo algorithm, HMC. The HMC method is a more efficient sampler than other MCMC algorithms, such as Metropolis-Hastings or Gibbs[19] (Gelman et al., 2020). As discussed above, we use `Stan` as a computational tool

[18] These adjustments to MCMC, such as' number of iterations', 'number of chains' and 'thin', are researchers' definitions.
[19] The Gibbs sampler and Metropolis algorithm are inefficient regarding their random walk behaviour, which requires parameter adjustments and other rules to improve their efficiency (Gelman et al., 2020).

based on the probabilistic programming language to define a log density function conditioned on data to estimate our UCM-Bayesian models (Gelman et al., 2020; Stan Development Team, 2019). Specifically, we use `rStan` (Stan Development Team, 2019), the `Stan` interface for R (RStudio Team, 2020). With `rStan` we fit the UCM-Bayesian model in and generate MCMC posteriors draws for each specified parameter (e.g., the parameters stated in Eq. (2.46), Eq. (2.47) and Eq. (2.48)). In Appendix A.1.1, we show the `Stan` code that represents our base model specification. The `rStan` output computes summary statistics, estimates and diagnostic indicators such as $\hat{R}$ statistic[20] to measure if the MCMC samples have converged to the *posterior* and evaluate that the posterior draws are distributed in a stationary manner.

2.7. Results

This section aims to outline the results using the data (see section 2.5) and methods (see section 2.6) described in the last sections. Firstly, we shall present our measures for the skill premium and the relative supply of skilled labour for 1980-2018. Secondly, we shall summarize the main findings of our VECM implementation. Thirdly, we outline the results of our UCM-Bayesian strategy.

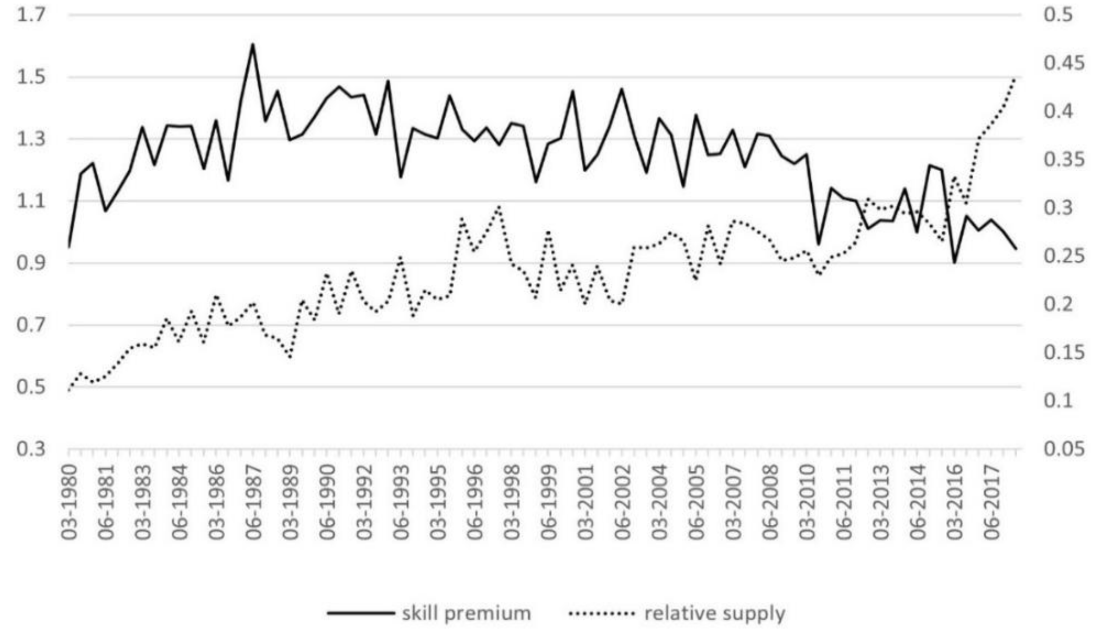
2.7.1. Estimation of the skill premium and the relative supply

This section outlines the results from estimating the skill premium and the relative supply of skilled labour following the strategies detailed in section 2.5. Figure 2.2 displays the evolution of our measures for both variables over 1980-2018. The skill premium shows an inverted U-shaped pattern, growing up to the late 1980s and then reducing after the 1990s, although with fluctuations. On average, the skill premium increased from 1.27 in the 1980s to 1.34 in the 1990s. In turn, in the 2000s, it decreased to 1.29 and, in the 2010-2018 span, to 1.06. This pattern over time, i.e., an increase followed by a decrease of the skill premium, is consistent with previous works (Gallego, 2012; Murakami, 2014; Parro & Reyes, 2017).

The relative supply of skilled labour shows an increasing pattern over the sample period, as shown in Figure 2.2 (secondary axis), with fluctuations as conspicuous as those in the skill premium but ending at a very different point. On average, this ratio grew from 0.16 in the 1980s to 0.22 and 0.24 in the 1990s and 2000s, respectively. In the span 2010-2018 it reached 0.30. These findings also are consistent with past studies (Gallego, 2012; Murakami, 2014; Murakami & Nomura, 2020; Parro & Reyes, 2017).

[20] The $\hat{R}$ diagnostic is known as the potential scale reduction factor. It compares the variation between the MCMC posterior samples or chains to the variation within the chains. It is expected $\hat{R} < 1.1$ for all parameters as indicator of convergence, i.e., if all chains converged on the same sampling region with similar behaviour, then the variance between them should be approximately equal to the average variance within chains (Gelman et al., 2020; Muth et al., 2018).

Figure 2.2. Evolution of the skill premium and the relative supply (secondary axis), 1980-2018



2.7.2. VECM Results

This section presents the results for our estimation using VECM. First, we outline the results from stationarity, lag order and cointegration testing. Then, we show our VECM estimates.

2.7.2.1. Stationarity testing

We determine the presence of unit roots and stationarity applying the ADF and KPSS tests (see section 2.6.1.1 for details), respectively, on the skill premium and the relative supply individually. Table 2.1 displays the ADF results, which show the presence of unit roots at levels for both variables. For example, the results for the skill premium at levels in without and with time trend indicate the presence of unit roots in this variable since we cannot reject the null hypothesis of unit roots at the 1% significance level.

Table 2.1. ADF test results

Variable		Case	Lags	t-critical 1%	t-statistics
Level	Log Skill Premium	with Constant	2	-3.520	-1.011
	Log Relative Supply		1	-3.519	-1.248
	Log Skill Premium	with Constant and Trend	2	-4.085	-2.511
	Log Relative Supply		1	-4.083	-2.973
First difference	Log Skill Premium	with Constant	1	-3.519	-11.209*
	Log Relative Supply		0	-3.517	-17.187*
	Log Skill Premium	with Constant and Trend	1	-4.083	-11.427*
	Log Relative Supply		0	-4.079	-17.067*

Note: Lag order selection using criterion BIC (max was 4). (*), (**) and (***) denotes a rejection of null of presence unit roots at 1%, 5% and 10% significance level, respectively.

In Table 2.2, we present the results for the KPSS test, which shows that the null hypothesis of stationarity is rejected for both variables. To illustrate, the skill premium at levels in without and with time trend shows that this variable is non-stationary since we reject the null hypothesis of stationarity at the 1% significance level. Therefore, based on our ADF and KPSS, there are unit roots in the skill premium and the relative supply, which implies that both series are non-stationary, and their order of integration is $I(1)$.

Table 2.2. KPSS test results

Variable		Case	Lags	t-critical 1%	t-statistics
Level	Log Skill Premium	No trend	2	0.731	1.229*
	Log Relative Supply		1		3.158*
	Log Skill Premium	Trend	2	0.215	0.468*
	Log Relative Supply		1		0.251*
First difference	Log Skill Premium	No trend	2	0.731	0.284
	Log Relative Supply		1		0.070
	Log Skill Premium	Trend	2	0.215	0.041
	Log Relative Supply		1		0.068

Note: Lag order selection as in the ADF test (see Table 2.1). (*), (**) and (***) denotes a rejection of the null of stationarity at 1%, 5% and 10% significance level, respectively.

2.7.2.2. Optimal lag order

Table 2.3 shows the results for optimal lag order estimation regarding the proper model specification for the VAR[21] specification (see details of this strategy in section 2.6.1.2). Our results show that the optimal number of lags to include is two, based on the minimized values of the respective information criteria.

[21] The generally accepted protocol is to include enough autoregressive lags in the VAR or VECM to neutralize the bias that would result from failure to control for the autoregression, since lags eliminate residual autocorrelation in the VAR model.

Table 2.3. Optimal lag order for the VAR using the BIC and HQC information criteria.

VAR system	lags	AIC		BIC		HQC	
With constant and trend	1	-3.324655		-3.075567		-3.225291	
	2	-3.606501	*	-3.232868	*	-3.457454	*
	3	-3.584954		-3.086778		-3.386226	
	4	-3.551588		-2.928868		-3.303177	

Note: Results estimated from VAR systems of order 1 to max. lag order 4[22]. (Log Skill premium and Log Relative Supply as endogenous variables. The results including control variables are the same) The asterisks indicate the best (that is, minimized) values for the respective information criteria.

2.7.2.3. Cointegration testing

Table 2.4 shows the results for the Johansen Cointegration tests (see section 2.6.1.3). The null hypothesis of cointegration rank = 0 is rejected for both the trace and the max eigenvalue value, at 5% significance level. However, rank=1 is not rejected; therefore, there is evidence that the two series are cointegrated.

Table 2.4. Johansen Cointegration statistical tests results.

Rank	Eigenvalue	Trace test	p-value	Lmax test	p-value
$r = 0$ (None)	0.23767	29.463	0.0151**	20.625	0.0299**
$r = 1$ (At most 1)	0.10978	8.8381	0.1957	8.8381	0.1956

Notes: Number of equations = 2; Lag order = 2; Estimation period: 2:1 - 39:2 (T = 76); Johansen approach's Case 4: Restricted trend, unrestricted constant[23]. (*), (**) and (***) denote a rejection of null ($= 0$ or $r = 1$) at 1%, 5% and 10% significance level, respectively.

2.7.2.4. VECM estimation results

The cointegration rank testing results from Table 2.4 suggest that the skill premium and the relative supply are cointegrated. This section presents the findings related to the coefficients that rule the cointegration relationship between both variables applying the VECM approach (see section 2.6.1.3). Table 2.5 displays the VECM estimation results, both cointegration vector coefficients and the VECM equation coefficients, with the skill premium as the target variable.

[22] BIC and HQC are sensible to choose maximum lag order, Therefore, this testing was also performed using 6 and 8 lags with the same results in terms of optimal lag order for all cases/models (with constant, without trend and with constant and trend).

[23] Following Johansen (1995), the modelling has included the presence of trends at level data and in the cointegrating equations. This research specifies the case of "unrestricted constant and restricted trend" or Case 4, which considers that the cointegration equation includes a trend, but the first difference of the series does not. Also, Cases 2 and 3 were analysed, for "restricted constant" and "unrestricted constant", respectively, with similar results, as follows:

	Rank	Eigenvalue	Trace test	p-value	Lmax test	p-value
Case 2	$r = 0$ (None)	0.21020	20.362	0.0468	17.934	0.0211
	$r = 1$ (at most 1)	0.031443	2.4280	0.6946	2.4280	0.6934
	Rank	Eigenvalue	Trace test	p-value	Lmax test	p-value
Case 3	$r = 0$ (None)	0.19311	16.494	0.0336	16.307	0.0214
	$r = 1$ (at most 1)	0.0024552	0.18683	0.6656	0.18683	0.6656

Table 2.5. VECM estimations results for the empirical base and extended models

Base model, Eq. (2.13) $\ln\omega_t = \beta_0 + \beta_1 t - \beta_2 \ln\left(\frac{S}{U}\right)_t + \beta_3 Ch98_t + e_t$

Extended model, Eq. (2.14) $\ln\omega_t = \beta_0 + \beta_1 t - \beta_2 \ln\left(\frac{S}{U}\right)_t + \beta_3 Ch98_t + \beta_4 Unem_t - \beta_5 MinW_t + e_t$

Estimated coefficients	Base model	Extended model
Cointegration vector		
$\ln\omega_{t-1}$ (Skill premium)	1.0000	1.0000
	(0.0000)	(0.0000)
t (Trend)	0.0095574	0.011545
	(0.071423)	(0.0026041)
$\ln(S/U)_{t-1}$(Relative supply)	-0.41398	-0.50134
	(0.16574)	(0.17140)
$Ch98_{t-1}$ (Change year 98)	-0.081560	-0.13380
	(0.0025162)	(0.076770)
$Unem_{t-1}$ (Unemployment)		0.17358
		(0.11089)
$MinW_{t-1}$ (Minimum wage)		0.030110
		(0.28368)
VECM equation with $\Delta\, ln\, \omega_t$ as target variable		
Constant	0.62033*	0.67643*
	(0.13080)	(0.13778)
$\Delta\ln\omega_{t-1}$	0.24901**	-0.28248*
	(0.09898)	(0.09501)
$\Delta\ln(S/U)_{t-1}$	-0.21019*	-0.21504*
	(0.07114)	(0.07096)
ect_{t-1}	−0.530232*	−0.501203*
	(0.111274)	(0.101650)
R^2	0.43	0.45

Note: The VECM system was estimated with cointegration rank = 1, lag order =2 and, Johansen's Case 4: restricted trend, unrestricted constant. Data are of biannual frequency (March and June) from 1980 to 2018. Standard errors are reported below the coefficients in parentheses (). (*), (**) and (***) denotes a rejection of the null hypothesis of zero coefficients at 1%, 5% and 10% significance level, respectively

From the VECM estimation procedure developed in section 2.6.1.3.4, the estimated coefficients for the cointegration vector (upper rows in Table 2.5) expressed as the ect_{t-1} for the base model (see Eq. (2.25)) yield

$$-0.5302ect_{t-1} = 1.0000\ln\omega_{t-1} - 0.0095t - 0.4139\ln\left(\frac{S}{U}\right)_{t-1} - 0.0815Ch98_{t-1} \quad (2.50)$$

reordering Eq. (2.50) on the skill premium (as shown in Eq. (2.26)),

$$\ln\omega_{t-1} = 0.0095t + 0.4139\ln\left(\frac{S}{U}\right)_{t-1} + 0.08156Ch98_{t-1} - 0.5302ect_{t-1}. \quad (2.51)$$

The result in Eq. (2.51) implies a wrong sign for the relative supply, $\left(\frac{S}{U}\right)$, coefficient, i.e., a positive sign and, consequently, an unfeasible theoretically result. Besides, following the computation of the elasticity of substitution as the reciprocal of this positive coefficient multiplied by -1 yields a negative elasticity: $-\left({}^{1}/_{0.4139}\right) = -2.42$ (see Eq. (2.12 related statements). The same unfeasible results are obtained for the extended model. Therefore, we cannot prove the expected negative relationship between the skill premium and the relative supply of skilled labour as posited in the conceptualization of the RBET model (see section 2.2.1.3), specifically, the inverse relationship between both variables discussed in the statements related to Eq. (2.8). Past studies like Murakami (2014) and Robbins (1994b) also estimated positive coefficients for the relative supply of skilled labour in some models, concluding that this variable did not contribute to the skill premium. For our own part, we will proceed cautiously and, being aware of the theoretically unfeasible results of our application of VECM and the resulting findings, we should not make assumptions under the RBET model[24].

2.7.3. UCM-Bayesian results

This section outlines the results from our UCM-Bayesian implementation detailed in section 2.6.2. Table 2.6 shows the statistics (mean, standard deviation and confidence intervals) that summarize the *posterior* distribution for all parameters given our observed data, chosen priors distributions, and assumed data generating process in our base and extended models. In particular, the 2.5% confidence interval, CI, and 97.5% CI show the bounds of the 95% central interval of the *posterior* probability distribution for a given parameter. Also, we display the $\hat{R}$ statistic results, which shows $\hat{R} < 1.1$ for all variables, implying that the MCMC samples have converged to the posterior (see footnote 20). More details on parameters convergence diagnostics (e.g., trace plot) and posterior full distribution plots are in Appendix A.1.2.

[24] We explore some different specifications of our VECM approach, such as additional lags, but the results remain unchanged.

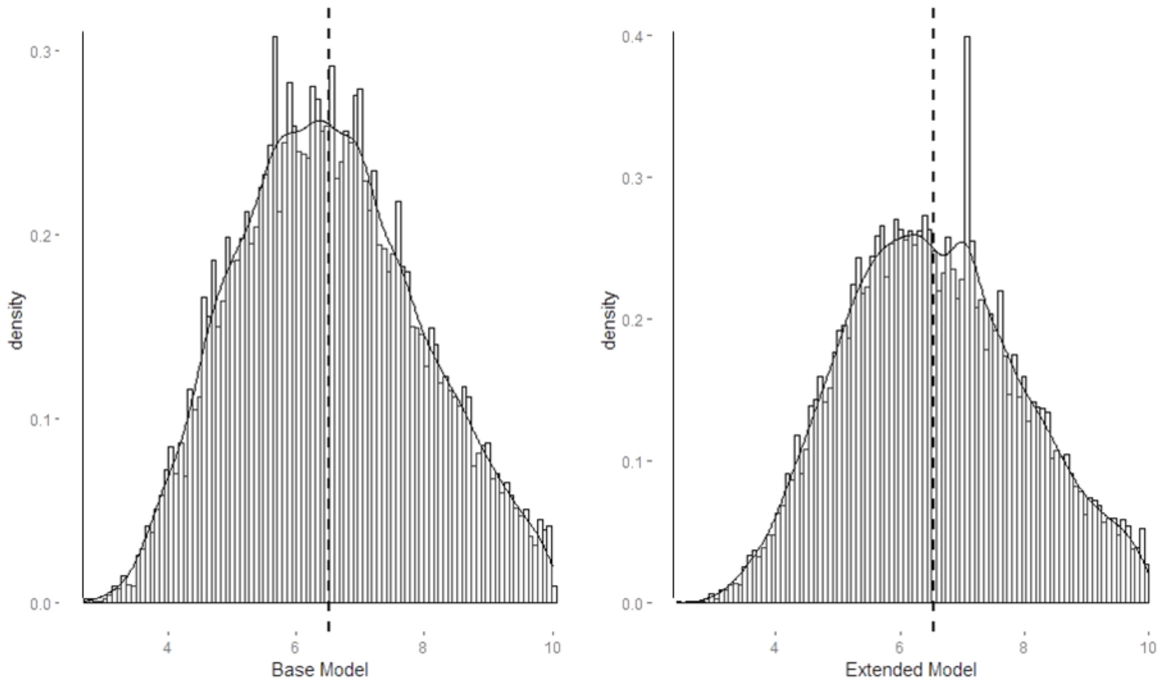
Table 2.6. Posterior summary statistics of UCM-Bayesian estimation

	Base model (see Eq. (2.46), (2.47) and (2.48). Priors in (2.49)					Extended model				
	$ln\ \omega_t \sim N\left(\mu_t - \frac{\ln\left(\frac{S}{U}\right)_t}{\sigma_{SU}} + \alpha Ch98_t + \gamma S_t, \sigma_\varepsilon^2\right)$ $\mu_t \sim N(\mu_{t-1} + \upsilon_{t-1}, \sigma_\eta^2)$ $\upsilon_t \sim N(\upsilon_{t-1}, \sigma_\zeta^2)$					$ln\ \omega_t \sim N\left(\mu_t - \frac{\ln\left(\frac{S}{U}\right)_t}{\sigma_{SU}} + \alpha Ch98_t + \gamma S_t + \delta Unem_t - \epsilon MinW_t, \sigma_\varepsilon^2\right)$ $\mu_t \sim N(\mu_{t-1} + \upsilon_{t-1}, \sigma_\eta^2)$ $\upsilon_t \sim N(\upsilon_{t-1}, \sigma_\zeta^2)$				
Parameters	mean	St dev	2.5% CI	97.5% CI	$\hat{R}$	mean	St dev	2.5% CI	97.5% CI	$\hat{R}$
Elasticity (σ_{SU})	6.51	1.42	3.97	7.50	1.00	6.54	1.42	3.98	7.52	1.00
Ch98 (α)	-2.09	5.79	-13.3	1.52	1.00	-2.53	5.78	-14.1	1.15	1.00
Seasonality(γ)	0.74	1.71	-2.63	1.89	1.00	0.65	1.69	-2.63	1.79	1.00
Unemployment (δ)						0.01	0.05	-0.10	0.04	1.01
Minimum wage (ϵ)						-0.07	0.11	-0.15	0.14	1.00
σ_η^2	1.66	0.97	0.37	2.23	1.01	1.66	1.06	0.24	2.30	1.02
σ_ζ^2	0.41	0.21	0.12	0.51	1.01	0.40	0.23	0.09	0.51	1.03
σ_ε^2	7.56	0.71	6.29	8.01	1.00	7.62	0.73	6.30	8.09	1.00
μ_0	-3.15	5.79	-15.1	0.87	1.00	-15.2	18.3	-51.4	-3.32	1.00
υ_0	2.49	1.35	0.20	3.31	1.00	2.55	1.35	0.19	3.29	1.00
μ_1	-3.17	5.48	-14.43	6.96	1.00	-15.28	18.25	-51.05	20.70	1.00
...	…	…	…	…	…	…	…	…	…	…
μ_{76}	9.24	7.47	4.35	24.69	1.00	20.58	18.11	-16.38	54.65	1.01
υ_1	2.49	1.27	1.62	5.27	1.00	2.55	1.28	0.25	5.44	1.00
...	…	…	…	…	…	…	…	…	…	…
υ_{76}	-0.46	1.18	-2.81	0.22	1.00	-0.27	1.22	-2.78	2.15	1.00

Notes: 1) To conserve space, note that in the case of the trend term level and trend term slope equations, μ and υ, we only display parameters for the initial conditions and the first (e.g., μ_1) and last estimates (e.g., μ_{76}).
2) Data are of biannual frequency (March and June) from 1980 to 2018.
3) The inference for the Stan model consisted of 12 chains, each with iter=25000; warm-up=20000; thin=5; post-warm-up draws per chain=1000, total post-warmup draws=12000). Glossary from Stan Manual (Stan Development Team, 2019): **iter** specifies the number of iterations for each chain (including warm-up), **warm-up** specifies the number of warm-up (also known as burn- in) iterations per chain to discard non-representative samples produced by early stages of sampling process, **thin** specifies the period for saving samples i.e., how often we store our post-warm-up iterations (thin=5 implies to store every fifth).

Regarding our direct estimate for the elasticity of substitution, σ_{su}, our point estimate or the mean of the *posterior* distribution is 6.51 with 95% *posterior* confidence intervals CI = [3.97, 7.50]. Similar results are obtained in our extended model. In Figure 2.3, we can visualize the *posterior* distribution of σ_{SU} for both models. The plots show that the probability mass for the elasticity of the substitution parameter is away of the bounds imposed in our parameter and *priors* modelling, suggesting that our results are not entirely driven by the constraints imposed on the parameters.

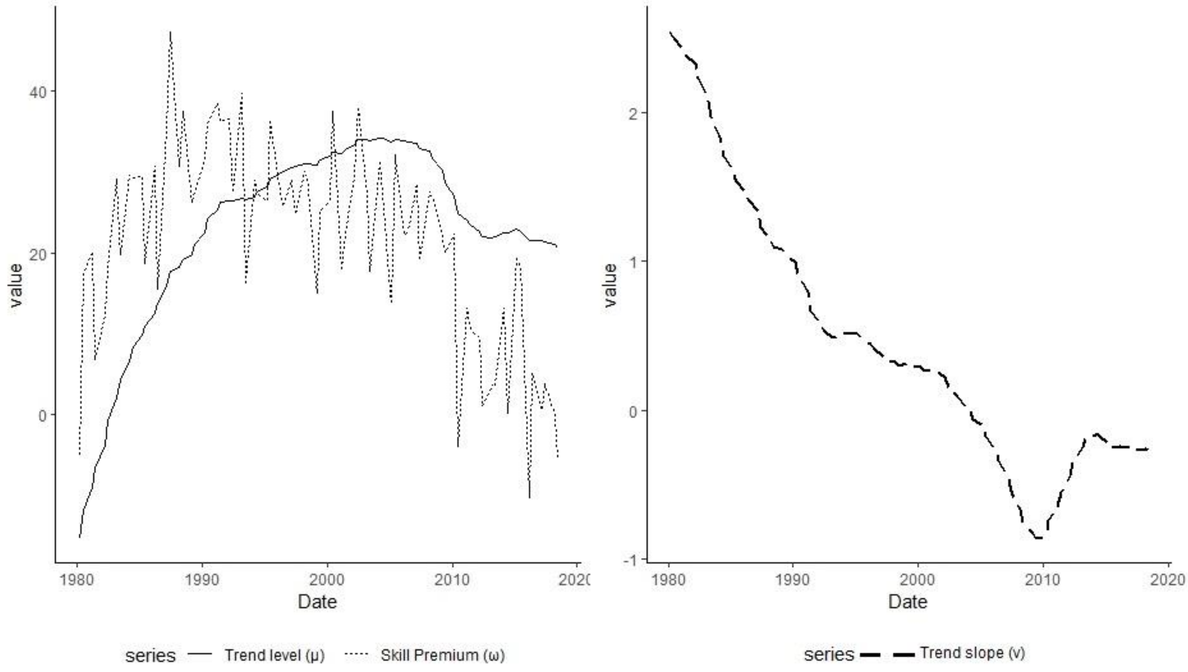
Figure 2.3. Posterior distribution of the elasticity of substitution parameter, σ_{SU}, for the base (left-side) and extended models. The dashed line shows the point estimate of the posterior mean



In terms of the RBET model conceptualization and predictions, this non-zero value of the elasticity shows that changes in the relative supply of skilled workers contributed to the evolution of the skill premium during 1980-2018 and, satisfies the inverse relationship between both variables as specified in our empirical models following the conceptual statements related to of Eq. (2.8). Furthermore, our estimated elasticity implies that both groups of workers are gross and imperfect substitutes: that is, the relative availabilities of each labour are not related to changes in wages (see section 2.2.1.1). Therefore, we reject the idea of perfect substitution between skilled and unskilled labour.

Related to our estimates for the time trend parameter μ that stand for the SBTC in the RBET model, Table 2.6 displays the results for its initial conditions $\mu_0$and υ_0 and first and last datapoints, but we evaluate the results visually, using the posterior mean for these parameters' series over time. Figure 2.4 displays the trend level (and the skill premium) and the trend slope in the left-hand and right-hand plots, respectively.

Figure 2.4. The skill premium and the trend level (left side) and the trend slope (right side)



Our estimates for the trend level in Figure 2.4 show an upward movement until the first half of the 2000s. Then, we see an enormous decline towards the beginning of the 2010s. Since this parameter captures the increases in the relative demand for skilled labour coming from technology, we can assume that the SBTC effect drove the skill premium intensively between 1980 and the first half of the 2000s. In most of the rest period, the trend displays a downward pattern suggesting a lesser importance of the SBTC. These patterns reflect the positive and negative slopes of the trend during the upward and downward periods, respectively, as seen in the right-hand plot in Figure 2.4.

Regarding our extended model, unemployment and minimum wages show results as expected. Despite a near-zero magnitude, the former is positive, showing that changes in the unemployment rate can explain the evolution of the skill premium. This result suggests that most of the unemployed are unskilled workers. Regarding minimum wages, our results show a negative impact on the skill premium evolution, which is expected given that these labour policies decrease the gap between skilled and unskilled labour' wages

2.8.Discussion

In this section, the results presented in the last section will be discussed. For clarity, the discussion has been divided into three parts. Firstly, we analyse and explain the main implications of our findings regarding the estimation of the skill premium and the relative supply of skilled labour. Secondly, we discuss the theoretically unfeasible results from VECM and how the UCM-Bayesian approach helped bring the data to the RBET Model. Thirdly, we discuss the UCM-Bayesian estimates and their implications under the RBET predictions. Also, we discuss the limitations of the RBET model and some ideas for future research.

2.8.1. The skill premium and the relative supply

Our estimation showed an inverted U-shaped pattern for the evolution of the skill premium (see

Figure 2.2). After reaching a peak of 1.34 on average in the 1990s, it decreased to 1.06 on average during 2010-2018. These results are consistent with previous studies (Beyer et al., 1999; Gallego, 2012; Murakami, 2014; Parro & Reyes, 2017). As a labour outcome that reflects the relative price of skills, these results imply both a rise and fall in demand for qualified workers during recent decades. On the one hand, Beyer et al. (1999), Gallego (2012) and Robbins (1994a) suggested that the increase in the relative demand for skilled labour in the 1980s and 1990s is related mainly to trade liberalization implemented in Chile in the pre-2000 period. One of the implications of this trade openness was the absorption of foreign technologies biased towards skilled labour, suggesting an SBTC effect leading to the increasing skill premium.

On the other hand, since Chile had already implemented these structural reforms, the significant increase in educational attainment in recent decades has been considered one of the critical forces behind the skill premium fall (Azevedo et al., 2013; Murakami & Nomura, 2020). In this regard, for countries like Chile, which recently became a high-income economy, this lower premium for skills might affect current economic status since the demand for skilled labour is an essential feature of their economic development (Gallego, 2012). Furthermore, it should be noted that the post-2000 skill premium decline may be attributed not only to improvements in the educational attainment of the workforce but also to other demand-side factors, such as changes in the demand for less-skilled workers due to structural changes or a commodities boom. In the 2000s, in most Latin American countries, this decline was partly driven by an expansion in the relative demand for less-skilled workers, mainly due to the expansion of the low-skilled intensive sector, e.g. services (Guerra-Salas, 2018). In Chile, the commodity price boom observed in the 2000s (in particular, copper) increased unskilled workers' wages (Pellandra, 2015). The role of these other factors in explaining the skill premium decline represents an opportunity for future research.

Our findings related to the relative supply of skilled labour are consistent with past studies reporting the faster growth in the Chilean tertiary educational system (Murakami & Nomura, 2020; Parro & Reyes, 2017). On average, our estimations show that this ratio increased significantly from 0.16 in the 1980s to 0.30 in the 2000-2018 period. Furthermore, the official statistics report that enrolment in tertiary education sextupled between 1984 and 2018 (INE, 2017; MINEDUC, 2020). This evolution reflects the gradual increase of skilled labour in the labour market and the exit of the older and less educated cohorts (Parro & Reyes, 2017). Additionally, beyond the endogenous response of agents to the increase in returns to education in the 1980s and 1990s, the changes in educational attainments were also fuelled by educational reforms designed to expand and diversify the Chilean tertiary educational system (Gallego, 2012; Murakami & Nomura, 2020; Parro & Reyes, 2017). Thus, the relative supply of skilled labour is suggested as a critical driver pushing the skill premium down in recent decades. In this regard, we found evidence of this relationship within the RBET model along with the effect of SBTC, as discussed below (see section 2.8.3).

2.8.2. On the applied methods

This section discusses the applied methods, highlighting the theoretically unfeasible results from VECM and arguing that the UCM-Bayesian is perhaps better suited to estimate the relationships posited by the RBET model. Recapitulating from our VECM method specification (see section 2.6.1.3), this technique has been shown to be a powerful tool to estimate cointegrated relationships (Gonzalo, 1994). However, it is generally a high dimensional model, and the cointegration approach entails imposing many auxiliary assumptions in terms of how we specify trends and lags. For example, VECM assume only linear trends within the cointegration equation and statistically significant lags are required (Von Brasch, 2016). These limitations might make cointegration techniques unsuitable for testing causal relationships (see, e.g., Guisan, 2001; Moosa, 2017). In our study, our estimated coefficient standing for the relative supply of skilled labour is positive. It does not show the theoretically negative sign that would be expected from the RBET conceptualization. The results here are not unique. Past studies also estimated positive signs (Murakami, 2014; Robbins, 1994b). For example, Murakami obtained positive coefficients in some of the models using cointegration techniques (Murakami, 2014, p. 93). More generally, researchers have warned about some limitations of cointegration approaches to testing causal relationships in Economics and Econometrics (see. E.g., Guisan, 2001; Moosa, 2017)

Researchers have also warned of several estimation and data difficulties when the skill premium evolution changes (Acosta et al., 2019). For Chile, our results show an inverted U-shaped pattern for the skill premium through time (see section 2.8.1 and

Figure 2.2). Most of the studies using data until 2000 obtained results as expected, i.e., the negative coefficient for the relative supply (see, e.g., Beyer et al., 1999; Gallego, 2012). By contrast, Murakami (2014) extended the period until 2007, where we can observe an incipient decline in the skill premium. As discussed earlier, standard cointegration might be not well equipped to model these changing patterns, particularly in the sense that they generally employ only linear trends within the cointegrating equation. Hence, we encourage future research on more flexible cointegration approaches that address the problems encountered in estimating the RBET model. However, cointegration is not a necessary condition for the validity of the RBET model. In this regard, the strategy combining UCM, and Bayesian estimation is a method that we believe should be further explored and seemed to be employed successfully for the data employed here.

Compared to other approaches, our UCM-Bayesian estimation offers us a better way to tackle two main difficulties related to cointegration strategies. Firstly, the modelling and estimation of the elasticity directly instead of as a reciprocal represent an advantage since direct methods might yield more precise estimates (Havranek et al., 2020). Secondly, this method is able to deal simultaneously with stochastic time trends (as proxies for SBTC) and the presence of unit roots in the data. This feature has been encouraged by Razzak & Timmins (2008) as a promising area for estimation methods research. Importantly, UCM neither requires cointegrated data nor makes use of stationarity assumptions. Also, the inclusion of current beliefs or *priors* about the elasticity of substitution values allows us to be explicit about the role of the *consensus* or the expected values for this parameter. This enables us to update our beliefs according to the observed data and to a specification that represents our RBET empirical modelling. In this sense, also it has been suggested whether imposing restrictions on the elasticity parameters might drive the results. However, as shown in Figure 2.3, the probability mass for the elasticity of the substitution parameter is away of the bounds imposed in our parameter and *priors* modelling. Thus, our results are not entirely driven by the imposing restrictions.

Regarding *priors,* although with more extensive data samples, the role of these parameters may become negligible, that is not the case with this kind of analysis evaluating long-run dynamics since annual or biannual data have become more available in recent decades. Remarkably, using the UCM-Bayesian strategy allows us to handle the task of understanding the skill premium dynamics under the RBET model. This task has been described by Acosta et al. (2019) as "*a difficult task, plagued by all sorts of methodological and data problems*" when we are focusing on Latin American countries due to mainly the changing pattern of the skill premium in recent decades.

2.8.3. On the evidence of the RBET model from UCM-Bayesian estimates

Our results support the empirical evidence for the RBET model for Chile. We found that demand and supply factors explain the evolution of the skill premium in Chile during 1980 – 2018. On the one hand, the time trend that captures the SBTC shows an increasing relative demand coming from technology, which shows an increasing pattern mainly in the pre-2000 period. On the other hand, we found evidence for the expected inverse relationship between our measure of the relative supply of skilled workers and the skill premium. Our estimate of the elasticity of substitution between skilled and unskilled labour is 6.5, which implies that both kinds of workers are imperfect substitutes. This result is consistent with past studies for Chile (Beyer et al., 1999; Gallego, 2012) but disagree with other studies due to "improbable estimation results" (Murakami, 2014; Robbins, 1994b).

Our result for the elasticity of substitution between skilled and unskilled labour is close to estimates from countries in the same region and to the few studies for Chile. For example, elasticities out of the *consensus* were reported for pools of Latin American countries with estimates between three and four (Acosta et al., 2019; Manacorda et al., 2010) and values above 10 for the crucial maquiladora industry in Mexico (Varella & Ibarra-Salazar, 2013). For Chile, Sánchez-Páramo & Schady (2003) estimated values around 10, although they reported them as "implausible", and Gallego (2012) estimated values between one and two. The distance between our values and Gallego's estimates might partly be explained by data granularity and the features of the period analysed, as reviewed in the literature (see section 2.4.2). For instance, it has been suggested that higher than annual data granularity (in our case, it is bi-annual) might expand the elasticity, which can be due to measurement error associated with a higher frequency of data (Havranek et al., 2020). Higher elasticities (e.g., four and above) were also linked to periods that witnessed a more rapid SBTC, but the evidence is inconclusive since values between one and two have featured in periods of slow SBTC growth (Acemoglu, 1998; Katz & Murphy, 1992). The relative demand estimated by Gallego (2012) showed an increasing pattern within the analysed period (1965 – 2000) but an elasticity in the range [1, 2]. In this regard, more research is needed to evaluate how the elasticity value responds to the analysis of sub-periods (e.g., decades).

The interpretation of larger elasticities is scarce in the RBET literature, but some implications emerge. Firstly, the possibility of switching between skilled and unskilled workers is higher. Therefore, our results suggest that skilled and unskilled labour are more substitutable than Gallego (2012) reported for Chile (elasticity of substitution between one and two). Following our conceptualization discussed in section 2.2.1.1, a higher substitution grade might imply that skilled labour might be located in less-skilled and less-productive job positions with lower wages. An additional implication from these potential movements to unskilled positions might be the lower demand for unskilled workers resulting in higher unemployment within this group. Secondly, a larger elasticity might also suggest that the impact on the skill premium for an observed relative supply time series will be more negligible than relative demand or SBTC (Katz & Murphy, 1992; Varella & Ibarra-Salazar, 2013), as discussed in our RBET conceptualization (section 2.2.1.3 and Eq. (2.7) related statements). However, from the 2000s, the SBTC effect is not enough to compensate for the strong growth in the relative supply of skills, resulting in the observed decline in the skill premium. As discussed earlier, the relative supply not only increased due to the endogenous response of agents but also was fuelled by policies promoting educational expansion (see, e.g., Murakami & Nomura, 2020; Parro & Reyes, 2017; Schneider, 2013; Valiente et al., 2020). In this regard, larger elasticities might co-occur with non-negligible impacts from the supply factor. Also, larger elasticities might imply that the market size of skilled workers drives the design and implementation of skill-biased technologies (Acemoglu, 2002). In this regard, examining the RBET model under a specification that assumes endogeneity between the skill premium and the SBTC parameter may be an attractive topic for future research.

In the context of the race between both forces, our results suggest that deciding on a given *winner* or dominant factor will depend on the analysed period. Before 2000, the dominant factor was the relative demand attributable to SBTC. Conversely, after 2000, this demand decreased, surpassed by the workforce's increases in educational attainment, mainly promoted by government policy. Thus, it seems that the relative supply of skilled labour has grown fast enough to meet the increased relative demand attributable to SBTC and thus to induce a declining trend in the skill premium as posits in our RBET conceptualization (see Eq. (2.7) and (2.8) related statements). Hence, in the post-2000 period, the new dominant factor is education. This story, with technology as an early dominant contributing to the skill premium increase, which then is moderated and reversed by the relative supply of skilled labour, coincides with Parro & Reyes (2017), who used a measure of hourly wage inequality to analyse the rise and fall in income inequality between 1990 and 2011. Thus, our study contributes to the evidence for the wage differential drivers under the RBET model.

Regarding our results on unemployment and minimum wages as factors representing labour markets conditions that can explain the skill premium, these results are as expected. The positive influence of the unemployment rate suggests that most of the unemployed are unskilled workers. Past studies also reported a positive but statistically no significant relationship between the skill premium and unemployment (Gindling & Robbins, 2001; Murakami, 2014). Thus, the small influence captured by our estimation might not have been captured with past estimation methods. With regard to minimum wage, our results show a negative impact on the skill premium evolution. Since this kind of labour policy mainly affects unskilled labour, it is expected to decrease the gap between skilled and unskilled workers' wages, a point that has been discussed earlier in the statements related to Eq. (2.14). In this sense, our result coincides with findings from Murakami (2014), who reported evidence about this inverse relationship between minimum wages and the skill premium (although this evidence comes from models yielding unfeasible results as in our VECM implementation). Other studies have reported similar findings, but they were not statistically significant (see e.g., Gallego, 2012; Gindling & Robbins, 2001). Thus, our results contribuye to the evidence on how conditions of labour markets such as those discussed here drive the skill premium.

Our findings supply some policy implications. First, investments in higher education are essential to achieve a reasonable income distribution in countries like Chile, where there is marked social inequality in the population. Furthermore, these investments have been essential for the expected transfer of knowledge and skills to jobs, resulting in a boost to Chile's economic development. (Schneider, 2013; Valiente et al., 2020). However, apparently, these investments do not consider the economy's capacity to absorb the observed greater availability of better-educated workers. This greater availability resulted from significant enrolment in tertiary education. For example, in 1984, the 18–24 age group enrolled in tertiary education grew from 11% of this age group (189,151 enrolments) to above 67% of this age group (above 1.2 million enrolments) in 2018 (INE, 2017; MINEDUC, 2020). In this sense, Chile does not have institutional mechanisms for creating relationships between firms and education suppliers: where workers' skills are concerned, and some have suggested that this lack of mechanisms has resulted in a disconnection between supply and demand, among other pervasive effects (Valiente et al., 2020). Furthermore, some suggested that the Chilean labour market compounds by a huge proportion of jobs linked to low levels of skills and technology; therefore, it might not require intensive use of skills provided by better-educated workers resulting in high rates of overqualification and over skilling (Sevilla & Farías, 2020).

Moreover, our findings show that skilled and unskilled workers are more substitutable than previously reported. This higher substitutability is also in line with the lower demand for skilled labour observed in recent decades, and it might imply that technologies are not suitable for Chilean skilled labour or that more technology-related training is required. As Gallego (2012) reported, most technologies biased toward skilled labour came from abroad. In this sense, technology might be being underexploited due to a lack of proper skills or workers in STEM[25] fields. For example, only 3% of students in tertiary education graduate with degrees in ICT, and only 1% with degrees in natural sciences, mathematics, and statistics, placing Chile in the lowest positions of all OECD countries (OECD, 2018). Therefore, policies to correct the mismatch between the supply and demand related to skills, the development of regulations such as intellectual property rights (Acemoglu, 2003), technological re-training or promoting technologies better suited to Chilean skilled labour, and the improving of graduating rates of fields of study like ICT and STEM may be required. Some strategies have been implemented recently to improve the coordination between demand and supply, such as the development of the National Qualification Framework (Fuentes et al., 2020; MINEDUC & CORFO, 2017; Sevilla & Farías, 2020) and the Job Prospection Policy Committee (Ministerio del Trabajo y Previsión Social, 2021).

Regarding some theoretical limitations of the RBET model, the absence of an upper threshold in conceptualising the three values for the elasticity, i.e., zero, one or ∞ (see Figure 2.1) difficult to interpret elasticities greater than one. As a result, we can only discuss *more* or *less* substitution between both groups of workers without theoretical support for this view. Also, the model is sensitive to ancillary (and untestable) assumptions about differences within skilled groups, complicating its use as a method for determining, for example, whether graduates and university dropouts are substitutes (Borjas et al., 2012). These limitations should be addressed in future studies focusing on theoretical aspects of the RBET model.

Finally, some important caveats to the study that deserve mention include difficulties in testing the RBET model. Although this conceptual view supplies a coherent viewpoint from which to analyse the effect of demand and supply factors on the skill premium evolution, its technical implementation requires assumptions that can lead to abandoning it. For example, researchers have preferred using alternative specifications without the use of time trends as proxies for relative demand in the case of Latin American countries (see, e.g., Acosta et al., 2019). However, the elasticity estimates are sensitive to the means by which the relative demand is specified (Borjas et al., 2012; Fernández & Messina, 2018). Regarding cointegration methods, our theoretically unfeasible results confirm some limitations

[25] STEM is abbreviation for fields of study such as Science, Technology Engineering and Mathematics.

of this approach to testing causal relationships as warned by some researchers (see. E.g., Guisan, 2001; Moosa, 2017). In this regard, the technical implementation of the RBET can be problematic. Therefore, it is convenient to give more emphasis to alternative estimation methods like UCM-Bayesian, which can handle the assumptions imposed by the RBET conceptualization.

2.9.Conclusion

The empirical testing of the RBET model has been a much-debated topic in recent decades. This conceptual idea has successfully explained the influence on the skill premium of changes in demand and supply factors. However, its implementation has challenged researchers, given the framework's assumptions and the changing nature of the data, particularly in Latin American countries. Since some researchers have abandoned or rejected the RBET model due to its problematic implementation or "estimation difficulties" such as the computation of unfeasible theoretically results leading to the computation of negative elasticities, the UCM-Bayesian approach offers a way to tackle these issues. In the case of larger elasticities that are outside the *consensus*, it is important to note that these values do not imply that the RBET model is invalid, as positive elasticities have no upper limit. While huge elasticities may raise questions about the RBET model's applicability, it is the estimation of theoretically implausible coefficients standing for the supply factor and, consequently, negative elasticities that are, therefore, incompatible with the underlying model that should deserve more attention.

This study of the Chilean labour market during 1980-2018 can help us understand the implicit race between technology and education over time. Most of the previous research analysed the period before the 2000s, which witnessed an important growth in the relative demand for skilled labour resulting in an upward pattern in the skill premium. In contrast, after 2000, this wage differential declined, and researchers testing the RBET model under this changing pattern reported theoretically unfeasible results using cointegration techniques. Our cointegration results also yielded unfeasible results, while our alternative UCM-Bayesian strategy has allowed us to estimate results consistent with the RBET model. In this regard, we gave empirical evidence for the relationships posited by the RBET model for Chile using bi-annual data from 1980 to 2018. We have shown that either the relative demand and/or the relative supply influence the skill premium evolution. Our direct estimate for the elasticity of substitution between skilled and unskilled labour is 6.5, showing that these forms of labour are imperfect substitutes. Previous research that has fallen within the *consensus* range has generally generated elasticities between one and two. Therefore, our larger estimate suggests that skilled and unskilled labour are more substitutable than commonly thought.

Our results could indirectly support some policy implications. First, the apparent mismatch between supply and demand requires the urgent implementation of corrective policies, since Chile lacks institutional mechanisms to coordinate stakeholders such as firms and education suppliers. Second, policies promoting technologies adapted to the abilities of Chilean skilled labour may be required in a context where most technical advancements come from abroad without consideration of the specific characteristics of Chile's labour force.

From the perspective of the race between technology and education over time, our findings suggest that in the 1980s and 1990s, the dominant factor was the relative demand attributable to SBTC, given its contribution to the skill premium. In this period, the growth of the relative supply of skilled labour was just starting, fuelled mainly by policies focussed on the tertiary education expansion: consequently, this factor was not capable of counterbalancing the SBTC effect. However, in the 2000s and 2010s, the vigorous educational expansion resulted in increases in the supply factor, which grew rapidly to meet the increasing demand attributable to SBTC. As a result, the provision of skills is winning the race, suggesting that this factor has been driving the skill premium decline in recent decades. This phenomenon might be a case where the lack of mechanisms for coordinating the supply of skills with the labour markets' needs has been underestimated, given Chile's inability to absorb the skilled labour in its workforce.

3. Essay II: The task-content of jobs and skills of workers as drivers of the skill premium: Evidence from online job ads for Chile 2009 – 2018

We evaluate the influence on the skill premium of the task-content of jobs and certain workers' abilities by exploiting the text data from online job posting ads covering 2009-2018 (over 189,000 ads) published by one of the main Chilean online job portals (www.trabajando.com) to capture demand for tasks and skills. Our task-related analysis tests the expected complementarity between skilled labour and non-routine cognitive (analytical and interactive) and routine cognitive tasks. In our skills-related analysis we evaluate the influence on the skill premium of cognitive and social abilities Our results show weak evidence of the influence on the skill premium of non-routine cognitive tasks. Furthermore, our findings do not show evidence of either cognitive or social abilities, or both together, explaining the skill premium evolution. Some implications arise from this apparent decrease in the importance of the tasks that skilled workers typically perform, and the abilities required of them, such as inefficient educational investment or unwanted changes in the occupational ladder.

Keywords: task-content, cognitive skills, social skills, technological change, skilled labour
JEL Classification: I26, J23, J24, J31, O15, O33

3.1.Introduction

The relationship between technological progress and labour markets has been the subject of a long-standing debate, as technology is a significant force driving employment and earnings (Autor et al., 1998). As a result, some see improvements in wages and labour productivity (Acemoglu, 2002; Acemoglu & Autor, 2011; Van Ark & O'Mahony, 2016), and others perceive *technological unemployment*, similar to past technology waves (Acemoglu & Restrepo, 2019; Elsafty & Elzeftawy, 2021; Frey & Osborne, 2017). In recent decades, computer-based technologies (e.g. Information and Communications Technologies (ICT), automation, and robotics) have further fuelled this controversy since it has been suggested that much of the technological change in production is driven by these advancements (Acemoglu & Autor, 2011; Almeida et al., 2020). For example, it has been suggested that ICT has contributed to the decline of the labour's share of GDP in recent decades (O'Mahony et al., 2021). Thus, this debate has motivated a vast literature examining the interactions between technology and labour markets where labour outputs such as the skill premium, i.e., the gap between skilled and unskilled labour wages,[26] have attracted attention. The skill premium is particularly important as a measure showing how the relative prices of skills evolve (Acemoglu & Autor, 2011), and it has been suggested that the increase in demand for more educated workers due to technological advancements is one of the main determinants that exacerbate this gap (Acemoglu & Autor, 2011; Goldin & Katz, 2008).

