University of Reading

Inference in functional factor models with applications to yield curves

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Horváth, L., Kokoszka, P., VanderDoes, J. and Wang, S. ORCID: https://orcid.org/0000-0003-2113-5521 (2022) Inference in functional factor models with applications to yield curves. Journal of Time Series Analysis, 43 (6). pp. 872-894. ISSN 1467-9892 doi: 10.1111/jtsa.12642 Available at https://centaur.reading.ac.uk/102538/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1111/jtsa.12642

Publisher: Wiley-Blackwell

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

University of Reading

Central Archive at the University of Reading

Reading's research outputs online

Check for updates

JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. (2022)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12642

ORIGINAL ARTICLE

INFERENCE IN FUNCTIONAL FACTOR MODELS WITH APPLICATIONS TO YIELD CURVES

LAJOS HORVÁTH,[a] PIOTR KOKOSZKA,[b] JEREMY VANDERDOES[c] AND SHIXUAN WANG[d]

[a]*Department of Mathematics, University of Utah, Salt Lake City, UT, USA*
[b]*Department of Statistics, Colorado State University, Fort Collins, CO, USA*
[c]*Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada*
[d]*Department of Economics, University of Reading, Reading, UK*

This article develops a set of inferential methods for functional factor models that have been extensively used in modelling yield curves. Our setting accommodates both temporal dependence and heteroskedasticity. First, we introduce an estimation approach based on minimizing the least-squares loss function and establish the consistency and asymptotic normality of the estimators. Second, we propose a goodness-of-fit test that allows us to determine whether a specific model fits the data. We derive the asymptotic distribution of the test statistics, and this leads to a significance test. A simulation study establishes the good finite-sample performance of our inferential methods. An application to US and UK yield curves demonstrates the generality of our framework, which can accommodate both sparsely and densely observed yield curves.

Received 27 March 2021; Revised 5 December 2021; Accepted 25 January 2022

Keywords: Yield curve; functional time series; functional factor model; estimation; goodness-of-fit test; heteroscedasticity.

MOS subject classification: 62F03, 62F05, 62F10, 62F12, 60F05.

1. INTRODUCTION

For three decades, models akin to the Nelson–Siegel model have been used to quantify the term structure of various economic and financial variables, including yields, spot rates, and futures, with hundreds of papers using it in various contexts. The general modelling paradigm is well known, but we recall it to be able to explain several issues. At each period i, which generally corresponds to day, week, or month, we observe a curve $X_i(t)$, where the argument t corresponds to the time to the expiration of a contact or to the maturity of a fixed-income investment. Nelson and Siegel (1987) introduced the following three curves:

$$f_1(t;\lambda) = 1, \quad f_2(t;\lambda) = \frac{1-e^{-\lambda t}}{\lambda t} \quad \text{and} \quad f_3(t;\lambda) = \frac{1-e^{-\lambda t}}{\lambda t} - e^{-\lambda t},$$

known as the 'level', the 'slope', and the 'curvature'. The essence of the modelling approach is that each curve X_i can be in some sense approximated by a linear combination of the above three curves: $X_i(t) \approx \sum_{\ell=1}^{3} b_{i,\ell} f_\ell(t;\lambda)$. This is definitely not the only approach to modelling the term structure. At about the same time, Bliss and Fama (1987) proposed a fully non-parametric approach which fits a smooth curve to the data. However, the Nelson–Siegel paradigm has become much more popular, chiefly because of its ability to reduce term structure to the vector $[b_{1,\ell}, b_{2,\ell}, b_{3,\ell}]$. Various formal or informal statistical approaches are easier to apply to three-dimensional vectors than to curves. It has been, however, immediately recognized that such a reduction is possible only if the value of the shape parameter λ is available. Nelson and Siegel (1987), in fact, used different values of λ for different periods, but it is now a common practice to fix λ over the time period under consideration. This fixed value is

* Correspondence to: Shixuan Wang, Department of Economics, University of Reading, Reading, RG6 6EL, UK.
Email: shixuan.wang@reading.ac.uk

© 2022 The Authors. *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

either based on past experience, for example, the value given in Diebold and Rudebusch (2013), or is estimated in some way prior to further analysis. It is well known that with slightly different 'fixed' values of λ, the coefficients $b_{i,\ell}$ can take on very different, often erratically behaved, values, limiting their usefulness and interpretability. Such findings are discussed in Vermani (2012) and Annaert *et al.* (2013), among others. The Nelson–Siegel model is highly nonlinear in its four parameters, which may cause estimation difficulties.