In Chile, as in most Latin American countries, the skill premium has been considered the main force driving the observed rise and fall of income inequality in recent decades (Acosta et al., 2019; Guerra-Salas, 2018; Parro & Reyes, 2017). There is a consensus about the inverted U-shaped pattern shown by the skill premium evolution during the last five decades. It grew considerably since the mid-1970s, peaked in the 1980s, then held steady over the 1990s and the first half of the 2000s, and it has been declining since the second half of the 2000s (Gallego, 2012; Murakami, 2014; Murakami & Nomura, 2020; Parro & Reyes, 2017). In our previous essay (see Chapter 2), we not only give evidence of this pattern but also suggest that the decreasing pattern of the skill premium continues through the 2010s.

The evolution of the skill premium provides opportunities to examine how economic forces (in particular, technological change) may influence the demand for highly qualified workers. This demand for skilled labour is an essential feature of economic development (Gallego, 2012). We have analysed and discussed the role of supply and demand factors such as education and technology, respectively, on the skill premium evolution in recent decades in Chile (see Chapter 2). However, the task-content of jobs and workers' skills endowments, beyond their formal qualifications, have recently attracted attention, and these aspects of the situation are also relevant to research attempting to explain and

[26] According to the conventional distinction based on the educational attainment of workers, skilled and unskilled individuals are those workers with post-secondary education and secondary or less schooling, respectively. This distinction facilitates international comparison. Another differentiation is "blue-collar" versus "white-collar", a functional distinction based on work tasks performed by both groups.

understand the dynamics between labour and technology (Acemoglu & Autor, 2011; Ehrenberg & Smith, 2018; Markowitsch & Plaimauer, 2009). Thus, in this essay, we evaluate the influence on the skill premium of tasks and specific workers' abilities or skills,[27] focusing on skilled labour.

We examine how measures representing work activities being performed primarily by most-educated or skilled workers such as cognitive tasks (e.g., reasoning, problem-solving, persuasion) drive the skill premium. Also, we evaluate how some aptitudes associated with skilled labour, such as cognitive and social abilities, explain the ratio of the skilled-unskilled wage. Beblavý et al. (2016) noted that analysing both aspects of jobs and workers is complementary. Thus, we study the technological change impact on the skill premium by evaluating the task composition of labour occupations under Autor, Levy and Murnane (2003) model or the ALM model. As discussed above, it has been suggested that much of the technological change in production is driven by computer-based technologies, and the ALM model enables us to evaluate the differentiated impact of these technologies on different kinds of labour. The analysis of skills and their influence on the skill premium relies on human capital theory, i.e., the complementarity between skilled labour and cognitive skills, and recent research on the increasing importance of social abilities in explaining the demand for most-educated labour (see, e.g., D. Deming, 2017).

The ALM model posits that technical progress is *biased* towards routine or codifiable work tasks. Formally, it expresses the production function in terms of tasks that different skills or machines can perform based on their comparative advantage. In this regard, the model allows an explicit distinction between tasks and skills. Acemoglu & Autor (2011) define a *task* as "a unit of work activity that produces the output", and *skill* is a "worker's endowments of capabilities for performing various tasks". Thus, the ALM model accounts for the interactions among skills, labour, task-content of jobs and technologies. These interactions are examined following a two-fold classification: *routine* with its opposite *non-routine* and *manual* versus *cognitive*, which divides into *analytical* and *interactive* (see Table 3.1 for details). This distinction tries to separate tasks that can be potentially programmable based on their degree of routineness. The model assumes that routine tasks can be expressed as programmable rules or as codifiable; in that case, they could be executed by computer-related technologies.

Examining the impact of technology under the ALM model allows us to understand how technology affects different kinds of labour. These labour groups perform occupations according to their skills level, and the ALM model enables us to evaluate the task composition of occupations. To illustrate, middle-skilled labour (see footnote 1) usually performs jobs rich in routine tasks, both manual and cognitive (e.g., clerical workers, assemblers). Alternatively, non-routine tasks cannot be easily codifiable, and under the manual/cognitive classification, we can identify the labour groups located in the poles of the skills distribution. On the one side, *manual* considers tasks whose demands include situational flexibility, visual and language appreciation, and in-person interactions. We typically

[27] In this research we use "skills" and "abilities" interchangeably.

observe less skilled or unskilled occupations involving non-routine manual tasks (e.g., food preparation and serving, cleaning, security services). On the other hand, *analytical* and *interactive* refer to cognitive tasks involving problem-solving aptitudes, intuition, and a capacity for persuasion, among other skills. These cognitive tasks mainly feature in skilled occupations (e.g., professionals, managers, associate professionals or technicians) requiring workers with specific knowledge or abilities provided by tertiary education (Autor, 2015). Details about how we can classify work activities according to the ALM model tasks groups are given in our conceptual framework (see section 3.2.1, Table 3.2).

The ALM model predicts, under the assumptions discussed above, the differentiated impact of computer-based technologies on different kinds of labour. We can describe three main potential interactions of this impact. First, technologies can become substitutes for human workers doing jobs with an intensive demand for routine tasks, both cognitive and manual. Secondly, technologies can complement non-routine cognitive activities, both analytical and interactive. Thirdly, in the case of non-routine manual tasks, technologies might have a limited role as substitutes or complements for labour. Thus, technologies' impact will depend on the task composition of labour occupations. Implications include the possibility that computer-based technologies might be detrimental to labour outputs (e.g., demand and, consequently, wages) of jobs with an intensive demand for routine tasks. On the other hand, skilled labour jobs abundant in non-routine analytical and interactive tasks might benefit from higher demand and productivity, resulting in skill premium improvements.

Some have empirically tested the differentiated technological impact predicted by the ALM model. Mainly, there is evidence of the complementarity between skilled workers and jobs with an intensive demand for non-routine cognitive tasks, both analytical and interactive (Autor et al., 2003; Goos & Manning, 2007; Goos et al., 2014; Sebastian, 2018; Spitz-Oener, 2006). Also, researchers focused on the pervasive effect of the differentiated impact of technology on different kinds of labour known as *job polarisation* (see, e.g., Autor, 2010; Nchor & Rozmahel, 2020; Spitz-Oener, 2006). As pointed out above, middle-skilled workers are usually distributed across jobs abundant in routine tasks. Under the assumption that these tasks can be programmable, the demand for this labour decreases with technology adoption. Conversely, non-routine tasks grow due to their complementarity with technology. Given that non-routine intensive jobs, cognitive and manual, are usually performed by skilled and unskilled workers, respectively, the polarisation is due to the growth at opposite ends of the skills distribution. For Chile, research analysing recent decades discards the job polarization hypothesis. Notably, an implication is that there is no increase in the relative demand for jobs consisting of non-routine cognitive tasks (Zapata-Román, 2021). In this regard, more evidence is required to understand these potential contradictions to the ALM model for the Chilean case.

In Chile, studies employing the ALM model are recent, and the evidence contradicts its main predictions. For example, Almeida et al. (2020) analysed the impact of complex software as proxies for computer-based technologies in Chilean firms between 2007 and 2013, finding that these technologies encouraged routine and manual tasks. Simultaneously, the software provided a substitute for analytical tasks, resulting in a displacement of skilled labour to less-skilled positions. The results of Almeida et al. (2020) are in line with recent studies showing a broader class of jobs at risk due to the potential displacement role of frontier technologies (Arntz et al., 2016; Frey & Osborne, 2017), such as robotics and artificial intelligence which can automate non-routine analytical or interactive tasks (Autor, 2015). Also, the results of Almeida et al. (2020) might represent the view that skill bias of technologies may be seen in the short run adoption since firms can replace highly educated labour with lesser educated individuals once new technologies are completely incorporated into the production process (O'Mahony et al., 2008). Also, it has been suggested that the routine content of jobs plays an important role in earnings (Zapata-Román, 2021). This finding, based on four waves of Chilean household data from 1992 to 2017, shows that the technological change in production, which is assumed abundant in computer-based technologies, would encourage jobs abundant in routine tasks: this contradicts ALM model predictions.

The lack of evidence to support the ALM model in the case of Chile contrasts with studies supporting this model in other high-income countries (Autor et al., 2003; Goos & Manning, 2007; Goos et al., 2014; Sebastian, 2018; Spitz-Oener, 2006). In this sense, an open question is whether these contradictions or lack of evidence persist under new analyses such as this essay. Also, another open question based on past Chilean findings, as discussed above, is whether the expected complementarity between non-routine cognitive tasks and skilled labour can also be extended to routine cognitive tasks. Responding to these questions led to some policy issues. For instance, the potential reallocation of skilled workers to routine-intense positions, or *downward movements*, might imply that skilled labour push middle-skilled to lower-skilled positions. In turn, these low-skilled workers might be pushed even further down the occupational ladder or forced to leave the labour market altogether. Some have termed this phenomenon a *de-skilling process* (Beaudry et al., 2016) when referring to employment structure changes. From a policy perspective, more attention needs to be paid to these changes in the context of the technological change in production since the unwanted changes in the occupational ladder, such as *downward movements*, can affect wages, resulting in the deterioration of educational and job prospects.

This essay examines the influence on the skill premium of measures standing for non-routine analytical and non-routine interactive tasks. Since occupations are bundles of tasks, we rely on standard occupational classifications to construct our task-content measures. In this regard, we expect that most of our sample standing for skilled labour correspond to professional, technical, and managerial occupations, given that these positions demand and employ labour with higher education and cognitive capability. We call the examination of the influence on the skill premium of our task-content measures "the task-content analysis" (section 3.5.1.2 gives details about the task measures construction).

Recapitulating, we also aim to evaluate the influence on the skill premium of specific workers' abilities or skills, focusing on skilled labour to complement our analysis. The human capital theory states that skilled labour poses specific competencies beyond formal qualifications or knowledge (Heckman et al., 2006). In particular, cognitive skills (also called abstract or analytical skills) such as learning, problem-solving, handling information, and intuition, complement the tasks performed by more-educated workers, resulting in increases in their relative demand (Acemoglu & Autor, 2011; Autor et al., 2003; Beblavý et al., 2016; Borghans et al., 2014). However, recent studies have warned about a decline in demand for cognitive skills and tasks related to new technologies implementation (Beaudry et al., 2016; Deming & Noray, 2020; O'Mahony et al., 2021; vom Lehn, 2018). At the same time, some have noted a rising complementarity between cognitive and social skills (e.g., communication, cooperation with others) with a positive impact on wages (Deming, 2017; Edin et al., 2017).

Social skills become critical in handling the resulting organisational change due to intense computer-based technological progress (Deming & Kahn, 2018). For instance, social skills in conjunction with cognitive ones can lead to a more conspicuous ability to implement new human resources practices (Bartel et al., 2007), to manage complex organisations (Edin et al., 2017), to promote innovation (Allen et al., 2020), and to obtain leadership positions (Deming, 2017), among others. Therefore, the joint demand for both cognitive and social skills might result in better labour outcomes for skilled labour. In this study, we examine the extent to which cognitive abilities, social abilities, and the simultaneous demand for both together influence the skill premium. To the writer's knowledge, the complementarity between cognitive and social skills in the case of Chile has not been studied. Also, we evaluate the influence of software skills as a proxy for specific computer-based abilities. We name the examination of the influence on the skill premium of these skills-related measures as "the skill-related analysis" (section 3.5.1.3 gives details about the measures' construction).

This essay examines the influence on the skill premium of the task-content of jobs and certain workers' abilities. Our task-related analysis tests some of the predictions suggested by the ALM model, beginning with the expected complementarity between skilled labour and non-routine analytical and interactive tasks. Secondly, we study the relationship between the skill premium and routine cognitive tasks since recent research for Chile supports this interaction, as introduced and discussed above. Therefore, we expect a positive influence not only on the skill premium of non-routine analytical and interactive tasks but also on that of routine cognitive ones. Regarding our skills-related analysis, we anticipate a positive influence on the skill premium of cognitive and social skills. Consequently, for the simultaneous demand for cognitive and social skills, we also expect a positive impact. We also expect a positive influence on the skill premium where software skills are concerned, assuming that this ability is a desirable quality in skilled workers since these kinds of technology would drive much of the current technical change.

However, despite these expectations, we will proceed cautiously because of recent trends in the skill premium evolution and past research on Chile. As introduced above, the skill premium shows a declining trend from the 2000s. In the first essay (see Chapter 2), we suggest that this decrease has persisted through the 2010s, and our task-related and skills-related analysis focus on most of this decade. Moreover, in Chapter 2, we have shown that the skill premium has declined due to the substantial expansion of Chilean tertiary education resulting in greater availability of skilled labour. Remarkably, our findings on the elasticity of substitution support the idea that skilled and unskilled labour are more substitutable than commonly assumed. This view coincides with suggested upward and downward movements in the Chilean occupational ladder during the post-2000 period. On the one hand, Zapata-Román (2021) suggested the displacement of workers from less-skilled occupations (e.g., craft and trade-related occupations) towards professional or technicians occupations. These upward movements might imply that workers in skilled positions are not performing non-routine cognitive tasks since it is assumed that less-skilled workers do not have the ability to perform that kind of task. However, they can perform routine-intensive tasks. On the other hand, Almeida et al. (2020) reported a reallocation of skilled workers toward middle-skilled positions. These downward movements suggest that the importance of non-routine cognitive tasks or skills is declining. Consequently, our results might not support our expectations.

We use online job posting ads as the data source. Our sample covers all job ads posted between 2009 and 2018 by the Chilean online job portal www.trabajando.com (over 189,000 ads). Previous studies examined wage determinants (e.g. skills) and job search behaviour, among other topics, by using this dataset (Banfi et al., 2019, 2020; Banfi & Villena-Roldán, 2019; Ramos et al., 2013). We chiefly rely on analysing the open text data (the job title, job description, and job requirements), wages and educational data.

We apply a set of analytical techniques to process and classify our data based on its text features, offered wages and educational level required alongside the use of standard occupational classifications and skills dictionaries. Thus, we build a monthly time series standing for the skill premium, the skilled labour and our task-related and skills-related metrics. We examine the influence on the skill premium of our task and skills metrics using a Vector Autoregressive framework, VAR. We follow the conventional steps to test our proposed empirical models and interpret our results by examining the Granger-Causality and Impulse Response Functions, IRF, results.

Our results show weak evidence of the influence on the skill premium of non-routine cognitive tasks for Chile during 2009-2018. Therefore, our results do not strongly support the ALM model prediction on the complementarity between non-routine cognitive tasks, both analytical and interactive, and skilled labour. Also, our findings do not show evidence of cognitive, social abilities or both explaining the skill premium evolution. As in other recent research on Chile, we suggest a potential *de-skilling process* of the Chilean labour market during the 2010s. This process aligns with the skill premium decline in this recent decade. Also, we speculate on the implications of *downward movements* in the occupational ladder. Consequently, our results supply some policy issues, given the suggested decline in the importance of tasks or abilities that skilled workers typically perform, such as efforts stimulating coordination between educational suppliers and labour markets' needs.

The essay is structured as follows. The following section presents our conceptual framework. We present the ALM model in depth and give insights regarding the theory supporting the importance of cognitive and social skills as part of skilled labour endowments. After this review, we develop the empirical models for our task-related and skills-related examinations, followed by the description of the data and the methods. Next, we present and discuss our results. The final section provides our conclusion.

3.2. Conceptual framework

This section focuses on the ALM model as a motivating theory. Also, we discuss the theoretical foundations supporting the complementarity between cognitive and social skills.

3.2.1. The ALM model

Table 3.1 replicates the classification of tasks, predictions and examples proposed by the ALM model. The two-fold classification generates five categories: 1) *routine cognitive*, which involves activities regarding the processing of information defined by explicit rules which can easily be programmable; 2) *non-routine analytic* and 3) *non-routine interactive*, both under the *cognitive* category, capture labour tasks involving reasoning skills and interactive abilities (e.g., communication and managerial skills), respectively; 4) *routine manual* and 5) *non-routine manual* refer to repetitive and non-repetitive physical work activities.

Table 3.1. Tasks classification, examples, and computer impact predictions (Autor et al., 2003)

	Routine tasks	**Non-routine tasks**
	Cognitive tasks (Analytic and Interactive)	
Examples	• Record-keeping • Calculation • Repetitive customer service (e.g., bank teller)	• Forming/testing hypotheses • Medical diagnosis • Legal writing • Persuading/selling • Managing others
Computer impact	○ Substantial substitution	○ Strong complementarities
	Manual tasks	
Examples	• Picking or sorting • Repetitive assembly	• Janitorial services • Truck driving
Computer impact	○ Substantial substitution	○ Limited opportunities for substitution or complementarity

The ALM model predicts a differentiated impact of computer-based technologies (see "computer impact" row in Table 3.1). On the one hand, computer-based technologies can perform jobs with an intensive demand for routine tasks, both cognitive and manual. On the other, new technologies complement non-routine cognitive activities, while non-routine manual tasks provide limited substitution or complementarity opportunities. Therefore, the impact of technological advancements depends on the task composition of occupations, with computer-based technologies biased towards jobs where non-routine tasks are abundant. Conversely, these technologies can substitute for human workers in routine-intensive jobs (manual or cognitive).

For the sake of clarity, Table 3.2 shows how we can allocate work activities to the five task categories discussed above, according to the literature (see, e.g., Atalay et al., 2018; Autor et al., 2003; Dengler et al., 2014; Mihaylov & Tijdens, 2019; Spitz-Oener, 2006). The first column presents the task category and a general definition. In the second column, we see the most representatives work activities featuring each kind of task. This dictionary matching tasks with work activities allow researchers to examine the task content of jobs.

Table 3.2. Examples of the assignment of work activities to task categories

Classification and definition	Job Tasks examples
Non-routine cognitive analytic: non-repetitive work activities involving quantitative reasoning, critical thinking and problem solving	Researching, analysing, evaluating and planning, making plans/constructions, designing, sketching, working out rules/prescriptions, using and interpreting rules, examining patients, using advanced software, drawing up agreements, among others
Non-routine cognitive interactive: non-repetitive work activities involving creativity and complex communication	Negotiating, lobbying, coordinating, organising, teaching or training, selling, buying, advising customers, advertising, entertaining or presenting, employing or managing personnel, pleading in courts of law, interviewing, among others
Routine cognitive: repetitive work activities regarding the processing of information	Calculating, book-keeping, correcting texts/data, and measuring length/weight/temperature, operating systems and networks, operating laboratory and office computer equipment, inspection and quality control, reading and processing information, among others
Routine manual: repetitive and physical work activities. (ICT and machines can automate them)	Operating, controlling or monitoring stationary machines and equipping machines (e.g., metal processing, chemical, rubber), making standardised products (e.g., clothes), assembling prefabricated parts or components, sorting and storing produce, among others
Non-routine manual	Repairing or renovating houses, apartments, machines, vehicles, restoring art and monuments, serving, or accommodating, operating non-stationary and mobile equipment (e.g., cranes), driving, guarding, protecting, sports (e.g., training), among others

Source: Adapted from past studies (e.g., Atalay et al., 2018; Autor et al., 2003; Dengler et al., 2014; Mihaylov & Tijdens, 2019; Spitz-Oener, 2006).

We use the ALM model as described in this section to examine the interactions between the skill premium and the task content of jobs performed by skilled labour. In the following section, we review the motivating theories regarding our skill-related analysis.

3.2.2. On cognitive and social skills and their complementarity

As introduced above, the human capital theory states that skilled labour offers specific skills and competencies such as abstract or analytical skills (e.g., learning, problem-solving, handling information, intuition). These abilities complement the tasks performed by these better-educated workers, resulting in increases in their relative demand (Acemoglu & Autor, 2011; Autor et al., 2003; Beblavý et al., 2016; Borghans et al., 2014). However, recent studies have warned of a decline in the demand for cognitive skills and tasks, related to the implementation of new technologies (Beaudry et al., 2016; Deming & Noray, 2020). At the same time, some have noted a rising complementarity between cognitive and social skills (e.g. communication, cooperation with others) with a positive impact on wages (Deming, 2017; Edin et al., 2017)

The decline in the demand for cognitive skills and tasks has been linked to stages of technology implementation (e.g., adoption versus consolidation) and to the current technological advances moving to a broader set of tasks. With reference to the effect on this decline of technology adoption stages, Beaudry et al. (2016) observe that these adoption stages increase the demand for workers with bundles of cognitive skills. Conversely, fewer of these workers are required in technology consolidation stages (e.g., maintenance activities), or they can be replaced by lesser-skilled workers once new technologies have been implemented by firms (O'Mahony et al., 2008). For the US, evidence supports the view that technology produced a boom and bust in demand for workers able to perform cognitive tasks around the 2000s (Beaudry et al., 2016). Remarkably, Deming & Noray (2020) show that technological change is an essential contributor to the erosion of skills needed to operate advancements located in the technological frontier (e.g. software skills). In consequence, skilled workers in STEM careers have witnessed the rapid obsolescence of their high analytical cognitive skills.

Recent research has examined the decreasing demand for cognitive skills in Chile. For example, Almeida et al. (2020) reported that a firm's adoption of complex software led to the relocation of workers with high attainment of cognitive skills towards less cognitive-intensive positions within that firm. This displacement suggests a decreasing importance of cognitive skills as a contributor to the relative demand for skilled workers. In this regard, it has been suggested that social skills can complement cognitive abilities, for example, by promoting innovative work practices or organisational changes occurring after the implementation of technological advancements. Therefore, both types of skills might be demanded to handle the current technical change, which implies some level of complementarity between them.

Deming (2017), Edin et al. (2017), Borghans et al. (2014), and Weinberger (2014), among others, show evidence of the rising complementarity between cognitive and social skills. According to the ALM model predictions, it is uncontroversial to admit that social interactions are difficult to automate or replace with technology (Autor, 2015; Deming, 2017). Alternatively, social skills can complement the innovative work practices or organisational changes occurring after implementing computer-based technologies, contributing to improvements in demand for both cognitive and social skills. For instance, Ehrenberg & Smith (2018) show that cognitive-intense jobs without much social interaction (e.g. engineering-related occupations) have lower employment growth in comparison with occupations rich in both cognitive and social skills. Bartel et al. (2007) noted that the adoption of ICT coincides with increases in technical skills and the ability to manage the new human resources practices resulting from the adoption of new technologies.

At the occupations level, Deming (2017) reported that non-cognitive skills had a positive effect because of their contribution to attaining leadership positions over time. Furthermore, social skills are required to manage complex organisations (Edin et al., 2017), and significant returns have been reported for socioemotional skills related to innovation (Allen et al., 2020). For Chile, Aedo et al. (2013) reported the rising importance of interpersonal skills in their cross-country study on skill intensities. Therefore, the demand for highly skilled workers includes abilities and endowments beyond the traditional cognitive aptitudes, with social skills gaining importance over time.

3.3. Empirical models

This section presents our empirical modelling to test the influence on the skill premium of our measures representing the task content of jobs (our task-content analysis) and workers' skills (our skill-related analysis)[28]. Since we do have not an explicit theoretical model, we estimate one equation for the task-content analysis and another for the skill-related examination, focusing on skilled labour as suggested by motivating theories developed in the conceptual framework (see the previous section, 3.2).

In the task-content analysis, we examine the influence of non-routine cognitive, non-routine interactive and routine cognitive tasks. Our specification for the task-content analysis is:

$$\omega_t = \beta_0 + \beta_1 TM_{NRA,t} + \beta_2 TM_{NRI,t} + \beta_3 TM_{RC,t} + \varepsilon_t, \tag{3.1}$$

where ω_t is the skill premium at time (month) t. In Eq. (3.1), TM_{NRA}, TM_{NRI} and, TM_{RC} are measures of task-content related to non-routine analytical, non-routine interactive and routine cognitive tasks, respectively. ε_t is a residual term. Section 3.5.1.1 and 3.5.1.2 describes the construction of the skill premium and task-content measures, respectively.

To evaluate the impact of cognitive skills, social skills, the combination between cognitive and social skills, and software skills categories on the skill premium, we specified the following equation

$$\omega_t = \gamma_0 + \gamma_1 SM_{Cogni,t} + \gamma_2 SM_{Social,t} + \gamma_3 SM_{CogniSoc,t} + \gamma_4 SM_{Soft,t} + \varepsilon_t, \tag{3.2}$$

[28] In the empirical modelling and estimation of models with the skill premium as the output variable, some models include exogenous variables, such as institutional controls (e.g., the unemployment rate and minimum wages). For example, see the first essay of this thesis, section 2.3, and past studies (see, e.g., Gallego, 2012; Murakami, 2014). We do not include them in our specifications, due to differences in frequency and availability. For example, in the case of unemployment, this variable has a quarterly basis or quarter moving average and we cannot construct it from job posting data. Similarly, minimum wages have a low frequency (mostly yearly) compared with our monthly data.

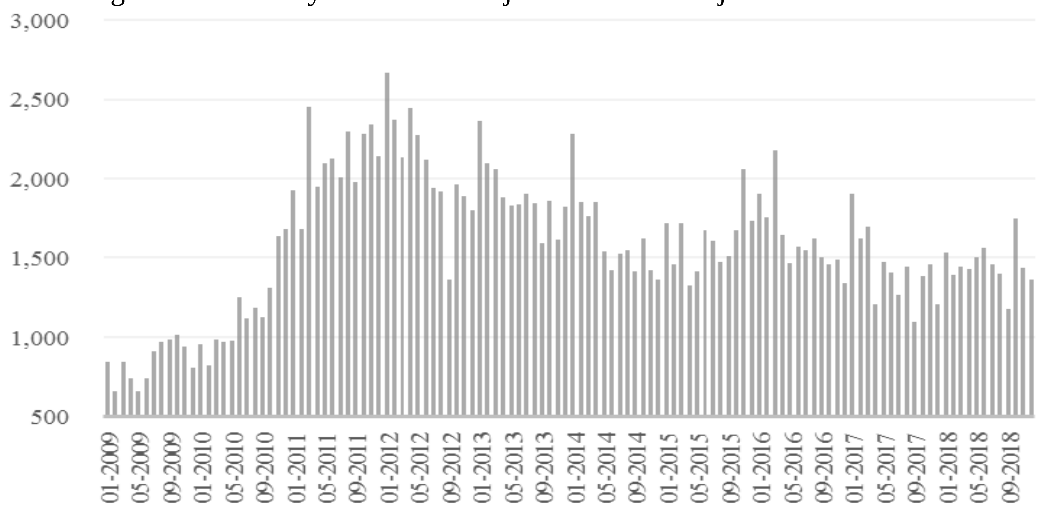
where ω_t is as in Eq. (3.2) and the variables SM_{Cogni}, SM_{Social}, $SM_{CogniSoc}$ and, SM_{Soft} are skill-measures representing the skills-related content of jobs for cognitive, social, cognitive and social, and software skills (as a proxy for computer-based technologies), respectively, required in job ads. ε_t is a residual term. We explain the construction of these skills measures in section 3.5.1.3.

Furthermore, Eq. (3.1) and (3.2) refer to static representations, but we analyse them under the VAR framework. In this regard, both equations will include lags of the variables following our estimation strategy as detailed in section 3.5.2.

3.4.Data

Our data cover all job ads posted between January 1, 2009, and December 31, 2018, for the Chilean online job portal www.trabajando.com.[29] After some cleaning[30], our sample consists of 189,986 unique job ads, some of them with missing fields. In this study, we analyse the job ads grouped by month i.e., we construct and examine 120 data points. Our motivation to use time series rely on the evaluation variables such as the skill premium over time which has shown a changing pattern over time (see Essay I, Chapter 2). The monthly disaggregation has been selected for proper variation in data comparing to quarterly or annual data. We show this monthly distribution in Figure 3.1. The average of job ads by month is 1,583 (standard deviation of 433), and the minimum and maximum frequencies are 657 and 2,670, respectively.

Figure 3.1. Monthly distribution of job ads from trabajando.com 2009-2018



[29] Although our dataset include data for 2008, we discarded this year because of lack of information enabling us to discriminate between educational levels, which is a critical variable for estimating the quantities of skilled and unskilled labour, and, consequently, the skill premium.

[30] We exclude job ads using some criteria, following the example of past studies (Banfi & Villena-Roldán, 2019): (i) monthly wages below CLP (Chilean Pesos) 150,000 (minimum wage at the start of the period) or above CLP 5,000,000 (unfeasible), (ii) work experience above 30 years (less probable).

For our purposes, the key advantages of this data are the detailed requirements stipulated by firms. For instance, some requirements such as formal qualifications (to identify job posting demanding skilled workers), offered wages (to build the skill premium), work activities to be performed (to build our task-content measures), and the candidate's required skills (to obtain our skill-related measures).

Past studies have used data from www.trabajando.com, which is considered as the principal internet labour market intermediary in Chile over the 2000s (Ramos et al., 2013), to examine the impact on wages of job skills, job search behaviour, among other aspects of labour markets (Banfi et al., 2019, 2020; Banfi & Villena-Roldán, 2019; Ramos et al., 2013). In this regard, the proper representativeness of data from www.trabajando.com, for the purposes of our research arise from the sizeable proportion of higher level technicians and professionals looking for work (some 13% according to Ramos et al., 2013). Regarding the offered wages, we note that this information is required for all firms posting a job ad, although they can choose whether this information is published in the job ad or not. Despite this feature, Banfi & Villena-Roldán (2019) shows that these hidden wages are reliable measures of salaries that firms expect to pay.

Table 3.3 shows some descriptive statistics and features of our job ads sample. These statistics are similar to Banfi & Villena-Roldán (2019), using the same data source but different periods (January 2008- June 2014).

Table 3.3. Summary of features for Trabajando.com 2009-2018 job ads

Required years of experience (%)	
0	16%
1	30%
2 to 3	39%
4 to 20	15%
average years of experience (SD)	2.09 (1.89)
Required education level (%)	
Primary/secondary/technical secondary	36%
Technical tertiary	29%
College (tertiary)/graduate	34%
Other	1%
Sectors (%)	
Manufacturing	17%
Electricity/gas/water	2%
Commerce	19%
Transportation	5%
Communication	9%
Financial/business/personal service	27%
Other	21%
Offered wage (%)	
CLP <= 300,000	22%
CLP 300,001-600,000	39%
CLP 600,001-1,000,000	23%
CLP>1,000,000	16%
Average CLP offered wage (SD)	690,839 (542,024)
Observations	189,986

Note: CLP stand for Chilean Pesos

3.5.Methods

This section discusses the techniques used to build the skill premium and task-content and skill-related measures and the VAR estimation method used to examine the empirical models specified in Eq. (3.1) and Eq. (3.2).

3.5.1. Construction of variables

In this section, we detail how we build the skill premium and our measures standing for the task-content of jobs and worker's skills endowments.

3.5.1.1. The skill premium estimation

The skill premium construction adopts the strategies generally used in studies that examine this variable (e.g., Autor et al., 2008; Card & Lemieux, 2001; Ciccone & Peri, 2005). For Chile, see, e.g., Gallego (2012), Murakami (2014), Beyer et al. (1999) and the first essay (section 2.5). We define skilled labour as college or tertiary graduates and unskilled labour as high-school or secondary education graduates, or those who have had even less education, according to the educational requirements in job ads. To estimate the skill premium, we regress the monthly offered wage on typical wage determinants available in the data following a Mincer regression strategy using all the job ads for a given month. Then, using the differences for predicted wages between skilled and unskilled workers, we construct our estimate for the skill premium month by month. We focus on job ads offering full-time positions. To adjust for compositional changes, we use weighted averages from education construction by experience subgroups. The skill premium estimation consists of the following three steps:

Step 1) Construction of education by experience sub-groups to adjust for compositional labour changes (e.g., different skills levels) within each sub-group using the educational level and experience specified in the job ads. We define four educational categories as our measure of schooling for different workers' school attainments: college graduates, some college, high school graduates and less educated (primary and high school dropouts). There are three experience subgroups: 0-2, 3-5 and 6-30 years. Combining the education and experience categories, we construct 12 education-by-experience sub-groups. We use the total hours worked monthly for each sub-group as weights, assuming that full-time positions correspond to 193.5 working hours per month (45 hours per week * 4.3 weeks per month).

Step 2) Estimating the predicted wages for skilled and unskilled workers regressing a Mincer type equation. We regress the wages for each monthly sample of job ads estimating the next standard wage equation[31]:

$$log(W_{i,t}) = cons + educ\_cat'_{i,t}\alpha_j + \beta_1 exp_{i,t} + \beta_2 exp^2_{i,t} + X'_{i,t}\delta \qquad (3.3)$$

where $W_{i,t}$ is the monthly offered (log) wage for job ad i in month t, expressed in December 2018 Chilean pesos (CLP) using the Unidad de Fomento as a deflator[32]. $educ\_cat$ are j educational categories defined in Step 1) with "less educated" as the base category. exp is the required work experience. X is a vector containing additional determinants, such as the economic sector of the firm posting the job ad (eight industries such as agriculture, mining, construction, etc., with manufacturing as the base category) and the firm size (big, medium, small, with micro as the base category). We use these regression results to compute the predicted wages for skilled and unskilled workers as detailed in Step 3).

Step 3) Estimation of the predicted average wage for skilled and unskilled groups and computation of the skill premium. We estimate the predicted log wages using regression results from Step 2) evaluated at the correspondent experience level (1, 4, or 10 years based on experience categories) and at base categories included in vector X. We compute the predicted log wages difference between the college graduates and high-school graduates as our proxy for the skill premium. We use the sum of monthly hours worked for each of the education x experience sub-groups built in Step 1) as weights. Thus, we quantify the difference between skilled and unskilled wages as our skill premium measure for a given month, t, which we denoted as ω_t, following the notation from our empirial models represented by Eq. (3.1) and Eq. (3.2).

3.5.1.2. Estimation of task-content measures

Our strategy of building task-content indicators from job ads data relies on the quantification and classification of tasks proposed by the ALM model. Our measures show the prevalence of each category of tasks across the total of job postings demanding skilled labour by allocating job posting to standard occupations. These standard occupations give detailed work activity descriptions, which we can classify according to the ALM model's task categories (see Table 3.1 and Table 3.2). However, job ads do not follow standard national or international labour classifications. Besides, the task descriptions are specific to the offered jobs, resulting in a lack of information about additional general tasks. To tackle this difficulty, we developed a strategy consisting of three steps. It starts with the manual classification of work activities that feature each occupational group into the categories proposed by the ALM model,

[31] This methodology allows to control of the labour supply by other demographic characteristics which are not related to the education premium.
[32] The Unidad de Fomento (UF) is a Chilean unit of account. The exchange rate between the UF and the Chilean peso is constantly adjusted for inflation.

using national and international classifications of occupations as statistical tools and dictionaries from the literature. The second step categorises each job ad according to its standard occupational groups, but the data does not contain references to standardised classes of occupations, so we infer that information using the text data from the job ads by applying a classifier algorithm. The third step corresponds to the mapping between the task analysis of occupations in step one and the classification of job ads from step two, and the construction of measures to represent the task content of occupations. We detail these steps as follows.

3.5.1.2.1. Step One: Examining the task content of standard occupational groups

To evaluate the task content of standard occupational groups, we rely on the task descriptions for occupations documented in the Chilean Classification of Occupations, CIUO08-CL (INE, 2018). In turn, CIUO08-CL relies on the current International Standard Classification of Occupations, ISCO-08 (ILO, 2012). To ensure reliability, CIUO08-CL is prepared and published by the government agency in charge of national statistics for the labour sector, in Spanish, *Instituto Nacional de Estadísticas*, INE (INE, 2018). Like ISCO-08, the CIUO08-CL structure is hierarchical, with standard occupations organised into one of the 444 unit groups at the most exhaustive level of the classification hierarchy. From the top down, ten major groups are composed of 44 sub-major groups, containing 129 minor groups. The 129 minor groups contain 444 unit groups, defined by their members' primary occupations. In terms of coding, 1-digit, 2-digit, 3-digit, and 4-digit codes represent the major, sub-major, minor and unit groups, respectively. To illustrate the CIUO08-CL structure, Table 3.4 presents an example of the hierarchy and tasks descriptions.

Table 3.4. Examples of CIUO08-CL and ISCO-08 structure

Groups	Codes	Occupational groups	Tasks descriptions
Major Group	2	Professionals	Conducting research and analysis, developing concepts, applying knowledge related to sciences, providing various businesses, legal and social services
Sub-major Group	25	Information and Communications Technology Professionals	Conducting research, planning, designing and providing advice for information technology systems, hardware, software, web applications.
Minor Group	251	Software and Applications Developers and Analysts	Evaluating, planning, and designing hardware or software configurations for specific applications, designing, writing, and maintaining software for specific requirements, consulting with users
Unit groups	2511	System Analysts (Computer scientists, Information systems analysts)	Consulting with users to formulate document requirements, identifying and analysing business processes, recommending optimal businesses and system functionalities

2512	Software Developers (Programmer analysts, Software designers, Software engineers)	Researching, designing, and developing software systems, directing software programming and development of documentation, consulting with customers concerning software systems
2513	Web and Multimedia Developers (Animation programmers, Internet developer, Website	Analysing, designing, and developing Internet sites, designing, and developing digital animations, imaging, presentations, games, assisting in analysing Internet strategies, web-based methods

Source: Own from ILO (2012) and INE (2018)

Since the analysis of task content for 4-digit and 1-digit groups can result in excessively narrow or broad descriptions of occupational duties, respectively, we analyse the task content for the 2-digit or sub-major occupational groups. We exclude three occupational groups representing Armed Forces occupations since CIUO08-CL does not detail their tasks. Therefore, we examine the task content of 41 2-digit occupations (Occupations codes and names in Spanish and English in Appendix A.2.1).

To enrich our analysis, we evaluate the job tasks according to the aggregation of 3-digit level groups. CIUO08-CL at 3-digit hierarchy reports 845 work activities (803 unique). We assign these job tasks manually to the five ALM model categories: routine cognitive (NR), non-routine analytic (NRA), non-routine interactive (NRI), routine manual (RM) and non-routine manual (NRM) (see section 3.2.1). We support this task's classification process using translated work-tasks dictionaries (English to Spanish) from the literature (see Table 3.2). Once we have done the classification, we compile tasks shares and routine and cognitive prevalence index for the 41 2-digit occupational groups following Autor et al. (2003) and Autor & Dorn (2013)[33].

Our task shares computation aims to show the relative importance of each task category j for the occupation k. We compute the share of work activities for a given task category over the total of work activities as follow:

$$TS_{j,k} = \frac{n\ of\ work\ activities\ in\ task\ category\ j\ in\ occupation\ k}{n\ of\ work\ activities\ in\ occupation\ k} \tag{3.4}$$

where TS is the Task Share with j referring to each of the five ALM model categories, $j = \{NRA, NRI, RC, RM, NRM\}$, as defined above. The term k represents each of the 41 2-digit occupations. As a result, we obtain five TS measures: TS_{NRA}, TS_{NRI}, TS_{RC},TS_{RM} and TS_{NRM}, which sum one and characterize each k occupation. These TS metrics measure the variation in intensity across the occupations. To illustrate, occupations with higher values for TS_{NRA} correspond to occupations with an intense demand for NRA tasks.

To show the prevalence of routine and cognitive tasks using the TS shares, we follow the approach of Autor & Dorn (2013) to compute the routine intensity index, RII, as follows:

$$RII_k = TS_{RC,k} + TS_{RM,k} - TS_{NRA,k} - TS_{NRI,k} - TS_{NRM,k} \tag{3.5}$$

[33] Several researchers have applied these kinds of metrics in past studies (e.g. Antonczyk et al., 2009; de Vries et al., 2020; Goos et al., 2014; Mihaylov & Tijdens, 2019; Perez-Silva & Campos, 2021).

where RII_k is the routine intensity index for occupation k. TS_{NRA}, TS_{NRI}, TS_{RC},TS_{RM} and TS_{NRM} are the TS measures described in Eq. (3.4). RII takes positive and negative values for occupations with an intense demand for routine and non-routine tasks, respectively, in the range [-1, 1]. Similarly, we build the cognitive intensity index, CII, as follow:

$$CII_k = TS_{NRA,k} + TS_{NRI,k} + TS_{RC,k} - TS_{RM,k} - TS_{NRM,k} \tag{3.6}$$

where CII_k is the cognitive intensity index for occupation k. The CII_k metric is rising in the prevalence of cognitive tasks, both non-routine and routine, and declining in manual tasks, both routine and non-routine. CII will take positive values for those occupations with a high prevalence of cognitive tasks with values near one for those with high intensity. Conversely, CII will take negative values for those occupations without intense demand for cognitive tasks or with a high demand for manual tasks. With the values of CII and RII we can map each occupation to the ALM model categories as shown in Table 3.5, following, e.g., Autor & Dorn (2013) and de Vries et al. (2020).

Table 3.5. Mapping 2-digits occupations to task-content intensities CII and RII

	Routine	Non-routine
Cognitive (interactive/ analytical)	Occupations with intense demand for cognitive and routine tasks $(CII+; RII+)$	Occupations with intense demand for cognitive non-routine tasks $(CII+; RII-)$
Manual	Occupations with intense demand for manual and routine tasks $(CII-; RII+)$	Occupations with intense demand for manual non-routine tasks $(CII-; RII-)$

3.5.1.2.2. Step Two: Classification of job ads into the 41 2-digit occupations

This step aims to classify our job ads sample according to the 41 2-digit occupations described by the Chilean standard classification system of occupations. The inputs for performing the classification are our open text variables (job title, job description and job-specific requirements) and the educational level. We apply a flow of techniques and algorithms to these inputs to obtain a 2-digit occupation label as a new variable for each job ad. This stage is a two-sub-stage flow compound. First, we pre-process the text data (e.g., cleaning, normalisation) and construct the document-term representation, DTM, based on our job ads *corpus*. The DTM is needed since algorithms do not deal directly with text data, but they perform on specific text features such as words or groups of words (Welbers et al., 2017). Secondly, we “train” and evaluate our classifier algorithm, Support Vector Machines, SVM (Cortes & Vapnik, 1995) using a training dataset. Then, we apply our SVM to our unlabelled dataset. The techniques described are implemented using R packages like `Quanteda` (Benoit et al., 2018) and the Python library `Scikit-learn` (Pedregosa et al., 2011).

3.5.1.2.2.1. Pre-processing and DTM representation

This sub-stage starts with the concatenation of the three open-text variables. We perform a set of standard techniques on the new concatenated variable following literature analysing text data (see, e.g., Welbers et al. 2017). In general, these techniques refer to converting words to lower case, removal of Spanish stop words, punctuation and special symbols, *tokenisation*, and stemming (words being reduced to their word *stem*). The tokenisation allows the text content to be split into words (unigrams) and groups of two consecutive words (bigrams) denoted as *tokens* or *features*. In this study, we use unigrams and bigrams. Bigrams can be more representative for job titles composed of two words (e.g., job titles preceded by generic words like, in Spanish, "Ingeniero" ("Engineer" in English) such as "Ingeniero Informático" ("Informatics Engineer" in English).

Based on these features, we build a DTM, which shows the collection of job ads or documents represented in the vector space model. In the DTM, job ads and tokens are rows and columns, respectively. The DTM represents the *corpus* as a bag of words and is usually sparse; it is the primary input for SVM. We applied these pre-processing and DTM techniques to, firstly, our training sample and, secondly, our unlabelled observations to classify the job ads against the Chilean standard classification system of occupations CIUO08-CL.

3.5.1.2.2.2. The SVM application

The SVM, initially known as *support-vector networks,* is an algorithm developed by Cortes & Vapnik (1995). SVM is a real-world oriented application (Nalepa & Kawulok, 2019; Smola & Scholkopf, 2004) which researchers have successfully performed on classification analysis in multiple fields due to its capacity to learn from data to attain the best separation between classes or groups of data (Gil & Johnson, 2011). SVM used as a classifier algorithm has shown reasonable accuracy in text classification using job ads (see, e.g., Guerrero & Cabezas, 2019, for a Chilean application and an Italian exercise in Boselli et al., 2018). In our analysis, we applied SVM to classify labour data into occupational categories similar to past studies (Guerrero & Cabezas, 2019; Javed et al., 2014, 2015; Lovaglio et al., 2018; Nahoomi, 2018).

Overall, the SVM algorithm predicts occupational category labels according to a subset of training data already labelled with their 2-digit occupational group or the *training dataset* (we detailed the training sample construction in section 3.5.1.2.2.2.2 below). SVM uses a set of functions to convert the training data into a high dimensional space to find one or multiple optimal separating hyperplanes. An ideal hyperplane separates one class from another based on the *support vectors,* which refers to the critical training instances that define its margins; therefore, they give the most information about the classification (Han et al., 2011). This hyperplane should also stay as far away from the nearest training instances as possible (Gerón, 2017). We evaluate the predictive capability of our SVM using measures developed for this purpose as detailed below (see section 3.5.1.2.2.2.3).

3.5.1.2.2.2.1. SVM theoretical overview and implementation

To illustrate how, theoretically, an SVM classifier achieves its goal, we will assume a two-class problem with a dataset D linearly separable. D refers to $(X_1, y_1), \ldots., (X_{|D|}, y_{|D|})$ where X_i corresponds to the set of training instances labelled according to their class, y_i, which can take the values of +1 or -1. Since our data is linearly separable, graphically, we can draw infinite straight lines between the two classes. The SVM searches a separating hyperplane, the *maximum marginal hyperplane*, to discriminate between the classes in a high dimensional space. Simultaneously, the margins refer to the shortest distance between the hyperplane and the closest training instance of either class (Gerón, 2017; Gil & Johnson, 2011; Han et al., 2011). Following Han et al. (2011), we can write the separating hyperplane as:

$$W \cdot X + b = 0 \tag{3.7}$$

where W is a row vector of weights, $W = (w_1, w_2, \ldots, w_n)$ and n is the number of attributes. In our assumption, we have two classes, denoted by a column vector $X = (x_1, x_2)$ where x_1 and x_2 are the values of attributes. b is a scalar usually associated with bias. We re-write Eq. (3.7) as:

$$b + w_1x_1 + w_2x_2 = 0 \tag{3.8}$$

Therefore, any data point located above or below the separating hyperplane satisfies Eq. (3.9) and Eq. (3.10), respectively

$$b + w_1x_1 + w_2x_2 > 0, \tag{3.9}$$

$$b + w_1x_1 + w_2x_2 < 0. \tag{3.10}$$

We can define the sides of the maximal margin using new hyperplanes, h_1 and h_2, based on the adjustment of weights as follow:

$$h_1: b + w_1x_1 + w_2x_2 \geq 1 \text{ for } y_i = +1, \tag{3.11}$$

$$h_2: b + w_1x_1 + w_2x_2 \leq 1 \text{ for } y_i = -1. \tag{3.12}$$

If any training data point falls on or above h_1 it will belong to class $+1$ while any training data point that falls on or below h_2 will belong to class -1. The combination of inequalities from Eq. (3.11) and Eq. (3.12) yields:

$$y_i(b + w_1x_1 + w_2x_2) \geq 1, \forall i. \tag{3.13}$$

The *support vectors* will be any training data point that falls on the sides of the maximal margin; this is the hyperplanes h_1 and h_2. Since the *support vector*s satisfy the Eq. (3.13) and are located equally near the separating hyperplane, we can use this expression to find the maximal margin between h_1 and h_2 (Gil & Johnson, 2011). By definition, the distance from a point (x_0, y_0) to a line $a_x + b_x + c = 0$ is $|a_x + b_x + c = 0| / \sqrt{a^2 + b^2}$, therefore, the distance from any *support vector* on h_1 to the separating hyperplane is $|W \cdot X + b|/\|W\|$ which is equal to $\frac{1}{\|W\|}$. $\|W\|$ is the Euclidean distance from the origin to W, that is $\sqrt{W \cdot W}$ (recalling from Eq. (3.7), $W = \{w_1, w_2, \ldots, w_n\}$, then $\sqrt{W \cdot W} = \sqrt{w_1^2, w_2^2, \ldots, w_n^2}$). Since this distance is the same from any *support vector* on h_2 to the separating hyperplane, the maximal theoretical margin possible is $\frac{2}{\|W\|}$. Therefore, to maximize the separating hyperplane, the value of $\|W\|$ need to be minimised with the condition given by Eq. (3.13) to avoid training data points falling between h_1 and h_2. We can re-write the problem as a quadratic programming problem

$$\min_{W,b} \frac{\|W\|^2}{2}, \tag{3.14}$$

concerning the constrain represented by Eq. (3.13). This formulation is usually known as the *primal form* (Nalepa & Kawulok, 2019), which is the problem to be solved applying SVM. SVM performs a set of mathematical functions and procedures and transformations, the so-called "fancy math tricks" (Han et al., 2011), to find the separating hyperplane and the *support vectors*. These mathematical approaches include Lagrangian formulations, Karush-Kuhn-Tucker conditions, and kernel functions (e.g. linear, polynomials) to handle linearly inseparable data, among others (Gil & Johnson, 2011; Nalepa & Kawulok, 2019).

To implement our SVM strategy, we use the `Scikit-learn` software library via an interface in Python (Pedregosa et al., 2011). We apply the linear support vector classification, LinearSVC, to solve our multi-class optimization problem. LinearSVC applies the one-vs-the rest strategy to fit one classifier per class. Thus, to obtain knowledge about a particular class, we evaluate only its computed classifier. In parameter tuning, we control the balance between maximising the margin and reducing misclassification using the C parameter, as explained below, and the potential imbalance between classes using weights. LinearSVC formulates Eq. (3.14) equivalently as:

$$\min_{W,b} \frac{1}{2} W^T W + C \sum_{i=1}^{n} \max\left(0, 1 - y_i(W^T \phi(x_i) + b)\right), \tag{3.15}$$

where ϕ is the identity function, and C is a real and positive constant. The LinearSVC algorithm uses the math tricks noted above with a linear kernel to optimize Eq. (3.15) with C as a tuning parameter. According to Gerón (2017), C controls the balance between keeping the maximal margin as wide as possible and limiting the misclassifying i.e., hen instances fall in the middle of the margin or even on the wrong side. We use $C = 1$, which is the recommended value for LinearSVC (Pedregosa et al., 2011). The same parameter has been used by Guerrero & Cabezas (2019) in their study classifying occupations for Chile from national labour surveys. We also experiment with alternative values to analyse impacts on classifier performance. A lower value for C gives more regularization if the data contains a high number of noisy observations, and higher values for C (e.g., 10, 100) result in a lower generalization ability of the classifier. This impact on generalization means that SVM may classify appropriately on the training stage but its performance on new samples would be poor (Auria & Rouslan, 2008).

We also control for the expected unbalance between classes due to the natural distribution of labour. For instance, clerical workers have a significantly higher representation than managers. We apply weights to optimize the classifier performance on less represented classes. In LinearSVC, we include class weights inversely proportional to the class frequencies using the formula $weight_j = n/(k * n_j)$ where $weight_j$ is the weight to class j, n is the number of observations in the dataset, k is the number of classes and n_j is the number of observations in class j.

3.5.1.2.2.2.2. The training dataset construction

The SVM, as noted above, requires a training dataset, i.e., a job ads sample already labelled with their 2-digit occupational group code. We start by selecting the most frequent job titles. We filter 3,359 job titles whose frequencies vary between 2,000 and 20. At this point, our training sample is a subset of 67,656 job ads, i.e., 35% of our whole job ads dataset described in section 3.4. The rest of our whole dataset, i.e., the 122, 330 unlabelled job ads, will be labelled using SVM. The distribution of our training sample in terms of industry and educational category is similar to the whole dataset distribution (See Appendix A.2.2). We manually label each job ad from our training sample according to the Chilean classification CIUO08-CL, supporting this labelling by observing the educational category, economic area and job tasks descriptions reported by CIUO08-CL. We also support our labelling process by examining the training dataset employed by Guerrero & Cabezas (2019), which was prepared by domain experts from the Chilean National Institute of Statistics (in Spanish *Instituto Nacional de Estadísticas*).

As observed above, since algorithms do not deal directly with text data but perform on specific text features, we apply the procedures detailed in section 3.5.1.2.2.1 to obtain our DTM representation of the training data. We use 80% of the training sample to train the SVM and the rest (20%), our testing dataset, to evaluate the SVM performance.

3.5.1.2.2.2.3. SVM evaluation and prediction

To evaluate the SVM classification performance, we use metrics (e.g., $accuracy$, $precision$) based on four outputs by comparing the labelled categories with those predicted by SVM using our testing dataset. These outputs are true negatives, TN, when the observation is negative and predicted negative; false negatives, FN, when the observation is positive but predicted negative; true positives, TP, when the observation is positive and predicted positive; and false positives, FP, when the observation is negative but predicted positive. Typically, the classifier accuracy is calculated as the ratio of all correct predicted observations to the total number of observations, as follows:

$$accuracy = \frac{TN+TP}{TN+FN+TP+FP}. \tag{3.16}$$

We also examine $precision$, $recall$, and $F1\ score$, which are standard metrics used to evaluate the classifier performance at the global and class level. The $precision$ and $recall$ measures refer to ratios to measure the ability of SVM to avoid labelling as positive an observation that is negative and to find all the positive observations, respectively and $F1\ score$ corresponds to the weighted harmonic mean of both metrics (Pedregosa et al., 2011). Formulation of these metrics is:

$$precision = \frac{TP}{TP+FP}, \tag{3.17}$$

$$recall = \frac{TP}{TP+FN}, \tag{3.18}$$

$$F1\ score = 2 * \frac{(precision * recall)}{(precision + recall)}. \tag{3.19}$$

Intuitively, $precision$ counts for the number of observations correctly classified among that class and $recall$ quantifies the number of cases for a given class found by the classifier over the total number of class cases. These metrics can also evaluate global performance by calculating averages ($avg\ precision$ or $avg\ accuracy$) which can take classes' imbalance into account. We also examine $macro\ avg\ F1\ score$, the unweighted mean of $F1\ score$, which results in higher penalization if the classifier does not perform appropriately with less represented classes since all classes have the same weight. We also examine $weighted\ avg\ F1\ score$, which uses as weights the number of true positives for each class. This weighted version adjusts $macro\ avg\ F1\ score$ to account for class imbalance. However, according to Pedregosa et al. (2011), it can result in an $F1\ score$ different from that described by Eq. (3.19). The results from these metrics are presented in the form of a classification report which outlines the results from these metrics at the global and class level.

We expect averages values for $precision$ and $accuracy$ of around 0.85 according to past studies (see, e.g., Guerrero & Cabezas, 2019) as a measure of the proper performance of SVM. Some adjustments related to tokens or features (e.g., number, frequency) or balancing between classes can impact these measures. Once we consider a proper SVM performance, we apply our SVM algorithm to the 122,330 unlabelled observations. Thus, we classify our whole data set of job ads according to the Chilean standard classification system of occupations CIUO08-CL.

3.5.1.2.3. Step Three: Construction of task-content time series variables from results in Step One (section 3.5.1.2.1) and Step Two (section 3.5.1.2.2)

This section describes the construction of time series representing our task-related measures on a monthly basis. Similar strategies followed past studies using Chilean data from household surveys (Perez-Silva & Campos, 2021). Since we are interested in the impact on the skill premium, we compute measures using only job postings requiring skilled labour, JPS, by examining the educational level required by firms (see section 3.4). We name our task-related measures as TM, and we compute them for all j ALM model categories (see section 3.5.1.2.1, Eq. (3.4)) over all months, t, according to our data sample, t={1,2,….120}(see section 3.4). Thus, our $TM_{j,t}$ measures stand for the share of job postings devoted to j task category relative to all job postings in t considering only JPS. For instance and following notation from Eq. (3.4), the $TM_{NRA,t}$ stand for the proportion of job postings devoted to non-routine analytical task category, NRA, relative to summing all JPS in a given t.

Three sub-steps compound this stage. First, using the output from Step Two above, i.e., our labelled job ads dataset with the 2-digit occupations (see section 3.5.1.2.2), we obtain skilled job posting frequencies for each k occupation in each t month, i.e., $JPS_{k,t}$ where $k = \{1,2,3 \ldots 41\}$. Secondly, we distribute each $JPS_{k,t}$ into the five j task categories using the computed $TS_{j,k}$ metrics (see section 3.5.1.2.1) as weights. Notably, $TS_{j,k}$ does not depend on time since we assume that the task content of occupations is constant over time (Reijnders & de Vries, 2018). Thirdly, we compute the numerator of our $TM_{j,t}$ by summing the weighted quantities, i.e., the product $JPS_{k,t} * TS_{j,k}$, for a given j task category over all the k occupations and the denominator by summing all JPS over all the k in t.We represent our $TM_{j,t}$ measure as follows:

$$TM_{j,t} = \frac{\sum_k (JPS_{k,t} * TS_{j,k})}{\sum_k JPS_t} \tag{3.20}$$

where TM is the task measures as explained earlier. As a result, we obtain our five TM measures standing for each of the ALM model categories: $TM_{NRA,t}$, $TM_{NRI,t}$, $TM_{RC,t}$,$TM_{RM,t}$ and $TM_{NRM,t}$. These metrics measure the prevalence of a given task category over t periods based on the task content of occupations. The use of TS as weights allows us to consider the variation in intensity for a given task category across the occupations. Since this research focuses on how cognitive tasks drive the relative demand for skilled labour, we evaluate the influence on the skill premium of TM_{NRA}, TM_{NRI}, and TM_{RC}.

3.5.1.3. Estimation of skills-related measures

We code a job ad as having a specific skill category if we find keywords or phrases at least once. To perform this task, we build a dictionary linking our skills categories to text data. Our dictionary extends the skills categorization of Deming & Kahn (2018) by adding words and phrases from Spanish versions of the European Dictionary of Skills and Competencies (3s Unternehmensberatung, 2020) and the Occupational Information Network, O*NET (National Center for O*NET Development, 2020). We present a brief English version of our dictionary categories and related keywords and phrases in Table 3.6 for illustration purposes. Appendix A.2.3 details the complete Spanish version of our dictionary.

The inputs used to classify each job ad according to the skills category are the job title, job description and job-specific requirement variables. We concatenate these three open text variables and use the R package `Quanteda` developed by Benoit et al. (2018) to apply our three-step dictionary analysis. First, we create a *corpus*, our library of original job postings text stored with the job postings identifier. Second, we apply *tokenization* to the *corpus* to identify each word individually along with other pre-processing text operations (e.g., removing symbols, numbers, Spanish stop words). We also evaluate the addition of bigrams and trigrams, groups of two or three consecutive words, respectively, since some phrases on specific skills are longer than one word. Thirdly, we apply the dictionary to the tokens data to obtain the matching results, which correspond to the numbers of keywords or phrases found in the job postings text data according to our dictionary categories.

Table 3.6. Dictionary of Skills Categories

Job Skills Categories	Keywords and Phrases
Cognitive	Problem-solving, research, analytical thinking, critical thinking, math, statistics, logical thinking, resourcefulness, self-assessment, technical understanding, intellectual curiosity, powers of discernment, among others.
Social	Communication, teamwork, collaboration, negotiation, presentation, communication in foreign languages, establishing contacts, fostering contacts, intercultural competences, among others
Character	Organised, detail-oriented, multitasking, time management, meeting deadlines, energetic, courage, personal initiative, judgement, discretion, ability to cope with pressure, punctuality, among others
Writing	Writing, clear writing style, elegant writing style, writing drafts, writing technical information and documents, among others
Customer service	Customer and customer orientation, sales talent, client and client orientation, patient
Project management	Project management, project manager
People management	Supervisory, leadership, management (not project), mentoring, staff, human resources management, among others
Financial	Budgeting, accounting, finance, cost
Computer (general)	Computer, spreadsheets, standard software (e.g., Microsoft Office, Microsoft Excel), internet user, among others
Software (specific)	List of 175 ICT technologies categorised as *hot technologies* regarding programming language and/or specialised software (e.g., Java, SQL, Python, Amazon Web Services)

Source: Adapted from literature and specialized dictionaries (3s Unternehmensberatung, 2020; Deming & Kahn, 2018; National Center for O*NET Development, 2020)

Also, in our procedure linking job ads to each skills category, a given job posting can contain more than one skill category and, we select the most prevalent, using the largest matching number. However, let us suppose we place a job posting in two or more categories due to equal or higher prevalence (e.g., the same number of keywords codified for the given categories). In that case, we cannot distinguish the dominant category, which excludes the job postings from the analysis.

Once we have categorized job postings according to the skills categories, we build measures to represent the prevalence of Cognitive ($Cogni$), Social (Soc), job postings requiring Cognitive and Social ($CogniSoc$), and Software ($Soft$) skills. As pointed out earlier, we are examining the impact on the skill premium, therefore, our measures are based on frequencies of job postings requiring skilled labour or JPS over t months. We represent our skills-related metrics, $SM_{z,t}$, for $z = \{Cogni, Soc, CogniSoc, Soft\}$ as follows:

$$SM_{z,t} = \frac{JPS_{z,t}}{JPS_t} \tag{3.21}$$

where SM refers to the share of skilled job postings frequency classified in a given z skill category relative to the total frequency of skilled job postings in month t. Thus, our skills metrics are SM_{Cogni} SM_{Soc}, $SM_{CogniSoc}$ and, SM_{Soft} for each t.

3.5.2. VAR modelling and econometric estimation

To test our empirical models from Eq. (3.1) and Eq. (3.2), we model our time series data interactions using the VAR framework (Sims, 1980). This modelling relies on an autoregressive model applied to a series vector. This modelling allows us to treat each variable symmetrically (Enders, 2015). Thus, every variable is specified as endogenous and, in essence, dependent on all other lagged variables. Employing standard forms of inference within the VAR specification depends on the assumption of the stationarity of the variables, where, among other things, it is assumed that "unit roots" are not present. Testing for unit roots is required since estimation and inference in VARs become non-standard in the presence of unit roots in the data. Also, we perform lag order testing to estimate the optimal lag order for our VAR specification. Once we have performed these diagnostics and optimal lag selection, we estimate our VAR parameters. These parameters allow us to identify whether changes in a given variable causes changes in our target variable, the skill premium, by Granger causality testing and impulse-response function analysis, IRF (Granger, 1969).

Also, since we analyse monthly data, like other labour outputs, our variables are natural candidates to be seasonal. In this sense, this data evolves in 12-month rounds; then, there is a potential serial correlation at the 12th lag. Therefore, we test and control for seasonality alongside our estimation strategy testing and including relevant seasonal dummies.

We detail our VAR strategy stages below. We perform these analyses using `Gretl` as statistical software (Baiocchi & Distaso, 2003; Cottrell & Lucchetti, 2021).

3.5.2.1. Stationarity and optimal lag order testing

We apply the same stationarity and optimal lag order testing approaches as in essay 1 (see sections 2.6.1.1 and 2.6.1.2 in Chapter 2). For stationarity, we conduct ADF and KPSS testing to modelling cases, including deterministic terms such as constant and linear trend. Also, we can add a quadratic trend, an available strategy in ADF but not in our KPSS due to software limitations[34]. We also examine seasonality by including seasonal dummies.

Regarding optimal lag order selection, the typical approach is estimating VAR models with different lag orders beginning with higher-order lags. Since we use monthly data, our higher-order lag

[34] We also found this KPSS limitation in other common statistical software like `EViews`, `Stata`, and R packages like `tseries`.

is 12. The selected lag order relies on inspecting minimum values of statistical information criteria such as BIC and HQC that penalize overfitted models (see section 2.6.1.2 for details). In this research, we also examine how seasonal dummies might affect the BIC and HQC through improvements in the information criteria values designed to select the optimal VAR considering seasonality[35].

3.5.2.2. VAR specification and estimation

To illustrate our VAR specification, let us suppose we are interested in capturing interactions between two economic variables, $x_{1,t}$ and $x_{2,t}$. According to Patterson (2000), in the VAR representation of this bivariate problem, $x_{1,t}$ is related to both its own lagged values and those of $x_{2,t}$, and equivalently $x_{2,t}$ is linked to its own lagged values and those of x_{1t}. Thus, two dimensions feature in a VAR model: the lag order in the autoregression, p, and the number of variables, k. In a two variables application, $k = 2$, a first-order VAR, $p = 1$, is

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \pi_{1,1} & \pi_{1,2} \\ \pi_{2,1} & \pi_{2,2} \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \tag{3.22}$$

where μ are deterministic terms (e.g., a constant, a deterministic trend or both), ε_t are error terms, and t is time. A multivariate VAR generalization with order p and n variables is (Enders, 2015)

$$X_t = \mu_t + A_1 X_{t-1} + A_2 X_{t-2} + \cdots + A_p X_{t-p} + \varepsilon_t \tag{3.23}$$

where X_t is an $(n \times 1)$ vector containing the n variables involved in the VAR, μ_t is an $(n \times 1)$ constant vector or deterministic function of time, A_i are the $(n \times n)$ matrices of coefficients and ε_t is a $(n \times 1)$ vector of i.i.d. multivariate normal error terms. To generalise the model in Eq. (3.23), we may add exogenous variables as explanatory variables, and the constant term μ_t might instead represent a polynomial in time.

Our empirical specifications from Eq. (3.1) and Eq. (3.2) modelled under the VAR framework yield the following representations. We assume that our VARs are first-order though additional lags will be included in the lag selection phase and also possibly in the selection of the optimal model. For our task-related analysis, our VAR model with a (4×1) vector of endogenous variables and, assuming a $p = 1$, is:

$$\begin{bmatrix} \omega_t \\ TM_{NRA,t} \\ TM_{NRI,t} \\ TM_{RC,t} \end{bmatrix} = \begin{bmatrix} \mu_{1,t} \\ \vdots \\ \mu_{4,t} \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,4} \\ \vdots & \ddots & \vdots \\ \beta_{4,1} & \cdots & \beta_{4,4} \end{bmatrix} \begin{bmatrix} \omega_{t-1} \\ TM_{NRA,t-1} \\ TM_{NRI,t-1} \\ TM_{RC,t-1} \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{s-1} \rho_{1,i} D_{1,i,t} \\ \vdots \\ \sum_{i=1}^{s-1} \rho_{4,i} D_{4,i,t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{4,t} \end{bmatrix} \tag{3.24}$$

where ω is the skill premium, and $TM_{NRA,}$, TM_{NRI} and TM_{RC} are the task-content related variables (see section 3.5.1.2.3 for construction details), t is time and μ is the deterministic trend component. $\beta_{i,j}$

[35] The addition of seasonal dummy variables prevents the optimal lag being equal to the seasonal period (e.g., 12 for monthly data) at the lag selection stage, since the high additive seasonality might otherwise induce a high autocorrelation at the 12th lag.

stands for the elements of the matrix of coefficients of lagged variables, and $\sum_{i=1}^{s} \rho_i D_{i,t}$ stands for our $s-1$ seasonal dummies D[36] (11 for our 12-month data periodicity).

Similarly, for the skill-related analysis, our VAR model with the (5×1) vector of endogenous variables is:

$$\begin{bmatrix} \omega_t \\ SM_{Cogni,t} \\ SM_{Soc,t} \\ SM_{CogSoc,t} \\ SM_{Soft,t} \end{bmatrix} = \begin{bmatrix} \mu_{1,t} \\ \vdots \\ \mu_{5,t} \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,5} \\ \vdots & \ddots & \vdots \\ \beta_{5,1} & \cdots & \beta_{5,5} \end{bmatrix} \begin{bmatrix} \omega_{t-1} \\ SM_{Cogni,t-1} \\ SM_{Soc,t-1} \\ SM_{CogSoc,t-1} \\ SM_{Soft,t-1} \end{bmatrix} + \begin{bmatrix} \sum_{k=1}^{s} \rho_{1,k} D_{1,k,t} \\ \vdots \\ \sum_{k=1}^{s} \rho_{5,k} D_{5,k,t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{5,t} \end{bmatrix} \tag{3.25}$$

where SM_{Cogni}, SM_{Soc}, $SM_{CogniSoc}$ and SM_{Soft} are our skill-related measures (see section 3.5.1.3). The rest of the parameters are as described below Eq. (3.24).

The econometric estimation of our VAR representations described by Eq. (3.24) and Eq. (3.25) can be estimated using OLS on each equation. This technique is possible because all regressions have identical right-hand side variables, and the error terms are assumed serially uncorrelated with constant variance (Enders, 2015). Since the parameters are unconstrained, the multivariate OLS delivers a consistent and asymptotically efficient estimator of all the parameters that coincides with Maximum Likelihood estimation under the assumption of multivariate normal i.i.d. errors (Cottrell & Lucchetti, 2021).

We focus below on parameter estimation from the equation with the skill premium, ω, as the target variable. We specified the deterministic component μ as a linear time trend (constant, μ_0, and trend, μt) since ω shows a trend, as noted by past studies (Gallego, 2012; Murakami, 2014), and potentially, this trend might imply non-stationarity. By adding this trend component, we can detrend the series to obtain a stationary process (Wooldridge, 2009). Also, we apply logarithms to all variables. Thus, our equation of interest for the task-content analysis is

$$\ln \omega_t = \mu_0 + \mu t + [\beta_{1,1} \ldots \beta_{1,4}] \begin{bmatrix} \ln \omega_{t-1} \\ \ln TM_{NRA,t-1} \\ \ln TM_{NRI,t-1} \\ \ln TM_{RC,t-1} \end{bmatrix} + \sum_{k=1}^{s} \rho_{1,k} D_{1,k,t} + \varepsilon_{1,t} \tag{3.26}$$

Similarly, for the skill-related analysis, we have

$$\ln \omega_t = \mu_0 + \mu t + [\beta_{1,1} \ldots \beta_{1,5}] \begin{bmatrix} \ln \omega_{t-1} \\ \ln SM_{Cogni,t-1} \\ \ln SM_{Soc,t-1} \\ \ln SM_{CogSoc,t-1} \\ \ln SM_{Soft,t-1} \end{bmatrix} + \sum_{k=1}^{s} \rho_{1,k} D_{1,k,t} + \varepsilon_{1,t} \tag{3.27}$$

where μ_0, μt, $\beta_{i,j}$, and $\rho_{,1k}$ are our parameters of interest to be examined and interpreted.

[36] Since our data cover the January 2009 - December 2018 period, D_1 shows we are in the first month, i.e., it takes on the value one in January and zero otherwise. D_2 applies to February and so on.

3.5.2.3. Granger causality

The VAR estimation parameters in our last stage allow us to perform Granger causality testing (Granger, 1969). Mainly, we evaluate the null hypothesis that no lags of j variable is significant in the regression for variable i. If lagged values of j improve the prediction of i, then it is said that j 'Granger causes' i. Therefore, we can estimate the direction of causality and whether causality is bi-directional. From our bi-variate VAR illustration deployed in Eq. (3.22) under Granger testing, we test the following hypotheses:

- lags of $x_{1,t}$ do not explain $x_{2,t}$, which implies the restriction $\pi_{2,1} = 0$
- lags of $x_{2,t}$ do not explain $x_{1,t}$, which implies the restriction $\pi_{1,2} = 0$

We test these hypotheses by evaluating the F test statistics performed for each variable. If x_1 causes x_2, the x_1 lagged values should be significant in x_2 equation. In this case, we can say that x_1 'Granger causes' x_2.

We evaluate the Granger causality statistics for an equation where the skill premium is the dependent variable from the VAR specification of our empirical modelling. For example, for the model represented in Eq. (3.24), we state null hypotheses such as 'lags of TM_{NRA} do not Granger-cause the skill premium, ω', and then they can be rejected or not based on F statistics (Fstatistic compared to F-value and resulting p-value). Thus, in our example we assume the Granger-causality of TM_{NRA} towards the skill premium whether the coefficients estimated on the lagged TM_{NRA} in Eq. (3.26) are statistically different zero as a group.

3.5.2.4. Impulse-response function, IRF, analysis

To enrich our understanding of the interaction between the variables in our VAR specification, given that the Granger causality statistics may not tell us the complete story, we apply the IRF analysis (Lutkepohl, 2005; Neusser, 2016). The IRF allow us to examine the *response* of our dependant variable through time, the skill premium, to an *impulse* in another variable specified in our VAR representations described by Eq. (3.24) and Eq. (3.25).

Formally, let us assume that the error term ε_t, from our multivariate VAR generalization with p order and k variables represented by Eq. (3.23) can be expressed as a linear function of a vector of *shocks* represented by u_t (Cottrell & Lucchetti, 2021). If the elements of u_t have unit variance and are mutually uncorrelated, then $V(u_t) = I$. Assuming that the errors in the VAR can be expressed as $\varepsilon_t = Ku_t$, we can write $\sum = Vcov(\varepsilon_t) = KK'$. From this configuration we have the following sequence of matrices C_k[37], in the following equation:

[37] This sequence of matrices is also called the moving average representation or VMA representation. It refers to the fact that every stationary VAR process has an infinite order vector moving average representation (Cottrell & Lucchetti, 2021).

$$C_k = \frac{\partial y_t}{\partial u_{t-i}} = \Theta_k K. \tag{3.28}$$

From our VAR generalization represented by Eq. (3.23), we can derive the IRF of variable i to shock j. This IRF will be the sequence of the elements in row i and column j of the sequence of matrices C_k given by Eq. (3.28). Using the notation given by Cottrell & Lucchetti (2021), the IRF represented by symbols is:

$$\zeta_{i,j,k} = \frac{\partial y_{i,t}}{\partial u_{j,t-k}}. \tag{3.29}$$

The IRF can be plotted graphically as a function of k to observe and interpret the occurrence of transmission from one specific variable to our dependant variable of interest through time. In practical terms, the scale of the IRF plots refers to the sizing of the "shock" at one standard deviation of the estimated innovations in the variable stated as the origin of the impulse. The responses are given in units of the target variable, which in our research refers to months. Since these results are estimations of each IRFs interaction, they are endowed with confidence intervals. Our `Gretl` implementation computes these intervals using bootstrap techniques, considering the construction of an artificial dataset with resampled residuals and evaluated by repetitive sampling (Cottrell & Lucchetti, 2021). In our IRF plots analysis, we set the following: the bootstrap confidence interval at $1 - \alpha = 0.95$, 1,999 bootstrap iterations (by default value) and a forecast horizon of 24 months.

Also, in the estimation process, we compute K, which is considered a known parameter in the formula given by Eq. (3.28). Following standard procedures in the literature (see, e.g., Lutkepohl, 2005), `Gretl` estimates K as the Cholesky decomposition of $\sum = Vcov(\varepsilon_t) = KK'$ since it is assumed that K is lower triangular (Cottrell & Lucchetti, 2021). However, the Cholesky decomposition is not unique because it depends on the ordering of the variables within the vector y_t i.e., our vector of endogenous variables. This ordering is critical since K is also the matrix of IRF at lag 0, and the assumed triangularity implies that the first variable in the vector y_t responds contemporaneously only to shock number one, the second variable only to shocks one and two, and so on. Therefore, the order of our variables is meaningful where the independent variables must be placed before our target variable, the skill premium, in the variables list. As a result, the shock in the independent variables affects our target variable instantaneously, but not vice versa. In terms of interpretation, since our variables are in logs, we can say that a 1% unexpected shock or increase in an independent variable one, two, three, etc., periods back is an increase/decrease (expressed in percentage) in the skill premium today.

3.6. Results

This section aims to show our results in three main subsections. First, we detail the results from our estimation of the dependant and explanatory variables. Second and third, we present the findings of our

VAR strategy, including the Granger-causality and IRF output preceded by stationary and optimal lag order testing results, for the task-content and skill-related analysis, respectively.

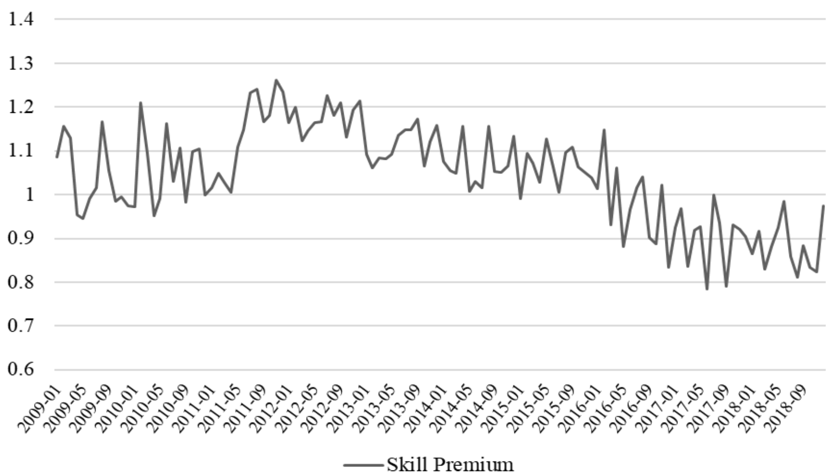
3.6.1. Estimation of variables

This section outlines the results from estimating the skill premium (see section 3.5.1.1 for construction details), the task-content of jobs and skill-related measures, following the strategies explained in section 3.5.1.2 and section 3.5.1.3, respectively.

3.6.1.1. The skill premium

Figure 3.2 displays the monthly evolution of our measure for the skill premium over 2009-2018. The skill premium shows an inverted U-shaped pattern, growing to a peak of 1.26 in November 2011 and then reducing, although with fluctuations. Over two years on average, the skill premium variable increased from 1.05 in 2009-2010 to 1.16 in 2011-2012. In turn, in 2013-2014, 2015-2016 and 2017-2018, it decreased to 1.09, 1.02 and 0.9, respectively. This pattern, composed by a reversal during most of the 2010s, has also been noted by past studies using different data sources, such as labour and households representative surveys (see, e.g., Murakami, 2014; Parro & Reyes, 2017). Also, in the first essay, we also add evidence on this skill premium evolution (see Chapter 2, section 2.7.1).

Figure 3.2. The skill premium monthly evolution Jan 2009- Dec 2018



3.6.1.2. The task-content measures

This section presents the results of our three-step strategy designed to build our task-content measures as described in section 3.5.1.2. First, we start by showing the findings of our manual classification of work activities into the categories proposed by the ALM model for 41 2-digit standard occupational groups. Secondly, we present the results obtained by applying our SVM algorithm to classify and label each job ad against the 41 2-digit occupations. Third, we outline our findings on constructing measures to represent the task content of job ads over time, focusing on the sample of job ads requiring skilled labour.

3.6.1.2.1. Estimation of the task content for the 41 2-digit occupations

Here we describe the output from the manual classification of work activities for the 41 2-digit occupational groups under analysis. On average, an occupation consists of around 20 work activities (min=5 and max= 45). We analyse and classify 848 work activities (803 unique) according to the five ALM model task categories. We build task-content percentage shares, TS, the RII, and CII indicators (See Eq. (3.4), (3.5), and (3.6), respectively) by occupational groups.

Table 3.7. shows the global distribution of tasks percentage shares across occupational groups. Each row represents one of the 41 2-digit occupational groups and columns, TS_{NRA}, TS_{NRI}, TS_{RC},TS_{RM} and TS_{NRM} depict the tasks shares of the five task types per occupational group. The score in columns TS_{NRA}, TS_{NRI}, TS_{RC},TS_{RM} and TS_{NRM} in Table 3.7 ranges between zero and one. A zero score implies that a given occupational group does not contain any work activity in that task category. Alternatively, scores equal to one show that all work activities for a given occupational group belong to a unique task category. For instance, the first row of Table 3.7 displays the distribution of task categories for the "Chief executives, senior officials, and legislators" group. Only non-routine analytical and non-routine interactive tasks feature this occupation, given the 0.54 and 0.46 scores for the TS_{NRA} and TS_{NRI} shares, respectively. In contrast, occupations such as 2-digits codes 82, 83, 91, among others, show a score equal to zero for TS_{NRA} and TS_{NRI}.

For the sake of clarity, and since we are interested in skilled occupations performing non-routine analytical and interactive tasks, TS_{NRA} and TS_{NRI}, respectively, Figure 3.3 depicts visually the global distribution shown by Table 3.7 sorted according to the higher TS_{NRA} score. The x-axis plots the 41 occupations (each bar stands for one group), and the y-axis shows the task categories' scores (each colour stands for one task category). The occupations with the two highest TS_{NRA} scores (bar's blue segment), i.e.., values over 0.75 or at least with ¾ of their task-content composed only of non-routine analytical tasks, are "ICT professionals" and "Science and engineering professionals" (the first two occupations in the x-axis). In the case of TS_{NRI} (bar's grey segment), some examples of occupations with high values are "Administrative and commercial managers" and "Production and specialized services managers" (see the second and third rows of Table 3.7). We see similar scores of TS_{NRI} for some occupations in the generic category of "associate professionals or technicians" (e.g., the 2-digit codes 35 and 36 in Table 3.7). These results are as expected since non-routine analytical and interactive work activities, such as researching, evaluating, designing, and managing, usually feature occupations performed by managers, professionals and some associate professionals or technicians. These workers are primarily highly-educated or skilled labour, given that post-secondary education provides and promotes specific knowledge and abilities, respectively. We give more insights on this relationship between non-routine analytical and interactive work activities and occupations employing skilled labour in our categorization of job ads according to the occupational classification in the next section, 3.6.1.2.2.

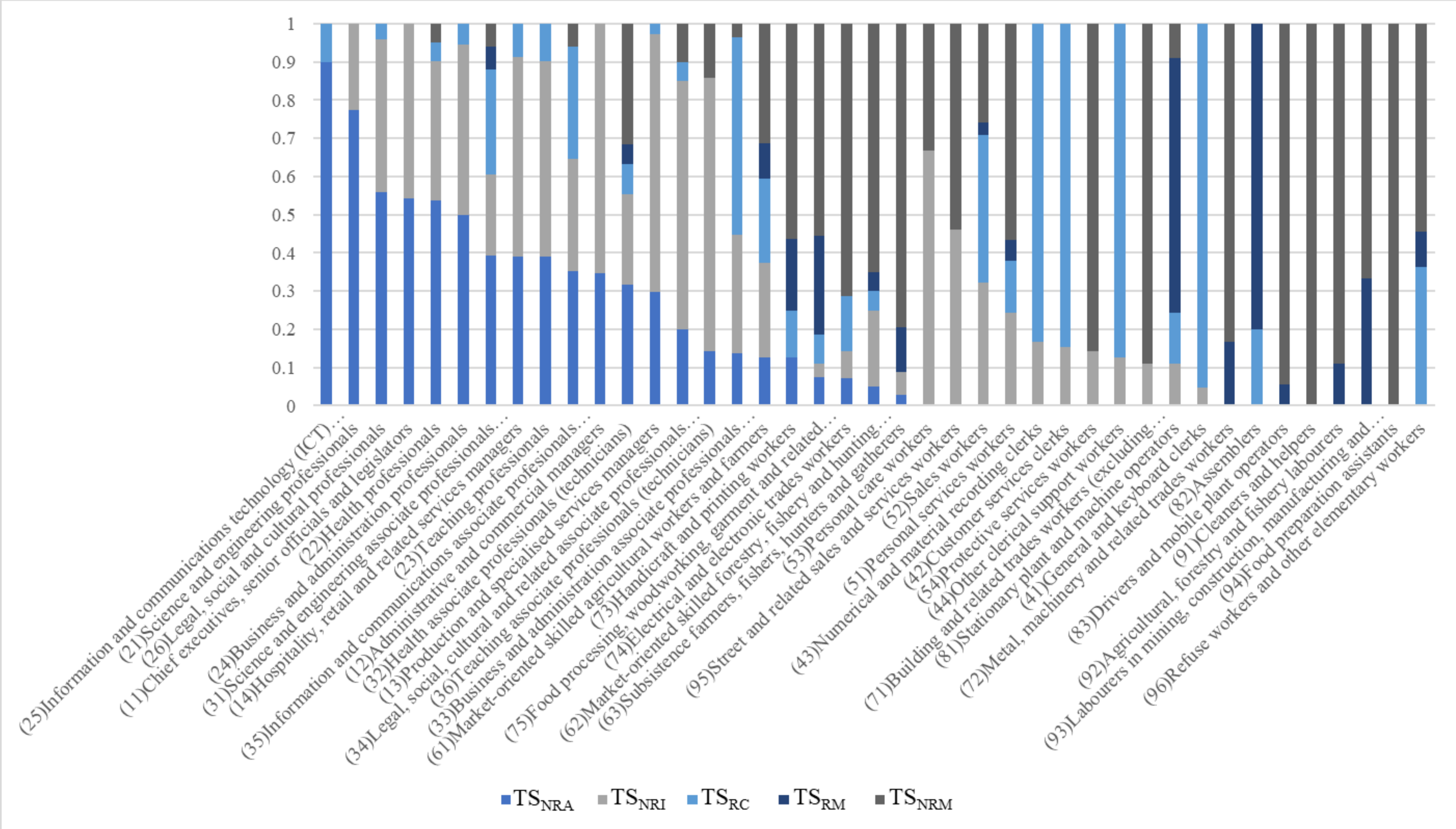
Turning to the RII and CII columns in Table 3.7, we see that ten occupations have an RII of -1 (e.g., 2-digits codes 11, 12, 21), and only “Assemblers” has an RII equal to one, which implies that these groups contain non-routine and routine tasks only, respectively. Regarding CII, 13 occupations have a CII equal to one (e.g., 11, 12, 13), which implies that these occupations consist of cognitive tasks only, routine and/or non-routine.

Table 3.7. Task-content shares for the 41 2-digit occupational groups

2-dig Occupation Code	2-dig Occupation Name	TS_{NRA}	TS_{NRI}	TS_{RC}	TS_{RM}	TS_{NRM}	RII	CII
11	Chief executives, senior officials, and legislators	0.54	0.46	0.00	0.00	0.00	-1.00	1.00
12	Administrative and commercial managers	0.35	0.65	0.00	0.00	0.00	-1.00	1.00
13	Production and specialised services managers	0.30	0.68	0.03	0.00	0.00	-0.95	1.00
14	Hospitality, retail, and related services managers	0.39	0.52	0.09	0.00	0.00	-0.83	1.00
21	Science and engineering professionals	0.77	0.23	0.00	0.00	0.00	-1.00	1.00
22	Health professionals	0.54	0.37	0.05	0.00	0.05	-0.90	0.90
23	Teaching professionals	0.39	0.51	0.10	0.00	0.00	-0.80	1.00
24	Business and administration professionals	0.50	0.44	0.06	0.00	0.00	-0.89	1.00
25	ICT professionals	0.90	0.00	0.10	0.00	0.00	-0.80	1.00
26	Legal, social, and cultural professionals	0.56	0.40	0.04	0.00	0.00	-0.92	1.00
31	Science and engineering associate professionals (technicians)	0.39	0.21	0.27	0.06	0.06	-0.33	0.76
32	Health associate professionals (technicians)	0.32	0.24	0.08	0.05	0.32	-0.74	0.26
33	Business and administration associate professionals (technicians)	0.14	0.31	0.52	0.00	0.03	0.03	0.93
34	Legal, social, cultural, and related associate professionals (technicians)	0.20	0.65	0.05	0.00	0.10	-0.90	0.80
35	ICT associate professionals (technicians)	0.35	0.29	0.29	0.00	0.06	-0.41	0.88
36	Teaching associate professionals (technicians)	0.14	0.71	0.00	0.00	0.14	-1.00	0.71
41	General and keyboard clerks	0.00	0.05	0.95	0.00	0.00	0.90	1.00
42	Customer services clerks	0.00	0.15	0.85	0.00	0.00	0.69	1.00
43	Numerical and material recording clerks	0.00	0.17	0.83	0.00	0.00	0.67	1.00
44	Other clerical support workers	0.00	0.13	0.88	0.00	0.00	0.75	1.00
51	Personal services workers	0.00	0.24	0.14	0.05	0.57	-0.62	-0.24
52	Sales workers	0.00	0.32	0.39	0.03	0.26	-0.16	0.42
53	Personal care workers	0.00	0.67	0.00	0.00	0.33	-1.00	0.33
54	Protective services workers	0.00	0.14	0.00	0.00	0.86	-1.00	-0.71
61	Market-oriented skilled agricultural workers and farmers	0.13	0.25	0.22	0.09	0.31	-0.38	0.19
62	Market-oriented skilled forestry, fishery and hunting workers	0.05	0.20	0.05	0.05	0.65	-0.80	-0.40
63	Subsistence farmers, fishers, hunters, and gatherers	0.03	0.06	0.00	0.12	0.79	-0.76	-0.82
71	Building and related trades workers (excluding electricians)	0.00	0.11	0.00	0.00	0.89	-1.00	-0.78
72	Metal, machinery, and related trades workers	0.00	0.00	0.00	0.17	0.83	-0.67	-1.00
73	Handicraft and printing workers	0.13	0.00	0.13	0.19	0.56	-0.38	-0.50
74	Electrical and electronic trades workers	0.07	0.07	0.14	0.00	0.71	-0.71	-0.43
75	Food processing, woodworking, garment, and related trades workers	0.07	0.04	0.07	0.26	0.56	-0.33	-0.63
81	Stationary plant and machine operators	0.00	0.11	0.13	0.67	0.09	0.60	-0.51
82	Assemblers	0.00	0.00	0.20	0.80	0.00	1.00	-0.60
83	Drivers and mobile plant operators	0.00	0.00	0.00	0.06	0.94	-0.89	-1.00
91	Cleaners and helpers	0.00	0.00	0.00	0.00	1.00	-1.00	-1.00
92	Agricultural, forestry and fishery labourers	0.00	0.00	0.00	0.11	0.89	-0.78	-1.00
93	Labourers in mining, construction, manufacturing, and transport	0.00	0.00	0.00	0.33	0.67	-0.33	-1.00
94	Food preparation assistants	0.00	0.00	0.00	0.00	1.00	-1.00	-1.00
95	Street and related sales and services workers	0.00	0.46	0.00	0.00	0.54	-1.00	-0.08
96	Refuse workers and other elementary workers	0.00	0.00	0.36	0.09	0.55	-0.09	-0.27

Abbreviations: TS_{NRA}= non-routine analytical tasks share, TS_{NRI}= non-routine interactive tasks share, TS_{RC}= routine cognitive tasks share, TS_{RM}= routine manual tasks share and, TS_{NRM}= non-routine manual tasks share. RII is Routine Intensity Index and CII is Cognitive Intensity Index

Figure 3.3. Distribution of task shares by occupation groups (occupation codes in parentheses).



Notes: Groups sorted from higher non-routine analytical task share, TS_{NRA} (See Table 3.7). Abbreviations: TS_{NRA}= non-routine analytical tasks share, TS_{NRI}= non-routine interactive tasks share, TS_{RC}= routine cognitive tasks share, TS_{RM}= routine manual tasks share and, TS_{NRM}= non-routine manual tasks share

3.6.1.2.2. Classification of job ads into the 41 2-digit occupational groups

This section outlines the results of our strategy of classifying our job ads sample against the 41 2-digit occupations. This strategy corresponds to the Step Two of our estimation of task-content measures. As detailed in section 3.5.1.2.2, we apply our SVM algorithm to obtain as an output a 2-digit occupation label as a new variable for each job ad.

The DTM representation of our training sample (see section 3.5.1.2.2.2.2) corresponds to a matrix of 67,656 documents and 210,689 features (unigrams and bigrams). Once we have "trained" our SVM algorithm (LinearSVC) using $C = 1$ and balanced class weights as detailed in section 3.5.1.2.2.2.1, we evaluate the SVM prediction following the metrics $precision$, $recall$ and $f1 - score$, as described by section 3.5.1.2.2.2.3. Table 3.8. displays the classification report from the results of the SVM evaluation.

Table 3.8. Classification report for the SVM (LinearSVC) application

2-dig Code	precision	recall	f1-score	N support (80% training sample)
11	0.91	0.54	0.68	143
12	0.81	0.76	0.78	571
13	0.76	0.48	0.59	219
14	0.83	0.37	0.51	94
21	0.86	0.88	0.87	3,867
22	0.93	0.97	0.95	1,418
23	0.95	0.93	0.94	533
24	0.94	0.94	0.94	8,894
25	0.89	0.88	0.88	1,661
26	0.91	0.95	0.93	1,030
31	0.9	0.86	0.88	2,589
32	0.96	0.96	0.96	805
33	0.97	0.97	0.97	12,437
34	0.86	0.8	0.83	571
35	0.86	0.92	0.88	1,653
36	0.91	0.7	0.79	211
41	0.9	0.86	0.88	3,970
42	0.84	0.87	0.85	1,420
43	0.89	0.93	0.91	2,379
44	0.93	0.86	0.89	578
51	0.92	0.86	0.89	772
52	0.91	0.96	0.93	3,748
53	0.75	0.78	0.77	99
54	0.97	0.99	0.98	902
61	0.9	0.52	0.66	86
71	0.88	0.68	0.77	117
72	0.84	0.87	0.85	885
73	0.93	0.93	0.93	75
74	0.81	0.78	0.8	497
75	0.68	0.69	0.68	106
81	0.81	0.66	0.73	479
83	0.93	0.98	0.95	634
91	0.92	0.99	0.95	530
93	0.93	0.52	0.67	50
94	0.93	0.85	0.89	102
global accuracy			0.92	54,125
macro average	0.88	0.81	0.84	54,125
weighted average	0.92	0.92	0.91	54,125

The first column in Table 3.8 refers to 35 2-digit occupational codes[38], and in the subsequent columns, we see the metrics $precision$, $recall$ and $f1 - score$ results at the occupation level. The last column shows the number of occurrences of the occupation in the training dataset. We can see that the predictive performance of SVM depends on the analysed occupational group with better results in occupations with higher representation in the sample. Overall, by observing the global evaluation of SVM in the bottom rows of Table 3.8, we see that $global\ accuracy$ is 0.92 and the $macro$ and $weighted$ averages for $precision$, $recall$ and $f1 - score$ fall between 0.81 and 0.92. These results are as expected and, in line with past studies, i.e., $global\ accuracy$ and $average\ precision$ around 85% (see e.g., Guerrero & Cabezas, 2019). In the following paragraph, we detail the output resulting from the application of our SVM algorithm on the unlabelled job posting data.

The SVM application labelled 122,330 job ads with an occupational group. This sample plus our training dataset (67,656 job ads) represents our whole dataset (189,986 job ads), as detailed in section 3.4. Our labelling procedure shows six missing occupational groups in our analysis since no job ad was distributed to any of them (see footnote 38). Consequently, our job ads sample is distributed across 35 2-digit occupational groups. In Table 3.9, we show the distribution of job ads by occupational group and year, focusing on the 19 most represented groups in the dataset (these 19 groups stand for 93% of the dataset). Two occupational groups, "Business and administration associate professionals (technicians)" and "Business and administration professionals", represent 34% of the sample. We can see at the bottom of the table those occupations that are 1% or a lower percentage of the sample (see row "Rest (16 Occupational Groups)").

Recapitulating, our measures of tasks-related measures aim to capture the distribution of task categories across skilled labour. In this regard, Table 3.10 shows the composition of our sample in terms of unskilled and skilled labour categories across occupational groups (see section 3.5.1.1 for details on how we define skilled and unskilled). Remarkably, we see in most represented occupations a clear differentiation between occupations requiring skilled or unskilled labour. Thus, in the case of occupations demanding skilled labour, they show percentages over 94% for a given occupation. As discussed in our last section 3.6.1.2.1, these results align with our expectation that most occupations filled by managers, professionals and associate professionals or technicians demand skilled labour (see, e.g., the 2-digit Code Occupations 33, 21, 31 in Table 3.10). In terms of our sample of interest to construct measures of task-content of jobs requiring skilled labour, the bottom row of Table 3.10 shows that our sample of job ads is 120,970.

[38] Unlike to the task-content analysis in section 3.6.1.2.1 examining 41 2-digit occupational groups, our training sample is composed of only 35 occupational groups. We cannot allocate job ads to any of the following six groups (code in parentheses): (62) Market-oriented skilled forestry, fishery and hunting workers, (63) Subsistence farmers, fishers, hunters and gatherers, (63) Assemblers, (92) Agricultural, forestry and fishery labourers, (95) Street and related sales and services workers, (96) Refuse workers and other elementary workers.

Table 3.9. Distribution (%) of Job ads by selected 2-digit occupations 2009-2018

2-dig Code	2- dig Name Occupation	Year										Total	
		2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	N	%
33	Business and administration associate professionals (technicians)	1,882	2,377	4,271	4,596	4,551	3,873	3,798	3,837	3,440	3,619	36,244	19.08
24	Business and administration professionals	2,098	2,838	4,315	3,759	3,157	2,737	2,761	2,410	2,413	2,792	29,280	15.41
52	Sales workers	712	1,058	2,202	2,261	2,341	1,722	1,818	1,738	1,391	1,492	16,735	8.81
21	Science and engineering professionals	854	1,422	2,370	1,966	1,522	1,283	1,599	1,467	1,201	1,442	15,126	7.96
41	General and keyboard clerks	608	941	1,570	1,575	1,511	1,192	1,100	1,121	1,001	788	11,407	6.00
42	Customer services clerks	595	593	1,517	1,398	1,234	1,868	1,602	1,217	764	510	11,298	5.95
31	Science and engineering associate professionals (technicians)	354	603	1,051	1,158	1,061	896	907	1,004	977	1,067	9,078	4.78
43	Numerical and material recording clerks	215	398	1,119	1,355	1,088	663	552	581	629	521	7,121	3.75
22	Health professionals	357	547	848	885	913	577	526	450	587	702	6,392	3.36
25	ICT professionals	604	727	1,143	790	543	447	462	550	368	540	6,174	3.25
35	ICT associate professionals (technicians)	356	401	859	811	667	534	513	649	451	478	5,719	3.01
72	Metal, machinery, and related trades workers	89	193	405	588	475	334	325	471	304	320	3,504	1.84
26	Legal, social, and cultural professionals	186	278	372	332	305	318	347	370	354	333	3,195	1.68
23	Teaching professionals	109	144	254	272	370	308	334	431	480	481	3,183	1.68
54	Protective services workers	79	233	550	485	336	218	298	321	216	188	2,924	1.54
51	Personal services workers	173	131	324	268	324	514	363	327	294	190	2,908	1.53
83	Drivers and mobile plant operators	54	109	261	327	356	275	288	364	364	283	2,681	1.41
32	Health associate professionals (technicians)	69	118	251	404	354	259	222	217	224	262	2,380	1.25
81	Stationary plant and machine operators	81	147	247	372	237	216	179	188	232	199	2,098	1.10
	Rest (16 Occupational Groups)	617	754	1,352	1,279	1,350	1,353	1,379	1,755	1,465	1,235	12,539	6.60
	Total	10,092	14,012	25,281	24,881	22,695	19,587	19,373	19,468	17,155	17,442	189,986	100

Notes: 2-dig codes of the 16 occupational groups in “Rest” category: 91,74,34,12,44,36,94,53,14,75,13,71,11,73,93, and 61.

Table 3.10. Job Ads distribution by occupations and skilled/unskilled definition

2-dig Code	2-dig Name Occupation	Unskilled		Skilled		Total
		N	%	N	%	
33	Business and administration technicians	25	0.07	36,219	99.93	36,244
24	Business and administration professionals	-	0.00	29,280	100	29,280
52	Sales workers	16,733	99.99	2	0.01	16,735
21	Science and engineering professionals	10	0.07	15,116	99.93	15,126
41	General and keyboard clerks	11,407	100	-	0.00	11,407
42	Customer services clerks	11,294	99.96	4	0.04	11,298
31	Science and engineering technicians	176	1.94	8,902	98.06	9,078
43	Numerical and material recording clerks	7,120	99.99	1	0.01	7,121
22	Health professionals	46	0.72	6,346	99.28	6,392
25	ICT professionals	21	0.34	6,153	99.66	6,174
35	ICT associate professionals (technicians)	315	5.51	5,404	94.49	5,719
72	Metal, machinery, and trades workers	3,503	99.97	1	0.03	3,504
26	Legal, social, and cultural professionals	28	0.88	3,167	99.12	3,195
23	Teaching professionals	19	0.60	3,164	99.40	3,183
54	Protective services workers	2,870	98.15	54	1.85	2,924
51	Personal services workers	2,859	98.31	49	1.69	2,908
83	Drivers and mobile plant operators	2,650	98.84	31	1.16	2,681
32	Health associate professionals (technicians)	71	2.98	2,309	97.02	2,380
81	Stationary plant and machine operators	2,098	100	-	0.00	2,098
	Rest (16 Occupational Groups)	7,771	61.97	4,768	38.03	12,539
	Total	69,016	36.33	120,970	63.67	189,986

Notes: 2-dig codes of the 16 occupational groups in “Rest” category: 91,74,34,12,44,36,94,53,14,75,13,71,11,73,93, and 61.

3.6.1.2.3. Estimation of the task-content measures

This section shows the results for estimating the task-content measures described by Eq. (3.20) (see section 3.5.1.2.3). Our sample is the 120,970 job ads requiring skilled labour (see section 3.6.1.2.2). Figure 3.4 displays the frequencies distribution for our 120 monthly data points (Jan 2009 – Dec 2018). The mean, standard deviation, min freq and max freq are 1,008.1, 247.4, 466 and 1,589, respectively. More general, the frequency of job ads under analysis in most data points (over 100) is in the range [700, 1400].

Figure 3.4. Histogram of monthly frequencies for job ads requiring skilled labour 2009-2018

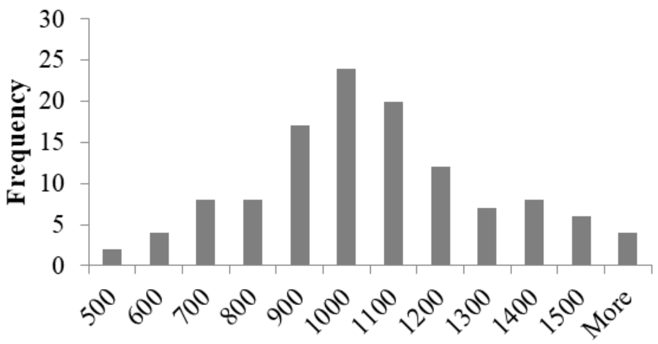
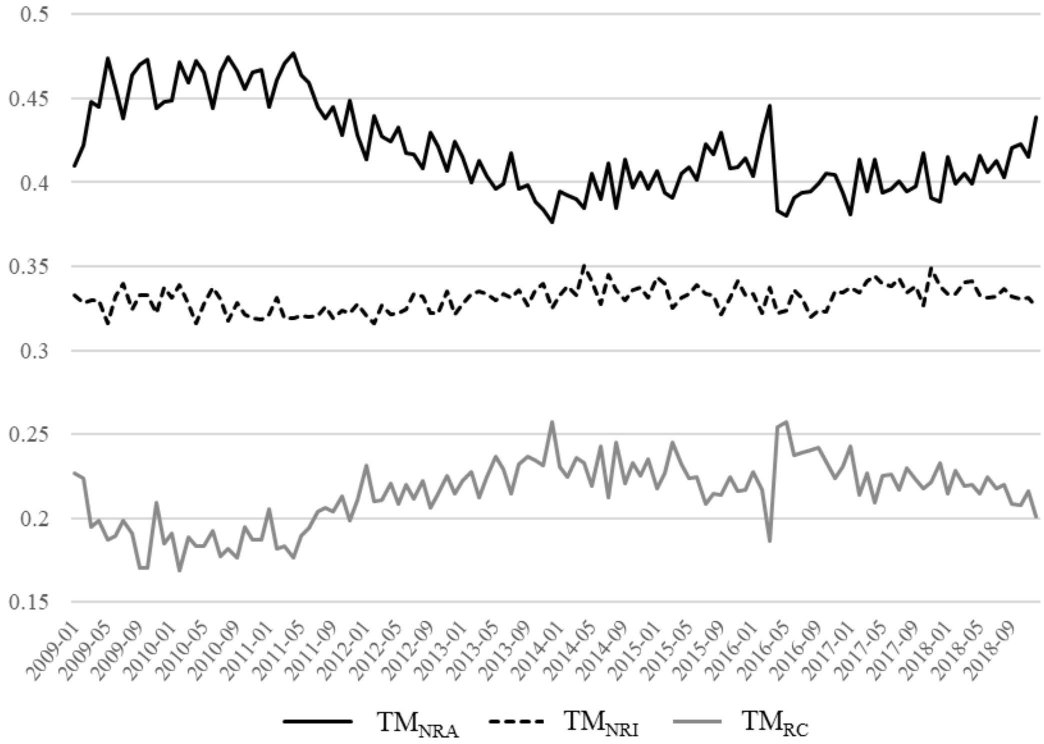


Figure 3.5 plots the series representing our task-content measures focusing on cognitive tasks. We focus on the task measures for the following ALM model categories: non-routine analytical (TM_{NRA}: solid black line on top of the plot), non-routine interactive (TM_{NRI}; black dashed line at the middle of plot) and routine cognitive (TM_{RC}; solid grey line at the bottom of the plot).

Figure 3.5. Task-content measures time series 2009-2018



Note: TM_{NRA}, TM_{NRI} and TM_{RC} stand for non-routine analytical, non-routine interactive and routine cognitive tasks, respectively.

In Figure 3.5, the TM_{NRA} series indicates that the intensity ratio of non-routine analytical tasks required by job ads for skilled labour fluctuates between 0.38 and 0.48 over the period. This measure shows an initial steady pattern, and then it decreases to grow up again, although with fluctuations. We also see a fluctuating pattern for the TM_{RC} measure between 0.17 and 0.2 6 but starting with an increasing trend and then a steady pattern. This measure shows that the intensity ratio of routine cognitive interactive task-content of jobs demanding skilled labour fluctuates. TM_{NRI} shows a narrower range, i.e. between 0.32 and 0.35, compared to the rest of the series. This pattern implies that the intensity of non-routine interactive tasks in job ads demanding skilled labour stays stable over the period.

3.6.1.3. The skills-related measures

This section outlines the output of the categorization of job ads according to the skills categories and the computation of the skills-related measures. See section 3.5.1.3 for construction details.

3.6.1.3.1. Classification of job ads according to the skills categories

Applying our skills categories dictionary to the job ads dataset resulted in 137,647 job ads categorized according to the ten proposed skills categories (see Table 3.6). Our classification relies on the frequency of keywords or phrases found in the job ads text across skills categories. We allocate each job ad in the most prevalent category, i.e., the skills category with the highest frequency of found keywords. However, some job ads have the same frequency of keywords for different categories, implying that we cannot decide which category is most prevalent. Therefore, we discard these observations and our sample, characterised by one prevalent skill category, results in 124,973 job ads.

We present the distribution of our 124,973 job ads sample across the ten skills categories over the years in Table 3.11. Globally, “Customer service” is the most represented category with over 53% of the sample, followed by “Financial” (15.54%) and “Social” (7.71%). The less represented categories are “Project management” and “Writing”, 0.55% and 0.45%, respectively, of our sample. Over time, although with fluctuations, the number of job ads requiring each category increases.

Regarding our measure to identify job ads demanding cognitive and social skills simultaneously, our sample consists of 7,905 observations. We display its distribution over the years in Table 3.12. The number of job ads requiring cognitive and social skills started with 358 in 2009 and reached a peak in 2011 with 1,031. Then it decreased to 727 in 2017, but it increased gain in 2018 to 832 observations.

Also, since we are interested in the skill categories across skilled labour, in Table 3.13, we show the distribution of our sample in terms of unskilled and skilled labour. We see that eight of the ten proposed skills categories are mostly demanding skilled labour (at least 70%). The first two rows show that Cognitive and Social represent over 73% and 74%, respectively, of job ads demanding skilled workers. This result is in line with our expectations of positions rich in cognitive (e.g., reasoning, evaluating) and social (e.g., communication ability) skills employing skilled labour. Our sample of job ads requiring skilled labour used to construct the skill-related metrics is 83,202, as shown in the bottom row of the Skilled column in Table 3.13.

Table 3.14 presents the job ads requiring simultaneously Cognitive and Social skills under the skilled-unskilled differentiation. We observe that over 80% of job ads requiring both Cognitive and Social skills demand skilled labour. In this regard, we construct our metric using the 6,401 job ads sample as shown by the Skilled column in Table 3.14.

Table 3.11. Job ads distribution (%) by skills category 2009-2018

Job Skills Categories	Year										Total	
	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	N	%
Cognitive	238	434	726	576	603	507	464	504	564	517	5,133	4.11
Social	497	727	1,151	1,192	1,108	908	1,013	961	999	1,076	9,632	7.71
Character	100	153	236	398	382	377	319	500	231	298	2,994	2.40
Writing	26	43	79	62	52	71	62	72	59	33	559	0.45
Customer service	3,714	4,702	8,730	8,782	7,590	7,680	7,240	6,652	5,435	5,923	66,448	53.17
Project management	28	50	81	77	67	75	63	75	91	75	682	0.55
People management	353	583	979	1,023	1,216	900	792	678	564	627	7,715	6.17
Financial	1,262	1,620	2,557	2,560	2,153	1,868	1,897	1,822	1,803	1,875	19,417	15.54
Computer (general)	321	451	752	708	627	467	448	492	477	496	5,239	4.19
Software (specific)	511	734	1,183	951	708	554	610	718	555	630	7,154	5.72
Total	7,050	9,497	16,474	16,329	14,506	13,407	12,908	12,474	10,778	11,550	124,973	100

Table 3.12. Distribution of jobs ads requiring Cognitive and Social skills simultaneously by year

	year										
	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
Job Ads with both Cognitive and Social skills	358	616	1031	993	853	871	876	748	727	832	7,905

Table 3.13. Distribution by skilled

Skills Categories	Unskilled		Skilled		Total
	N	%	N	%	
Cognitive	1,377	26.83	3,756	73.17	5,133
Social	2,468	25.62	7,164	74.38	9,632
Character	1,517	50.67	1,477	49.33	2,994
Writing	154	27.55	405	72.45	559
Customer service	28,475	42.85	37,973	57.15	66,448
Project management	7	1.03	675	98.97	682
People management	2,471	32.03	5,244	67.97	7,715
Financial	3,045	15.68	16,372	84.32	19,417
Computer (general)	1,541	29.41	3,698	70.59	5,239
Software (specific)	616	8.61	6,538	91.39	7,154
Total	41,671	33.34	83,302	66.66	124,973

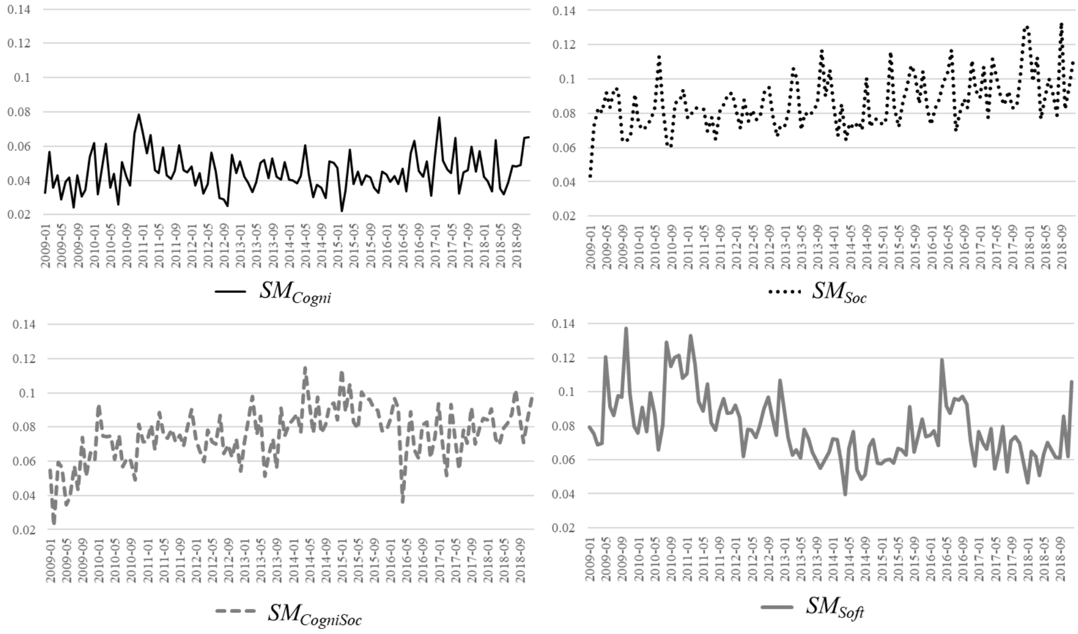
Table 3.14. Distribution job ads Cognitive social by skilled

Job Ads requiring Cognitive and Social skills	Unskilled		Skilled		Total
	N	%	N	%	
	1,504	19.03	6,401	80.97	7,905

3.6.1.3.2. Estimation of the skills-related measures

This section outlines the results for estimating our skill-related measures described by Eq. (3.21) (see section 3.5.1.3). Our metrics SM_{Cogni} SM_{Soc}, $SM_{CogniSoc}$ and, SM_{Soft} capture the prevalence of job ads requiring cognitive, social, cognitive and social simultaneously, and software skills, respectively, relative to the total frequency of job ads demanding skilled labour. We compute these metrics monthly from January 2009 to December 2018, resulting in the series displayed in Figure 3.6. In the top left-hand plot, we see the SM_{Cogni} variable, which shows a fluctuating pattern over the period without a clear increasing or decreasing trend. In the case of the SM_{Soc} variable, top right-hand plot, there is an increasing pattern over time, although with high fluctuations. The evolution of our metric $SM_{CogniSoc}$, is shown in the bottom left-hand plot. We can observe an initial increasing tend until a significant decline in the first half of 2016, followed by a recovery path. The bottom right-hand plot shows the evolution of the SM_{Soft} which fluctuates over time, showing similar values at the start and end of the period.

Figure 3.6. Skills-related measures. Monthly time series from Jan-2009 to Dec-2018.



Note: SM_{Cogni} SM_{Soc}, $SM_{CogniSoc}$ and SM_{Soft} are our metrics standing for Cognitive, Social, Cognitive & Social, and Software skills, respectively.

3.6.2. Task-content VAR analysis

This section aims to show the results of our VAR econometric estimation as described in section 3.5.2 using the estimated skill premium and task-related metrics, TM_{NRA}, TM_{NRI} and TM_{RC}, as presented in sections 3.6.1.1 and 3.6.1.2, respectively. First, we outline the results from stationarity, optimal lag order and VAR parameters estimation as detailed in sections 3.5.2.1 and 3.5.2.2. Secondly, we present the Granger-causality testing and IRF results (See sections 3.5.2.3 and 3.5.2.4, respectively).

3.6.2.1. Stationarity and optimal lag order testing and VAR estimation

We determine the presence of unit roots and stationarity by applying the ADF and KPSS tests, respectively, to all our endogenous variables (the skill premium, TM_{NRA}, TM_{NRI} and TM_{RC}) individually. First, we analyse the ADF results displayed in Table 3.15. Columns display the variable names, modelling case (constant, “C”, constant and linear trend, “C, T”, and adding a quadratic trend, “C, T, TT”). Next, seasonal dummies addition, lag order (selection using criterion BIC with max order=12 given our monthly data) and test-statistics and related level of significance. ADF results show that we cannot reject the null of unit roots when modelling includes only a constant and the linear trend (the “C, T” case rows in column “Modelling case”) for three of our variables (the skill premium, TM_{NRA}, and TM_{RC}). In the case of TM_{NRI}, we reject the null of unit roots for the “C, T” modelling case at 1% of significance. These results are consistent with the inclusion or not of seasonal dummies. The *modelling case* with a quadratic trend shows that we can reject the null of unit roots at the 1% of significance for all endogenous variables excepting TM_{NRA} variable (rejection is at 10% of significance). These ADF results with a quadratic trend are obtained with and without seasonal dummies. Overall, our ADF output implies that our endogenous series are stationary by detrending the series using a linear and a quadratic time trend.

Table 3.15. ADF results.

Variable	Modelling case	Seasonal dummies	Lag order	test-statistic	p-value
$\ln skill\ premium$	C, T	No	2	-2.4978	0.3292
	C, T, TT		2	-4.7600	0.0025*
	C, T	Yes	2	-2.1447	0.5201
	C, T, TT		2	-4.4575	0.0074*
$\ln TM_{NRA}$	C, T	No	2	-1.9262	0.6406
	C, T, TT		1	-5.1071	0.0006*
	C, T	Yes	2	-1.5797	0.8012
	C, T, TT		2	-3.6847	0.0729***
$\ln TM_{NRI}$	C, T	No	0	-8.0841	<0.001*
	C, T, TT		0	-8.0323	<0.001*
	C, T	Yes	0	-7.5479	<0.001*
	C, T, TT		0	-7.4966	<0.001*
$\ln TM_{RC}$	C, T	No	3	-1.6561	0.7706
	C, T, TT		0	-7.9026	<0.001*
	C, T	Yes	2	-2.055	0.5704
	C, T, TT		0	-7.0221	<0.001*

Note: ADF H_0 = the series has a unit root. Lag order selection using criterion BIC (max was 12). (*), (**) and (***) denotes a rejection of H_0 at 1%, 5% and 10% significance level, respectively. TM_{NRA}, TM_{NRI} and TM_{RC} stand for non-routine analytical, non-routine interactive and routine cognitive tasks, respectively.

With regard to KPSS stationarity testing, we show these results in Table 3.16. We consider the modelling case of a constant plus a linear trend. Columns display the variable names, use of seasonal dummies, lag order (the same as in the ADF test, i.e., selection using criterion BIC with max order=12 given our monthly data) and test-statistics and related level of significance. We reject the null of stationarity at 1% of the significance level for all the endogenous variables. These results are robust to the inclusion of seasonal dummies. We confirm our ADF results for the same modelling case, i.e., constant plus linear trend or “C, T”. We cannot compare the case “C, T, TT” since our KPSS implementation test the hypothesis of stationarity only around a linear trend (See section 3.5.2.1 for details).

Table 3.16. KPSS test results (the modelling case specifies a constant plus a linear trend)

Variable	Seasonal dummies	Lag order	test-statistic	p-value
$\ln skill\ premium$	No	2	0.7248	<0.01*
	Yes	2	0.7453	<0.01*
$\ln TM_{NRA}$	No	2	0.5622	<0.01*
	Yes	2	0.5746	<0.01*
$\ln TM_{NRI}$	No	0	0.2778	<0.01*
	Yes	0	0.2933	<0.01*
$\ln TM_{RC}$	No	3	0.4548	<0.01*
	Yes	2	0.4611	<0.01*

Note: KPSS H_0 =the series is stationary. Lag order is the same as in ADF (see Table 3.15). p-values as in `Gretl` output. (*), (**) and (***) denotes a rejection of H_0 at 1%, 5% and 10% significance level, respectively. TM_{NRA}, TM_{NRI} and TM_{RC} stand for non-routine analytical, non-routine interactive and routine cognitive tasks, respectively.

The stationarity results discussed above show that our variables are stationary around linear and quadratic trends. Therefore, we follow this modelling strategy in our VAR specification and estimation.

Regarding the optimal lag order testing, Table 3.17 shows the results for our VAR representation for our task-content analysis derived for the computation of BIC and HQC information criteria values.

Table 3.17 Optimal lag order for the VAR

Lags	BIC				HQC			
	Seasonal dummies				Seasonal dummies			
	No		Yes		No		Yes	
1	-17.016	*	-15.668	*	-17.429	*	-16.731	*
2	-16.644		-15.269		-17.294		-16.568	
3	-16.296		-15.040		-17.182		-16.576	
4	-15.792		-14.579		-16.914		-16.350	
5	-15.397		-14.205		-16.755		-16.213	
6	-14.905		-13.771		-16.499		-16.016	
7	-14.631		-13.537		-16.461		-16.018	
8	-14.215		-13.219		-16.282		-15.935	
9	-13.704		-12.844		-16.007		-15.797	
10	-13.382		-12.554		-15.922		-15.743	
11	-13.098		-12.250		-15.874		-15.675	
12	-12.592		-11.905		-15.604		-15.567	

Note: Results estimated from VAR systems of order 1 to max. lag order 12. The asterisks indicate the best lag order, that is, the minimized values of the respective information criteria. VAR model with constant, linear and quadratic trends and our four endogenous variables (the skill premium and the task-related measures).

Given our results from stationarity testing above, the tested VAR models from 1st to 12th lag order included constant, linear, and quadratic trends to estimate the best lag order VAR. In Table 3.17 we display the results differentiating because of the addition of seasonal dummies. Our results indicate that the optimal number of lags to include is one, based on the minimized values of the respective information criteria. However, more importantly, the addition of all seasonal dummies worsened the BIC and HQC values, i.e., a bigger penalization due to the increased number of parameters. In this regard, in our VAR estimation, we assess the addition only of the 12th lag to control the potential seasonality since the nature of our labour data can be highly seasonal alongside the first lag order following our BIC and HQC results.

Recapitulating from our stationarity and optimal lag order results above, our VAR representing the interactions between our four endogenous variables includes linear and quadratic trends and the first lag order variables. In our estimation process, we also test if the 12th variable lag (to control potential seasonality) favours the specification fitting. However, our testing cannot reject the null hypothesis that these regression parameters are zero for the 12th lag variables[39]. Therefore, we remove the 12th variable lags, which implies the estimation of a VAR with only the first lag or a VAR (1).

The results of our VAR estimation for the equation with the skill premium, ω, as the target variable are displayed in Table 3.18. In the first three rows, we can see that the constant, linear and quadratic

[39] We use the option given by `Gretl` to perform this test after the VAR estimation including the 12th lag. The Wald test statistics result was Chi-square = 14.7055 and p-value>0.1 (0.538628).

trend time influence the skill premium at 10%, 1% and 1% significance levels, respectively. We do not observe influence from the lagged skill premium. Regarding our lagged task-content measures, TM_{NRI} shows a positive and significant coefficient at a 5% level, and both TM_{NRA} and TM_{RC} are also positive but significant at 10%. The exposed results allow us to evaluate the dependency between variables. However, these results do not necessarily imply causality or infer how the skill premium responds to shocks in the task-content variables. Hence, we apply the Granger-causality to analyse if the explanatory variables *Granger-causes* the skill premium and IRF analysis to examine the *response* of the skill premium to an *impulse* in another variable. The results appear in the next section.

Table 3.18. VAR estimation, lag order 1. OLS estimates, observations 2009:02-2018:12 (T=119). Results for equation with the logged skill premium as the target variable. See Eq. (3.26)

Parameter	*Coefficient*	*Std. Error*	*t-ratio*	*p-value*
Constant	3.9656	2.0348	1.9490	0.0538***
Time	0.0046	0.0011	4.2140	5.11e-05*
Time^2	−4.99e-05	0.0000	−5.698	9.96e-08*
Skill premium, ω_{t-1}	0.1477	0.0964	1.5320	0.1282
$TM_{NRA,t-1}$	1.4507	0.7957	1.8230	0.0709***
$TM_{NRI,t-1}$	1.3069	0.6540	1.9980	0.0481**
$TM_{RC,t-1}$	0.8218	0.4518	1.8190	0.0716***
R^2	0.64			

Note: Recalling from Eq. (3.24) TM_{NRA}, TM_{NRI} and TM_{RC} stand for non-routine analytical, non-routine interactive and routine cognitive task-content, respectively. (*), (**) and (***) denote a rejection of H_0: the regression parameter is zero at 1%, 5% and 10% significance level, respectively. All the variables, except time and time^2, are in natural logs.

3.6.2.2. Granger-causality testing and IRF results

Table 3.19 outlines the Granger-causality testing results. The first column shows the stated null hypotheses and the results of F-statistic evaluation and significance level for the one lag and two lag models. We have included the results with an additional lag to show that a redundant lag still gives similar results but with some *p-value* changes. The TM_{NRA} variable shows significance at 10% level in both lag orders, while the TM_{NRI} is significant at 5% in both models. In the case of TM_{RC} our results show significance only in the lag one model (at 10% significance level). From these results, we assume the Granger-causality of our task-content measures towards the skill premium only for the TM_{NRA} and TM_{NRI} variables at 10% and 5% of significance level, respectively.

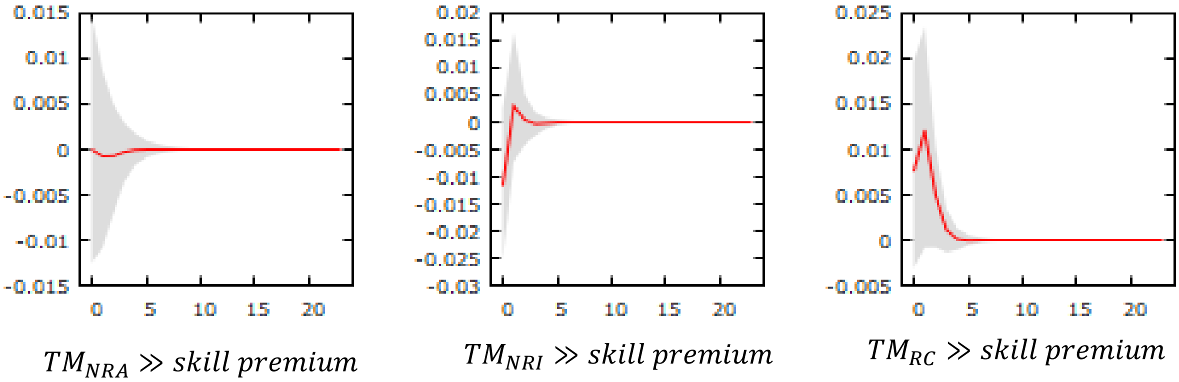
Table 3.19. Granger-causality testing results

Null hypothesis	Lag Order 1 (N=112)		Lag order 2 (N=107)	
	F Statistic	*p-value*	*F Statistic*	*p-value*
All lags of TM_{NRA} do not Granger-cause ω	3.3243	0.0709***	2.5586	0.0821***
All lags of TM_{NRI} do not Granger-cause ω	3.9939	0.0481**	4.4575	0.0138**
All lags of TM_{RC} do not Granger-cause ω	3.3079	0.0716***	2.3304	0.1022

Note: (*), (**) and (***) denote a rejection of H_0 at 1%, 5% and 10% significance level, respectively. ω is the skill premium and, TM_{NRA}, TM_{NRI} and TM_{RC}stand for non-routine analytical, non-routine interactive and routine cognitive task-content, respectively.

Related to our IRF analysis, Figure 3.7 displays the IRF plots where the scale refers to the sizing of the "shock" at one standard deviation of the estimated innovations in the variable stated as the origin of the impulse. In the list specification of our analysis, we have placed our explaining variables, TM_{NRA}, TM_{NRI} and TM_{RC}, before the skill premium (see section 3.5.2.4 for details on the importance of ordering). Since our variables are in logs, we can say that a 1% unexpected shock or increase in a task-content variable one, two, three, etc., periods back is an increase/decrease (expressed in percentage) in the skill premium today. In this regard, the left-hand plot suggests that a 1% unexpected increase in TM_{NRA} around 2-3 months back is a negligible decline in the skill premium today. In the case of TM_{NRI} and TM_{RC}, middle and right-hand plots, respectively, we see negligible increases as the response of the skill premium.

Figure 3.7. IRF plots for the response of the skill premium to a shock in TM_{NRA} (left-hand plot), TM_{NRI} (middle plot) and TM_{RC} (right-hand plot)



Note: TM_{NRA}, TM_{NRI} and TM_{RC}stand for non-routine analytical, non-routine interactive and routine cognitive task-content, respectively

3.6.3. Skills-related VAR analysis

This section outlines, firstly, the results from stationarity, optimal lag order, VAR estimation, Granger-causality testing and impulse-response analysis following steps described in section 3.5.2 using the estimated skill premium and skills-related metrics, SM_{Soc}, $SM_{CogniSoc,}$ and SM_{Soft}, presented in sections 3.6.1.1 and 3.6.1.3, respectively.

3.6.3.1. Stationarity and optimal lag order testing, and VAR estimation

Table 3.20 displays our ADF results for stationarity testing. The columns refer to the variable names, i.e., the skill premium, SM_{Cogni} SM_{Soc}, $SM_{CogniSoc}$, and SM_{Soft}, modelling case (with constant, C, constant and linear trend, “C, T”, and adding a quadratic trend, “C, T, TT”), seasonal dummies addition, lag order (selection using criterion BIC with max order=12 given our monthly data), test-statistics and the level of significance. Our results show that we cannot reject the null of the presence of unit roots using only a constant and the linear trend (the “C, T” case rows in the modelling case column) in the case of the skill premium. However, adding a quadratic trend (the “C, T, TT” case), we can reject the null of unit roots at a 1% significance level, implying stationarity. The rest of our endogenous variables shows stationarity detrending them just with the linear trend. These results are consistent throughout the addition or not of seasonal dummies in the ADF testing.

Table 3.20. ADF results

Variable	Modelling case	Seasonals	Lag order	test-statistic	p-value	
$\ln skill\ premium$	C, T	No	2	-2.4978	0.3292	
	C, T, TT		2	-4.7600	0.0025	*
	C, T	Yes	2	-2.1447	0.5201	
	C, T, TT		2	-4.4575	0.0074	*
$\ln SM_{Cogni}$	C, T	No	0	-9.094	9.46e-012	*
	C, T, TT		0	-9.1273	5.897e-011	*
	C, T	Yes	0	-8.8136	3.144e-011	*
	C, T, TT		0	-8.8652	1.822e-010	*
$\ln SM_{Soc}$	C, T	No	0	-9.5481	1.435e-012	*
	C, T, TT		0	-9.7281	4.77e-012	*
	C, T	Yes	0	-8.8467	2.727e-011	*
	C, T, TT		0	-9.0224	9.244e-011	*
$\ln SM_{CogniSoc}$	C, T	No	0	-8.2854	3.163e-010	*
	C, T, TT		0	-9.2921	2.926e-011	*
	C, T	Yes	0	-7.7795	3.001e-009	*
	C, T, TT		0	-8.6838	4.013e-010	*
$\ln SM_{Soft}$	C, T	No	2	-3.1754	0.08935	***
	C, T, TT		2	-3.5838	0.09304	***
	C, T	Yes	2	-2.9297	0.1529	
	C, T, TT		2	-3.3167	0.1667	

Note: ADF H_0 = the series has a unit root. Lag order selection using criterion BIC (max was 12). (*), (**) and (***) denotes a rejection of H_0 at 1%, 5% and 10% significance level, respectively. SM_{Cogni} SM_{Soc}, $SM_{CogniSoc}$ and SM_{Soft} are our metrics standing for Cognitive, Social, Cognitive & Social, and Software skills, respectively.

Regarding the KPSS testing results, in Table 3.21, we see similar results of our ADF testing, i.e., rejecting stationarity when the series includes a constant and a linear trend, except for SM_{Soc}. The result for this variable shows that it is stationary in levels. As noted earlier, we cannot compare the case with quadratic trends between ADF and KPSS since our KPSS implementation tests only the linear trend case.

Table 3.21. KPSS results (the modelling case specifies a constant plus a linear trend)

Variable	Seasonal dummies	Lag order	test-statistic	p-value
$\ln skill\ premium$	No	2	0.7248	<0.01*
	Yes	2	0.7453	<0.01*
$\ln SM_{Cogni}$	No	0	0.1518	0.048**
	Yes	0	0.1658	0.040**
$\ln SM_{Soc}$	No	0	0.0736	>0.10
	Yes	0	0.0783	>0.10
$\ln SM_{CogniSoc}$	No	0	0.4999	<0.01*
	Yes	0	0.5256	<0.01*
$\ln SM_{Soft}$	No	2	0.2552	<0.01*
	Yes	2	0.2704	<0.01*

Note: KPSS H_0 =the series is stationary. Lag order as in ADF testing above (see Table 3.20). p-values as in `Gretl` output. (*), (**) and (***) denote a rejection of H_0 at 1%, 5% and 10% significance level, respectively. SM_{Cogni} SM_{Soc}, $SM_{CogniSoc}$ and SM_{Soft} are our metrics standing for Cogntive, Social, Cognitive & Social, and Software skills, respectively.

Table 3.22 displays the output for our optimal lag order analysis. Our results indicate that the optimal number of lags to include is one, based on the minimized values of the respective information criteria. Remarkably, adding all seasonal dummies in the VAR representation shows a deterioration in the BIC and HQC values, i.e., a bigger penalization due to an increase in the number of parameters. Therefore, we assess the addition only of the 12th lag to control the potential seasonality in our VAR estimation.

Table 3.22 Optimal lag order for the VAR using the BIC and HQC information criteria.

Lags	BIC		HQC	
	Seasonal dummies		Seasonal dummies	
	No	Yes	No	Yes
1	-3.512*	-1.743*	-4.103*	-3.146*
2	-2.692	-0.940	-3.652	-2.712
3	-2.209	-0.648	-3.538	-2.789
4	-1.470	0.113	-3.168	-2.397
5	-0.642	0.888	-2.709	-1.992
6	-0.035	1.525	-2.471	-1.723
7	0.727	2.244	-2.078	-1.374
8	1.374	2.851	-1.801	-1.136
9	2.114	3.554	-1.430	-0.802
10	2.513	3.860	-1.400	-0.864
11	3.038	4.113	-1.244	-0.981
12	3.568	4.385	-1.083	-1.078

Note: Model with constant, linear and quadratic trends and five endogenous variables (the skill premium and the four skill-related measures). Results estimated from VAR systems of max. lag order 12. The asterisks indicate the best (that is, minimized) values of the respective information criteria.

Given the results from our stationarity and lag order testing, we estimate our VAR considering the addition of a constant, trend and quadratic trend. We also add the 12th lag to control the potential seasonality in our data[40]. Table 3.23 presents the results of our skill-related VAR specification for the equation with the skill premium. We see only an influence of the 12th lagged SM_{Cogni} and the 1st lagged $SM_{CogniSoc}$ variables on the skill premium at 10% and 5% of significance level, respectively. Therefore, we do not observe influence from the rest of the endogenous variables. The following section shows the Granger-causality and IRF results from the VAR parameters estimation discussed here.

[40] We test the inclusion of this 12th lag in our estimation process. Our result based on the Wald test shows that the inclusion of the 12th variables lag favours the specification fitting since we reject the null hypothesis that the 12th lag variables regression parameters are zero at 1% of significance level (*p-value* = 0.00031).

Table 3.23. VAR estimation, lag order 1 and 12. OLS estimates, observations 2010:01-2018:12 (T=108). Results for equation with the logged skill premium as the target variable. See Eq. (3.27)

Parameter	*Coefficient*	*Std. Error*	*t-ratio*	*p-value*	
Constant	0.2916	0.2745	1.0620	0.2908	
Time	0.0029	0.0016	1.8370	0.0693	***
Time^2	−3.675e-05	0.0000	−2.933	0.0042	*
ω_{t-1}	0.0696	0.1018	0.6838	0.4958	
ω_{t-12}	0.1841	0.1004	1.8330	0.0700	***
$SM_{Cogni,t-1}$	−0.0381	0.0284	−1.343	0.1824	
$SM_{Cogni,t-12}$	0.0458	0.0272	1.6820	0.0958	***
$SM_{Soc,t-1}$	−0.0619	0.0440	−1.407	0.1626	
$SM_{Soc,t-12}$	0.0424	0.0411	1.0310	0.3052	
$SM_{CogSoc,t-1}$	0.0840	0.0414	2.0260	0.0456	**
$SM_{CogSoc,t-12}$	−0.0180	0.0342	−0.5272	0.5993	
$SM_{Soft,t-1}$	−0.0079	0.0359	−0.2207	0.8258	
$SM_{Soft,t-12}$	0.0541	0.0354	1.5290	0.1297	
R^2	0.70				

Note: Recalling Eq. (3.27), ω is the skill premium, and SM_{Cogni} SM_{Soc}, $SM_{CogniSoc}$ and SM_{Soft} are our metrics standing for Cognitive, Social, Cognitive & Social, and Software skills, respectively. All variables except the time terms are in Natural Logs. (*), (**) and (***) denote a rejection of H_0: the regression parameter is zero at 1%, 5% and 10% significance level, respectively. All the variables, except time and time^2, are in natural logs.

3.6.3.2. Granger-causality testing and IRF estimation

Table 3.24 outlines the Granger-causality testing results for our VAR skills-related. We detail the stated null hypotheses, the results of F-statistic evaluation and *p-values* for the one lag model and the redundant model with two lags. Our results show that we cannot reject any of the stated hypotheses. Therefore, we cannot assume some Granger-causality of our skills-related measures towards the skill premium.

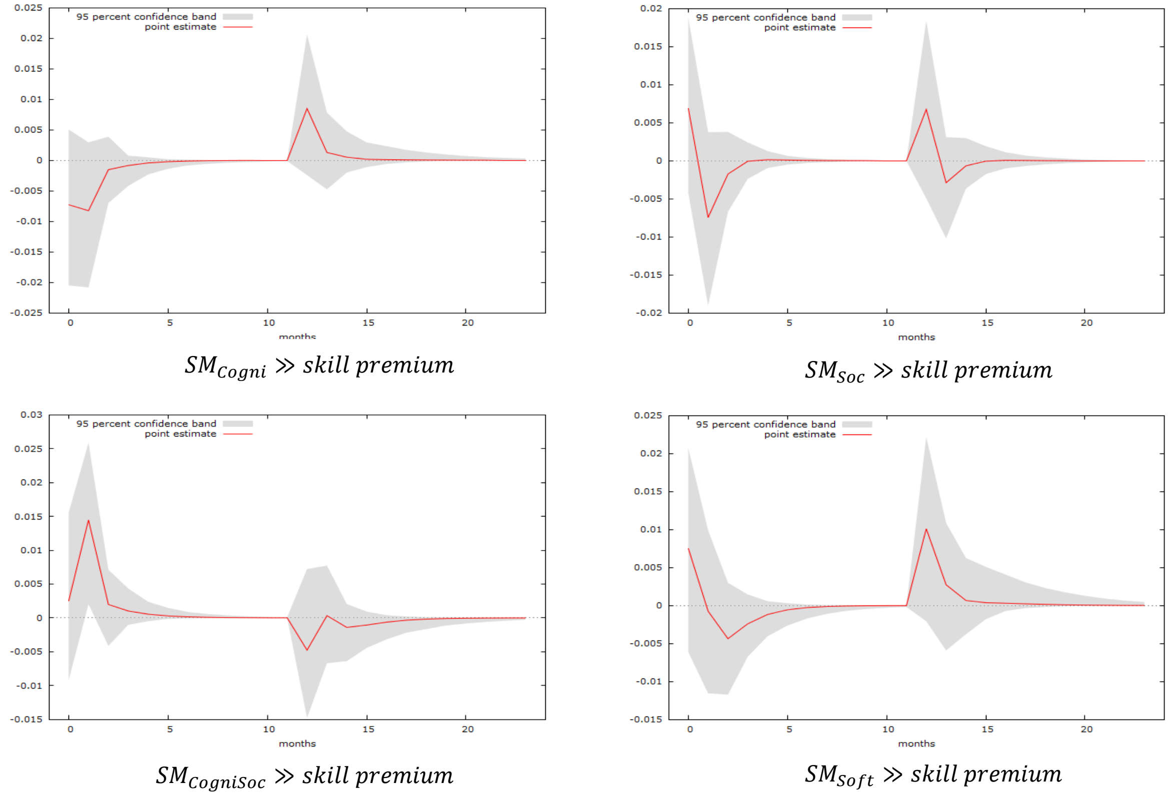
Table 3.24. Granger-causality testing results.

Null hypothesis	Lag Order 1,12 (N=95)		Lag order 2,12 (N=95)	
	F Statistic	*p-value*	*F Statistic*	*p-value*
All lags of SM_{Cogni} do not Granger-cause ω	2.0503	0.1343	1.3931	0.2533
All lags of SM_{Soc} do not Granger-cause ω	1.5051	0.2272	0.67135	0.5134
All lags of $SM_{CogniSoc}$ do not Granger-cause ω	2.1575	0.1212	0.10295	0.9023
All lags of SM_{Soft} do not Granger-cause ω	1.1701	0.3148	0.63244	0.5335

Note: (*), (**) and (***) denote a rejection of H_0 at 1%, 5% and 10% significance level, respectively. ω is the skill premium and SM_{Cogni} SM_{Soc}, $SM_{CogniSoc}$ and SM_{Soft} are our metrics standing for Cogntive, Social, Cognitive & Social, and Software skills, respectively.

Figure 3.8 displays the results of our IRF analysis. Each plot evaluates the skill premium responses to the “shock” at one standard deviation in our skills-related variables. In our specification, we have placed our skills-related variables before the skill premium (see section 3.5.2.4 on the importance of ordering). Because our variables are in logs, we can say that a 1% unexpected shock in a skill-related variable one, two, etc., periods (months) back is a contemporaneous increase/decrease (expressed in percentage) in the skill premium.

Figure 3.8. IRF plots for the skill-related analysis. Response of the skill premium to a shock in SM_{Cogni} (top left), SM_{Soc} (top right), $SM_{CogniSoc}$ (bottom left) and SM_{Soft}(bottom right).



Note: SM_{Cogni} SM_{Soc}, $SM_{CogniSoc}$ and SM_{Soft} are our metrics standing for Cogntive, Social, Cognitive & Social, and Software skills, respectively.

The top left plot in Figure 3.8 suggests that a 1% unexpected increase in SM_{Cogni} has a small negative effect that then disappears before having a small positive effect in the twelfth period again because of the 12th lag included in our VAR specification (see Table 3.23). Similar results appear for SM_{Soc} and SM_{Soft}, top right and bottom right plots, respectively. In the case of $SM_{CogniSoc}$, , bottom left side plot, the skill premium response to a 1% unexpected increase in this variable is a negligible increase at the beginning that then disappears, followed by a small negative effect in the twelfth period.

3.7.Discussion

This section discusses the results outlined in the last section throughout four sub-sections. First, we examine our findings from estimating measures representing the skill premium, the task-content of jobs and the skills required by jobs using job ads postings as data. Secondly and thirdly, we examine the VAR task-content and skill-related outputs, respectively. Fourthly, we give a general discussion, emphasizing policy issues and further research.

3.7.1. On the estimation of the skill premium and task-content and skills-related measures

The skill premium for Chile shows a decline in most of the 2009-2018 period. We estimate a peak of 1.26 in November 2011, then reducing, although with fluctuations. This pattern is consistent with previous works analysing similar periods (Murakami, 2014; Parro & Reyes, 2017). We have discussed the implications and potential drivers of, mainly, the fall of the skill premium over most of the 2010s in the first essay (see Chapter 2, section 2.8.1 for details).

Also, our estimated magnitude and pattern for the skill premium using online job ads is similar to estimations using recurrent sources like the Employment and Unemployment Survey for Greater Santiago data (in Spanish, *Encuesta de Ocupación y Desempleo del Gran Santiago*), EOD. The EOD has been carried out by the University of Chile since 1956 (University of Chile, 2020); this labour survey is generally used as source material when estimating the skill premium in Chile (see, e.g., Beyer et al., 1999; Gallego, 2012; Murakami, 2014; Robbins, 1994b, 1994a). We used the EOD in our first essay (see Chapter 2 section 2.5), and we estimated that average values for the skill premium for 2010-2018 were around 1.06 (see section 2.7.1). For the same period, our estimation using job postings is 1.03 (see a plot comparing these skill premium measures from both data sources over time in Appendix A.2.4). The similarity between our results and those using EOD shows the reliability of our estimations using job posting ads data. This kind of data has raised increasing interest in labour and development economics (see, e.g., D. Deming & Kahn, 2018; Hershbein & Kahn, 2018; Kureková et al., 2015; Marinescu, 2017).

Our strategy for task-content measures relies on examining the task content of standard occupations, the classification of our job posting ads according to the occupational classification, and the construction of our task-content measures as time series (see Figure 3.5). Our series standing for the ALM model category named as TM_{NRA} or non-routine analytical tasks required by job ads for skilled labour fluctuates between 0.38 and 0.48 over 2009-2018. These findings imply that a significant portion of skilled jobs, between a third and a half, involve non-routine analytical tasks. Examples of these work tasks are researching, evaluating, and managing, which usually feature skilled occupations (see Figure 3.3 for details and occupation examples). More generally, our results show a higher prevalence of non-routine analytical tasks in skilled occupations classified generically as managers, professionals and associate professionals or technicians in line with our expectations and previous literature (Mihaylov & Tijdens, 2019; Perez-Silva & Campos, 2021; Reijnders & de Vries, 2018). Like our non-routine analytical measure, our measure standing for non-routine interactive tasks or TM_{NRI} also shows a high prevalence in these skilled occupations. Over time, TM_{NRI} fluctuated between 0.32 and 0.35 implying a more stable pattern compared to TM_{NRA}.

Our measure standing for the routine cognitive content of skilled jobs, or TM_{RC}, started with an increasing trend and then a steady pattern over the period. In line with expectations and previous studies

(Mihaylov & Tijdens, 2019), our results show that these kinds of tasks are less prevalent in skilled jobs, with ratios fluctuating between 0.17 and 0.26. Our motivation for including this measure arose from previous studies on Chile, suggesting the relocation of skilled workers to less skilled positions due to complex software adoption as proxies for computer-based technologies (Almeida et al., 2020). Since less skilled or middle-skilled positions are more abundant in routine cognitive tasks, as proposed by the ALM model, we might see some relationship between skilled labour and the skill premium. We come back to this point in the next section on our findings on the influence of our task-content measures on the skill premium.

Regarding our estimation of skills-related measures, we found that most job advertisements distributed in the Cognitive and Social categories primarily demand skilled labour, as expected. This result aligns with the view on the endowment of these workers being characterized by cognitive (e.g., reasoning, evaluation) and social skills (e.g., communication skills). Our metrics capturing the prevalence of cognitive, social, cognitive & social, and software skills, SM_{Cogni}, SM_{Soc},, $SM_{CogniSoc}$ and SM_{Soft} respectively, show steep fluctuations with monthly values between 0.02 and 0.14 (see Figure 3.6). In terms of evolution, SM_{Cogni} shows a pattern without a clear increasing or decreasing trend over the period while SM_{Soc} depicts an increasing pattern. Regarding $SM_{CogniSoc}$, we can see an increasing trend in most of the period, probably fuelled by the increase in SM_{Soc}. This growth in social skills is in line with previous literature for countries like the US (see, e.g., Deming, 2017).

3.7.2. On the influence on the skill premium of our task-related measures

This section discusses the results from our task-related VAR estimations. On the one side, our findings from Granger-causality testing and IRF analysis support weakly the empirical evidence of the influence of our task-related measures on the skill premium.

We found the Granger-causality of our task-content measures towards the skill premium for the TM_{NRA} and TM_{NRI} variables only at 10% (*p-value* = 0.0709) and 5% (*p-value* = 0.0481) of significance level, respectively (see Table 3.19). Similarly, our IRF analysis shows negligible increases as the response of the skill premium to unexpected shocks in these task-related metrics. Although our results are weak, they support the ALM prediction about complementarity between computer-based technologies and non-routine cognitive tasks, both analytical and interactive, given the positive influence of TM_{NRA} and TM_{NRI}on the skill premium. In this regard, we show that changes in non-routine cognitive tasks may imply a greater demand for skilled labour and, consequently, a skill premium improvement. Conversely, past studies on Chile do not support this ALM prediction since they found substitution effects of computer-based technologies instead of complementarity (Almeida et al., 2020). Warnings of this substitution effect have also emerged from evidence showing a broader class of jobs at risk due to the potential ability of frontier technologies (e.g., robotics and artificial intelligence) to automate non-routine analytical or interactive tasks (Arntz et al., 2016; Autor, 2015; Frey & Osborne, 2017).

In the case of TM_{RC}, our Granger-causality testing is not robust to the addition of one lag (see Table 3.19); therefore, we do not assume that the routine cognitive content of jobs influences the skill premium at some significant level. This finding is not in line with our expectations based on recent research for Chile carried out by Zapata-Román (2021), who suggested that routine tasks had an important role in explaining earnings variation. Relevant differences between our approaches to data and methods might explain this discrepancy[41]. However, our results agree with previous and prominent literature for other countries (e.g., Autor et al., 2003; Goos & Manning, 2007; Goos et al., 2014; Sebastian, 2018; Spitz-Oener, 2006). In this regard, more research is needed to understand the interactions between the skill premium and the task-related metrics analysed here, given the weakness or absence of our evidence and the potential already revealed by incipient research in this field for Chile.

As limitations of our task-related analysis, we consider some characteristics of our data. For example, although we examine monthly data, the low number of observations (120 data points over 2009-2018) might be not enough to capture an adequate data variation. Additionally, the categorization of our global and skilled labour samples according to their occupational groups is not well balanced. These unbalanced data imply an over-representation of groups related to *Business and administration* (2-dig Code Occupations 33 and 24 in Table 3.10) characterized by medium or low content of non-routine cognitive analytical and interactive tasks (see 2-dig Code Occupation 33 and 24 in Table 3.7). Thus, there is less representation of observations standing for non-routine cognitive task content. This potential bias towards particular occupational groups needs to be considered in future research.

3.7.3. On the influence on the skill premium of cognitive and social skills

According to the human capital theory, skilled labour displays specific abilities beyond those acquired through formal qualifications (Heckman et al., 2006). In particular, prominent literature shows that cognitive skills complement the tasks performed by better-educated workers resulting in increases in their relative demand and wages (Acemoglu & Autor, 2011; Autor et al., 2003; Beblavý et al., 2016; Borghans et al., 2014). Alternatively, we do not find evidence on this complementarity from our Granger-causality testing, while our IRF analysis yielded negligible impacts. Like the lack of strong influence on the skill premium of non-routine cognitive tasks discussed above, our findings show that cognitive skills do not play a role in this wage differential. Our results disagree with past studies for Chile (Ramos et al., 2013) and other countries (Beblavý et al., 2016; Borghans et al., 2014) but agree with studies showing the lower importance of cognitive skills. Since a particular skill is implicitly associated with a particular task to be executed, our lack of support for the influence on the skill

[41] Zapata-Román (2021) used a Chilean household survey in four waves between 1992 and 2017, with decomposition methods to observe changes in occupational structure.

premium of cognitive skills coincides with our weak evidence of analytical tasks, as discussed in the last section.

Similarly, our results do not show Granger-causality evidence between social skills and the skill premium. Our findings disagree with past studies for Chile (Aedo et al., 2013; Ramos et al., 2013) and other countries like the U.S. (Deming, 2017), where social skills have been found explaining the skilled labour wages. Equally, our metric for cognitive & social skills does not explain the skill premium evolution. Therefore, we cannot posit that the demand for highly skilled workers relies on cognitive or social skills, singly or in combination, as expected.

A limitation of our study, as in the case of the task-related analysis, is that the number of our observations might be not enough to capture proper variability in the data. Our skills categorization, based on ten groups, shows that most of the sample refer to categories of *Customer Service,* demanding sales talent or client orientation abilities, or *Financial,* calling for skills in accounting or budgeting. In contrast, fewer job ads refer to *Cognitive* and *Social* skills (see section 3.6.1.3.1). Consequently, our sample would be biased towards data mainly describing abilities and endowments seen typically in middle-skilled positions (e.g., clerical workers, sales workers).

3.7.4. General discussion, policy issues and further research

Our results from our task-related analysis show weak evidence for non-routine cognitive tasks driving the skill premium. Regarding our skills-related analysis, we do not find support for the expected relationships between cognitive abilities, their combination with social skills, and the skill premium. We can speculate on the reasons for our lack of strong evidence on the expected role of cognitive tasks and skills, mainly in the context of the skill premium decline observed in recent decades by past studies (see, e.g., Murakami, 2014; Murakami & Nomura, 2020; Parro & Reyes, 2017; Also, see section 2.7.1 in the first essay) since our analysis focuses on this period. First, some suggest that the skill premium decrease has been driven by the drop in returns to skilled labour due to the substantial expansion of Chilean tertiary education (Murakami & Nomura, 2020; Parro & Reyes, 2017; Also, see section 2.7.3 in the first essay). If the return to higher education, which gives knowledge and stimulates cognitive skills to perform analytical tasks, is falling, then it would be expected that this knowledge and ability has little influence on skilled labour wages. Secondly, researchers have recently reported *downward* movements in the occupational ladder in the post-2000 period, such as the reassignment of skilled workers to less skilled positions (Almeida et al., 2020; Zapata-Román, 2021). These downward movements could explain the declining importance of cognitive tasks and skills in explaining the wages of skilled workers.

Some policy implications beyond our results, similar to those discussed in our first essay (see section 2.8.3), emerge. First, the lack of a strong relationship between the skill premium and cognitive tasks and skills might imply an unanticipated impact of technology adoption underestimated by the expected coordination between policies examining labour markets (demanding skills) and education (supplying skills). In this case, instead of the expected complementarity between technology and skilled labour, we might be seeing a neutral or substitution effect, leading to changes in the demand not only for skilled but also unskilled labour. Recently, Almeida et al. (2020), examining Chilean data at the firm level, suggested that because of the adoption of advanced technologies like complex software, the demand for unskilled workers was growing faster than for skilled labour: the lower demand for cognitive tasks and skills might be a potential explanation. Hence, the lower demand for skilled workers due to technology adoption may not be responding to the significant growth in supply resulting from the strong expansion of the tertiary education system starting in the 1980s and 1990s, as discussed above.

The inability to absorb the skilled labour in the workforce may have been underestimated by policymakers[42] implementing the expansion of the supply of skills or the "supply-shock" to boost the economic development of Chile through the expected transfer of knowledge and skills to jobs (Schneider, 2013; Valiente et al., 2020). Although, as suggested by Didier (2018), the transfer of learning towards labour markets could work in the short term, in the long run, it depends on policy coordination between demand and supply. However, as discussed in our first essay (see section 2.8.3), Chile does not have institutional mechanisms connecting firms with education suppliers (Valiente et al., 2020), and the institutional monitoring of mismatches between labour demand and supply is recent and promoted, mainly, by labour policymakers[43]. Also, and more specifically relevant to the supply of skills, Chile lacks an articulated educational and training system, viewed from a lifelong perspective, which is controlled by education policymakers (Didier, 2018). In this regard, systems monitoring the mismatch between labour demand and supply could promote the adoption of coordinated educational and labour policies if the government agenda were to support alliances among inter-sector policymakers. As suggested by Valiente et al. (2020), various actors in the education and labour sector are demanding higher levels of institutional coordination. Also, and more generally, the use of online job postings like the data used in this study can help policymakers to track labour markets' mismatches since they provide real-time information on features of demand (job offers) and supply (job seekers).

A second policy implication arising from our lack of strong evidence about the complementarity between skilled labour and cognitive tasks and skills could be the potential displacement of skilled labour to lesser skilled positions. In other words, unwanted changes in the occupational ladder. For instance, displacements of skilled labour to middle-skilled positions rich in cognitive but intensive in

[42] There is a consensus on that in Chile the push for expanding the supply of skills in recent decades came more from policy than from firms or business (see e.g., Schneider, 2013; Valiente et al., 2020).
[43] For example, the Labour Observatory (on Spanish, *Observatorio Laboral*) of the Ministry of Work and Pensions was created in 2015 and it has the mission of producing knowledge about gaps between supply and demand for occupations and job skills (SENCE, 2022). Also, the Job Prospection Committee was established in 2021: this is a policy council with objectives such as balancing the knowledge, skills and competencies of workers with the demand for human capital from the various productive sectors; promoting the labour trajectories of the workers; adapting to the constant variations of the labour market; and creating and sustaining skills that meet the needs of the labour market in the future (Ministerio del Trabajo y Previsión Social, 2021).

routine tasks (Almeida et al., 2020) would push middle-skilled workers to lower or unskilled skilled positions; in turn, these less-skilled workers can be pushed further down the occupational ladder, even affecting their chances of employment participation. These sequential *downward movements* represent unwanted changes in the occupational ladder for the educational and job prospects of workers. Therefore, policymakers need to predict these unwanted movements and mitigate its potential pervasive effects, especially among most affected employers. As pointed out above, policy efforts stimulating better coordination between higher education institutions and industry can support the development of specific skills or training systems to mitigate potential negative effects.

Some caveats arise concerning further research. First, to what extent do changes in the task-content of jobs affect the skill premium evolution? This evaluation might face some data challenges since it requires higher granularity at the occupation level, which we do not observe in labour surveys, given the sample sizes. Besides, in the case of Chile and other LAC countries like Colombia (Servicio Nacional de Aprendizaje, 2020), the current classification is based on international occupational hierarchies dating from 2008 (ILO, 2012) and relating it to previous versions (e.g., 1988) can be difficult. In this regard, and as pointed out above, the online job posting data is an attractive source for tracking labour market features. But if further research is to be conducted on this topic, it will be necessary to overcome some problems concerning the collection and treatment of data, such as the lower number of observations compared to official labour surveys (Spiezia, 2018) and the classification of job postings following official classifications for industries and occupations.

Secondly, more attention needs to be paid to middle or low-skilled groups—for instance, the bundle of tasks performed by these groups. Even more importantly, we need to know whether these tasks place a premium on the wages. This is also true for low or unskilled labour. In this regard, technologies could play a role as a complement instead of a substitute for this kind of labour, contradicting some ALM model predictions. Hence, technological advancements would receive the premium resulting in turn in an endogenous technical change (Acemoglu, 2002). This endogeneity might imply that technological change is biased by profit incentives where the market size of these labour groups drives the technologies' creation and adoption. In this regard, less-skilled labour might also encourage the technological innovation market. From a conceptual view, past studies for Chile using different data and methods show similar outputs, suggesting that, although the ALM model predictions were fulfilled in other high-income countries, this would not be the case with the Chilean labour market. Therefore, we encourage developing and applying "local" versions of the ALM model, suggesting alternative predictions or new conceptual or theoretical insights establishing the interrelationships between tasks, skills, and technology.

3.8. Conclusion

The evolution of the skill premium supplies opportunities to examine how economic forces (in particular, technological change) may influence the demand for skilled labour. Research on the task-content of jobs and workers' skill endowments provides material for relevant contributions to explanations of the dynamics between labour and technology, particularly the expected complementarity between cognitive tasks and skills and more educated workers (Acemoglu & Autor, 2011; Ehrenberg & Smith, 2018; Markowitsch & Plaimauer, 2009). We examine how measures standing for cognitive work activities employing mainly skilled workers, such as reasoning, problem-solving, and persuasion, drive the skill premium. Also, we evaluate the role of high-level labour skills such as cognitive and social abilities. However, our analysis focuses on a period witnessing a declining trend in the skill premium when cognitive tasks and skills might be less important. In this regard, our results support only weakly the ALM model prediction of the complementarity between non-routine cognitive tasks and skilled labour. Moreover, we do not find evidence of cognitive and social abilities driving the skill premium.

Like Almeida et al. (2020) and Zapata-Román (2021), we contribute to the recent strand of literature examining the ALM model predictions in the case of countries like Chile that have recently graduated from middle to high-income status. The lack of strong support for the complementarity between cognitive tasks and skilled labour is a key contribution of this study. In this regard, from a policy perspective, we encourage higher levels of institutional coordination between education and labour policymakers. If the premium for analytical capability is becoming less important, it might imply mismatches between skills demand and supply. Therefore, the adoption of coordinated educational and labour policies to correct these mismatches is needed. Also, our lack of strong support for the view on the complementarity between cognitive tasks, skilled labour and technology would imply a role for technological progress that would be potentially neutral or become a substitute for skilled labour. In this case, skilled workers would perform cognitive but routine tasks, which are typically performed by middle-skilled workers. In turn, this middle-skilled labour would be filling lower-skilled positions. Relocating better-educated workers to less-skilled positions might imply an inefficient educational investment and produce other unwanted impacts, like the deterioration of workers' prospects. Again, we highlight the importance of coordinated education and labour policies to predict and mitigate unwanted effects of technology adoption.

Regarding further research, more understanding of changes in the employment structure or occupational ladder is needed. It is especially important to discover the extent to which these changes in the occupational ladder have affected the skill premium evolution, given the decline observed in recent decades. We also encourage the development of "local" versions of our motivating theories and conceptual frameworks. Typically, the ALM model predictions are fulfilled in the case of high-income countries like the U.S. and some countries in Europe but have limited application to countries like Chile and other nations in Latin America and the Caribbean region.

4. Essay III: Exploring with text the demand for ICT labour as a proxy for technological replacement in the aftermath of disasters

Over the past few decades, an extensive literature has been developed on the economic impact of natural disasters. However, evidence of specific impacts on labour markets has received less attention. Using a massive earthquake (above 8.0 M_w) that struck Chile's Central Region in 2010, the 27th February Biobío earthquake, as a natural experiment, we assess whether disasters can accelerate the adoption of equipment compatible with Information and Communications Technologies, ICT, which drive much of the technical change in production. We examine changes in demand for ICT labour as a proxy for technological upgrading. Our data are the open text from a collection of 4,136 online job postings published between 2008 and 2012 in the most severely affected regions. We implement a structural topic model to estimate the difference in the prevalence of ICT and Construction labour topics by comparing periods two years before and two years after the earthquake. Our results show that the prevalence of the ICT labour topic does not change. In contrast, the prevalence of the Construction labour topic was significantly different after the disaster, suggesting that reconstruction activities lead to employment differences. Our results suggest that there was no substantive technological replacement following the 27th February Biobío earthquake.

Keywords: technological upgrading, creative-destruction, ICT labour, natural disasters
JEL Classification: J20, Q54, O33

4.1.Introduction

The study of social and economic impacts caused by natural disasters has become increasingly important due to the higher exposition of the global population to these shocks. However, research on the impacts on labour is less abundant (Kirchberger, 2017), with most research documenting impacts on aggregated labour outputs such as unemployment, participation rates, and wages (e.g. Brown et al., 2006; Kirchberger, 2017; Xiao & Feser, 2014; Zissimopoulos & Karoly, 2010). Research focusing on aggregated labour might hide impacts on particular labour sub-groups or sectorial labour (How & Kerr, 2019; Zissimopoulos & Karoly, 2010). For example, little attention has been paid to other labour subgroups, like labour employed by the Information and Communications Technologies, ICT, sector in the context of the technological change in production that is supposed to be mainly driven by ICT and other computer-based technologies (see, e.g., Acemoglu & Autor, 2011; Almeida et al., 2020; Hwang & Shin, 2017). In this regard, some suggest that disasters can be considered as episodes or substantial events affecting the pace of technological change (Crespo Cuaresma et al., 2008; Okuyama, 2003; Okuyama et al., 2004; Skidmore & Toya, 2002). Nevertheless, no definitive answer has yet been given to the question of whether disasters can accelerate the pace of the current technological progress associated with ICT, assuming that updated and ICT-compatible equipment replaces the machinery destroyed by recent catastrophes. In turn, this faster rate of technology adoption would lead to increases in demand for ICT-related labour or workers employed in the ICT sector. This study aims to explore these interactions using the analysis of the Solow-Swan model with technical change under a disaster situation to conceptualize the expected increase in the pace of technological change due to technological replacement. In turn, this technical upgrading may lead to improvements in demand for ICT-related labour (see section 4.2.2 for details).

Examining how natural disasters might accelerate the ICT-intense technical change rate proxied by changes in demand for ICT is relevant to countries like Chile. First, Chile supplies an environment that is particularly suitable for studying the impacts of disasters like earthquakes. Ten of the most destructive earthquakes, i.e., 8 M_w and above (See footnote 4 for M_w definition), hit Chile in the past century (Barrientos & CSN Team, 2018). In the last decade, three earthquakes over this magnitude affected different Chilean regions in 2010, 2014 and 2015, characterizing Chile as a site of recurring earthquakes. Secondly, technical change has been an important driver of the economic development seen by Chile in the last 40-50 years. Although the importance of this technological change has declined over time for outputs like the skill premium, as shown in our first and second essays (see Chapters 2 and 3), it has continued operating in Chile in recent decades. Remarkably, indicators covering assets like hardware, telecommunications and software shows that the share of ICT in total investment for Chile has been growing, resulting in important ICT capital formation (ECLAC, 2013). In this regard, examining the impacts of disasters and how they are related to technical change is an added step towards understanding changes in demand for employment, especially subgroups like ICT labour.

Conceptually, examining the interactions between natural disasters, labour markets, and technological change has relied on extensions of growth models like the Solow-Swan model (Solow, 1956; Swan, 1956) and a more literal explanation of the Schumpeterian creative-destruction hypothesis (Aghion & Howitt, 1990; Schumpeter, 1976). An extended Solow-Swan model provides insights on resource allocation involving labour and capital for economic recovery in the aftermath of disasters (Okuyama, 2003). It can compare the effects resulting from the destruction and subsequent upgrading of capital goods on the steady-state of the economy and eventual recovery. The main assumption is that older and outdated capital goods are more prone to be damaged by a catastrophe because of vulnerabilities, including weaker structure, mechanical fatigue due to age, and outdated regulations, from which updated equipment is free (Okuyama, 2003). Related to the creative-destruction hypothesis, originally, this conceptual idea gives prominence to the effects of competition between new consumer goods, new markets, and new technologies. These dynamics incessantly transform the economic structure *from within*; that is, the creative-destruction process permanently destroys the old and creates the new. In the natural disasters literature, the concept refers literally to the process of *technology replacement after a catastrophe* (Crespo Cuaresma et al., 2008). This sudden turnover of capital might represent a positive *jump* in technological improvement.

We develop our conceptual framework (see section 4.2) according to the extended Solow-Swan model considering the expected technology replacement after a catastrophe. We examine how natural disasters can positively affect the pace of technical change, resulting in positive impacts on growth and employment. The model encompassing disaster impact evaluation responds to researchers' attempts to develop conceptual and theoretical foundations such as the works of Okuyama (2003) and Okuyama et al. (2004). We do not test any post-disaster theoretical predictions since there is no comprehensive theory in this literature, and assumptions regarding expected impacts from the aftermath of disasters are many and varied (Coffman & Noy, 2011). However, the analysis of the impact of disasters using the Solow-Swan model with technical change helps us to conceptualize and predict the pace of technological change and, in turn, the effects on labour.

As observed above, we assume that ICT and similar technologies drive much of the technological change in production. Consequently, we assume a technical change embodied in ICT capital goods covering assets like hardware equipment, telecommunications, software, among others (Bassanini & Scarpetta, 2002; ECLAC, 2013). In this sense, the expected *jump* in technological adoption would imply that much of the technology replacement after a disaster will be based on ICT capital goods. This rapid move towards equipment compatible with ICT might result in improvements in demand for ICT labour. We have defined ICT labour as those occupations involved in the provision of goods and services related to the ICT sector[44] , and we expect positive changes in its demand due to the recent occurrence of disasters.

Studies show inconclusive evidence regarding natural disasters as forces affecting technological upgrading and, consequently, economic and labour outputs. Some show that replacing damaged capital goods with updated equipment in the aftermath of disasters can improve economic growth (Benson & Clay, 2004; Crespo Cuaresma et al., 2008; Loayza et al., 2012; Toya & Skidmore, 2007). Disasters can lead to increased industrial growth (Loayza et al., 2012) and increased physical capital accumulation (Leiter et al., 2009). However, others reported that natural disasters do not significantly affect subsequent economic growth (Cavallo et al., 2013). In addition, benefits from capital upgrading have been linked to countries with higher levels of development because of better institutions, policy, and financial systems, among other factors (Crespo Cuaresma et al., 2008; Toya & Skidmore, 2007). In this regard, technology upgrades in post-disaster scenarios usually face financial and time constraints (Benson & Clay, 2004; Di Pietro & Mora, 2015). More importantly, some analyses of various disasters from a pool of countries have suggested significant adverse impacts of disasters on technological innovation, measured by the number of patent applications (Chen et al., 2021). Therefore, we cannot establish that disasters are unequivocally a source of adjustment for technological change and, consequently, for changes in demand for labour.

Employment adjustments can result from reconstruction efforts unrelated to technological improvements. For instance, when labour is a substitute for damaged or missing physical equipment, a disaster will lead to positive employment impacts (e.g., more demand), especially in the construction sector (Belasen & Polachek, 2009; Skidmore & Toya, 2002). Also, Leiter et al. (2009) reported employment growth, given the higher physical capital accumulation in regions affected by disasters. However, even if a catastrophe promotes a more significant capital stock, it does not necessarily imply positive impacts on labour participation. Tanaka (2015) found a negative impact on employment, despite over-investment in physical capital. Tanaka speculates that a decreased population in the affected area may be a possible reason. A lower population might result from direct impacts on labour (e.g., death, injuries) or indirect, like forced displacements. The extent to which workers can stay in the labour market after a disaster also influences potential technological replacements.

With regard to our stated prediction that disasters may positively affect the demand for ICT employment because of subsequent higher levels of ICT technology adoption, it must be borne in mind that no existing study has examined the role of natural disasters in explaining changes in demand for ICT labour. Most studies have analysed changes in aggregated labour, hiding impacts on particular sub-groups (How & Kerr, 2019; Zissimopoulos & Karoly, 2010). Overall, the natural disasters literature emphasises the importance of ICT technologies in coping with problems in the aftermath of

[44] Following the International Classification of Occupations (ILO, 2012) some examples of these ICT occupations are Systems Analyst, Software Developers, Database Designers and Administrators, Computer programmers. Computer Network and Systems Professionals and Technicians.

catastrophes, where ICT plays a vital role in reducing disaster fatalities, managing recovery costs and dealing with other aspects of disaster management (Benali & Feki, 2018; Toya & Skidmore, 2015; Walker, 2012). Yet, more attention has been paid to ICT labour in the context of other shocks, like recessions and pandemics. It has been suggested that recessions affect ICT employment negatively (Holm & Østergaard, 2015). Conversely, the recent COVID-19 pandemic has affected the ICT workforce relatively less than other occupations, given the prevalence of teleworking in their sector and their lower exposure to social or face-to-face interactions (Pouliakas & Branka, 2020; Redmond & Mcguinness, 2020). Yet the overall lack of studies on the role of disasters in explaining differences in ICT employment impedes our understanding of disasters' impacts on sub-groups or specialized labour. More importantly, this particular research field of natural disasters' impacts on labour requires a cumulative number of cases to support an explanatory framework strong enough to enable us to understand how employment is affected (Jara & Faggian, 2018).

But the insufficiency of studies is a general problem in the literature examining the interaction between labour and natural disasters. According to Jiménez et al. (2020), between 1900 and November 2019, only 118 articles on the effects of disasters on labour were published in indexed journals. Most of them refer to Japan, the US and China. For Chile, only two studies appeared: Jiménez & Cubillos (2010) and Jiménez et al. (2020). Some additional research can be found in other sources, with Jara & Faggian (2018) and Sanhueza et al. (2012) as the only studies referencing impacts on labour. As noted above, the lack of published studies might also be attributed to a *publication bias*, whereby significant findings generate higher chances of publication (Klomp & Valckx, 2014). Additionally, since disasters' interruption of economic activities is usually only temporary, most past research has focused on shorter-term impacts since it is more difficult to identify long-term effects (Jiménez et al., 2020). We contribute to this limited literature focusing on ICT labour.

We explore the impact of disasters on technological replacement proxied by changes in demand for ICT labour. This involves examining the text content of a collection of 4,136 online job postings published two years before and two years after the event in most Chilean regions affected by the 27 February 2010 Biobío earthquake (M_w 8.8) (see section 4.3 for details). Our online job postings correspond to a sub-sample from the www.trabajando.com data used in the second essay (see Chapter 3, section 3.4). Our pre-disaster period is represented by data available from January 2008; the decision to use post-disaster material two years after the earthquake is based on the assumption that the economic scenario in the second year after the disaster might provide a more stable basis for making technological replacements decisions.

The 27th February 2010 Biobío is considered the second most severe in Chile's history and one of the ten strongest worldwide since these events have been recorded by instruments (Barrientos & CSN Team, 2018; Contreras & Winckler, 2013; M. Jiménez et al., 2020; Sanhueza et al., 2012). The earthquake, and subsequent tsunami, affected several regions in the central and south areas of the country that are inhabited by approximately 80% of the Chilean population. The estimated destruction included 500,000 damaged houses, 12,000 injured people, over 400 deaths and an economic cost of US$30,000 million (NOAA, 2019). This earthquake has been used as a natural experiment in other studies examining the link between natural disasters and labour. The topics of studies on the earthquake have included its impact on perceived stress and job satisfaction (A. Jiménez & Cubillos, 2010) and on the probability of employment, unemployment and lack of access to social security (M. Jiménez et al., 2020; Karnani, 2015; Sanhueza et al., 2012; Sehnbruch et al., 2017). Most of this evidence suggests that the earthquake negatively affected the labour market in the short run. However, in the long term, it has been suggested that these negative impacts are attenuated by the recovery process, which is facilitated by the government's efforts and other institutional factors (M. Jiménez et al., 2020). We add to this literature by considering the potential role of the 27 February 2010 Biobío to explain changes in workforce sub-groups like ICT labour.

We apply a set of techniques based on the text data of our collection of job postings to evaluate changes in ICT labour after our natural disaster. However, our sample lacks a variable to filter ICT-specific job postings. Besides, ICT occupations or job titles can vary widely. In this regard, we are not able to apply recurrent methodologies used in this kind of analysis (e.g., differences-in-differences) since there is a not a variable for selecting our sample of interest i.e., ICT-related jobs. Hence, our modelling and estimation strategy relies on the Structural Topic Model, STM, developed by Roberts et al. (2016, 2013). As a topic model, STM uncovers word co-occurrence patterns across a collection of documents, i.e., our sample of job posting ads, to estimate a set of word clusters or *topics*. Next, we identify the ICT-related topic that best represents ICT labour, and we examine changes in its prevalence by applying a treatment effect estimation. We identify whether the job postings were published before or after the disaster, where the post-disaster period corresponds to our treated period. In this regard, different to recurrent topic model approaches, STM incorporates document metadata i.e., the date of job posting publication to *structure* the document collection (see section 4.4 for details). In terms of results, we expect a higher prevalence of ICT labour after the disaster because of the rapid adoption of equipment compatible with ICT. Also, as pointed out in the literature, we expect that our natural experiment positively influences topics standing for Construction labour, given the recovery and reconstruction activities.

Our results show that the prevalence of the topic representing ICT labour does not significantly change after the earthquake. Conversely, the Construction labour topic prevalence is significantly different after the disaster, i.e., the prevalence increased. These findings suggest that reconstruction activities lead to differences in Construction employment while we do not observe changes in ICT labour. Thus, our results do not support the view on substantial technological replacements occurring after the 27th February 2010 Biobío earthquake of a kind that impacted the labour market, particularly the demand for ICT labour.

Some policy recommendations emerge from our results. For example, although most of the policy on the recovery process was focused on returning to the normal or pre-disaster circumstances, policymakers can take advantage of recovery activities, considering more technology upgrading initiatives. Also, our findings on improvements in Construction employment raise some policy issues, given the temporary nature of reconstruction activities and the predominance of low or unskilled workers in this sector. Consequently, a policy is needed to promote transitions to permanent jobs or training for workers (most of them vulnerable) to mitigate the eventual lack of income once the reconstruction finishes.

This essay structure is as follows. We begin by presenting our conceptual framework. Then, we describe the data and the STM as our methodological strategy. Next, we present and discuss our results. In the final section, we recapitulate our argument and resulting policy recommendations.

4.2. Conceptual framework

We offer some insights into the impact of technological replacements and, in turn, on ICT labour, based on extensions of the basic neoclassical model of Solow-Swan (Solow, 1956; Swan, 1956). This model helps us to conceptualize our assumption of a higher rate of technological change during the recovery process in the aftermath of a disaster due to the potential replacement of destroyed capital with updated equipment. Since this conceptualization assumes a component standing for labour-augmenting technology, which we suppose is intensive in ICT and related technologies, its higher rate might lead to the faster growth of labour, particularly ICT labour.

The Solow-Swan model in its original version evaluates economic growth based on the shape of the neoclassical production function. In the first essay, we test the RBET model, which is also based on this production function (see section 2.2.1).

We demonstrate the application of the Solow-Swan model in a disaster situation in per capita terms following Okuyama (2003), who applies the model with labour-augmenting technological progress described by Barro & Sala-i-Martin (2004; pp. 54-56). This proposed model has been extensively used in natural disasters impacts literature (see, e.g., Crespo Cuaresma et al., 2008; Hallegatte & Dumas, 2009; Leiter et al., 2009; Lynham et al., 2017; Panwar & Sen, 2019).

In the following, we firstly describe the basic Solow-Swan model fundamentals applied to a disaster situation without technological progress. Then we add a term representing a labour-augmenting technological change, emphasizing how its pace changes because technological replacement affects the model's dynamics, particularly, the growth of labour.

4.2.1. The basic Solow-Swan model in a disaster situation

Let us assume that the aggregated production function neglecting technological change is:

$$Y = F(K, L) \tag{4.1}$$

where Y, K and L are the total output of the economy, the level of capital accumulated in the economy and the level of labour population, respectively. Assuming that the production function is homogeneous of degree one, we can express the Eq. (4.1) in its *intensive form*, i.e., in per capita or per worker form, as follows:

$$y = f(k) \tag{4.2}$$

where $k \equiv K/L$ is the capital per worker and $y \equiv Y/L$ is the output per worker. Eq. (4.2) implies that the output produced by each worker is determined by the amount of capital each person can access and, assuming k is constant, changes in the number of workers do not affect the total output per capita. In other words, the production function shows no "scale effects" (Barro & Sala-i-Martin, 2004).

The change in per capita capital stock over time, setting as constants the terms s, δ and n which stand for the saving rate, the capital depreciation, and the population growth rate, respectively, becomes as follows:

$$\dot{k} = s \cdot f(k) - (n + \delta) \cdot k \tag{4.3}$$

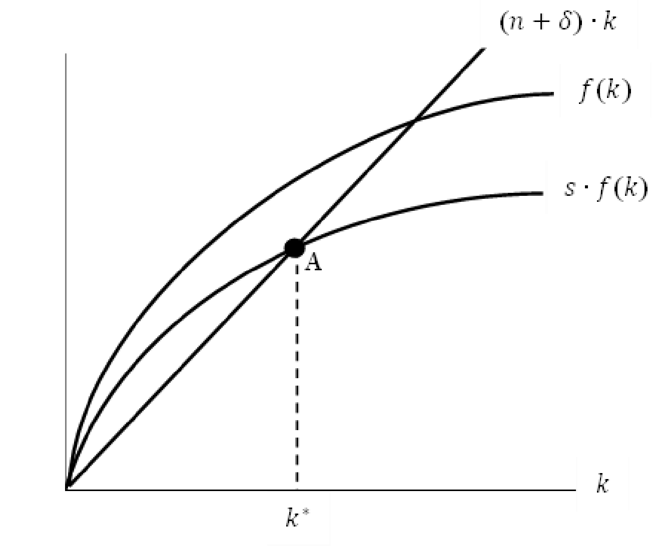
where $\dot{k} = \partial k(t)/\partial t$ following the convention that a dot over a variable denotes differentiation concerning time as used by Barro & Sala-i-Martin (2004). The nonlinear Eq. (4.3) depends only on k, and is the fundamental differential equation of the Solow-Swan model. From this fundamental representation we can assume that the term $n + \delta$ stands for the effective depreciation rate for the capital per worker, $k \equiv K/L$. If the saving rate s equals to zero, then the capital-labour ratio k would partially decrease both by the depreciation of capital at rate δ and the population growth at rate n (Barro & Sala-i-Martin, 2004). We can also examine the steady-state (or long-run) and the transitional dynamics (short-run) of the Solow-Swan model from the stated relationships represented by Eq. (4.3).

The steady-state in the Solow-Swan model refers to $\dot{k} = 0$ in Eq. (4.3). In this state, the quantities of the factors grow at constant rates implying a steady-state level of capital accumulation. This steady value of k is termed k^* and, algebraically, k^* satisfies the following condition:

$$s \cdot f(k^*) = (n + \delta) \cdot k^* \tag{4.4}$$

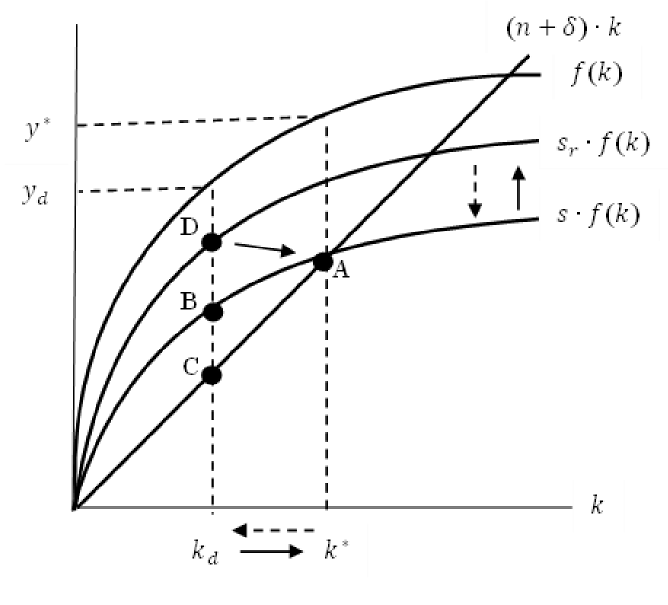
The workings of Eq. (4.3), along with the condition standing for the steady-state k^* as shown in Eq. (4.4), are graphically represented in Figure 4.1. The upper curve is the production function $f(k)$ and it is proportional to the curve $s \cdot f(k)$ which is like $f(k)$ except for the multiplication by the positive term s. The effective depreciation rate for the capital per worker, k, is given by the straight line from the origin $(n + \delta) \cdot k$ with the positive slope $n + \delta$. The change in k over time is determined by the vertical distance between the curve $s \cdot f(k)$ and the line $(n + \delta) \cdot k$ while the steady-state level of capital accumulation, k^*, is found at the point where both shapes intersect. We named this intersection **A** for purposes of our following disaster situation application of the Solow-Swan model.

Figure 4.1. The Solow-Swan model (adopted from Okuyama, 2003)



To show the impact of a disaster under the Solow-Swan model, Figure 4.2 reproduces the graphical representation given by Figure 4.1 but adds the effect of a decline in the capital accumulation k because of the destructive power of the disaster.

Figure 4.2. The Solow-Swan model and the disaster impact (reproduced from Okuyama, 2003)



First, let us suppose that the economy is in the steady-state condition or at point **A**, as shown in Figure 4.2. When the disaster hits the economy, we assume that the capital accumulation is massively damaged, but there is no damage to the workforce level. We see the capital decline in the displacement of k^* to the deteriorated level k_d (the dashed line arrow in the x-axis), therefore $k_d < k^*$. Consequently, in the y-axis we can see how the economy's output level decreases from the steady-state level y^* to the level in a disaster situation or decreased output y_d, where $y_d < y^*$. The displacements of y^* and k^*

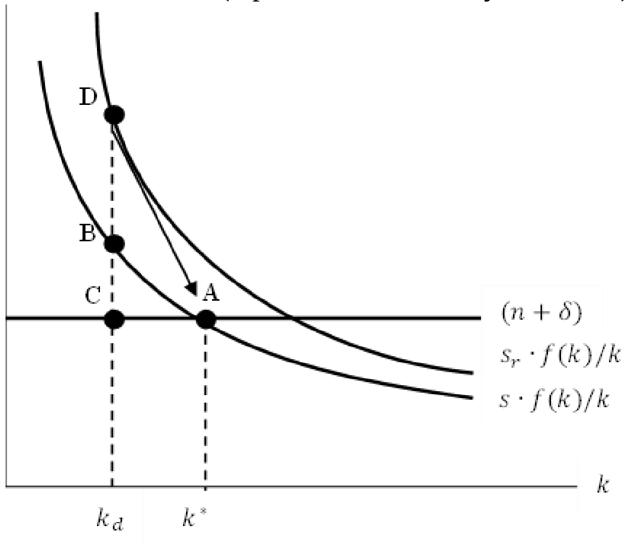
imply that the economy is out of its steady-state level given by **A**. Therefore, the economy needs to return to this steady level where the distance between **B** and **C** corresponds to the per capita capital accumulation that needs to be recovered. In this regard, the recovery process is the required increase from k_d toward k^* (the solid line arrow in the x-axis). One implication of this recovery stage is a greater allocation of resources to reconstruction activities compared to the pre-disaster situation. As a result, the saving rate may increase during the recovery process. We set this recovery saving rate as s_r, where $s_r > s$ as shown by the displacement of the curve $s \cdot f(k)$ toward $s_r \cdot f(k)$ (see the vertical solid line arrow in the right-hand side of Figure 4.2). The recovery saving may encourage the speed of the recovery process, given the greater distance between points **D** and **C** compared to that between **B** and **C**. However, as the recovery process progresses, s_r should go back to s (see the dashed line arrow in the right-hand side of Figure 4.2). We represented the economy recovery toward its pre-disaster level of capital accumulation or steady-state level k^* by the arrow from the point **D** towards **A.**

As noted above, we can also examine the transitional dynamics (short-run) of the Solow-Swan model from the stated relationships represented by Eq. (4.3). These dynamics show how the economy output converges toward its steady-state level, as discussed earlier in the recovery process context. Therefore, this analysis gives us further understanding of the recovery process by using the growth rate of k. Division of both sides of Eq. (4.3) by k results in the growth rate of k, γ_k, as follows:

$$\gamma_k \equiv \dot{k}/k = s \cdot f(k)/k - (n + \delta). \tag{4.5}$$

Eq. (4.5) shows that $\dot{k}/k$ equals the difference between the *saving curve*, $s \cdot f(k)/k$, and the *depreciation curve* $(n + \delta)$. Following the notation of Okuyama (2003), we plot the saving and depreciation curves to indicate the transitional dynamics around the steady-state of the Solow-Swan model in Figure 4.3. The vertical distance between these two curves gives the growth rate of k, which becomes zero at the steady-state k^*due to the intersection of both curves, where $s \cdot f(k)/k = (n + \delta)$.

Figure 4.3. Transitional dynamics of recovery under the Solow-Swan model (reproduced from Okuyama, 2003)



Recalling our disaster situation, Figure 4.3 shows that the level of k turns into k_d due to the damaged capital. Since $k_d < k^*$, the growth rate of k is positive (the space between the points **B** and **C** in Figure 4.3), implying that k approaches k^* as the recovery process operates. Given the intensity of reconstruction activities, i.e., the economy encouraging the allocation of resources to return to pre-disaster levels, the saving rate may become higher temporally[45], s_r, as represented by the curve *recovery saving rate*, $s_r \cdot f(k)/k$ in Figure 4.3. As $s < s_r$, the growth rate of k also rises, i.e., the distance between the points **D** and **C** is higher than the distance between **B** and **C**. As the recovery process progresses over time, the growth rate declines and approaches 0 as k approaches k^*. These recovery dynamics are represented by the arrow from **D** to **A** in Figure 4.3. Since more resources are relocated for recovery, the reconstruction activities encourage capital re-accumulation more rapidly (Okuyama, 2003). Now we turn to the situation with technological change.

4.2.2. The Solow-Swan model with technological change in a disaster situation

We suppose now that our production function includes technological progress: more specifically, the level of labour-augmenting technical change, i.e., technology that increases output in the same way that the stock of labour increases[46] (Barro & Sala-i-Martin, 2004). The inclusion of the level of technology over time, $A(t)$, as factor in the primary production function represented by Eq. (4.1) yields:

$$Y = F[K, L \cdot A(t)] \tag{4.6}$$

where $A(t)$ appears as a multiple of L due to the assumption of labour-augmenting technology. Also, $A(t)$ grows at a constant rate, x. We turn to this rate later.

The change in per capita capital stock over time represented by Eq. (4.3), including $A(t)$, becomes

$$\dot{k} = s \cdot f[k, A(t)] - (n + \delta) \cdot k \tag{4.7}$$

where the output per capita now depends on the level of technology, $A(t)$.

The analysis of the transitional dynamics of the Solow-Swan model with labour-augmenting technical progress requires the rewriting of the model in terms of variables staying constant in the steady-state. In this regard, k and $A(t)$ grow in the steady-state at the same rate, so we can work with the ratio

$$\hat{k} \equiv k/A(t) = K/[L \cdot A(t)] \tag{4.8}$$

[45] We assume a temporal change in the saving rate due to the shock generated by the disaster. This framework also allows evaluation of permanent changes (e.g., changes in consumption, policy impacts) generating an alternative steady-state level of capital accumulation (see Barro & Sala-i-Martin, 2004, pp. 41)

[46] The assumption of a labour-augmenting technical change is based on the consideration of constant rates of technological progress. Given that in the Solow-Swan model the population grows at a constant rate, only a labour-augmenting technological change is consistent with the existence of a steady-state, i.e., constant rates of growth of factor quantities in the long term (Barro & Sala-i-Martin, 2004).

where $L \cdot A(t) \equiv \hat{L}.\hat{L}$ is often named the *effective amount of labour*. This terminology is convenient since the economy works as if its labour input were $\hat{L}$, i.e., the labour population, L, multiplied by its efficiency, $A(t)$. As a result, $\hat{k}$ in Eq. (4.8) refers to the capital accumulation per unit of effective labour. Then, the output per unit of effective labour is given by

$$\hat{y} \equiv f(\hat{k}). \tag{4.9}$$

We can obtain the production function in intensive form replacing y and k by $\hat{y}$ and $\hat{k}$, respectively. Following the same procedures to write Eq. (4.3) and Eq. (4.5), but now, using the information that $A(t)$ grows at the rate x, as discussed earlier, Eq. (4.3) becomes

$$\dot{\hat{k}} = s \cdot f(\hat{k}) - (x + n + \delta) \cdot \hat{k} \tag{4.10}$$

where the term $x + n + \delta$ is now the effective depreciation rate for $\hat{k} \equiv K/\hat{L}$. In the case of Eq. (4.5), the growth rate of $\hat{k}$ is

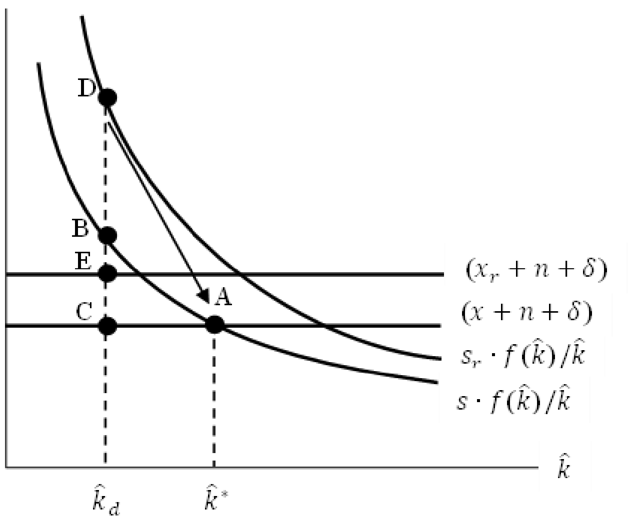
$$\gamma_{\hat{k}} \equiv \dot{\hat{k}}/\hat{k} = s \cdot f(\hat{k})/\hat{k} - (x + n + \delta). \tag{4.11}$$

As in the argument discussed in section 4.2.1, at the steady-state, the growth rate of $\hat{k}$ becomes zero in Eq. (4.11). This steady value of $\hat{k}$ is termed $\hat{k}^*$ and, algebraically, $\hat{k}^* = 0$ satisfies the following condition:

$$s \cdot f(\hat{k}^*)/\hat{k}^* = (x + n + \delta). \tag{4.12}$$

The transitional dynamics of $\hat{k}$ are similar to those in the model without technological change. As in Figure 4.3, we plot these dynamics in Figure 4.4 following the notation of Okuyama (2003) with the x-axis involving $\hat{k}$ to analyse the disaster situation, but now with technical change.

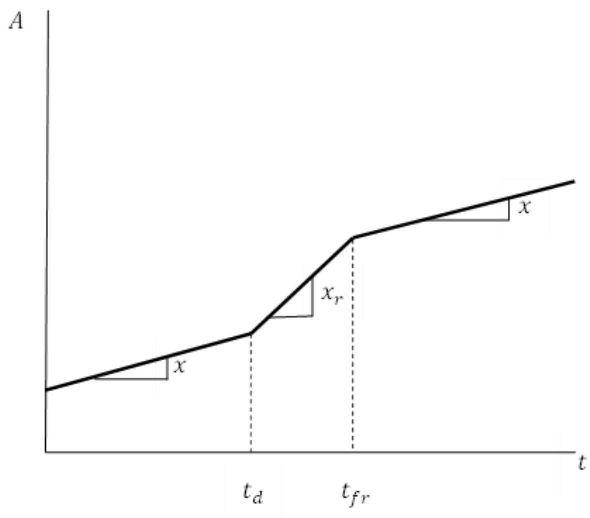
Figure 4.4. Transitional dynamics of recovery under the
Solow-Swan model with technological change
(reproduced from Okuyama, 2003)



As in the last section, 4.2.1, in Figure 4.4, the steady-state level $\hat{k}^*$ went down to $\hat{k}_d$, which stands for the capital damaged by the disaster. Here, the growth rate of the recovery process is the space between **B** and **C** in a scenario where the economy does not allocate resources in some particular way. The recovery speed might be encouraged by increases in the saving rate to favour a higher allocation of resources to reconstruction or capital re-accumulation. We term this recovery saving rate as s_r and given that $s < s_r$, the growth rate of $\hat{k}$ also increases, i.e., the distance between the points **D** and **C** is greater than the distance between **B** and **C**. These dynamics are practically the same as in the model without technical change, as described above in section 4.2.1. However, the displacement of $(x + n + \delta)$ to $(x_r + n + \delta)$ shows our assumption of a higher rate of technological change during the recovery process, x_r, due to the potential replacement of destroyed capital by updated equipment where $x_r > x$.

For the sake of clarity, we reproduce the plot of Okuyama (2003, p. 17) in Figure 4.5 (the y-axis is the level of technology, A, and the x-axis is time, t) to show visually our assumption of $x_r > x$ between the moment of disaster occurrence, t_d, and the full recovery, t_{fr}.

Figure 4.5. The rate of technological change with and without disaster (adopted from Okuyama, 2003)



Recapitulating the dynamics in Figure 4.4 according to our assumption of $x_r > x$, the distance between **D** and our new point **E** is the new growth rate of $\hat{k}$. The distance between **D-E** is shorter than between **D-C,** implying a slightly slower growth rate of $\hat{k}$ during the recovery process compared to the model without technological change discussed in section 4.2.1. This slower growth of $\hat{k}$ is due to the more rapid technological change which leads to the faster growth of effective labour under the assumption of labour-augmenting technology.

The analysis of the Solow-Swan model with technical change under a disaster situation helps us to conceptualize the expected increase in the pace of technological change due to technological replacement. This technical upgrading may lead to improvements in demand for specialized labour. Assuming that much of the technical change in production is driven by ICT-related advancements, we can predict that a rapid ICT-related change during a recovery process might lead to improvements in demand for ICT labour. In the following sections, we present the data and our empirical strategy based on topic modelling to test this prediction.

4.3.Data

Our data corresponds to a sub-sample from the online job ads dataset www.trabajando.com used in our second essay (see section 3.4 for details). We filter the Chilean regions considered to be most affected by the 27 February 2010 Biobío earthquake, i.e., the regions (in Spanish) *VI de O'Higgins*, *VII del Maule*, *VIII del Biobío* (ECLAC, 2010; Sanhueza et al., 2012)[47]. We use the job posts published from January 2008 to March 2012. Using the job postings publication date, we create a dummy indicating whether the job post was published after the disaster (treated period), $27F$, which is specified as follows:

$$27F = \begin{cases} 1 & if\ the\ post\ is\ published\ between\ March\ 2010 - \ March 2012 \\ 0 & if\ the\ post\ is\ published\ between\ January\ 2008 - Februrary\ 2010 \end{cases} \tag{4.13}$$

Our pre-disaster period is represented by data available from 2008 and the occurrence of the disaster on 27 February 2010. The post-disaster definition relies on short-run impacts, considering not only the first year after the disaster's occurrence but also the second year. Unlike past studies evaluating only one post-disaster year (see, e.g., Karnani, 2015), we consider that one year might be a very short period for considering decisions on technological replacements and potential ICT labour hiring. Besides, firms might be coping with several potential restrictions (e.g., financial and labour shortages) during the first post-disaster year. We consider that the economic scenario in the second year after the disaster might supply a more stable basis for making these decisions. Also, we have not considered more years in the post-disaster span to balance properly the number observations between pre- and post-disaster span.

After filtering by affected regions and periods before and after the disaster, our sample consists of 4,136 online job posts. Table 4.1 shows the distribution of our sample according to pre-and post-disaster periods.

Table 4.1. Distribution of online job ads in the most affected regions by pre and post-disaster periods

Period	Number of job post ads
Pre-Disaster (January 2008 – February 2010)	1,720
Post-Disaster (March 2010 – March 2012)	2,416
Total	4,136

[47] Other studies include some additional regions such as Región Metropolitana, V de Valparaíso and the IX de La Araucanía (M. Jiménez et al., 2020; Karnani, 2015) but these regions were less affected (ECLAC, 2010).

From our collection of job posts, we concatenate three open text variables (job title, job description and job-specific requirements). These concatenated text variables, along with the date of publication, correspond to our input for performing our estimation strategies, as detailed in the next section, 4.4.

4.4. Structural topic modelling, STM

The probabilistic or statistical topic models, TM, pioneered by Latent Dirichlet Allocation, LDA (Blei et al., 2003), are tools designed for analysing and understanding large text corpora based on words' co-occurrence. TM are known as "unsupervised techniques" since they infer topics' content from a collection of texts or *corpus* rather than assume them as supervised techniques that require ex-ante definitions of topics (Roberts et al., 2014). Since we only observe the documents, TM aim to infer the *latent* or *hidden* topics by applying Bayesian and non-Bayesian estimation strategies (see details on Bayesian analysis in the first essay, section 2.6.2.2). By specifying a Bayesian model, we can evaluate how a document is generated by estimating how words are distributed in topics and topics in documents. Conceptually, we refer to topics as distributions or mixtures of words that belong to a topic with a certain probability or weight. These weights indicate how important a word is in a given topic. In this context, documents are distributions over topics where a single document can be composed of multiple topics and words can be shared across topics. Thus, we can represent a document as a vector of proportions that shows the share of words belonging to each topic (Roberts et al., 2014).

TM allow us to evaluate the importance of topics in the documents. The sum of shares of topics across all topics in a document, the so-called *document-topic proportions*, is one. Equally, the sum of the word probabilities or *topic-word distributions* for a given topic is also one (Roberts et al., 2019). The input for TM is the collection of our raw job postings transformed into a *document-term matrix* representation, DTM. DTM represents the *corpus* of our words or terms as a *bag of words* or terms[48]. DTM is usually sparse and allows us to analyse the data using vectors and matrix algebra to filter and weigh the essential features of our documents collection (see additional details on these procedures applied in the second essay, section 3.5.1.2.2.1). Also, a critical input is the number of topics to be considered in the model. The researcher must choose this number based on some criterion (e.g., the held-out log likelihood proposed by Wallach et al., 2009) or it can be estimated following strategies developed for this purpose (e.g., the Anchor Words algorithm developed by Lee & Mimno, 2014).

Most of TM assume that document collections are unstructured since all documents arise from the same generative model without considering additional information (Roberts et al., 2014). Instead, in this study, we implement the STM developed by Roberts et al. (2016, 2013). STM incorporates document metadata into the standard TM approach to *structure* the document collection, i.e., STM accommodates corpus structure through document-level covariates affecting topical prevalence. This

[48] We use "words" and "terms" as interchangeable concepts which can refer to a unique word or unigram, two words or bi-gram, and so on.

feature contrasts with other TM like LDA. Thus, the critical contribution of STM is to include the covariates into the prior distributions for *document-topic proportions* and *topic-word distributions*. These document-level covariates can affect the *topical prevalence,* i.e., the proportion of each document devoted to a given topic, and we can measure these changes (Roberts et al., 2013). Also, we can evaluate the *topical content,* which refers to the rate of word use within a given topic, but we do not implement this evaluation here.

In this study, we applied the STM topical prevalence model, which examines how much each topic contributes to a document as a function of explanatory variables or *topical prevalence covariates*. In our case, the covariate corresponds to our dummy $27F$ stated by Eq. (4.13), showing that our collection of job postings comes from the pre- and post-disaster periods. Next, we examine the topical prevalence variation between these two periods by applying a treatment effect regression.

In the next sections, we describe the specification and estimation of the STM topical prevalence model.

4.4.1. STM Topic-prevalence model specification

This section and the subsequent 4.4.2 follow the descriptions and technical guidelines detailed in Roberts et al. (2016, 2019, 2013, 2014) and Grajzl & Murrell (2019). As a model based on word counts, STM defines a data generating process for each document, and the observed data are used to find the most likely values for the parameters specified by the model.

The specification starts by indexing the documents by $d \in \{1 \dots D\}$ and each word in the documents by $n \in \{1 \dots N_d\}$ in our DTM representation. The observed words, $w_{d,n}$, are unique instances of terms from a vocabulary of size V (our corpus of interest) that we indexed by $v \in \{1 \dots V\}$. Regarding the addition of covariates for examining the topical prevalence, a designed matrix denoted by $\boldsymbol{X}$ holds this information. Each row defines a vector of document covariates for a given document. $\boldsymbol{X}$ has dimension $D \times P$ (where p indexes the covariates in the design matrix $\boldsymbol{X}$, $p \in \{1 \dots P\}$). The rows of $\boldsymbol{X}$ are represented by $\boldsymbol{x}_d$. Finally, the specification of the number of topics K is indexed by $k \in \{1 \dots K\}$.

Overall, the generative process considers each document, d, as beginning with a collection of N_d empty positions, which are filled with terms[49]. The filling process starts with the number of topics chosen by the researcher (details below in section 4.4.2.2) to build a vector of parameters of dimension K of a distribution that produces one of the topics $k \in \{1 \dots K\}$ for each position in d. This vector is the so-called *topic-prevalence vector* since it contains the probabilities that each of the k topics is assigned to a singular empty position. STM models the topic-prevalence vector as a function of the covariates to estimate the document properties' influence on topic-prevalence. The process continues with selecting

[49] Since our data is represented as a DTM or *bag of words* representation we can assume that, for a given document, all positions are interchangeable. Thus, the choice of topic for any empty position is the same for all positions in that document (Grajzl & Murrell, 2019)

terms from the V vocabulary to generate a k-specific vector of dimension V, which will contain the probabilities of each term to be chosen to fill an empty position.

Formally, the generative process for each d, given the vocabulary of size V and observed words $\{w_{d,n}\}$, the number of topics K, and the design matrix $\boldsymbol{X}$, for our STM Topic-prevalence model specification can be represented as a four-step method. First, we draw the topic-prevalence vector from a logistic-normal generalised linear distribution (Roberts et al., 2019), with a mean vector parameterized as a function of the vector of covariates. This specification allows the expected topic proportions to vary as a function of the document-level covariates, as follows:

$$\vec{\theta}_d | X_d\gamma, \Sigma \sim \text{LogisticNormal}(X_d\gamma, \Sigma), \tag{4.14}$$

where $\vec{\theta}_d$ is the topic-prevalence vector for document d, X_d is the 1-by-p vector, and γ is the p-by-$(K-1)$ matrix of coefficients. Σ is a $(K-1)$ -by- $(K-1)$ covariance matrix that allows for correlations in the topic proportions across documents. The covariates' addition into the model allows the observed metadata to influence the frequency of discussion in the corpus for a given topic. In our specification, the covariate corresponds to the $27F$ dummy stated by Eq. (4.13).

Secondly, given the topic-prevalence vector $\vec{\theta}_d$ from Eq. (4.14), for each n word within document d, which is the process of filling the empty positions $n \in \{1 \dots N_d\}$, a topic is sampled and assigned to that position from a multinomial distribution as follows:

$$z_{d,n} \sim \text{Multinomial}(\vec{\theta}_d), \tag{4.15}$$

where $z_{d,n}$ is the topic assignment of words based on the document-specific distribution over topics, where the k^{th} element of $z_{d,n}$ is one and the rest are zero for the selected k.

Thirdly, we form the document-specific distribution over terms representing each topic k choosing specific vocabulary words v as follows:

$$\beta_{d,k,v} | z_{d,n} \propto \exp(m_v + k_{k,v}), \tag{4.16}$$

where $\beta_{d,k,v}$ is the probability of drawing the v-th word in the vocabulary to fill a position in document d for topic k. m_v is the marginal log frequency estimated from the total word counts of term v in the vocabulary V, representing the baseline word distribution across all documents. $k_{k,v}$ is the topic-specific deviation for each topic k and term v over the baseline log-transformed rate for term v. $k_{k,v}$ represents the importance of the term, given the topic. The logistic transformation of m_v and $k_{k,v}$ converts their sum into probabilities for use in the subsequent and final step, which refers to drawing an observed word conditional on the chosen topic.

Fourthly, the observed word $w_{d,n}$ is drawn from its distribution over the vocabulary V to fill a position n in document d as follows:

$$w_{d,n} \sim \text{Multinomial}(\beta_{d,k,1}, \dots, \beta_{d,k,V}) \tag{4.17}$$

Also, default regularizing prior distributions are used for γ in Eq. (4.14) and k in Eq. (4.16). The regularizing prior distributions refer to zero mean Gaussian distribution with shared variance parameter i.e.$\gamma_{p,k} \sim Normal\left(0, \sigma_k^2\right)$ and $\sigma_k^2 \sim Inverse - Gamma(1,1)$ (Roberts et al., 2016), where p and k indexes the covariates and topics, respectively, as shown above.

4.4.2. STM Topic-prevalence model and effect estimation

This section outlines the techniques used to process our text data, to estimate the number of topics, the parameters inference of our STM Topic-prevalence model and, based on these parameters, to estimate the effect of our natural experiment on topic-prevalence. For our purposes, we use R packages like `Quanteda` (Benoit et al., 2018) to manage and analyse text data. The STM specification and estimation, as well as the treatment effect analysis, is performed using the `Stm` R package (Roberts et al., 2016, 2019, 2020).

4.4.2.1. Pre-processing and DTM representation

We perform standard pre-processing procedures on our collection of 4,136 job postings (see section 4.3 for details). As pointed out above, since our analysis does not deal directly with text data but is performed on specific text features such as word frequencies, we construct a DTM representation (Welbers et al., 2017). We apply cleaning, tokenization and stemming as the pre-processing procedures to construct our DTM following techniques applied in the second essay (see section 3.5.1.2.2.1). We use unigrams (unique words) and bigrams (two consecutive words) as *tokens* or *features*. The use of bigrams allows us to capture text structure or context that we cannot see using single words. For example, in the case of some job titles with generic words like "Engineer", including bigrams might make tokens more comprehensible since we are observing terms like "Software Engineer", "Construction Engineer", etc. We also apply the removal of infrequent terms by dropping features that do not appear in at least ten documents.

4.4.2.2. Estimating the number of topics, K, and the STM topic prevalence model parameters

We estimate K by applying the Anchor Words algorithm (Lee & Mimno, 2014). This technique infers K by finding an approximated convex hull or the smallest convex polygon in a multi-dimensional word co-occurrence space given by our DTM representation. The central assumption of the Anchor Words algorithm is *separability*, i.e., each topic has a specific term that appears only in the context of that particular topic. This separability assumption implies that the terms corresponding to vertices are anchor words for topics. Alternatively, the non-anchor words correspond to the point within the convex hull. We expect a K between 5 and 50, which is the range suggested for a small collection of documents, i.e., a few hundred to a few thousand (Roberts et al., 2020, pp. 65), like our dataset (See section 4.3).

Also, since there is no *true* K parameter (Lee & Mimno, 2014; Roberts et al., 2016, 2019), we apply a K data-driven search as *confirmatory analysis*. Therefore, we conduct an examination across different topic numbers to select the proper specification from the computation of diagnostics, such as the held-out log likelihood (Wallach et al., 2009) and residuals analysis (Taddy, 2012). The held-out log likelihood test evaluates the prediction of words within the document when those words have been removed from the document to estimate the probability of unseen held-out documents (given some training data). For the best specification, on average, we will observe a higher probability of held-out documents indicating a better predictive model. In practical terms, we plot the number of topics and their held-out likelihood to look for some breaks in this relationship as a diagnostic showing that additional topics are not improving this likelihood much. Related to the residual analysis, it evaluates the variance overdispersion of the multinomial described by Eq. (4.15) within the data generating process. An appropriate number of topics will restrict this dispersion. In a plot showing K and their estimated dispersion or residuals level, we are interested in the number of topics with lower values.

Regarding the STM Topic-prevalence model estimation, the strategy takes the DTM, K and the covariate and returns fitted model parameters. To put it differently, given the observed data, K and our $27F$ dummy, it estimates the most likely values for the model parameters specified by maximizing the posterior likelihood (see section 4.4.1). As a result, we can examine the proportion of job postings devoted to a given topic, or topical prevalence, over the $27F$ dummy. However, as occurs in this kind of probabilistic model, the STM posterior distribution is intractable. Therefore, we apply the approximate inference method implemented by Roberts et al. (2019). This method, the so-called partially-collapsed variational expectation-maximization algorithm, posterior variational EM, gives us, upon convergence, the estimates of our STM Topic-prevalence model. We discuss our convergence evaluation below.

Another complexity that follows from the intractable nature of the posterior is the starting value of the parameters: in our case, this is the initial mixture of words for a given topic. This complexity is known as initialization, and our estimation depends on how we approach it. We specified the initialization method using the default choice named "Spectral"[50]. The spectral algorithm is recommended for a large number of documents like ours (Roberts et al., 2020, pp. 65). The described estimation is executed with a maximum number of 200 posterior variational EM iterations subject to meeting convergence. Convergence is examined by observing the change in the approximate variational lower bound. The model is considered converged when the change in the approximate variational lower bound between the iterations becomes very small (default value is 1e-5). We use functionalities included in the R package `Stm` (Roberts et al., 2020) to perform the estimation of K and STM topic-prevalence model parameters.

[50] The spectral initialization is based on the technique of moments and it employs a spectrum decomposition (non-negative matrix factorization) of the word co-occurrence matrix (Roberts et al., 2016, 2020).

In practical terms, the STM Topic-prevalence estimation described above allows us to measure how much a given topic contributes to each of our online job postings. We interpret our result by inspecting the estimated mixture of terms associated with topics. We include the most important terms for each topic using metrics like the highest probability terms and the FREX terms (Roberts et al., 2019). FREX[51] measures the exclusivity of that term to a given topic. This association between terms, documents and topics is the result of the estimated model. However, for the sake of clarity, we name each topic according to our interpretation of the set of terms that motivates each of them. Thus, we can find topics associated with ICT labour. Since we specified the topical prevalence as a function of the $27F$ dummy (see Eq. (4.14) related statements), we can measure the ICT labour topic prevalence variation between the pre- and post-disaster periods. We detail this effect treatment estimation in the next section.

4.4.2.3. Treatment effect estimation and evaluation

Once we have estimated our STM Topic-prevalence model, the fitted parameters allow us to estimate a regression using the online job postings as units or documents, d, to evaluate the influence of our dummy $27F$ defined by Eq. (4.13) on topic-prevalence for a topic j (Roberts et al., 2019). Since $27F$ indicates whether the job posting was published in the period before the earthquake impact or after, i.e., in the post-disaster or "treated" period (see section 4.3), we can study how the prevalence of topics changes in the aftermath of the disasters. In other words, we evaluate the "treatment effect" of the disaster on the topical prevalence by examining changes in topics' proportions over our sample of job postings published after the earthquake. The effect estimates are analogous to Generalized Linear Models, GLM, coefficients (Roberts et al., 2013).

We compute the topic proportions from the θ matrix where each column is the topic-prevalence vector for document d, $\vec{\theta}_d$ (see Eq. (4.14)), and rows are d. Thus, each element $\theta_{d,j}$ is the probability of job posting d being assigned to topic j. As an illustration, in a model with only two topics, we consider the probability of each job posting for each of these two topics. In this example, for job posting d we can denote its proportions over the two topics as $\theta_{d,1}$ and $\theta_{d,2}$ where $\theta_{d,1} + \theta_{d,2} = 1$. Thus, the regression to evaluate the treatment effect where the topic proportions for a given topic are the outcome variable can be represented as

$$\vec{\theta}_d = \alpha + \beta * 27F_d \tag{4.18}$$

where α is the intercept and β is the coefficient to be estimated. A significant β can be interpreted as changes (positive or negative) in topical prevalence because of our dummy standing for the post-disaster period.

---

[51] FREX terms corresponds to labelled terms using a variation of the Frequency-Exclusivity algorithm (Bischof & Airoldi, 2012) available in the `Stm` R Package.

The effect estimation procedure in the `Stm` R package relies on simulated draws of topic proportions from the EM variational posterior (see section 4.4.2.2) to compute the coefficients. We use the default value of 25 simulated draws to compute an average over all the results. In other words, the procedure randomly samples topic proportions from the estimated topic proportion distributions for each job posting repeatedly to estimate any given effect. Also, as suggested by the software's authors, we include estimation uncertainty of the topic proportions in uncertainty estimates, or "Global" uncertainty, using the method of composition (Roberts et al., 2019, pp. 19). Regression table results will display the various quantities of interest (e.g., coefficients, standard error, t-distribution approximation). The procedure uses 500 simulations (default value) to obtain the required confidence intervals in the standard error computation (draws from the covariance matrix of each simulation) and a t-distribution approximation (Roberts et al., 2020). We also show our results visually by displaying the contrast produced by the change in topical prevalence, shifting from the pre-disaster to the post-disaster periods, using the mean difference estimates in topic proportions.

Regarding the evaluation of our estimation, although the robustness of the treatment effect estimation implemented here in terms of *spurious effect*[52] has been validated by using several tests (e.g., Monte Carlo experiments) (Roberts et al., 2014), we still apply a permutation test[53] to evaluate the robustness of our findings. The procedure estimates our model 100 times, where each run applies a random permutation of our $27F$ dummy to the job postings or documents. Then, the largest effect on our topics of interest is calculated. We would find a substantial effect, regardless of how we assigned the treatment to documents, if the results connecting treatment to topics were an artefact of the model (Roberts et al., 2014). Alternatively, we would find a treatment effect only in the case where the assignment of our $27F$ dummy aligned with the true data. We present the results of our permutation tests by plotting the contrast between our permutated model and the true model for our topics of interest.

4.5. Results

This section outlines the results from our STM estimation, following the strategies detailed in section 4.4.2 in three subsections. First, we show the results from our pre-processing procedures and DTM construction. Second, we describe our estimation outputs for K and model parameters by characterizing the discovered topics and identifying the ICT labour topic. Third, we show the effect estimation findings and evaluation focusing on ICT labour topic changes.

[52] A *spurious effect* estimation refers to the model estimating an effect when the effect is actually zero.
[53] We apply the test available in the `Stm` R package. In this test, rather than using the true assignment of our $27F$ dummy, the $27F$ variable is randomly assigned to a job posting with probability equal to its empirical probability in the sample (Roberts et al., 2019).

4.5.1. Pre-processing and DTM representation

This subsection aims to show the pre-processing and DTM results by applying the techniques described in section 4.4.2.1. to our sample described in section 4.3. After applied cleaning, tokenization and stemming, our DTM matrix is compound by 4,136 documents, 63,038 features (99,9% sparse) and one covariate ($27F$ dummy). However, we find an important number of features belonging only to a few documents. In this regard, we remove infrequent terms by dropping features that do not appear in at least ten documents. As a result, our DTM now has 4,129 documents and 2,748 terms whose frequency is in the range [11, 2,095].

In Table 4.2, we show the 15 most frequent terms in our DTM representation. Overall, the terms refer to the most frequent words in job titles and job areas that characterize our collection of job postings, such as sales, customer service, commercial, and management. Also, in column "Document frequency", last column in Table 4.2, we can observe how frequent the features are allocated to documents. For example, in the second row, "client" in Spanish ("customer" in English) is the most represented feature since it is found in 1,210 of the job postings in our sample. We perform on DTM our STM Topic-prevalence model strategy, whose results are shown in the next section.

Table 4.2. The 15 most frequent DTM terms

Feature (stem word in Spanish)	Feature (in English)	Frequency	Rank	Document frequency
vent	sales	2,095	1	908
client	customer	2,083	2	1210
tecnic	technical	1,857	3	1102
manej	handling	1,644	4	1098
comercial	commercial	1,637	5	881
profesional	professional	1,557	6	1130
equip	team	1,400	7	1045
ingenier	engineering	1,397	8	774
servici	service	1,356	9	930
nivel	level	1,313	10	981
gestion	management	1,183	11	758
control	control	1,030	12	646
respons	responsibility	1,008	13	887
administr	management	1,004	14	617
administracion	management	999	15	610

Note: Own English translation of features considering the most probable Spanish stem word

4.5.2. Estimating K and STM Topic-prevalence model parameters

This section shows the findings from our estimation strategies detailed in section 4.4.2.2. The number of topics applying the Anchor Words algorithm yielded a K equal to 53. Our alternative data-driven search of K produces similar results, as shown in Figure 4.6. The left-hand plot corresponds to the held-out log-likelihood application. We see a "break" between 40 and 50 topics. After that point, we see more minor improvements in the log-likelihood by adding more topics. In the case of the residual analysis, the right-hand side plot of Figure 4.6 shows the lower dispersion levels between 50 and 60 topics. In this regard, we can validate our K equals 53 since this quantity falls approximately within the estimated ranges from both data-driven measures.

Figure 4.7 shows the distribution of the expected topic proportions for the 53 topics over our job posting distribution. The x-axis corresponds to the expected topic proportion, and topic labels highlight the three words of highest probability (stem words in Spanish).

Figure 4.6. Diagnostics values of held-out log-likelihood (left-hand plot) and residuals (right-hand plot) by number of topics

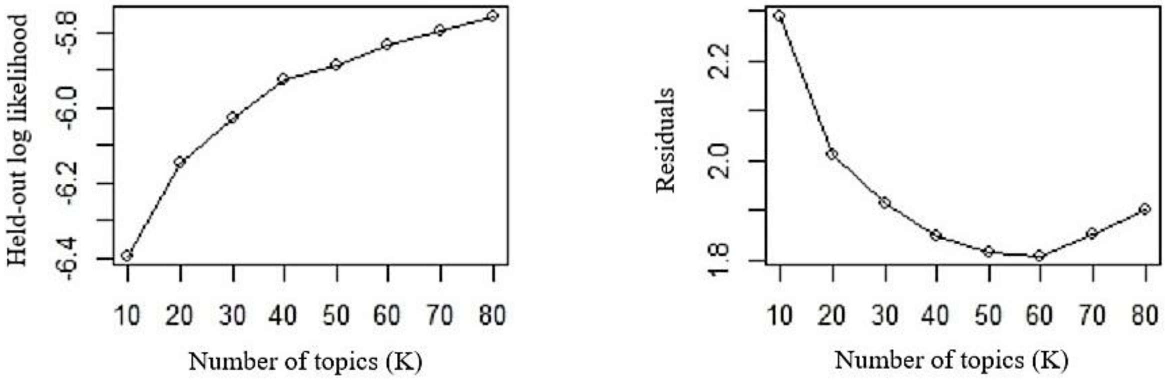
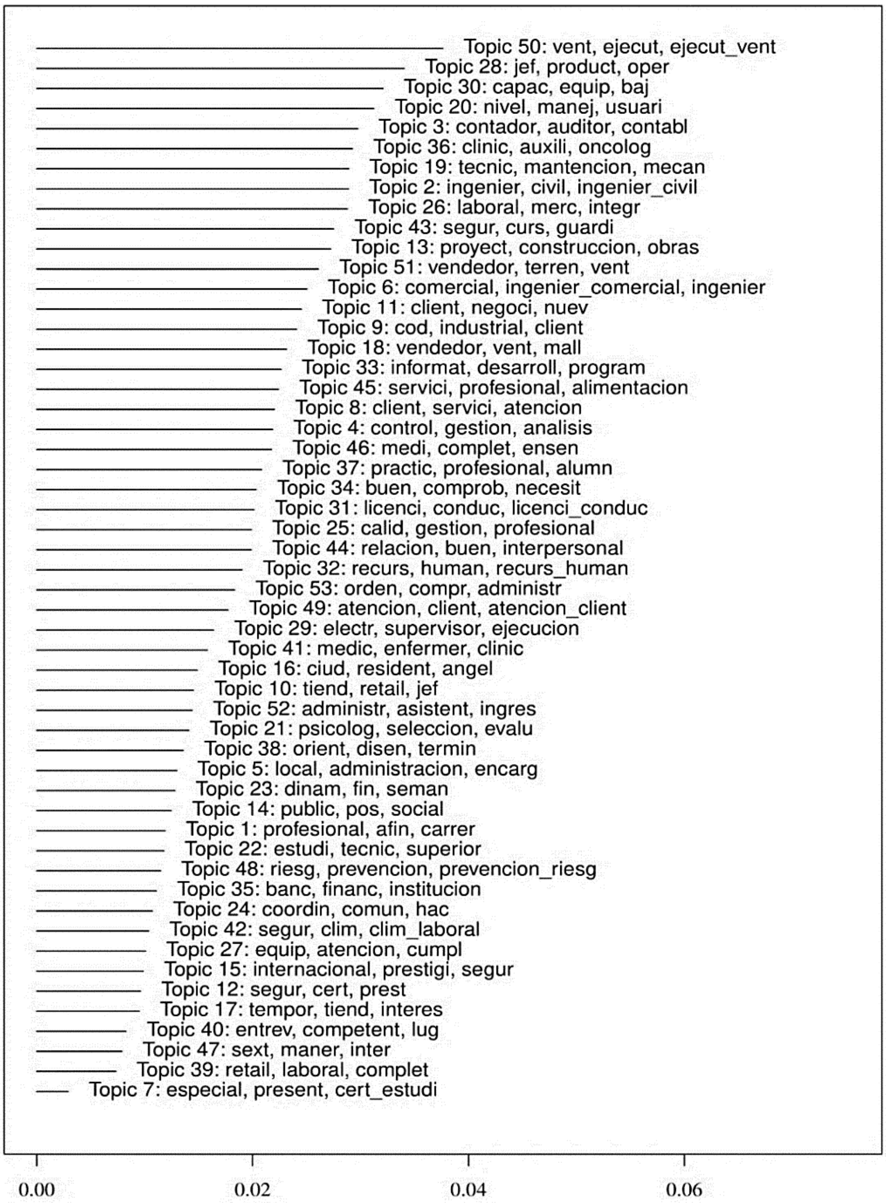


Figure 4.7. Expected topic proportions (x-axis) and the three highest probability words (in Spanish) for the 53 topics



The highest topic proportion in Figure 4.7 corresponds to Topic #50 with the associated terms "vent", "ejecut", and "ejecut_vent". Translated into English, these terms are sales, executive, and sales executive, respectively, implying that most of our collection of jobs is devoted to sales-related jobs.. We examine the 53 topics and name them based on the ten most probable words and FREX terms (See footnote 51). In Appendix A.3, we show the full details of high probability and FREX terms and our proposal of names for topics (in Spanish and English).

Returning to Figure 4.7, we look at topics standing for ICT labour. We find that Topic # 33 (top half of Figure 4.7) can be interpreted as an *ICT labour topic*, given that the most probable terms, i.e., stem words in Spanish, are "informat", "desarroll" and, "program". These words, as non-stem English words, would be *informatics*, *development* and *programming*, respectively. Additional FREX terms stand for English words like *data*, *support*, and *database* (see Topic #33 in Appendix A.3). Furthermore, software or programming languages belong to this topic (e.g., *SQL*, *PHP*). We do not observe other topics with similar terms, suggesting that only our topic of interest contains the expected mixture of ICT-related words.

We adopt the same approach to the interpretation of the rest of our topics: that is, analysing the higher probability and FREX top words. Overall, topics refer to occupational or economic areas (e.g., Sales, Accountancy, Logistics, Health, Education). Also, some of them correspond to specific job titles (e.g., Retail Store Manager, Management Assistants) and job posting sections (e.g., job posting rewards, job posting qualifications requirements). Furthermore, we cannot interpret some topics (we have denoted them as "Undefinable") since we do not see a clear concept emerging from the mixture of words.

In the next section, we examine the treatment effect of the disaster on the topical prevalence of our ICT labour topic. Also, for comparative purposes, we examine the Construction labour topic (Topic #13 in the top half of Figure 4.7) since it is expected that reconstruction activities after the earthquake would encourage the post-disaster prevalence of this topic.

4.5.3. Effect estimation of the earthquake

This section outlines the effect estimation results, as described in section 4.4.2.3. We focus on the prevalence of ICT labour and Construction labour topics. In Table 4.3, we present the results for the regression represented by Eq. (4.18).

Table 4.3. Effect treatment regression results for ICT labour and Construction labour topics prevalence

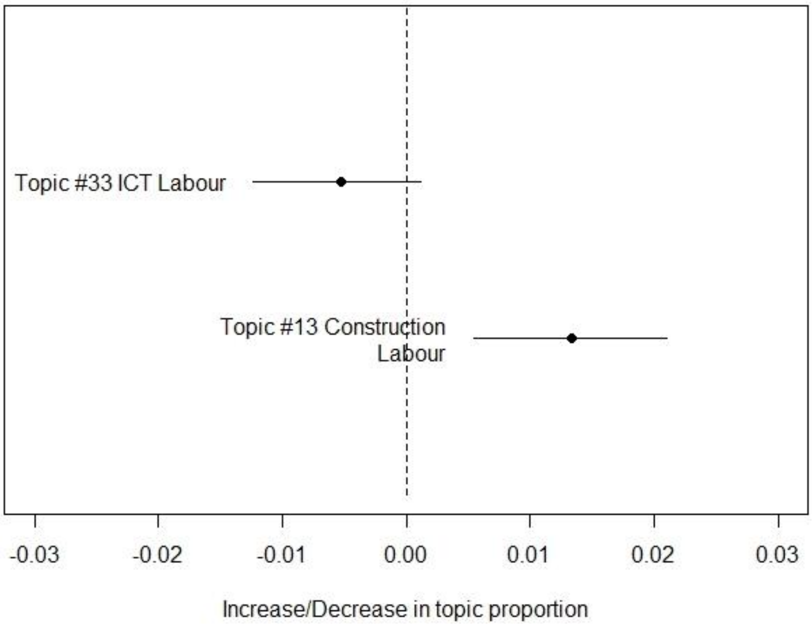
Topic	Variable	Estimate	Std. Error	t value	Pr(>|t|)	
#33 – ICT labour	$Intercept$	0.025684	0.002664	9.643	<2e-16	***
	$27F$	-0.00536	0.003446	-1.555	0.12	
#13 – Construction labour	$Intercept$	0.018878	0.00275	6.866	7.59e-12	***
	$27F$	0.013366	0.003864	3.459	0.000548	***

Note: ***, ** and, * denote significance at 1%, 5% and 10% level respectively.

The first two rows in Table 4.3 stand for the ICT labour topic coefficients. We can see that the $27F$ covariate is not statistically significant, using the ICT topical prevalence as the output variable. In contrast, $27F$ is significant (p-value < 0,01) and positive for the Construction labour topical prevalence. These findings show that the prevalence of the ICT labour topic does not change, suggesting that there is no difference in demand for ICT labour. Conversely, the Construction topic prevalence is significantly different and positive after the disaster, suggesting that reconstruction activities took place in the earthquake's aftermath.

Visually, Figure 4.8 shows that topical prevalence differed significantly and positively between the pre-disaster and post-disaster periods only for the Construction labour topic.

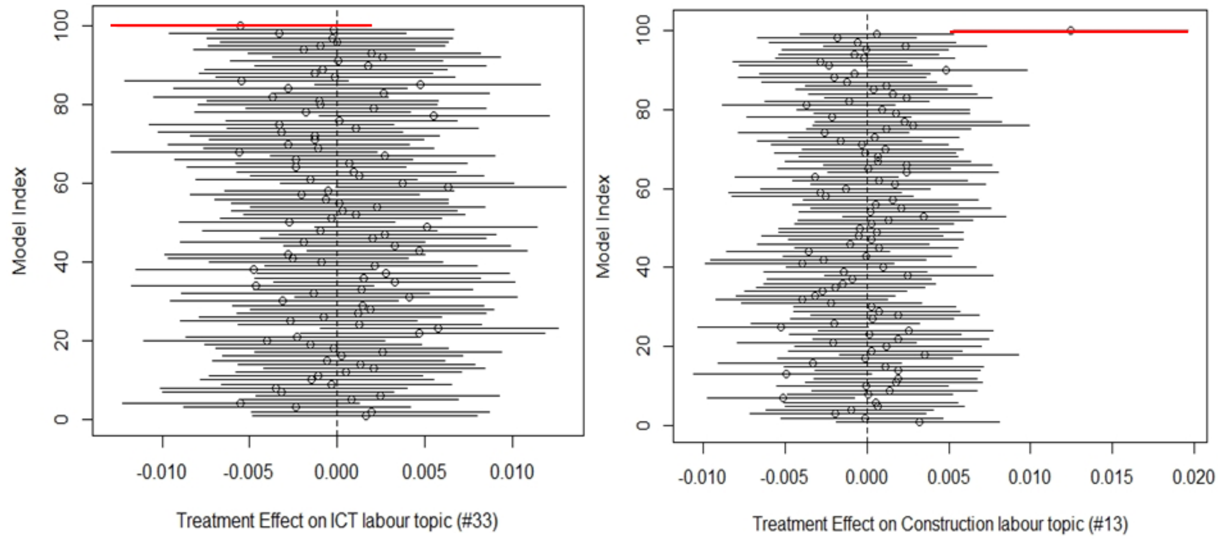
Figure 4.8. Difference in topic prevalence between pre-disaster and post-disaster periods for ICT and Construction labour topics.



Note: Negative and positive values indicate that the topic is more prevalent in pre- and post-disaster periods, respectively. (Confidence intervals at 95%)

Figure 4.9 shows the results of our permutation test (see section 4.4.2.3 for details). For the ICT labour topic (left-hand plot), the permutation output suggests that our results of no change in topic proportions are robust since the models with a random permutation of our $27F$ dummy and our model with the true assignment of our variable, shown by the red line on the top of the plot, have effect sizes around zero. In the case of the Construction labour topic (right-hand plot), most of the estimated models have effect sizes grouped around zero. However, the model including the true assignment of our $27F$ dummy, shown by the red line on the top of the plot, is a result that is far to the right of zero. Thus, the relationship between the treatment and examined topics arises within the sample, and it is not driven by the estimation method itself.

Figure 4.9. Permutation test results for the ICT labour topic (left-hand plot) and Construction labour topic (right-hand plot)



Note: Confidence intervals at 95%

4.6.Discussion

This study examined the impact of the 27th of February Biobío earthquake on demand for ICT labour as a proxy for technological replacement. We do not find evidence that this large earthquake (>8 M_W) influenced the demand for ICT labour, which was represented by a topic featuring ICT-related terms from our job postings collection. This ICT labour topic corresponds to one of the 53 discovered by the application of our STM-Topical Prevalence modelling and estimation strategy. Our number of topics is as expected, given the number of our job postings (Roberts et al., 2020, pp. 65) and the data-driven measures.

Our treatment effect regression results show that the ICT labour topic prevalence did not change in the earthquake's aftermath. This result suggests no substantive technological change in the most affected regions. We do not have enough data to measure region-specific impacts. This lack of evidence does not support our conceptual framework’s main prediction that the expected technological upgrading with ICT compatible equipment would lead to a faster growth in demand for ICT labour. Unlike other studies on shocks like pandemics and recessions, as far as we know, this is the first study that has attempted to link ICT labour with natural disasters. Most of the literature emphasises the importance of ICT as a means of coping with disaster prevention and disaster management.

We can speculate as to the reasons why we have not observed evidence of technological upgrading after the earthquake. First, there is the sectorial structure of the Chilean economy. Assuming that older and outdated physical assets are more prone to be damaged by an earthquake because of weaker structure, mechanical fatigue, and other vulnerabilities (Okuyama, 2003), there is a relatively low representation of the sectors accounting typically for these tangible physical assets, like the manufacturing industry. As Chile has grown, its economic development has been more concentrated in the services sector, which accounts for mainly intangible assets, while manufacturing and other sectors have declined (de la Torre et al., 2013; Parro & Reyes, 2017). In the Chilean GDP structure, the services sector accounts for more than half of GDP, whereas the manufacturing sector in GDP decreased from over 20% in the 1980s to 10% by 2010 (World Bank, 2022). Consequently, the potential negative impact of a disaster on an underrepresented sector like manufacturing might be untraceable. In addition, the predominance of the services sector also can explain the lack of evidence since it has been suggested that this sector, given the intangible nature of its assets and operations, does not suffer the impact of natural disasters as severe as, for example, manufacturing (Doytch, 2020).

Secondly, comparative studies also suggest Chile may be well equipped to cope with disasters due to factors including building policies and codes and economic conditions. Severe economic damage was expected in the aftermath of the 27th of February Biobío earthquake because it affected the central regions of the country, where most of the economic activity and population are concentrated. However, the detrimental effects on the economy were much less than those observed in low-income countries like Haiti, when it was hit by a less severe earthquake (7 M_W) in January 2010 (Cavallo & Noy, 2010; Congressional Research Service, 2010). Another possibility suggested by past studies is that economic innovations usually appear when the economy completely recovers from a disaster (Park et al., 2017). In this regard, a longer-term analysis could capture technological upgrading by observing changes in demand for ICT labour.

As for our findings on Construction employment, they are in line with our expectations and past studies (e.g., Belasen & Polachek, 2009; Skidmore & Toya, 2002). The positive impact on this labour sector suggests that reconstruction activities took place in the earthquake's aftermath. This positive influence might occur as labour is substituted for damaged or missing physical capital in this sector. Some consequences of reconstruction activities leading to growth in Construction employment may be a potential decrease of workers in other sectors, such as Agriculture, attracted by better salaries (Kirchberger, 2017). In this regard, some authors suggest that rebuilding activities favour unskilled and less-educated workers due to increases in demand for the Construction sector, which is a highly intensive employer of unskilled labour (Di Pietro & Mora, 2015). Less favoured groups, like migrants, can also see improvements in their labour outputs during recovery stages (How & Kerr, 2019). The analysis of these positive influences of disasters on labour is beyond the scope of this study, but it represents an opportunity for further research.

There are some caveats, mainly methodological, to the study that deserve mention. First, there is potential ambiguity in the discovered topics. We cannot interpret some of them. This difficulty might be greater for researchers with no prior knowledge of the data or who are analysing text in a foreign language. Secondly, given that STM is recent, both its utility and limitations are still developing. In the case of the treatment effect estimation implemented in this study, there have been some warnings about the modelling of topic proportions, such as that STM ignores the fact that proportions belong to the interval [0, 1] and the regression approach combining Bayesian and frequentist methods[54] (Schulze et al., 2021). Improvements in tackling these limitations should be implemented in future versions of STM.

As suggestions for future research, we would suggest focusing on a more disaggregated analysis, theoretical development and extending the post-disaster period under examination. The importance of research differentiating labour groups or other distributions of workers lies in its ability to facilitate the identification of the worst affected or most favoured workers, either in the aftermath of a disaster or during the economic recovery. Typically, aggregated analysis hides impacts on sub-groups (How & Kerr, 2019; Zissimopoulos & Karoly, 2010). Regarding theoretical developments, some authors have made economic generalizations about disaster dynamics, such as the conceptual framework proposed by Okuyama (2003) and reproduced in this study. However, much theoretical work remains to be done. Regarding the extension of the post-period examination, as pointed out above, technological replacements might be not only a short-run but also a middle or long-run decision.

Our analysis speculates on two potential policy implications. First, policymakers can take advantage of recovery activities, considering to a greater extent the potential for technological upgrading. This is of special interest for countries or regions exposed to disasters like Chile, where the lack of technological upgrading in the planning of recovery activities might explain why we cannot observe technological replacements. Policymakers usually emphasize aspects like disasters risk reduction to improve resilience, where upgrading is mainly planned for infrastructure since disasters are seen as a threat to sustainable development (Bello et al., 2021). But a recovery process promoting technological replacements for firms could exploit and encourage potential technological adoption after disasters (Benson & Clay, 2004; Doytch, 2020). For example, policies could promote the upgrading of firms through fiscal incentives (e.g., tax reductions, financial support). In the case of countries receiving greater inflows of external capital in the aftermath of disasters, such as foreign direct investment, FDI, this investment could be attracted by a focus on technological upgrading (Doytch, 2020). Other highly seismic countries, like Japan, supply abundant liquidity to mitigate the financial constraints on businesses located in affected areas (Okazaki et al., 2019).

In the case of Chile, as one of the region's strongest economies, after the 27th of February Biobío earthquake, it had a good chance of receiving support from international financial institutions (e.g., the World Bank, International Monetary Fund), not only for reconstruction (Congressional Research Service, 2010) but also for technological upgrading. But, to the best of our knowledge, there was no strategy in place to consider the issue discussed here. Therefore, we would encourage policymakers to take advantage of reconstruction activities promoting potential technological upgrading by means of, e.g., fiscal incentives, mitigation of financial restrictions, and policies targeting the replacement of

[54] The potential issue regarding mixing Bayesian and frequentist methods arises from the way in which each technique approaches its parameters. For example, while for the Bayesian method the parameters are random variables, the parameters for the frequentist framework are fixed.

industrial technology, as discussed above. In turn, this "forced" upgrading might lead to improvements in sub-groups of workers like ICT labour.

Secondly, more attention must be paid to disaggregated labour, for example, lesser favoured workers employed in recovery activities. These activities supply job opportunities for these workers that might not exist otherwise, which is desirable from a policy perspective. However, reconstruction activities typically employ low-skilled or unskilled workers, as usually occurs in the Construction sector (Rodríguez-Oreggia, 2013). In terms of wages, this unskilled labour appears at the lowest end of the Construction sector's wages (Sisk & Bankston, 2014). In addition, these low-paying jobs are often dangerous. For example, in the aftermath of Hurricane Katrina, it has been suggested that an undocumented and foreign-born labour force carried out the most unsafe reconstruction activities, like demolition (Trujillo-Pagan, 2012). Bearing this in mind, policymakers should promote strategies focused on these most vulnerable workers, such as improvements in workers' prospects by retraining to mitigate the eventual lack of income once the recovery process finishes. Also, more attention should be paid to work safety policies, since hard and hazardous jobs usually employ less favoured workers.

4.7. Conclusion

The impact on ICT employment derived from technological upgrading due to impacts of disasters has not received attention. Nevertheless, disasters can be an opportunity to accelerate technology adoption, which in turn can have a positive impact on demand for specialized labour like ICT labour. This impact, along with increasing demand for labour being used as a substitute for destroyed equipment or labour required for reconstruction activities, can mitigate the negative impact of disasters.

We explored the influence of the 27th of February Biobío earthquake on demand for ICT labour as a proxy for a technological replacement event. Our findings using open text data on jobs, alongside our topic modelling and treatment effect estimations, show that demand for ICT labour did not significantly change in the aftermath of the earthquake. Given these results, we would assert that in the most affected regions, there was no significant technological upgrading or replacements of destroyed equipment by capital goods compatible with ICT. However, we observed an increase in Construction labour. Therefore, and as expected, reconstruction activities featured strongly in the recovery process.

Our lack of support on the influence on ICT labour of shocks like the examined earthquake might reflect features characteristic of Chile, such as building policies, economic conditions, and the size of the manufacturing sector. Furthermore, technological replacements might occur in the medium term or long run or, possibly, when the recovery activities finish. In this regard, future research should examine periods beyond our post-disaster span of two years. Also, we encourage further research, analysing disaggregated labour and developing more theoretical foundations for a better conceptualization of interactions between disasters, labour, and technology.

Finally, we discussed some policy implications given our our lack of support for changes in demand for ICT labour and the increase in construction employment during the recovery process. On the one side, we encourage policies considering technology upgrading as part of recovery process planning, and on the other, we recommend that more should be done to improve the prospects and safety of lesser favoured workers employed in reconstruction activities.

5. Conclusion

Our examination of the interactions between labour markets, technological change and natural disasters in Chile will be concluded in four sections. To answer the research questions we presented, we first summarise and combine the key findings from the three essays, followed by a section noting the study's limitations and suggestions for further research. The third section discusses some potential implications for policymakers and the fourth section concludes with final remarks.

5.1. Results summary

In our first essay (see Chapter 2), we found empirical evidence for the RBET model for Chile, where demand and supply factors can explain the evolution of the skill premium during 1980 – 2018 using recurrent survey labour data. Our measures of the skill premium and the relative demand for skilled labour coming from technology or the SBTC effect show an inverted U-shaped pattern, while the relative supply shows an upward pattern. We support the view on the complementarity between the skill premium and the relative demand for skilled labour coming from technology. Also, we found evidence for the expected inverse relationship between the relative supply of skilled workers and the skill premium, as posited by the RBET conceptualization. Our estimate of 6.5 for the elasticity of substitution between skilled and unskilled implies that both kinds of workers are imperfect substitutes but more substitutable than suggested.

Past studies examining data between the 1960s and 2000s also supported the RBET evidence for Chile (Beyer et al., 1999; Gallego, 2012), but others found results inconsistent with the theoretical expectations of the RBET model due to "improbable estimation results" such as the computation of the wrong sign for the coefficient standing for the supply factor, i.e., a positive sign (Murakami, 2014; Robbins, 1994b). A positive coefficient contradicts the expected negative relationship between the relative supply and the skill premium posited by the RBET model. Besides, a positive coefficient leads to the computation of negative elasticities. Our VECM estimation also yielded the wrong signs. Alternatively, our UCM-Bayesian estimation supports the conceptualization and predictions of the RBET model for Chile in the period 1980-2018: the SBTC and the relative supply of skilled labour drive the evolution of the skill premium. In the context of a race between education and technology, our results support a story where the SBTC was the dominant factor in the pre-2000 period resulting in the increasing skill premium. However, after 2000, the SBTC effect was surpassed by the workforce's educational attainment, leading to the decline of the observed skill premium in recent decades.

In contrast to our first essay (see Chapter 2), in our second study (see Chapter 3), we found weak evidence supporting the ALM model for Chile using online job posting data for 2009-2018, particularly the expected complementarity between cognitive tasks and the skill premium. Furthermore, we did not find evidence of skilled labour abilities like cognitive and social skills driving the skill premium. As in the first essay, we also confirm the skill premium decline in recent decades using alternative data. In this regard, we speculate that a potential explanation for our weak evidence on the expected complementarity between cognitive tasks and skilled labour (and the lack of evidence regarding cognitive and social abilities) may lie in the period of our analysis. Most of our analysis focuses on the skill premium decline. In the first essay (see Chapter 2), our findings suggested the skill premium decreased in the post-2000 period due to the substantial expansion of Chilean tertiary education: this concurred with past studies (Murakami & Nomura, 2020; Parro & Reyes, 2017). The decreasing importance of cognitive tasks and skills in explaining the wages of skilled labour is also consistent with the reassignment of skilled workers to less skilled positions due to technological adoption, i.e., *downward* movements in the occupational ladder observed in the post-2000 period (Almeida et al., 2020; Zapata-Román, 2021).

Using the demand for ICT labour as a proxy for technological replacement, in the third essay (see Chapter 4), our results do not support our conceptual view on the 27th of February 2010 Biobío earthquake affecting the pace of technological change. Our STM estimation strategy using the text from a subsample of online job posting data used in the second essay (see Chapter 3) allowed us, firstly, to discover topics featuring some specialized labour like ICT and Construction labour. Then, we applied a treatment effect regression approach whose results show that the ICT labour topic prevalence did not change between the periods before and after the disaster. On the other hand, we found evidence of positive changes in the prevalence of the Construction labour topic after the earthquake, suggesting that reconstruction activities took place in the disaster's aftermath. Thus, our results do not support our conceptual framework's main prediction that the expected acceleration of ICT-related technological adoption in the aftermath of the earthquake would lead to increases in demand for ICT labour. Reasons that might explain our lack of evidence include the Chilean economy's sectorial structure and high ability to cope with disasters. The argument for the role of the sectorial structure relies on the assumption that the services sector would not suffer a more severe impact from a natural disaster given its intangible nature compared with, for example, manufacturing, which is a sector relying typically on tangible physical assets. These tangible physical assets are more prone to be damaged by an earthquake and potential tsunami that could occur later, but manufacturing only accounted for 10% of GDP in 2010 (World Bank, 2022). In contrast, the services sector accounted for more than 50% of GDP. Regarding the ability of Chile to cope with earthquakes, comparative studies have suggested that some features like strong building policies and economic conditions have been favourable (Cavallo & Noy, 2010; Congressional Research Service, 2010).

5.2. Limitations of the study and suggestions for further research

The main limitations of our work, including some theoretical and technical implementation caveats, as well as some further research areas that might be developed because of the work presented in this thesis, are described below.

i. *Addressing theoretical limitations of the RBET model*

The elasticity of substitution between skilled and unskilled conceptualizations in the RBET model only allows interpretation for values tending to zero, one or ∞ (see Figure 2.1): the absence of an upper threshold allows us to discuss only *more* or *less* substitution between both groups of workers without theoretical support. The elasticity estimates are also sensitive to the use of time trends to proxy the relative demand for skilled labour (Borjas et al., 2012; Fernández & Messina, 2018). Our theoretically unfeasible results in the first essay (see Chapter 2), produced by applying cointegration techniques, show that the technical implementation of the RBET can be difficult. These limitations should be addressed in future studies focusing on bringing data to the RBET model. In this regard, giving more emphasis to alternative estimation methods like the UCM-Bayesian model would be appropriate since it can better handle the technical implementation of the RBET model. More generally, the models employed in this area have assumed stable parameters such as cointegration approaches, which generally employ only linear trends. Bearing this in mind, it is possible that models which allow for the evolution of a wider set of parameters may better describe the underlying phenomena.

ii. *Data and conceptual limitations for testing the ALM model*

The 120 data points of monthly data used in our second essay (see Chapter 3) might be not enough to capture adequate data variation in both the task-content and skills-related analysis. Besides, there is a sub-representation of groups related to skilled labour since most of our sample is devoted to middle or low-skilled occupations, e.g., clerical workers in business and administration occupations, whose task content is less rich in non-routine cognitive tasks. Similarly, in our skills-related analysis, most of the sample refers to Customer Service or Financial abilities, and fewer observations have references to abilities required from skilled labour, like Cognitive and Social skills. This sub-representation of observations standing for cognitive tasks and skills might be a potential bias towards lesser skilled groups, which needs to be considered in future studies on skilled labour.

Conceptually, past studies using different data and estimation methods also suggest that the ALM model predictions would be not fulfilled so completely for Chile as in other high-income countries. Further research should develop “local” versions of the ALM model, suggesting alternative predictions on the interaction of technology and groups of labour based on their skills.

iii. *STM estimation caveats and suggestions for future research*

On the one hand, we were not able to interpret some of the topics discovered. Researchers have noted that topics are usually difficult to decode (Nanni et al., 2016; Schmidt, 2012). Besides, the topic's interpretation relies on researchers' intuitions (Chang et al., 2009) which might lead to biased interpretation[55] (Shadrova, 2021). Further research is needed to take these difficulties into consideration.

On the other hand, STM combines Bayesian and frequentist strategies in the treatment effect estimation. Although for some researchers, a hybrid technique can represent a combination of the best aspects of both strategies, for others, this combination might represent a weakness since the approaches rely on different assumptions. In this regard, we would suggest the implementation of a full Bayesian estimation of treatment effect in the STM technical implementation as proposed by Schulze et al. (2021).

5.3.Policy implications

From the viewpoint of policy design implications, we speculate on how some findings from this work can inform policymakers in two ways since our results could not directly support policy. However, they offer elements for further research and contribute to the current discussion on related policy issues. To illustrate, in the first and second essays (Chapter 2 and Chapter 3, respectively), we have discussed on the urgent need for coordination between the supply and demand for skilled labour and the anticipation of potential negative impacts due to the adoption of new technologies. In our third essay (see Chapter 4), we suggest strengthening the presence of technological replacements in the process of recovery planning, along with paying greater attention to lesser favoured workers employed in low-paying jobs during reconstruction activities.

i. *On the need for coordination between labour and educational policies and the anticipation of potentially pervasive effects due to technological adoption*

The last four decades witnessed strong investments in Chilean higher education. The 18–24 age group enrolled in tertiary education grew from 189,151 (11% of this age group) in 1984 to above 1.2 million (approximately 67% of this age group) in 2018 (INE, 2017; MINEDUC, 2020). These investments have been essential to boost Chile's economic development through the expected transfer of knowledge and skills to jobs (Schneider, 2013; Valiente et al., 2020). However, policymakers making these investments possible do not seem to have considered the economy's ability to absorb the greater availability of better-educated workers in recent decades. This greater supply surpassed the demand for skilled labour coming from technology, the SBTC effect, resulting in the skill premium decline. Besides, it has been suggested that the Chilean labour market might not require intensive use of advanced skills, which might result in high rates of over-qualification and over-skilling (Sevilla & Farías, 2020). In this sense, the nature of the qualifications obtained by better-educated workers also seems to require more careful consideration. For example, in Chile, only 3% of post-secondary students

[55] For example, biased interpretation of topics might result from *apophenia*, i.e., the human proclivity to interpret random groups of elements as meaningful patterns, and *confirmation bias*, i.e., the human tendency to choose patterns that match pre-existing beliefs (Shadrova, 2021).

graduate with degrees in ICT, and only 1% with degrees in STEM-related fields, placing Chile in the lowest position of all OECD countries (OECD, 2018).

Crucially, Chile lacks institutional mechanisms to coordinate the needs of firms with the educational system (Valiente et al., 2020). Some incipient strategies are pointing in the direction of greater coordination or examining the mismatching between demand and supply: these include the development of the National Qualification Framework, NQF (Fuentes et al., 2020; MINEDUC & CORFO, 2017; Sevilla & Farías, 2020) and the institutional monitoring of this mismatching by the Labour Observatory (*Observatorio Laboral* in Spanish) of the Ministry of Work and Pensions (SENCE, 2022). Another innovation implemented in 2021 was the Job Prospection Policy Committee, which aims to balance labour skills with the needs of the labour market, among other objectives (Ministerio del Trabajo y Previsión Social, 2021). Past efforts like the National Vocational Qualifications Framework[56] (*Marco de Cualificaciones Técnico Profesional,* MCTP, in Spanish) proposed in 2021 and with a focus on Chilean secondary and tertiary vocational educational levels, among others, provide an experience that oversees the development of qualifications in key economic sectors (e.g., ICT, mining). The MCTP strategy represents an experience of improving educational and occupational mismatch at the vocational education level that can extend to the university or college level. Also, UNESCO (2015) reviewed NQFs from a variety of countries. Among these countries are some that recently graduated as high-income countries, such as Chile (e.g., Portugal, Republic of Korea; See section 1.1 and footnote 3), which might supply insights on how they address the expected coordination between educational institutions and the industry needs. For example, in the case of the Republic of Korea, the NFQ integrates university and vocational qualifications, and it serves as a conduit for reacting to labour market needs as well as restructuring secondary and tertiary both academic and vocational educational curriculum (UNESCO, 2015).

Furthermore, our evidence on skilled and unskilled labour being more substitutable than commonly assumed, and the lack of a strong relationship between the skill premium and cognitive tasks and skills, would imply an unanticipated impact of technology adoption resulting in unwanted changes in the occupational ladder (e.g., downward movements). Some reported that most technologies biased toward skilled labour came from abroad (Gallego, 2012), which could imply that these technologies are not suitable for Chilean skilled workers. In this regard, innovation policies can support the development of technologies that specifically complement Chilean skilled labour. One good place to start is for policymakers to strengthen intellectual property regulations (Acemoglu, 2003) to promote local technologies. They could also fund practice-based testing of worker interactions with most of the technologies featuring the technological change in production (e.g., ICT, automation) to assess their alignment with Chile's labour market needs and to anticipate potentially pervasive effects. Besides, educational policies promoting ICT and STEM qualifications are required.

[56] Available in https://marcodecualificacionestp.mineduc.cl/ (in Spanish. Accessed 01-Apr-2022)

The issues discussed here highlight the urgent need for efforts to improve the balance between the needs of the labour market: evidence for this arises from our results on the effect of demand and supply factors on the skill premium evolution. Warnings of this urgency have already been raised by past studies and international institutions (OECD, 2018; Sevilla & Farías, 2020; Valiente et al., 2020). We also suggest policies anticipating the potentially pervasive effect of technological adoption, such as unwanted changes in the occupational ladder, i.e., *downward movements*, due to displacements of skilled labour to less skilled positions, as suggested by Almeida et al. (2020) examining the adoption of complex software in Chilean firms. Our results on the weak complementarity between skilled labour and cognitive tasks and skills might also imply that better-educated workers move to middle-skilled positions rich in cognitive but intensive in routine tasks. Bearing this in mind, we should consider that the problem to be addressed is not technology *per se* but failures to understand the interaction between labour markets and technological change or its misgovernance (Goos, 2018).

ii. On a planned recovery process promoting potential technological improvements and policy focused on reconstruction workers

The potential technological upgrading after disasters is of particular interest for countries or regions exposed to disasters like Chile. Nevertheless, Chile, despite being a country hit by recurrent earthquakes, lacks policies designed to take advantage of recovery activities by giving more consideration to technology upgrading. Some experiences show that policymakers might promote technological improvements. For example, these could include fiscal incentives (e.g., tax reductions, financial support), re-directing FDI towards technological upgrading (Doytch, 2020), and mitigation of financial constraints by supplying abundant liquidity for firms to recover their productive capacity by upgrading their equipment's technology (Okazaki et al., 2019). In the case of the 27th of February Biobío earthquake, Chile, as one of the region's strongest economies, was likely to receive financial support from international financial institutions (e.g., the World Bank, International Monetary Fund) not only for reconstruction (Congressional Research Service, 2010) but also for technological upgrading. To the best of our knowledge, no government report on planning considers specialized support for technological upgrading.

We also suggest more attention should be paid to lesser favoured workers. Our results suggest that reconstruction activities took place after the earthquake due to the positive change in the Construction labour prevalence. Reconstruction activities typically employ low-paid or unskilled labourers, as occurred in the Construction sector (Rodríguez-Oreggia, 2013). These activities are usually not only linked to low-paid but also to hazardous jobs like demolition (Trujillo-Pagan, 2012). Trujillo-Pagan, (2012) also suggested that inflows of less favoured workers (e.g., undocumented and foreign-born labour) were promoted by policymakers (e.g., by suspending labour-related policies) to accelerate the recovery process after Hurricane Katrina. However, as suggested by Trujillo-Pagan, (2012), this workforce carried out the most hazardous reconstruction activities. In this sense, reconstruction activities can represent an opportunity to increase employment participation for lesser favoured workers, but policy should not encourage inflows of vulnerable workers without consideration of work safety measures. Besides, once reconstruction or the recovery process has finished, there are other issues with policy implications, such as the potential lack of income of workers employ in reconstruction jobs. In this regard, retraining or strategies supporting the transition to new jobs might be required.

5.4.Final remarks

The work in this thesis contribute to our understanding of labour markets' response to both regular economic forces like technological change and unexpected shocks like natural catastrophes. We have explored the interactions between the skill premium and demand and supply factors by testing the RBET and ALM models' implications as well as the role of cognitive and social skills. We have learned that, in the context of a race between education and technology over 1980-2018 for Chile, the SBTC was the dominant factor in the pre-2000 period resulting in the increasing skill premium. However, after 2000, the SBTC effect was surpassed by the workforce's educational attainment, leading to the decline of the observed skill premium in recent decades. However, we found weak evidence supporting the ALM model for Chile for 2009-2018, particularly the expected positive influence on the skill premium of cognitive tasks and skills. Additionally, this work gives us insights on the question of whether natural disasters can be considered as substantial events affecting the pace of technological change pace proxied by demand for ICT labour. In this regard, we cannot support a conceptual view on recent natural disasters affecting the pace of technological change.

As limitations we have discussed conceptual limitations arising from the theories used as frameworks in this thesis. Also, an added limitation is that they might be considered as unrelated or separate frameworks. However, these theories respond to different research questions aimed at conceptually examine how labour markets respond to regular forces like technological change and unexpected shocks like catastrophes considering the potential technological upgrading. Similarly, and as illustration, we have highlighted policy implications about the lack of coordination between educational and labour policies. Although these implications might be not directly supported by our results, they can feed the policy discussion on the mismatch between educational and labour policies which recently started in Chile. In this regard, our results supply empirical evidence on labour demand and supply forces behind the skill premium evolution.

References

3s Unternehmensberatung. (2020). *DISCO European Dictionary of Skills and Competences*. http://disco-tools.eu/disco2_portal/index.php

Acemoglu, D. (1998). Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality. *The Quarterly Journal of Economics*, **113**(4), 1055–1089. https://www.jstor.org/stable/2586974

Acemoglu, D. (2002). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature*, **40**(1), 7–72. https://doi.org/10.1257/jel.40.1.7

Acemoglu, D. (2003). Patterns of Skill Premia. *Review of Economic Studies*, **70**(2), 199–230. http://www.jstor.org/stable/3648632

Acemoglu, D., & Autor, D. (2011). Skills , Tasks and Technologies : Implications for Employment and Earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 4B, pp. 1043–1171). Elsevier Science & Technology, Oxford, UK. https://doi.org/10.1016/S0169-7218(11)02410-5

Acemoglu, D., & Autor, D. (2012). What does human capital do? A review of Goldin and Katz's The Race between Education and Technology. *Journal of Economic Literature*, **50**(2), 426–463. https://doi.org/10.1257/jel.50.2.426

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, **33**(2), 3–30. https://doi.org/10.1257/jep.33.2.3

Acosta, P., Cruces, G., Galiani, S., & Gasparini, L. (2019). Educational upgrading and returns to skills in Latin America: evidence from a supply–demand framework. *Latin American Economic Review*, **28**, 1–20. https://doi.org/10.1186/s40503-019-0080-6

Aedo, C., Hentschel, J., Luque, J., & Moreno, M. (2013). *From occupations to embedded skills: A cross-country comparison* (No. 6560; Policy Research Working Paper).

Aghion, P., & Howitt, P. (1990). A Model of growth through creative desctruction. In *NBER Working Paper Series* (Issue 3223).

Allen, J., Belfi, B., & Borghans, L. (2020). Is There a Rise in the Importance of Socioemotional Skills in the Labor Market? Evidence From a Trend Study Among College Graduates. *Frontiers in Psychology*, **11**(July), 1–13. https://doi.org/10.3389/fpsyg.2020.01710

Almeida, R. K., Fernandes, A. M., & Viollaz, M. (2020). Software Adoption, Employment

Composition, and the Skill Content of Occupations in Chilean Firms. *Journal of Development Studies*, **56**(1), 169–185. https://doi.org/10.1080/00220388.2018.1546847

Antonczyk, D., Fitzenberger, B., & Freiburg, U. L. (2009). Can a task-based approach explain the recent changes in the German wage structure? *Jahrbucher Fur Nationalokonomie Und Statistik*, **229**(2–3), 214–238. https://doi.org/10.1515/jbnst-2009-2-309

Arntz, M., Gregory, T., & Zierahn, U. (2016). *The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis* (No. 189; OECD Social, Employment and Migration Working Papers). https://doi.org/https://doi.org/10.1787/1815199X

Askilden, J. E., & Nilsen, Ø. A. (2005). Apprentices and young workers: A study of the Norwegian youth labour market. *Scottish Journal of Political Economy*, **52**(1), 1–17. https://doi.org/10.1111/j.0036-9292.2005.00331.x

Atalay, E., Phongthiengtham, P., Sotelo, S., & Tannenbaum, D. (2018). New technologies and the labor market. *Journal of Monetary Economics*, **97**, 48–67. https://doi.org/10.1016/J.JMONECO.2018.05.008

Auria, L., & Rouslan, M. (2008). *Support Vector Machines (SVM) as a technique for solvency analysis* (No. 811; DIW Discussion Papers). http://hdl.handle.net/10419/27334

Autor, D. (2010). *The Polarization of Job Opportunities in the U.S. Labor Market Implications for Employment and Earnings*. https://www.brookings.edu/wp-content/uploads/2016/06/04_jobs_autor.pdf

Autor, D. (2015). Why are there still so many jobs? the history and future of workplace automation. *Journal of Economic Perspectives*, **29**(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Autor, D., & Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US Labor Market. *American Economic Review*, **103**(5), 1553–1597. https://doi.org/10.1257/aer.103.5.1553

Autor, D., Goldin, C., & Katz, L. F. (2020). Extending the Race between Education and Technology. *AEA Papers and Proceedings*, **110**, 347–351. https://www.aeaweb.org/articles?id=10.1257/pandp.20201061

Autor, D., Katz, L. F., & Kearney, M. S. (2006). The Polarization of the U.S. Labor Market. *American Economic Review*, **96**(2), 189–194. http://econ-www.mit.edu/files/5554

Autor, D., Katz, L. F., & Krueger, A. B. (1998). Computing Inequality: Have Computers Changed the Labor Market? *The Quarterly Journal of Economics*, **113**(4), 1169–1213.

Autor, D., Katz, L., & Kearney, M. (2008). Trends in U . S . Wage Inequality : Revising the Revisionists. *The Review of Economics and Statistics*, **90**(2), 300–323.

https://doi.org/10.1016/j.mycmed.2014.02.007

Autor, D., Levy, F., & Murnane, R. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, **118**(4), 1279–1333.

Azevedo, J. P., Dávalos, M. E., Diaz-Bonilla, C., Atuesta, B., & Castañeda, R. A. (2013). *Fifteen Years of Inequality in Latin America. How Have Labor Markets Helped ?* (No. 6384; Policy Research Working Paper).

Baiocchi, G., & Distaso, W. (2003). GRETL: Econometric software for the GNU generation. *Journal of Applied Econometrics*, **18**(1), 105–110. https://doi.org/10.1002/jae.704

Balleer, A., & Rens, T. Van. (2013). Skill-biased technological change and the business cycle. *The Review of Economics and Statistics*, **95**(4), 1222–1237.

Banco Central de Chile. (2020). *Base de datos estadísticos*. Tasa de Desocupación. https://si3.bcentral.cl/siete

Banfi, S., Choi, S., Menzio, G., Samaniego De La Parra, B., Eeckhout, J., Shi, S., Chang, Y., Langot, F., & Kaymak, B. (2019). *Deconstructing Job Search Behavior* (No. 92482; MPRA Working Paper).

Banfi, S., Choi, S., & Villena-Roldán, B. (2020). *Sorting On-line and On-time* (19/706; Discussion Paper). http://www.bristol.ac.uk/efm/media/workingpapers/working_papers/pdffiles/dp19706.pdf

Banfi, S., & Villena-Roldán, B. (2019). Do High-Wage Jobs Attract More Applicants? Directed Search Evidence from the Online Labor Market. *Journal of Labor Economics*, **37**(3), 715–746. https://doi.org/10.1086/702627

Barrientos, S., & CSN Team, N. S. C. (2018). The Seismic Network of Chile. *Seismological Research Letters*, **89**(2A), 467–474. https://doi.org/10.1785/0220160195

Barro, R. J., & Sala-i-Martin, X. (2004). *Economic Growth* (2nd ed.). The MIT Press, Cambridge MA. https://doi.org/10.1159/000241246

Bartel, A., Ichniowski, C., & Shaw, K. (2007). How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills. *The Quarterly Journal of Economics*, **122**(4), 1721–1758.

Bassanini, A., & Scarpetta, S. (2002). Growth, technological change, and ICT diffusion: Recent evidence from OECD countries. *Oxford Review of Economic Policy*, **18**(3), 324–344. https://doi.org/10.1093/oxrep/18.3.324

Beaudry, P., Green, D. A., & Sand, B. M. (2016). The great reversal in the demand for skill and cognitive tasks. *Journal of Labor Economics*, **34**(S1), S199–S247. https://doi.org/10.1086/682347

Beblavý, M., Fabo, B., & Lenaerts, K. (2016). *Skills requirements for the 30 most-frequently advertised occupations in the United States : an analysis based on online vacancy data*. https://www.ceps.eu/ceps-publications/skills-requirements-30-most-frequently-advertised-occupations-united-states-analysis/

Behar, A. (2009). *Directed technical change, the elasticity of substitution and wage inequality in developing countries* (No. 467; Discussion Paper Series). https://ora.ox.ac.uk/objects/uuid:e0c12d8f-9fbb-419d-86ed-ac281c5675e9

Belasen, A. R., & Polachek, S. W. (2009). How Disasters Affect Local Labor Markets: The Effects of Hurricanes in Florida. *Journal of Human Resources*, **44**(1), 251–276. https://doi.org/10.3368/jhr.44.1.251

Bello, O., Bustamante, A., & Pizarro, P. (2021). Planning for disaster risk reduction within the framework of the 2030 Agenda for Sustainable Development. In *Project Documents: Vol. LC/TS.2020*. https://repositorio.cepal.org/bitstream/handle/11362/46639/1/S2000452_en.pdf

Benali, N., & Feki, R. (2018). Natural disasters, information/communication technologies, foreign direct investment and economic growth in developed countries. *Environmental Economics*, **9**(2), 80–87. https://doi.org/10.21511/ee.09(2).2018.06

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, **3**(30), 774. https://doi.org/10.21105/joss.00774

Benson, C., & Clay, E. J. (2004). *Understanding the economic and financial impact of natural disasters* (No. 4; Disaster Risk Management Series). https://openknowledge.worldbank.org/handle/10986/15025

Bergström, V., & Panas, E. (1992). How Robust is the Capital-Skill Complementarity Hypothesis? *The Review of Economics and Statistics*, **74**(3), 540–546. https://www.jstor.org/stable/2109499

Beyer, H., Rojas, P., & Vergara, R. (1999). Trade liberalization and wage inequality. *Journal of Development Economics*, **59**(1), 103–123. https://doi.org/10.1016/S0304-3878(99)00007-3

Biblioteca del Congreso Nacional. (2020). *Leyes de Salario Mínimo*. https://www.bcn.cl/portal/

Bischof, J. M., & Airoldi, E. M. (2012). Summarizing topical content with word frequency and exclusivity. *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*, **1**, 201–208. https://dl.acm.org/doi/10.5555/3042573.3042578

Blankenau, W. F., & Cassou, S. P. (2011). Industry estimates of the elasticity of substitution and the rate of biased technological change between skilled and unskilled labour. *Applied Economics*, **43**(23), 3129–3142. https://doi.org/10.1080/00036840903476361

Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, **3**, 993–1022. https://dl.acm.org/doi/10.5555/944919.944937

Borghans, L., Weel, B. Ter, & Weinberg, B. A. (2014). People skills and the labor-market outcomes of underrepresented groups. *ILR Review*, **67**(2), 287–334. https://doi.org/10.1177/001979391406700202

Borjas, G. J., Grogger, J., & Hanson, G. H. (2012). Comment : On estimating elasticities of substitution. *Journal Ofthe European Economic Association*, **10**(1), 198–223. https://doi.org/10.1111/j.1542-4774.2011.01055.x

Boselli, R., Cesarini, M., Mercorio, F., & Mezzanzanica, M. (2018). Classifying online Job Advertisements through Machine Learning. *Future Generation Computer Systems*, **86**, 319–328. https://doi.org/10.1016/J.FUTURE.2018.03.035

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (3rd ed.). Springer International Publishing, AG, Switzerland. https://doi.org/10.1007/978-3-319-29854-2

Brown, S. P., Mason, S. L., & Tiller, R. B. (2006). The effect of Hurricane Katrina on employment and unemployment. *Monthly Labor Review*, **129**(August), 52–69. https://www.bls.gov/opub/mlr/2006/08/art5full.pdf

Buera, F. J., & Kaboski, J. P. (2012). The rise of the service economy. *American Economic Review*, **102**(6), 2540–2569. https://doi.org/10.1257/aer.102.6.2540

Cantore, C., Ferroni, F., & León-Ledesma, M. A. (2017). The dynamics of hours worked and technology. *Journal of Economic Dynamics and Control*, **82**, 67–82. https://doi.org/10.1016/j.jedc.2017.05.009

Card, D., & Lemieux, T. (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics*, **116**(2), 705–746.

Cavallo, E., Galiani, S., Noy, I., & Pantano, J. (2013). Catastrophic Natural Disasters and Economic Growth. *The Review of Economics and Statistics*, **95**(5), 1549–1561. https://doi.org/10.1162/REST_a_00413

Cavallo, E., & Noy, I. (2010). *The economics of natural disasters* (No. 124; IDB Working Paper Series). https://doi.org/10.4337/9781785365980

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems (NIPS 2009)*, **22**.

Chen, Y.-E., Li, C., Chang, C.-P., & Zheng, M. (2021). Identifying the influence of natural disasters on technological innovation. *Economic Analysis and Policy*, **70**(20), 22–36. https://doi.org/10.1016/j.eap.2021.01.016

Ciccone, A., & Peri, G. (2005). Long-run substitutability between more and less educated workers: evidence from U.S. States, 1950-1990. *The Review of Economics and Statistics*, **87**(4), 652–663. https://www.mitpressjournals.org/doi/pdf/10.1162%2F003465305775098233

Coffman, M., & Noy, I. (2011). Hurricane Iniki: measuring the long-term economic impact of a natural disaster using synthetic control. *Environment and Development Economics*, **17**, 187–205. https://doi.org/10.1017/S1355770X11000350

Congressional Research Service. (2010). *Chile earthquake: U.S. and international response* (CRS Report for Congress). https://www.everycrsreport.com/reports/R41112.html

Conte, A., & Vivarelli, M. (2011). Imported skill-biased technological change in developing countries. *The Developing Economies*, **49**(1), 36–65. https://doi.org/10.1111/j.1746-1049.2010.00121.x

Contreras, M., & Winckler, P. (2013). Pérdidas de vidas, viviendas, infraestructura y embarcaciones por el tsunami del 27 de Febrero de 2010 en la costa central de Chile. *Obras y Proyectos*, **14**, 6–19. https://doi.org/10.4067/s0718-28132013000200001

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, **20**, 273–297. https://doi.org/10.1109/64.163674

Cottrell, A., & Lucchetti, R. (2021). *Gretl User's Guide*. http://gretl.sourceforge.net/gretl-help/gretl-guide.pdf

Crespo Cuaresma, J., Hlouskova, J., & Obersteiner, M. (2008). Natural disasters as creative destruction? Evidence from developing countries. *Economic Inquiry*, **46**(2), 214–226. https://doi.org/10.1111/j.1465-7295.2007.00063.x

de la Torre, A., Levy Yeyati, E., & Pienknagura, S. (2013). *Latin America and the Caribbean as Tailwinds Recede: In Search of Higher Growth* (p. 60). https://doi.org/10.1596/978-0-8213-9975-0

de Vries, G. J., Gentile, E., Miroudot, S., & Wacker, K. M. (2020). The rise of robots and the fall of routine jobs. *Labour Economics*, **66**(October). https://doi.org/10.1016/j.labeco.2020.101885

Deming, D. (2017). The growing importance of social skills in the Labour Market. *The Quarterly Journal of Economics*, **132**(4), 1593–1640.

Deming, D., & Kahn, L. (2018). Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *Journal of Labor Economics*, **36**(4). https://www.journals.uchicago.edu/doi/10.1086/694106

Deming, D., & Noray, K. (2020). Earnings dynamics, changing job skills, and STEM careers. *The Quarterly Journal of Economics*, **135**(4), 1965–2005. https://doi.org/10.1093/qje/qjaa021.Advance

Dengler, K., Matthes, B., & Paulus, W. (2014). *Occupational Tasks in the German Labour Market - An alternative measurement on the basis of an expert database* (12/2014 EN; FDZ-Methodenreport). https://doku.iab.de/fdz/reporte/2014/MR_12-14_EN.pdf

Di Pietro, G., & Mora, T. (2015). The effect of the L'Aquila earthquake on labour market outcomes. *Environment and Planning C: Government and Policy*, **33**(2), 239–255. https://doi.org/10.1068/c12121r

Didier, N. (2018). About the (non) existence of a lifelong education system in Chile. *Journal of Adult and Continuing Education*, **24**(2), 229–249. https://doi.org/10.1177/1477971418818574

Doytch, N. (2020). Upgrading destruction?: How do climate-related and geophysical natural disasters impact sectoral FDI. *International Journal of Climate Change Strategies and Management*, **12**(2), 182–200. https://doi.org/10.1108/IJCCSM-07-2019-0044

Dupuy, A., & Marey, P. S. (2008). Shifts and twists in the relative productivity of skilled labor. *Journal of Macroeconomics*, **30**(2), 718–735. https://doi.org/10.1016/j.jmacro.2007.08.003

Durbin, J., & Koopman, S. J. (2012). *Time Series Analysis by State Space Methods* (2nd ed.). Oxford University Press, Oxford UK. https://doi.org/10.1093/acprof:oso/9780199641178.001.0001

ECLAC. (2010). *The Chilean earthquake of 27 February 2010: an overview*. https://www.cepal.org/en/publications/3161-chilean-earthquake-27-february-2010-overview

ECLAC. (2013). *The digital economy for structural change and equality*. https://repositorio.cepal.org/handle/11362/35954

Edin, P.-A., Fredriksson, P., Nybom, M., & Ockert, B. (2017). *The Rising Return to Non-Cognitive Skill* (No. 10914; IZA Discussion Paper, Issue 10914). www.iza.org

Eeckhout, J., & Jovanovic, B. (2007). *Occupational choice and development* (No. 13686; NBER Working Paper Series). https://doi.org/10.1016/j.jet.2011.01.002

Ehrenberg, R. G., & Smith, R. S. (2018). *Modern Labor Economics : Theory and Public Policy* (Thirteenth). Routledge, New York.

Elsafty, A., & Elzeftawy, A. (2021). The New Era of Digital Transformation and COVID-19 Effect on The Employment in Mobile Operators in Egypt. *Business and Management Studies*, **7**(1), 74. https://doi.org/10.11114/bms.v7i1.5145

Enders, W. (2015). *Applied Econometric Time Series* (4th ed.). John Wiley & Sons, Ltd, New York.

Engle, R. F., & Granger, C. (1987). Co-Integration and Error Correction: Representation, Estimation and Testing. *Econometrica*, **55**(2), 251–276.

Fernández, M., & Messina, J. (2018). Skill premium, labor supply, and changes in the structure of wages in Latin America. *Journal of Development Economics*, **135**, 555–573. https://doi.org/10.1016/j.jdeveco.2018.08.012

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, **114**, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Fuentes, S. C., Amar, F. K., Mayol, E. M., López, M. L., & Cerda, L. G. (2020). Pertinence of the graduation profiles of engineering programs in chile: Analysis of their consistency with the international agreements and the national qualifications framework for higher education. *Pensamiento Educativo*, **57**(2), 1–21. https://doi.org/10.7764/PEL.57.2.2020.10

Gabry, J., Stan Development Team, Andreae, M., Betancourt, M., Carpenter, B., Gao, Y., Gelman, A., Goodrich, B., Lee, D., Song, D., & Trangucci, R. (2018). *Package "shinystan."* http://mc-stan.org/, http://discourse.mc-stan.org

Galeano, S., & Gallego, P. (2018). Overcoming The " Middle-Income Trap ": Chile Case Study. *Revista Mundo Asia Pacífico*, **8**(13), 19–36. https://doi.org/10.17230/map.7.13.2

Gallego, F. A. (2012). Skill Premium in Chile: Studying Skill Upgrading in the South. *World Development*, **40**(3), 594–609. https://doi.org/10.1016/j.worlddev.2011.07.009

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**(3), 515–533.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2020). *Bayesian Data Analysis* (3rd Ed.). Taylor & Francis, New York.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, **2**(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Gerón, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O`Reilly Media, Inc., Sebastopol, CA.

Gil, D., & Johnson, M. (2011). Support Vector Machines in Medical Classification Tasks. In B. H. Boyle (Ed.), *Support Vector Machines: Data Analysis, Machine Learning and Applications*. Nova Science Publishers, Incorporated, New York.

Gindling, T. H., & Robbins, D. (2001). Patterns and sources of changing wage inequality in Chile and Costa Rica during structural adjustment. *World Development*, **29**(4), 725–745. https://doi.org/10.1016/S0305-750X(00)00117-0

Glitz, A., & Wissmann, D. (2017). Skill Premiums and the Supply of Young Workers in Germany. In *CESifo Working Paper Series* (Issue 6576). https://ideas.repec.org/p/ces/ceswps/_6576.html

Goldin, C., & Katz, L. F. (2008). *The Race Between Technology & Education*. The Belknap Press of Harvard University Press, Cambridge, MA.

Goldin, C., & Katz, L. F. (2009). *The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005* (No. 12984; NBER Working Paper Series).

Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics*, **60**(1–2), 203–233. https://doi.org/10.1016/0304-4076(94)90044-2

Goos, M. (2018). The impact of technological progress on labour markets: Policy challenges. *Oxford Review of Economic Policy*, **34**(3), 362–375. https://doi.org/10.1093/oxrep/gry002

Goos, M., & Manning, A. (2007). Lousy and Lovely Jobs : The Rising Polarization of Work in Britain. *The Review of Economics and Statistics*, **89**(1), 118–133.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring †. *American Economic Review*, **104**(8), 2509–2526. https://doi.org/10.1257/aer.104.8.2509

Grajzl, P., & Murrell, P. (2019). Toward understanding 17th century English culture: A structural topic model of Francis Bacon’s ideas. *Journal of Comparative Economics*, **47**(1), 111–135. https://doi.org/10.1016/j.jce.2018.10.004

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, **37**(3), 424–438. https://www.jstor.org/stable/1912791

Granger, C. W. J., & Newbold, P. (1974). Spurious Regressions in Econometrics. *Journal of Econometrics*, **2**, 111–120. https://doi.org/10.1002/9780470996249.ch27

Guerra-Salas, J. F. (2018). Latin America’S Declining Skill Premium: a Macroeconomic Analysis. *Economic Inquiry*, **56**(1), 620–636. https://doi.org/10.1111/ecin.12497

Guerrero, J., & Cabezas, J. (2019). *Clasificación automática de textos utilizando técnicas de text* *mining: Aplicación a las glosas de la Encuesta Nacional de Empleo (ENE)* (No. 6; Documentos de Trabajo). http://www.ine.cl/docs/default-source/documentos-de-trabajo/clasificacion-automatica-de-textos-utilizando-tecnicas-de-text-mining-aplicacion-a-las-glosas-de-la-encuesta-nacional-de-empleo.pdf?sfvrsn=1ddfe7d2_2

Guisan, M.-C. (2001). Causality and Cointegration Between Consumption and GDP in 25 OECD Countries: Limitations of the Cointegration Approach. *Applied Econometrics and International Development*, **1**, 39–61. http://ideas.repec.org/a/eaa/aeinde/v1y2001i1_2.html

Guvenen, F., & Kuruscu, B. (2009). A quantitative analysis of the evolution of the U.S. wage distribution: 1970-2000. *NBER Macroeconomics Annual*, **24**(1), 227–276. https://doi.org/10.1086/648297

Hallegatte, S., & Dumas, P. (2009). Can natural disasters have positive consequences? Investigating the role of embodied technical change. *Ecological Economics*, **68**(3), 777–786. https://doi.org/10.1016/j.ecolecon.2008.06.011

Han, J., Pei, J., & Kamber, M. (2011). Classification: Advanced Methods. Support Vector Machines. In *Data Mining: Concepts and Techniques*. Elsevier Science & Technology, Waltham, MA.

Hausman, J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *Journal of Economic Perspectives*, **15**(4), 57–67. https://doi.org/10.1257/jep.15.4.57

Havranek, T., Irsova, Z., Laslopova, L., & Zeynalova, O. (2020). *Skilled and Unskilled Labor Are Less Substitutable than Commonly Thought* (No. 29; IES Working Paper).

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, **24**(3), 411–482. https://doi.org/10.1086/504455

Hemous, D., & Olsen, M. (2022). The Rise of the Machines: Automation, Horizontal Innovation and Income Inequality. *American Economic Journal: Macroeconomics*, **14**(1), 179–223. https://doi.org/10.1257/mac.20160164

Hershbein, B., & Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? evidence from vacancy postings. *American Economic Review*, **108**(7), 1737–1772. https://doi.org/10.1257/aer.20161570

Holm, J. R., & Østergaard, C. R. (2015). Regional Employment Growth, Shocks and Regional Industrial Resilience: A Quantitative Analysis of the Danish ICT Sector. *Regional Studies*, **49**(1), 95–112. https://doi.org/10.1080/00343404.2013.787159

How, S. M., & Kerr, G. N. (2019). Earthquake Impacts on Immigrant Participation in the Greater Christchurch Construction Labor Market. *Population Research and Policy Review*, **38**(2), 241–269. https://doi.org/10.1007/s11113-018-9500-6

Hutter, C., & Weber, E. (2017). *Labour market effects of wage inequality and skill-biased technical change in Germany* (No. 5; IAB-Discussion Paper). http://hdl.handle.net/10419/162476www.econstor.eu

Hwang, W. S., & Shin, J. (2017). ICT-specific technological change and economic growth in Korea. *Telecommunications Policy*, **41**(4), 282–294. https://doi.org/10.1016/j.telpol.2016.12.006

ILO. (2012). *The International Standard Classification of Occupations (ISCO-08)*. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_172572.pdf#374

INE. (2017). *Resultados Censo de Población y Vivienda 1982, 1992, 2002 y 2017*. https://www.ine.cl/

INE. (2018). *Clasificador Chileno de Ocupaciones CIUO 08.CL*. https://www.ine.cl/docs/default-source/publicaciones/2018/ciuo-08.cl-clasificador-chileno-de-ocupaciones.pdf?sfvrsn=50bd58d2_4

Jara, B., & Faggian, A. (2018). Labor market resilience and reorientation in disaster scenarios. In *Resilience, Crisis and Innovation Dynamics* (pp. 153–168). Edward Elgar Publishing, Cheltenham UK. https://doi.org/10.4337/9781786432193

Javed, F., Luo, Q., McNair, M., Jacob, F., Zhao, M., & Kang, T. S. (2015). Carotene: A job title classification system for the online recruitment domain. *Proceedings - 2015 IEEE 1st International Conference on Big Data Computing Service and Applications, BigDataService 2015*, 286–293. https://doi.org/10.1109/BigDataService.2015.61

Javed, F., Mcnair, M., Jacob, F., & Zhao, M. (2014). *Towards a Job Title Classification System* (WSDM'14 - Workshop on Web-Scale Classification: Classifying Big Data from the Web). http://hadoop.apache.org

Jiménez, A., & Cubillos, R. (2010). Estrés percibido y satisfacción laboral después del terremoto ocurrido el 27 de Febrero de 2010 en la Zona Centro-Sur de Chile. *Terapia Psicologica*, **28**(2), 187–192. https://doi.org/10.4067/s0718-48082010000200007

Jiménez, M., Jiménez, M., & Romero-Jarén, R. (2020). How resilient is the labour market against natural disaster? Evaluating the effects from the 2010 earthquake in Chile. *Natural Hazards*, **104**(2), 1481–1533. https://doi.org/10.1007/s11069-020-04229-9

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (C. W.

J. Granger & G. E. Mizon (eds.)). Oxford University Press, Oxford, UK.

Johnson, G. E. (1997). Changes in Earnings Inequality: The Role of Demand Shifts. *Journal of Economic Perspectives*, **11**(2), 41–54. https://doi.org/10.1257/jep.11.2.41

Karnani, M. (2015). *Labor Shakes: Mid-Run Effects of the 27F Earthquake on Unemployment* (No. 68935; Munich Personal RePEc Archive). https://mpra.ub.uni-muenchen.de/68935/1/MPRA_paper_68935.pdf

Katz, L. F., & Autor, D. H. (1999). Changes in the wage structure and earnings inequality. In *Handbook of Labor Economics* (pp. 1463–1555). Elsevier Science & Technology, Oxford, UK. https://doi.org/10.1016/S1573-4463(99)03007-2

Katz, L. F., & Murphy, K. (1992). Changes in Relative Wages , 1963-1987 : Supply and Demand Factors. *The Quarterly Journal of Economics*, **107**(1), 35–78. https://doi.org/10.2307/2118323

Kearney, I. (1997). *Estimating the demand for skilled labour, unskilled labour and clerical workers: A dynamic framework* (No. 91; ESRI Working Paper Series).

Kirchberger, M. (2017). Natural disasters and labor markets. *Journal of Development Economics*, **125**, 40–58. https://doi.org/10.1016/j.jdeveco.2016.11.002

Klomp, J., & Valckx, K. (2014). Natural disasters and economic growth: A meta-analysis. *Global Environmental Change*, **26**(1), 183–195. https://doi.org/10.1016/j.gloenvcha.2014.02.006

Koop, G. (2003). *Bayesian econometrics*. Wiley, Chichester. https://doi.org/10.1057/9780230226203.0108

Kureková, L. M., Beblavý, M., & Thum-Thysen, A. (2015). Using online vacancies and web surveys to analyse the labour market: a methodological inquiry. *IZA Journal of Labor Economics*, **4**(1), 18. https://doi.org/10.1186/s40172-015-0034-4

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? *Journal of Econometrics*, **54**, 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

Larrañaga, O. (2001). Distribución de ingresos: 1958-2001. In R. Ffrennch-Davis & B. Stallings (Eds.), *Reformas, crecimiento y políticas sociales en Chile desde 1973*. CEPAL / LOM Ediciones.

Lee, M., & Mimno, D. (2014). Low-dimensional embeddings for interpretable anchor-based topic inference. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1319–1328. https://www.aclweb.org/anthology/D14-1138

Leiter, A. M., Oberhofer, H., & Raschky, P. A. (2009). Creative disasters? Flooding effects on capital, labour and productivity within European firms. *Environmental and Resource Economics*, **43**(3), 333–350. https://doi.org/10.1007/s10640-009-9273-9

Levy, B. F., & Murnane, R. J. (1996). What Skills Are Computers a Complement? *The American Economic Review*, **86**(2), 258–262. https://www.jstor.org/stable/2118133

Li, B. (2010). Multinational production and choice of technologies: New evidence on skill-biased technological change from China. *Economics Letters*, **108**(2), 181–183. https://doi.org/10.1016/j.econlet.2010.04.042

Loayza, N. V., Olaberría, E., Rigolini, J., & Christiaensen, L. (2012). Natural Disasters and Growth: Going Beyond the Averages. *World Development*, **40**(7), 1317–1336. https://doi.org/10.1016/j.worlddev.2012.03.002

Lovaglio, P. G., Cesarini, M., Mercorio, F., & Mezzanzanica, M. (2018). Skills in demand for ICT and statistical occupations: Evidence from web-based job vacancies. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, **11**(2), 78–91. https://doi.org/10.1002/sam.11372

Lutkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. In *New Introduction to Multiple Time Series Analysis*. Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-27752-8

Lynham, J., Noy, I., & Page, J. (2017). The 1960 Tsunami in Hawaii: Long-Term Consequences of a Coastal Disaster. *World Development*, **94**, 106–118. https://doi.org/10.1016/J.WORLDDEV.2016.12.043

Maddala, G., & Kim, I.-M. (1998). *Unit Roots, Cointegration and Structural Change*. Cambridge University Press, Cambridge, MA.

Mallick, S. K., & Sousa, R. M. (2017). The skill premium effect of technological change: New evidence from United States manufacturing. *International Labour Review*, **156**(1), 113–131. https://doi.org/10.1111/j.1564-913X.2015.00047.x

Manacorda, M., Sanchez-Paramo, C., & Schady, N. (2010). Changes in returns to education in Latin America: The role of demand and supply of skills. *Industrial and Labor Relations Review*, **63**(2), 307–326. https://www.jstor.org/stable/pdf/40649501.pdf?refreqid=excelsior%3A87d6589b000fd69f1a55f315d999f2a7

Marinescu, I. (2017). The general equilibrium impacts of unemployment insurance: Evidence from a large online job board. *Journal of Public Economics*, **150**, 14–29.

https://doi.org/10.1016/j.jpubeco.2017.02.012

Markowitsch, J., & Plaimauer, C. (2009). Descriptors for competence: Towards an international standard classification for skills and competences. *Journal of European Industrial Training*, **33**(8), 817–837. https://doi.org/10.1108/03090590910993652

McAdam, P., & Willman, A. (2018). Unraveling the Skill Premium. *Macroeconomic Dynamics*, **22**, 33–62. https://doi.org/10.1017/S1365100516000547

Medina, C. A., & Posso, C. M. (2010). *Technical Change and Polarization of the Labor Market: Evidence for Colombia, Brazil and Mexico* (No. 614; Borradores de Economía).

Mihaylov, E., & Tijdens, K. G. (2019). *Measuring the Routine and Non-Routine Task Content of 427 Four-Digit ISCO-08 Occupations* (TI 2019-035/V; Tinbergen Institute Discussion Paper). https://doi.org/10.2139/ssrn.3389681

MINEDUC. (2020). *Compendio histórico matrícula 2020 SIES*. https://www.mifuturo.cl/bases-de-datos-de-matriculados/

MINEDUC, & CORFO. (2017). *Marco De Cualificaciones Técnico-Profesional*. https://www.oitcinterfor.org/sites/default/files/file_publicacion/marcocualificacioneschile.pdf

Ministerio del Trabajo y Previsión Social. (2021). *Decreto 34. Crea comisión asesora ministerial de prospección laboral*. https://www.bcn.cl/leychile/navegar?i=1166947

Moosa, I. A. (2017). Blaming suicide on NASA and divorce on margarine: the hazard of using cointegration to derive inference on spurious correlation. *Applied Economics*, **49**(15), 1483–1490. https://doi.org/10.1080/00036846.2016.1218434

Murakami, Y. (2014). Trade liberalization and the skill premium in Chile. *México y La Cuenca Del Pacífico*, **3**(6), 77–101. https://doi.org/10.32870/mycp.v3i6.418

Murakami, Y., & Nomura, T. (2020). Expanding higher education and wage inequality in Chile. *Journal of Economic Studies*, **47**(4), 877–889. https://doi.org/10.1108/JES-12-2018-0445

Muth, C., Oravecz, Z., & Gabry, J. (2018). User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *The Quantitative Mehods for Psychology*, **14**(2). https://doi.org/10.20982/tqmp.14.2.p099

Nahoomi, N. (2018). *Automatically Coding Occupation Titles to a Standard Occupation Classification* [The University of Guelph]. https://pdfs.semanticscholar.org/be53/193184dfd665525249568487bcef5455c93d.pdf

Nalepa, J., & Kawulok, M. (2019). Selecting training sets for support vector machines: a review.

Artificial Intelligence Review, **52**(2), 857–900. https://doi.org/10.1007/s10462-017-9611-1

Nanni, F., Kümper, H., & Ponzetto, S. P. (2016). Semi-supervised Textual Analysis and Historical Research Helping Each Other: Some Thoughts and Observations. *International Journal of Humanities and Arts Computing*, **10**(1), 63–77. https://doi.org/10.3366/ijhac.2016.0160

National Center for O*NET Development. (2020). *O*NET OnLine*. https://www.onetonline.org/

Nchor, D., & Rozmahel, P. (2020). Job Polarization in Europe: Evidence from Central and Eastern European Countries. *DANUBE: Law, Economics and Social Issues Review*, **11**(1), 52–74. https://doi.org/10.2478/DANB-2020-0004

Neusser, K. (2016). *Time Series Econometrics*. Springer International Publishing Switzerland.

NOAA, N. G. D. C. (2019). *National Geophysical Data Center / World Data Service (NGDC/WDS): Significant Earthquake Database*. https://doi.org/doi:10.7289/V5TD9V7K

Nogueira, M. C., Afonso, Ó., & Soukiazis, E. (2017). Skill premium in Portuguese manufacturing industries. *Applied Economics Letters*, **25**(14), 1015–1018. https://doi.org/10.1080/13504851.2017.1391993

O'Mahony, M., Robinson, C., & Vecchi, M. (2008). The impact of ICT on the demand for skilled labour: A cross-country comparison. *Labour Economics*, **15**(6), 1435–1450. https://doi.org/10.1016/j.labeco.2008.02.001

O'Mahony, M., Vecchi, M., & Venturini, F. (2021). Capital Heterogeneity and the Decline of the Labour Share. *Economica*, **88**(350), 271–296. https://doi.org/10.1111/ecca.12356

OECD. (2018). *Chile Policy Brief. Education and skills* (Issue February). https://www.oecd.org/policy-briefs/Chile-Education-and-Skills-EN.pdf

OECD, Caribbean, E. C. for L. A. and the, & CAF Development Bank of Latin America European Union. (2019). Changing the lens : Development beyond income. In *Latin American Economic Outlook 2019* (pp. 65–94). OECD Publishing. https://doi.org/10.1787/g2g9ff18-en

Okazaki, T., Okubo, T., & Strobl, E. (2019). Creative destruction of industries: Yokohama City in the Great Kanto Earthquake, 1923. *Journal of Economic History*, **79**(1), 1–31. https://doi.org/10.1017/S0022050718000748

Okuyama, Y. (2003). *Economics of natural disasters: A critical review* (No. 12; Research Paper). https://researchrepository.wvu.edu/rri_pubs/131/

Okuyama, Y., Hewings, G. J. D., & Sonis, M. (2004). Measuring Economic Impacts of Disasters: Interregional Input-Output Analysis Using Sequential Interindustry Model. In Y. Okuyama & S.

Chang (Eds.), *Modeling Spatial and Economic Impacts of Disasters*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24787-6_5

Panwar, V., & Sen, S. (2019). Economic Impact of Natural Disasters: An Empirical Re-examination. *Margin-The Journal of Applied Economic Research*, **13**(1), 109–139. https://doi.org/10.1177/0973801018800087

Park, J. Y., Son, M., & Park, C. K. (2017). Natural disasters and deterrence of economic innovation: A case of temporary job losses by Hurricane Sandy. *Journal of Open Innovation: Technology, Market, and Complexity*, **3**(1). https://doi.org/10.1186/s40852-017-0055-2

Parro, F., & Reyes, L. (2017). The rise and fall of income inequality in Chile. *Latin American Economic Review*, **26**(3), 31. https://doi.org/10.1007/s40503-017-0040-y

Patterson, K. (2000). *An Introduction to Applied Econometrics, a time series approach*. Palgrave, New York.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.

Pelagatti, M. M. (2016). *Time Series Modelling with Unobserved Components*. Taylor & Francis, Boca Raton.

Pellandra, A. (2015). *The commodity price boom and regional workers in Chile: a natural resources blessing?*

Perez-Silva, R., & Campos, J. (2021). Agriculture 4.0? Studying the evidence for automation in Chilean agriculture. *International Journal of Agriculture and Natural Resources*, **48**(3), 233–247. https://doi.org/10.7764/ijanr.v48i3.2339

Pouliakas, K., & Branka, J. (2020). *EU Jobs at Highest Risk of COVID-19 Social Distancing: Will the Pandemic Exacerbate Labour Market Divide?* (No. 13281; Discussion Paper Series). https://doi.org/10.2801/968483

Ramos, J., Coble, D., Elfernan, R., & Soto, C. (2013). The Impact of Cognitive and Noncognitive Skills on Professional Salaries in An Emerging Economy, Chile. *The Developing Economies*, **51**(1), 1–33. https://doi.org/10.1111/deve.12000

Razzak, W., & Timmins, J. (2008). A macroeconomic perspective on skill shortages and the skill premium in New Zealand. *Australian Economic Papers*, **47**(1), 74–91. https://doi.org/10.1111/j.1467-8454.2008.00330.x

Redmond, P., & Mcguinness, S. (2020). *Who can work from home in Ireland?* (No. 87; Survey and Statistical Report Series). https://doi.org/10.26504/sustat87

Reijnders, L. S. M., & de Vries, G. J. (2018). Technology, offshoring and the rise of non-routine jobs. *Journal of Development Economics*, **135**(August), 412–432. https://doi.org/10.1016/j.jdeveco.2018.08.009

Robbins, D. (1994a). Relative wage structure in Chile, 1957 - 1992: changes in the structure of demand for schooling. *Estudios de Economía*, **21**, 49–78.

Robbins, D. (1994b). *Worsening Relative Wage Dispersion in Chile during Trade Liberalization: Supply or Demand driven Inequality?* (No. 122).

Robbins, D. (1994c). *Worsening relative wage dispersion in Chile during trade liberatization, and its causes : is supply at fault?* (No. 484; Development Discussion Paper).

Robbins, D. (1996). *HOS HITS FACTS: Facts win evidence on trade and wages in the developing world* (No. 557; Development Discussion Papers).

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association*, **111**(515), 988–1003. https://www.tandfonline.com/doi/full/10.1080/01621459.2016.1141684

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, **91**(2). https://doi.org/10.18637/jss.v091.i02

Roberts, M. E., Stewart, B. M., Tingley, D., & Airoldi, E. M. (2013). The Structural Topic Model and Applied Social Science. *NIPS 2013 Workshop on Topic Models: Computation, Application, and Evaluation*. https://scholar.princeton.edu/files/bstewart/files/stmnips2013.pdf

Roberts, M. E., Stewart, B. M., Tingley, D., & Benoit, K. (2020). *R Package "stm."* https://doi.org/10.1111/ajps.12103

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, **58**(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Rodríguez-Oreggia, E. (2013). Hurricanes and labor market outcomes: Evidence for Mexico. *Global Environmental Change*, **23**(1), 351–359. https://doi.org/10.1016/j.gloenvcha.2012.08.001

Rothwell, J. (2012). *Education, Job Openings, and Unemployment in Metropolitan America* (August 2012; Metropolitan Policy Program). http://hourglassfoundation.org/pdf/29_education_gap_rothwell.pdf

RStudio Team. (2020). *Rstudio: Integrated Development for R.* RStudio, PBC. http://www.rstudio.com/.

Sánchez-Páramo, C., & Schady, N. (2003). *Off and Running? Technology, Trade, and the Rising Demand for Skilled Workers in Latin America* (No. 3015; Policy Research Working Paper). http://econ.worldbank.org.

Sanhueza, C., Contreras, D., & Denis, Á. (2012). Terremoto y sus efectos sobre el bienestar: un análisis multidimensional. *Persona y Sociedad*, **26**(1), 43–66.

Schmidt, B. (2012). Words Alone: Dismantling Topic Models in the Humanities. *Journal of Digital Humanities*, **2**(1), 49–65.

Schneider, B. R. (2013). *Hierarchical capitalism in Latin America: Business, labor, and the challenges of equitable development*. Cambridge University Press, New York.

Schulze, P., Wiegrebe, S., Thurner, P. W., Heumann, C., Aßenmacher, M., & Wankmüller, S. (2021). Exploring Topic-Metadata Relationships with the STM: A Bayesian Approach. *ArXiv*, **abs/2104.0**. http://arxiv.org/abs/2104.02496

Schumpeter, J. (1976). The Process of Creative Destruction. In *Capitalism, Socialism and Democracy* (pp. 81–86). Routledge, London. https://doi.org/10.4324/9780203202050_chapter_vii

Sebastian, R. (2018). Explaining job polarisation in Spain from a task perspective. *SERIEs*, **9**(9), 215–248. https://doi.org/10.1007/s13209-018-0177-1

Sehnbruch, K., Agloni, N., Imilan, W., & Sanhueza, C. (2017). Social Policy Responses of the Chilean State to the Earthquake and Tsunami of 2010. *Latin American Perspectives*, **44**(4), 24–40. https://doi.org/10.1177/0094582X16648955

SENCE. (2022). *Observatorio Laboral*. Observatorio Laboral. https://www.observatorionacional.cl/

Servicio Nacional de Aprendizaje. (2020). *Clasificación Nacional de Ocupaciones de Colombia*. https://observatorio.sena.edu.co/Content/pdf/cno_version_2020_2.pdf

Sevilla, M. P., & Farías, M. (2020). Labour market mismatch in emerging countries: the case of Chile. *Compare: A Journal of Comparative and International Education*, **50**(2), 276–293. https://doi.org/10.1080/03057925.2019.1675495

Shadrova, A. (2021). Topic models do not model topics: epistemological remarks and steps towards best practices. *Journal of Data Mining & Digital Humanities*, **2021**. https://doi.org/10.46298/jdmdh.7595

Sims, C. (1980). Macroeconomics and reality. *Econometrica*, **48**(1), 1–48.

https://www.jstor.org/stable/1912017

Sisk, B., & Bankston, C. L. (2014). Hurricane Katrina, a Construction Boom, and a New Labor Force: Latino Immigrants and the New Orleans Construction Industry, 2000 and 2006–2010. *Population Research and Policy Review*, **33**(3), 309–334. https://doi.org/10.1007/s11113-013-9311-8

Skidmore, M., & Toya, H. (2002). Do natural disasters promote long-run growth? *Economic Inquiry*, **40**(4), 664–687. https://onlinelibrary.wiley.com/doi/pdf/10.1093/ei/40.4.664

Smola, A. J., & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, **14**, 199–222. https://alex.smola.org/papers/2004/SmoSch04.pdf

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, **70**(1), 65–94. https://www.jstor.org/stable/1884513

Solow, R. M. (1957). Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics*, **39**(3), 312–320. https://www.jstor.org/stable/1926047

Spiezia, V. (2018). Getting skills right : Measuring the demand for skills in the digital economy. In *Power from Statistics: data,information and knowledge* (pp. 173–181). European Union. https://ec.europa.eu/eurostat/cros/powerfromstatistics/OR/PfS-OutlookReport.pdf#page=173

Spitz-Oener, A. (2006). Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics*, **24**(2), 235–270. https://doi.org/10.1086/499972

Stan Development Team. (2019). *Stan Reference Manual V 2.23*. https://mc-stan.org/docs/2_23/reference-manual-2_23.pdf

Swan, T. (1956). Economic growth and capital accumulation. *Economic Record*, **32**(2), 334–361. https://doi.org/10.1111/j.1475-4932.1956.tb00434.x

Taddy, M. A. (2012). On estimation and selection for topic models. *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 1184–1193.

Tanaka, A. (2015). The impacts of natural disasters on plants' growth: Evidence from the Great Hanshin-Awaji (Kobe) earthquake. *Regional Science and Urban Economics*, **50**, 31–41. https://doi.org/10.1016/j.regsciurbeco.2014.11.002

Tinbergen, J. (1972). The impact of education of education on income distribution. *Review of Income and Wealth*, **18**(3), 255–265. https://doi.org/10.1111/j.1475-4991.1972.tb00865.x

Tinbergen, J. (1974). Substituion of graduate by other labour. *Kyklos (Basel)*, **27**(2), 217–226. https://doi.org/10.1111/j.1467-6435.1974.tb01903.x

Toya, H., & Skidmore, M. (2007). Economic development and the impacts of natural disasters. *Economics Letters*, **94**(1), 20–25. https://doi.org/10.1016/j.econlet.2006.06.020

Toya, H., & Skidmore, M. (2015). Information/communication technology and natural disaster vulnerability. *Economics Letters*, **137**, 143–145. https://doi.org/10.1016/j.econlet.2015.10.018

Trujillo-Pagan, N. (2012). Neoliberal disasters and racialisation: The case of post-Katrina Latino labour. *Race & Class*, **53**(4), 54–66. https://doi.org/10.1177/0306396811433986

University of Chile. (2020). *Encuesta de Ocupación y Desocupación en el Gran Santiago (EOD)*. http://documentos.microdatos.cl/Encuestas/Ocupacion/OcupacionDesocupacion?C=T1

Valiente, O., Zancajo, A., & Jacovkis, J. (2020). The coordination of skill supply and demand in the market model of skill formation: testing the assumptions for the case of Chile. *International Journal of Lifelong Education*, **39**(1), 90–103. https://doi.org/10.1080/02601370.2019.1678692

Van Ark, B., & O'Mahony, M. (2016). Productivity growth in Europe before and since the 2008/2009 economic and financial crisis. In *The World Economy: Growth or Stagnation?* https://doi.org/10.1017/9781316534502.004

Varella, A. (2008a). Relative wages, labor supplies and trade in Mexican manufacturing: Evidence from two samples. *Journal of International Trade and Economic Development*, **17**(2), 213–241. https://doi.org/10.1080/09638190701872673

Varella, A. (2008b). The Rise of the Skill Premium in Mexican Maquiladoras. *The Journal of Development Studies*, **44**(9), 1382–1404. https://doi.org/10.1080/00220380802265272

Varella, A. (2011). The world elasticity of labor substitution across education levels. *Empirical Economics*, **41**(3), 769–785. https://doi.org/10.1007/s00181-010-0398-z

Varella, A., & Ibarra-Salazar, J. (2013). Productivity Effects on the Wage Premium of Mexican Maquiladoras. *Economic Development Quarterly*, **27**(3), 208–220. https://doi.org/10.1177/0891242413478463

Violante, G. L. (2016). Skill-Biased Technical Change. In *The New Palgrave Dictionary of Economics* (pp. 1–6). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-349-95121-5_2388-1

vom Lehn, C. (2018). Understanding the decline in the U.S. labor share: Evidence from occupational tasks. *European Economic Review*, **108**, 191–220. https://doi.org/10.1016/j.euroecorev.2018.07.005

Von Brasch, T. (2016). Identifying the sector bias of technical change. *Empirical Economics*, **50**, 595–621. https://doi.org/10.1007/s00181-015-0938-7

Walker, D. N. (2012). *Communication technology in disaster management*. http://digitalcommons.wayne.edu/oa_theses

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. *Proceedings of the 26th International Conference On Machine Learning, ICML 2009*, *4*, 1105–1112. https://doi.org/10.1145/1553374.1553515

Weinberger, C. J. (2014). The increasing complementarity between cognitive and social skills. *The Review of Economics and Statistics*, **96**(5), 849–861. https://www.jstor.org/stable/43554962

Welbers, K., Van Atteveldt, W., & Benoit, K. (2017). Text Analysis in R. *Communication Methods and Measures*, **11**(4), 245–265. https://doi.org/10.1080/19312458.2017.1387238

Wooldridge, J. (2009). *Introductory Econometrics: A modern approach* (4th ed.). Cengage Learning, Mason, OH.

World Bank. (2013). China 2030: Building a Modern, Harmonious, and Creative Society. In *China 2030*. https://doi.org/10.1596/9780821395455_overview

World Bank. (2020). *World Development Indicators*. Countries Historical Classifications by Income. https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries

World Bank. (2022). *World Bank Open Data*. http://data.worldbank.org

Xiao, Y., & Feser, E. (2014). The unemployment impact of the 1993 US midwest flood: A quasi-experimental structural break point analysis. *Environmental Hazards*, **13**(2), 93–113. https://doi.org/10.1080/17477891.2013.777892

Zapata-Román, G. (2021). *The role of skills and tasks in changing employment trends and income inequality in Chile* (2021/48; Wider Working Paper).

Zissimopoulos, J., & Karoly, L. A. (2010). Employment and Self-employment in the wake of hurricane Katrina. *Demography*, **47**(2), 345–367. https://doi.org/10.1353/dem.0.0099

Appendices

A.1. Essay I

A.1.1. Stan Code

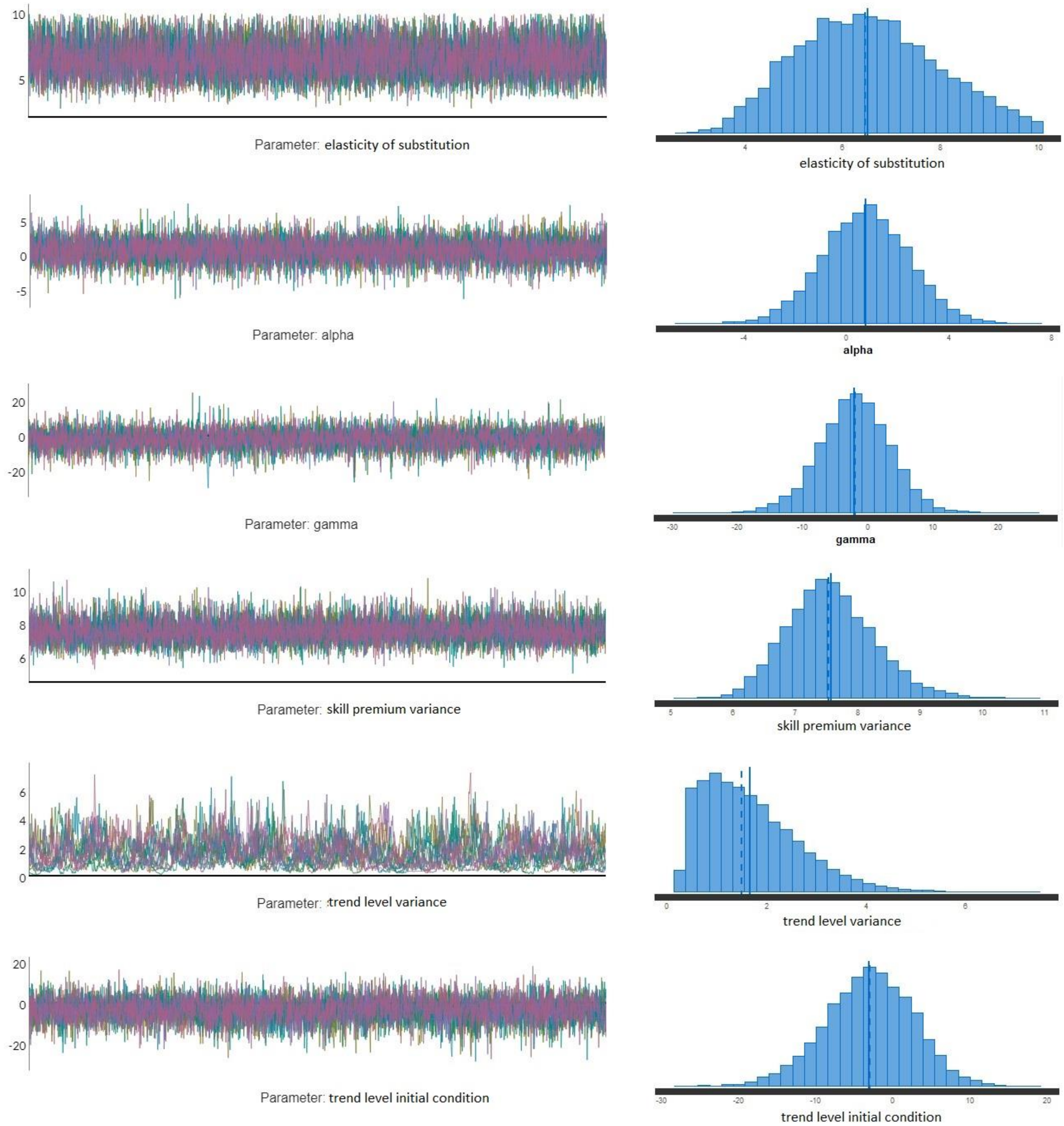
The Stan code representing our UCM-Bayesian specification (see section 2.6.2.4) is structured within “program blocks”: **data**, **parameters**, and **model**. The **data** block is for the declaration of variables that are read in as data. In the **parameters** block we indicate the parameters to be modelled. The variables declared here are the variables that will be sampled by Stan. The **model** block is where we specified the priors and likelihood, along with the declaration of any variables necessary. The code for the base model is as follows:

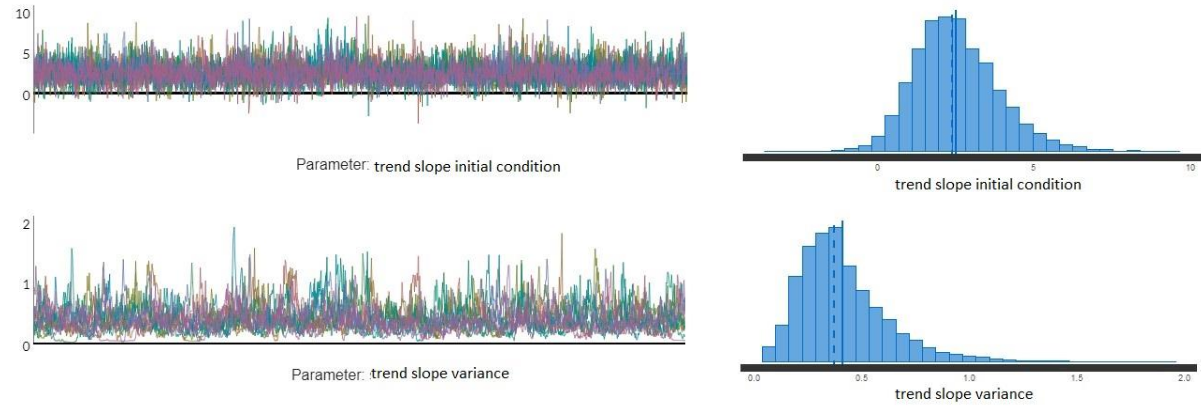
```
modelBase <- "
data {
int N;              //number of observations
vector[N] y;        //outcome vector for the outcome (the skill premium)
vector[N] x;        //predictor vector for the relative supply
vector[N] z;        //predictor vector for S (seasonality)
vector[N] d;        //predictor vector for Ch98 (dummy change year 1998)
}

parameters {
vector[N] u;        //level vector parameter
vector[N] v;        //slope vector parameter
real<lower=0.01> s_u;     //level white noise parameter with restriction
real<lower=0.01> s_v;     //slope white noise parameter with restriction
real<lower=0.01> s_y;     //outcome white noise parameter with restriction
real<lower=.1,upper=10> beta; //elasticity parameter restricted to [0.1-10]
real alpha;         //S predictor parameter
real gamma;         //Ch98 predictor parameter
real u0;            //level initial condition parameter
real v0;            //slope initial condition parameter
}

model {
alpha~cauchy(0,10);       //prior for S predictor parameter
gamma~cauchy(0,10);       //prior for Ch98 predictor parameter
s_u~cauchy(0,10);         //prior for level white noise parameter
s_v~cauchy(0,10);         //prior for slope white noise parameter
s_y~cauchy(0,10);         //prior for outcome white noise parameter
beta~normal(0.1,3);       //prior for elasticity parameter
u[1]~normal(u0,s_u);      //prior for the level initial condition
v[1]~normal(v0,s_v);      //prior for the slope initial condition

v[2:N] ~ normal(v[1:N-1], s_v);             //likelihood for the slope
u[2:N] ~ normal(u[1:N-1] + v[1:N-1], s_u);  //likelihood for the level
y ~ normal(u-x/beta+alpha*z+gamma*d, s_y);  //likelihood for the outcome
}"
```

A.1.2. UCM-Bayesian modelling diagnostics

Here we display graphical diagnostics of the main parameters from our UCM-Bayesian estimation presented in Table 2.6 using the *shinystan* R package (Gabry et al., 2022; Muth et al., 2018). The plots on the left correspond to the **trace plots**, a visual complement to the $\hat{R}$ statistics detailed in Table 2.6, to check convergence. In the trace plot the lines show different chains. If the chains are indistinguishable, this is a signal of convergence. The plots on the right show the parameters' **full *posterior*** to check how the mass probability is distributed, and the vertical dashed line indicates the *posterior* mean of the parameter.

Figure A.1.2. Trace plots (left-side) and full posterior plots (right-side) from UCM-Bayesian estimation





A.2. Essay II

A.2.1. Chilean Occupational Classification, CIUO08-CL for the 41 2-digit sub-groups

The next table shows the Chilean Occupational Classification, CIUO08-CL for the 41 2- digit sub-groups (names in Spanish from Clasificador Chileno de Ocupaciones (INE, 2018) and in English from the International Standard Classification of Occupations (ILO, 2012))

Table A.2.1. Chilean Occupational Classification, CIUO08-CL for the 41 2-digit sub-groups

Cod	CIUO-08 CL Name in Spanish	CIUO-08 CL Name in English
11	Miembros del Poder Ejecutivo y Legislativo, personal directivo de la administración pública y de otras organizaciones sociales y/o políticas, directores ejecutivos y gerentes generales.	Chief executives, senior officials and legislators
12	Directores y gerentes administrativos y de servicios comerciales.	Administrative and commercial managers
13	Directores, gerentes y administradores de producción y operaciones.	Production and specialised services managers
14	Directores, gerentes y administradores de hoteles, restaurantes, comercios y de otros servicios.	Hospitality, retail and related services managers
21	Profesionales de las ciencias y de la ingeniería.	Science and engineering professionals
22	Profesionales de la salud	Health professionals
23	Profesionales de la educación.	Teaching professionals
24	Profesionales de negocios y administración.	Business and administration professionals
25	Profesionales de tecnología de la información y las comunicaciones.	Information and communications technology (ICT) professionals

26	Profesionales en derecho, ciencias sociales y culturales.	Legal, social and cultural professionals
31	Técnicos de las ciencias y la ingeniería	Science and engineering associate professionals (technicians)
32	Técnicos de la salud	Health associate professionals (technicians)
33	Técnicos en operaciones financieras y administrativas.	Business and administration associate professionals (technicians)
34	Técnicos de servicios jurídicos, sociales, deportivos y culturales.	Legal, social, cultural and related associate professionals (technicians)
35	Técnicos de la tecnología de la información y las comunicaciones.	Information and communications associate professionals (technicians)
36	Técnicos en educación.	Teaching associate professionals (technicians)
41	Oficinistas.	General and keyboard clerks
42	Empleados en trato directo con el público.	Customer services clerks
43	Auxiliares y ayudantes de registros contables y encargados del registro de materiales.	Numerical and material recording clerks
44	Otro personal de apoyo administrativo.	Other clerical support workers
51	Trabajadores de los servicios a las personas.	Personal services workers
52	Vendedores.	Sales workers
53	Trabajadores de los cuidados personales.	Personal care workers
54	Personal de los servicios de protección y seguridad.	Protective services workers
61	Agricultores y trabajadores calificados de explotaciones agropecuarias cuya producción se destina al mercado.	Market-oriented skilled agricultural workers and farmers
62	Trabajadores forestales calificados, pescadores ycazadores cuya producción se destina al mercado	Market-oriented skilled forestry, fishery and hunting workers
63	Trabajadores agropecuarios, pescadores, cazadores y recolectores de subsistencia.	Subsistence farmers, fishers, hunters and gatherers
71	Operarios de la construcción (no incluye electricistas).	Building and related trades workers (excluding electricians)

72	Operarios de la metalurgia y operarios de máquinas herramientas; mecánicos de vehículos, maquinarias, aviones y bicicletas.	Metal, machinery and related trades workers
73	Artesanos y operarios de las artes gráficas.	Handicraft and printing workers
74	Trabajadores especializados en electricidad y electrónica.	Electrical and electronic trades workers
75	Operarios de procesamiento de alimentos, de la confección, ebanistas y otros oficios.	Food processing, woodworking, garment and related trades workers
81	Operadores de instalaciones fijas y máquinas.	Stationary plant and machine operators
82	Ensambladores.	Assemblers
83	Conductores de vehículos y operadores de equipos pesados y móviles.	Drivers and mobile plant operators
91	Auxiliares de aseo y trabajadores de casa particular.	Cleaners and helpers
92	Obreros agropecuarios, pesqueros y forestales.	Agricultural, forestry and fishery labourers
93	Obreros de la minería, la construcción, la industria manufacturera y el transporte.	Labourers in mining, construction, manufacturing and transport
94	Cocineros de comida rápida y ayudantes de cocina.	Food preparation assistants
95	Trabajadores ambulantes de servicios y vendedores ambulantes (excluyendo comida de consumo inmediato).	Street and related sales and services workers
96	Recolectores de desechos y otras ocupaciones elementales.	Refuse workers and other elementary workers

A.2.2. Distribution (in percentages) of educational and industry categories for the whole and training datasets

Table A.2.2. Distribution (in percentages) of educational and industry categories for the whole and training datasets

Educational and industry categories		Whole dataset	Training dataset
Educational category			
	Primary	2%	1%
	Secondary	20%	15%
	Secondary Education Technician	15%	16%
	Higher Professional Technician	29%	34%
	Graduate	34%	33%
	Postgraduate	1%	0%
	Number of observations	189,986	67,656
Industry			
	Agriculture and fishing	1%	1%
	Commerce	19%	16%
	Communication	9%	8%
	Construction	4%	5%
	Electricity, water and gas	2%	2%
	Financial services	6%	4%
	Industry	17%	17%
	Mining	2%	2%
	Other activity	5%	13%
	Other services	8%	8%
	Personal services	19%	18%
	Public Administration	1%	0%
	Restaurants and Hotels	2%	2%
	Transportation	5%	5%
	Number of observations	189,986	67,656

A.2.3. Dictionary of Skills (full Spanish version).

The next table shows our dictionary of Skills (full Spanish version). The “*” represents different variations for that word (e.g., singular-plural variations, gender variations, etc).

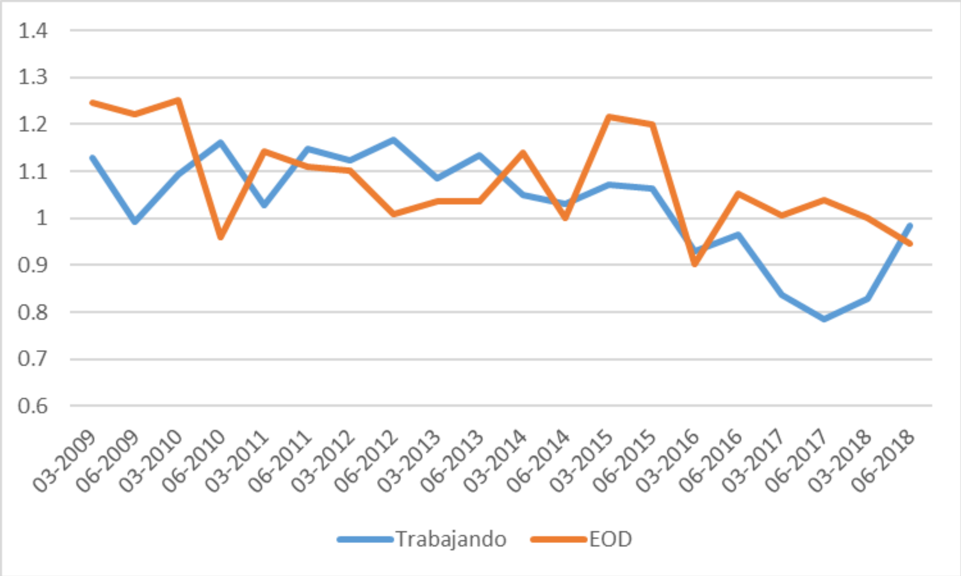
Table A.2.3. Dictionary of Skills (full Spanish version)

Skill Category	Phrases and Words
Cognitive (in Spanish “Cognitivo”)	- resolver_conflicto* - resolver_problema* - resolucion_conflicto* - resolucion_problema* - investigacion - analitic* - ensamiento_critico - matematica - estadistica - alfabetiza* - aplica*_leyes - aplica*_regla* - aplica*_directric* - aplica*_tecnica*_ rof* - aplica*_autoevaluacion - aprend*_nuevo*_idioma* - aprendizaje - concentra* - observa* - colec*_informacion* - ntorno*_trabajo* - aritmetic* - conocimiento*_tecnico* - curios*_intelectual - enfoque*_sistematico* - facilidad_numer* - calculo*_numerico* - memori* - tecnica*_general* - inventiva - pensamiento_logico - razona* - ideas - aprender - alcan*_compromiso* - alcan*_consenso* - evita*_conflictos - maneja*_situacion*_conflict*
Social (idem Spanish)	- comunicacional - comunicarse* - trabajo_equipo* - colabora* - negocia* - presentacion - pedagogic* - cliente - comunica*_profesional* - expres*_verbal* - intercultural* - compren*_oral* - comunica*_ingles - comunica*_idioma*extranjer* - cuestiona*_efectiv* - establec*_contacto* - foment*_contacto* - diplomatic* - motivador* - forma*_equipo* - lidera*_equipo* - retorica*
Character (in Spanish "Personalidad" o "Carácter")	- organizado - orienta*_detalle - multitarea - multifuncion - manejo_tiempo - administracion_tiempo - cumpl*_plazo* - energetic* - cortes* - credibilidad - discernimiento - discrecion - empati* - firmeza - iniciativa*_personal* - leal* - pruden* - puntual* - toleran*_estres - toleran*_frustracion* - toleran*_cambio* - toleran*_incertidumbre*
Writing (in Spanish "Escritura", "Escribir")	- escribir - escritura - expresion*_escrita* - elabora*_borrador* - escri*_clar* - escri*_creativa* - escri*_elegante - redac*_documento*_tecnico* - redac*_informe* - ortografi* - gramatica*

Customer Service (in Spanish "Atención al Cliente", "Servicio al Cliente"))	- cliente* - venta* - consumidor* - usuario* - paciencia - persua* - talento*_comercial* - orienta*_cliente* - orienta*_consumidor* - servicio*_cliente* - servicio*_consumidor*
Project management (in Spanish "Gestión de Proyectos")	- gestion*_proyect* - project_manag*
People management (in Spanish "gestión de los Recursos Humanos")	- supervis*_personal - supervis*_rrhh - gestion*_personal - gestion*_rrhh - gestion*_recurso*_humano* - liderazgo - capacita*
Financial (in spanish "Finanzas")	- contabilidad - contable - presupuest* - finan* - costo*
Computer (in Spanish "Computador", "Ordenador")	- computador - microsoft_office - ms_office - hoja*_calculo - excel - word - powerpoint - navega*_internet - software*_oficina - procesador*_texto - microsoft_word - microsoft_outlook - microsoft_powerpoint
Software (some examples from O*NET Hot technologies List which contains 175 items)	- adobe_aftereffects - adobe_creative_cloud - adobe_illustrator - adobe_indesign - adobe_photoshop - adp_workforce_now - abap - ajax - amazon_ec2 - amazon_redshift - amazon_web_services - ibm_cognos_impromptu - ibm_notes - ibm_spss_statistics - ibm_websphere - integrated_development_environment - ide - intuit_quickbooks - javascript - javascript_object_notation - json - jquery - junit - linuxce - microsoft_sql_server - microsoft_visio - microsoft_visual_basic - microsoft_windows_server - microstrategy - minitab - mongodb - mysql - nagios - oracle_fusion_middleware - oracle_hyperion - oracle_java - oracle_javaserver_pages_jsp - oracle_jd_edwards - oracle_jdbc - oracle_peoplesoft - oracle_pl/sql - oracle_primavera_enterprise - oracle - oracle_solaris - oracle_taleo - oracle_weblogic_server - palm_os

A.2.4. Comparing the skill premium estimation between jobs ads data from Trabajando.com and EOD

Figure A.2.4. Comparing the skill premium estimation between jobs ads data from Trabajando.com and EOD, for 2009-2018 (Trabajando.com data grouped bi-annually)



A.3. Essay III

A.3.1. Full details of 53 topics' interpretation and highest probability and FREX terms

Table A.3.1. Full details of 53 topics' interpretation and highest probability and FREX terms

Topic #	Topic Name in Spanish	Topic Name in English	Metric	10 top words (stem words in Spanish)
1	Sin definir	Undefinable	Highest Probability	profesional, afin, carrer, marketing, desempen, ing, viv, laboratori, zon, titul
			FREX	marketing, ing, carrer, laboratori, ant, alrededor, viv, afin, creativ, profesional_desempen
2	Ingeniería Civil	Civil Engineering	Highest Probability	ingenier, civil, ingenier_civil, ingles, industrial, ingenieri, idiom, civil_industrial, idiom_ingles, avanz
			FREX	habl, ingenier_civil, idiom_ingles, idiom, civil_mecan, ingles, civil, ingenier, ingenieri_civil, industrial_ingenier
3	Contabilidad & Auditoría	Accountancy & Audit	Highest Probability	contador, auditor, contabl, contador_auditor, contabil, general, administracion, manej, administr, tributari
			FREX	contador, auditor, contador_auditor, contador_general, tributari, contabl, contabil, softland, auditor_ingenier, auditori
4	Control y Gestión Financiera	Financial Control and Management	Highest Probability	control, gestion, analisis, anal, control_gestion, cobranz, proces, inform, presupuest, financier

			FREX	control_gestion, cobranz, analisis, anal, fluj, finanz, presupuest, fluj_caj, inform_gestion, balanc
5	Logística	Logistics	Highest Probability	local, administracion, encarg, personal, vehicul, supermerc, caden, reponedor, equip, funcion
			FREX	reponedor, local, vehicul, manan, dependent, caden, manan_tard, funcion, correct_funcion, tard
6	Ingeniería Comercial	Business Management	Highest Probability	comercial, ingenier_comercial, ingenier, sucursal, egres, profesional, desarroll, gerent, administracion, servici
			FREX	ingenier_comercial, comercial_ingenier, comercial_civil, comercial, estrategi_comercial, gerent, jef_comercial, ingenieri_comercial, implement, cercani
7	Sin definir	Undefinable	Highest Probability	especial, present, cert_estudi, diferent, gas, disciplin, enfasis, pesquer, vist, agenci
			FREX	cert_estudi, especial, gas, present, disciplin, enfasis, diferent, vist, pesquer, ambos
8	Servicio al Cliente	Customer service	Highest Probability	client, servici, atencion, atencion_client, orientacion, servici_client, orientacion_servici, sucursal, orientacion_client, product
			FREX	servici_client, orientacion_servici, client_mall, servici_financier, orientacion_client, brind, plataform, ejecut_servici, atencion_client, atencion
9	Sin definir	Undefinable	Highest Probability	cod, industrial, client, plant, cod_client, ocup, perfil, gestion, respons, integr
			FREX	cod_client, ocup_respons, cod_perfil, group, perfil_ingenier, cod, group_seleccion, seleccion_invit, client_industrial, indic_clar
10	Jefe Local / Tienda	Retail Store Manager	Highest Probability	tiend, retail, jef, vent, jef_local, jef_tiend, person, apertur, sub, ubic
			FREX	jef_tiend, jef_local, tiend, vent_tiend, sub, retail, reconoc_retail, apertur, vestuari, mercaderi
11	Servicio al Cliente	Customer service	Highest Probability	client, negoci, nuev, carter, respons, ejecut, product, comercial, manten, carter_client
			FREX	negoci, carter, zonal, ejecut_comercial, carter_client, nuev_client, nuev_negoci, nuev, oportun, segment
12	Guardias de Seguridad	Security Guards	Highest Probability	segur, cert, prest, guardi_segur, prest_servici, guardi, servici, ciud, hac, segur_prest
			FREX	cert, segur_prest, prest, prest_servici, comun_ciud, curs_present, deferent, servici_deferent, fin_especial, ciud_curs

13	Fuerza laboral en Construcción	Construction labour	Highest Probability	proyect, construccion, obras, civil, constructor, ingenier, obra, tecnic, terren, jef
			FREX	obras, constructor_civil, constructor, construccion, obra, proyect, autoc, dibuj, obras_civil, viviend
14	Servicio al Cliente	Customer service	Highest Probability	public, pos, social, valor, atencion_public, institucion, deb, priv, servici, postul
			FREX	atencion_public, valor, emprend, corre, social, public, public_priv, cuid_deb, corre_electron, pos_tecnic
15	Guardias de Seguridad	Security Guards	Highest Probability	internacional, prestigi, segur, mejor, guardi, iso, ohsas, mejor_guardi, iso_ohsas, deferent_comun
			FREX	internacional, deferent_comun, ohsas, iso_ohsas, mejor_guardi, iso, prestigi, prestigi_segur, segur_prestigi, certif
16	Educación	Education	Highest Probability	ciud, resident, angel, profesor, resident_ciud, basic, cont, aprendizaj, educ, ciud_angel
			FREX	resident_ciud, profesor, angel, ciud_angel, colegi, aprendizaj, ciud, matemat, resident, educ
17	Servicio al Cliente	Customer service	Highest Probability	tempor, tiend, interes, auxiliar, rend, trat, disposicion, buen_trat, medi_rend, edad
			FREX	tempor, person_interes, auxiliar, medi_rend, buen_trat, rend, interes, portal, trat, tiend_buen
18	Ventas	Sales	Highest Probability	vendedor, vent, mall, complet, full, comision, sueld, optic, retail, proactiv
			FREX	optic, full, mall, sueld_comision, vendedor_optic, vendedor, tiemp_complet, perfumeri, benefici_sueld, complet_mall
19	Mantención Industrial	Industrial maintenance	Highest Probability	tecnic, mantencion, mecan, equip, electron, industrial, manten, electr, maquinari, tecnic_electr
			FREX	maquinari, mantencion, tecnic_electron, electromecan, tecnic_electr, mecan, soldadur, tecnic_mecan, maquin, prevent_correct
20	Computación Nivel Usuario	Computer User Level	Highest Probability	nivel, manej, usuari, nivel_usuari, offic, excel, tecnic, computacional, conoc, intermedi
			FREX	offic, nivel_usuari, nivel_intermedi, usuari, intermedi, nivel, offic_nivel, computacion_nivel, excel, computacion
21	Psicólogos-Recursos Humanos	Psychologist-Human resources	Highest Probability	psicolog, seleccion, evalu, consultor, lanc, fre, fre_lanc, reclut, freelanc, psicolog_fre
			FREX	freelanc, seleccion, psicolog, psicolog_freelanc, reclut, psicolaboral, evalu, consultor, reclut_seleccion, entrev_baj

22	Abogados	Lawyers	Highest Probability	estudi, tecnic, superior, estudi_tecnic, especial_cert, estudi_superior, abog, min, proces, tecnic_superior
			FREX	estudi, especial_cert, abog, estudi_superior, min, estudi_tecnic, superior, penal, profesional_estudi, tecnic_superior
23	Días de trabajo en aviso de empleo	Job postings work arrangements	Highest Probability	dinam, fin, seman, fin_seman, part, canal, supermerc, festiv, inclu, client
			FREX	fin_seman, seman, festiv, dinam, canal, person_dinam, seman_festiv, fin_excluyent, inclu_fin, dinam_proactiv
24	Sin definir	Undefinable	Highest Probability	coordin, comun, hac, period, informacion, relacion, tod, organizacion, medi, hac_cert
			FREX	coordin, hac_cert, comun, tod, extension, especializacion, informacion, period, hac, aplicacion
25	Control de Calidad	Quality Control	Highest Probability	calid, gestion, profesional, control, sistem, quimic, iso, ambient, control_calid, norm
			FREX	sistem_gestion, control_calid, ambiental, ingenier_prevencion, prim, calid, medi_ambient, mader, prim_nivel, norm_iso
26	Beneficios en el aviso de empleo	Job posting rewards	Highest Probability	laboral, merc, integr, benefici, ambient, profesional, grat, estabil, desarroll, estabil_laboral
			FREX	estabil_laboral, atract, integr_sol, desarroll_profesional, grat, proyeccion, remuneracion, logr_objet, grat_ambient, estabil
27	Empleados restaurant de comida rápida	Fast food restaurants staff	Highest Probability	equip, atencion, cumpl, administr, administracion, cocin, control, dentr_principal, servici, retail
			FREX	food, fast, fast_food, principal_encuentr, dentr_principal, retail_fast, cumpl_estandar, equip_mejor, cocin, habil_gestion
28	Profesionales en Agricultura	Agricultural professionals	Highest Probability	jef, product, oper, plant, proces, produccion, agricol, agronom, supervis, respons
			FREX	agronom, exterior, ingenier_agronom, agricol, comerci_exterior, export, tecnic_agricol, produccion, comerci, jef_plant
29	Profesionales de la Electricidad	Electricity professionals	Highest Probability	electr, supervisor, ejecucion, ingenier_ejecucion, instal, manten, servici, ayud, distribucion, puest
			FREX	electr, maestr, ejecucion_electr, ingenier_electr, termic, instal, energi, ingenier_ejecucion, central_termic, puest
30	Habilidades requeridas en el aviso de empleo	Job posting soft skills requirements	Highest Probability	capac, equip, baj, presion, baj_presion, proactiv, orientacion, habil, alta, person
			FREX	capac_equip, capac_liderazg, capac, form_equip, orientacion_logr, presion, baj,

				maner_indefin, baj_presion, proactiv_capac
31	Requisito de licencia de conducir en aviso de empleo	Job postings driving licence requirements	Highest Probability	licenci, conduc, licenci_conduc, clas, oper, conduc_clas, telefon, equip, lid, cent
			FREX	licenci, licenci_conduc, conduc_clas, cent, call, call_cent, conduc, clas, conductor, transcom
32	Recursos Humanos	Human resources	Highest Probability	recurs, human, recurs_human, laboral, administracion, personal, remuner, gestion, legislacion, rrhh
			FREX	recurs_human, human, recurs, rrhh, legislacion_laboral, legislacion, remuner, administr_recurs, asistent_social, ley_laboral
33	Fuerza laboral en sector de las Tecnologias de la Información y Comunicaciones, TICs	ICT labour	Highest Probability	informat, desarroll, program, dat, sistem, ingenier, bas, soport, proyect, anal
			FREX	informat, sql, soport, bas, bas_dat, ingenier_informat, dat, ejecucion_informat, php, sql_serv
34	Sin definir	Undefinable	Highest Probability	buen, comprob, necesit, personal, gan, sector, trabaj, diccion, buen_diccion, sueld
			FREX	buen_diccion, gan, salud_compat, comprob, compat, biling, necesit, diccion, recomend, caracterist
35	Bancos e Instituciones Financieras	Banking & Financial institutions	Highest Probability	banc, financ, institucion, relacion, event, public, financier, sucursal, bancari, tecnic
			FREX	event, relacion_public, banc, oficin_indic, financ, anfitrion, variabl, institucion, cuerp, prestigi_institucion
36	Salud	Health	Highest Probability	clinic, auxili, oncolog, institut, clinic_oncolog, institut_clinic, postulacion, plaz, capacitacion, enfermeri
			FREX	oncolog, clinic_oncolog, institut_clinic, plaz_postulacion, estabil_capacitacion, tecnic_enfermeri, auxili, capacitacion_continu, continu_merc, falp
37	Prácticas profesionales	Apprenticeship	Highest Probability	practic, profesional, alumn, univers, academ, educacion, universitari, docent, alumn_practic, quimic
			FREX	practic, docent, academ, practic_profesional, univers, alumn, quimic_farmaceut, alumn_practic, educacion_superior, docenci
38	Sin definir	Undefinable	Highest Probability	orient, disen, termin, person, minim, material, orient_person, client, respons, estudi
			FREX	termin, grafic, merc_grat, disen, tare, material_construccion, orient_person, operari, disen_grafic, orient_tare
39	Habilidades requeridas en el aviso de empleo	Job posting skills requirements	Highest Probability	retail, laboral, complet, medi, orden, complementari, benefici, actitud, mejor, medi_complet

			FREX	complementari, complementari_salud, actitud, ambient_laboral, buen_servici, excelent_clim, pasion, concret_hac, concret, apasion
40	Sin definir	Undefinable	Highest Probability	entrev, competent, lug, fre_lanc, fre, lanc, entrev_competent, psicolog_fre, psicolog, plaz
			FREX	competent_lug, fisic_entrev, evalu_personal, entrev, lug, entrev_competent, lug_fisic, competent, escan, fre_lanc
41	Salud	Health	Highest Probability	medic, enfermer, clinic, salud, bio, equip, atencion, victim, enfermer_clinic, pacient
			FREX	victim, enfermer, equip_medic, bio, represent_vent, bio_bio, enfermer_clinic, medic, atencion_victim, delit
42	Beneficios en el aviso de empleo	Job posting rewards	Highest Probability	segur, clim, clim_laboral, credit, ejecut, vent, retail, sucursal, buen_clim, jef_sucursal
			FREX	clim_laboral, sector_retail, buen_clim, clim, jef_sucursal, vent_segur, credit, constant_crecimient, corredor, crecimient_buen
43	Calificaciones requeridas en el aviso de empleo	Job posting qualifications requirements	Highest Probability	segur, curs, guardi, guardi_segur, vigent, certific, curs_vigent, relator, vigil, segur_curs
			FREX	curs_vigent, segur_curs, carabiner, relator, segur_guardi, relator_curs, curs_guardi, vigil, curs, vigent
44	Habilidades requeridas en el aviso de empleo	Job posting skills requirements	Highest Probability	relacion, buen, interpersonal, relacion_interpersonal, manej, excelent, buen_relacion, capac, secretari, buen_manej
			FREX	buen_relacion, buen_manej, relacion_interpersonal, interpersonal, manej_relacion, excelent_manej, buen_nivel, relacion, conflict, capac_cumpl
45	Servicios de Alimentación	Catering	Highest Probability	servici, profesional, alimentacion, administracion, administr, manej, supervision, ingres, servici_alimentacion, casin
			FREX	servici_alimentacion, alimentacion, nutricion, casin, administr_servici, ase, administr_casin, ingres_pagin, casin_alimentacion, orient_profesional
46	Calificaciones requeridas en el aviso de empleo	Job posting qualifications requirements	Highest Probability	medi, complet, ensen, medi_complet, ensen_medi, cajer, person, curs, bancari, educacion
			FREX	medi_complet, ensen, ensen_medi, educacion_medi, complet, medi, cajer, cajer_bancari, curs_cajer, servipag
47			Highest Probability	sext, maner, inter, inter_desarroll, person_alta, desempen,

	Habilidades requeridas en el aviso de empleo	Job posting skills requirements		personal_desempen, period, desarroll, person
			FREX	person_alta, sext, presion_respons, maner, inter_desarroll, nivel_profesional, personal_desempen, inter, focaliz, profesional_integr
48	Prevención de Riesgos Laborales	Occupational risk prevention	Highest Probability	riesg, prevencion, prevencion_riesg, expert, administracion, ejecucion, salud, segur, practic, expert_prevencion
			FREX	prevencion_riesg, prevencion, riesg, expert, expert_prevencion, ingenieri_ejecucion, ejecucion_administracion, accident, administracion_ingenieri, tecnic_prevencion
49	Servicio al Cliente	Customer service	Highest Probability	atencion, client, atencion_client, ejecut, ejecut_atencion, telecomun, oficin, presencial, comercial, client_presencial
			FREX	client_presencial, atencion_oficin, modul_atencion, presencial, ejecut_atencion, telecomun_ejecut, atencion_client, oficin_comercial, presencial_modul, turism
50	Ventas	Sales	Highest Probability	vent, ejecut, ejecut_vent, met, zon, comercial, cumplimient, equip, terren, servici
			FREX	zon_sur, ejecut_vent, supervisor_vent, jef_vent, sur, vent, equip_vent, met_vent, telemarketing, fuerz_vent
51	Ventas	Sales	Highest Probability	vendedor, terren, vent, product, client, consum, vent_terren, masiv, consum_masiv, carter
			FREX	vendedor_terren, consum_masiv, product_consum, vendedor_tecnic, consum, masiv, vent_terren, terren, minim_vent, punt
52	Asistentes de Administración	Management assistants	Highest Probability	administr, asistent, ingres, cajer, caj, secretari, manej, document, bancari, recepcion
			FREX	asistent_administr, correspondent, diner, archiv, recaud, recib, secretari, caj, documentacion, caj_manej
53	Logística	Logistics	Highest Probability	orden, compr, administr, asistent, bodeg, control, principal, inventari, logist, material
			FREX	compr, bodeg, insum, adquisicion, orden, inventari, despach, proveedor, factur, abastec