The objective of this article is to develop estimation and testing theory and methodology for a broad class of yield curve models, which includes the Nelson–Siegel model. Our approach is based on established principles of statistical inference, which involve the specification of model parameters and structure of model errors. It allows us to formulate results on consistency and asymptotic distribution of parameter estimators and derive a goodness-of-fit test. This test permits us to evaluate the admissibility of a specific model over a specific time period within the Neyman–Pearson testing paradigm. Existing literature, including Diebold *et al.* (2006) and Diebold and Li (2006), focuses on practical aspects of yield curve fitting and forecasting. Our study is different, and complementary, because we provide a set of inferential methods based on a general statistical framework. It is hoped that the rigorous and general development presented in this article will be a useful addition to a large body of applied work on models of this type.

We consider functional observations $X_1(t), X_2(t), \ldots, X_N(t)$ defined on the interval $\mathcal{T}$. We assume that these observations follow the model

$$X_i(t) = \sum_{\ell=1}^{K} b_{i,\ell,0} f_\ell(t; \lambda_0) + \varepsilon_i(t), \quad \text{with } E\varepsilon_i(t) = 0, \quad t \in \mathcal{T}, \quad 1 \le i \le N, \tag{1.1}$$

where the random coefficients satisfy

$$b_{i,\ell,0} = c_{\ell,0} + e_{i,\ell} \quad \text{with } Ee_{i,\ell} = 0, \quad 1 \le \ell \le K \text{ and } 1 \le i \le N. \tag{1.2}$$

Under model (1.1) and (1.2),

$$EX_i(t) = \sum_{\ell=1}^{K} c_{\ell,0} f_\ell(t; \lambda_0), \quad t \in \mathcal{T} \text{ and } 1 \le i \le N,$$

that is, the mean of the observations can be written as a linear combination of the functions $f_1(t; \lambda_0)$, $f_2(t; \lambda_0), \ldots, f_K(t; \lambda_0)$, where the functions $f_1, f_2, \ldots, f_K$ are known, and $\lambda_0 \in R^d$ is the true value of an unknown parameter. The parameter vector specifying the model is

$$\mathbf{a} = (c_1, c_2, \ldots, c_K, \lambda^\top)^\top \in R^{K+d}, \tag{1.3}$$

where the subscript 0 indicates the true values. As explained above, in most current applications λ is considered as a fixed value obtained in some way. In subsequent analyses, the curves $f_k(t; \lambda_0)$ are treated as given deterministic functions. We will call these functions functional factors, or simply factors. In our approach, we focus on the estimation of the whole parameter vector $\mathbf{a}$.

To give an idea of the scope of the models in current use, and consequently the scope of the theory and methodology we develop, we list several examples. All functions of t are defined on a compact interval $\mathcal{T} \subset [0, \infty)$. While historical data are generally available at discrete subsets, which do not need to be regularly spaced, contracts of any maturities can be traded, so factor models of the type we study have traditionally been considered on a continuous domain $\mathcal{T}$. As we will see in Section 4, some data are available as densely observed curves.

Example 1.1. (Svensson 1994)**.** Compared to the Nelson–Siegel model, in this model $\lambda_0 = (\lambda_{0,1}, \lambda_{0,2})^\top$ is a two-dimensional parameter. There are four factor functions. The first three coincide with those of the

© 2022 The Authors.
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

DOI: 10.1111/jtsa.12642

Nelson–Siegel model, that is,

$$f_1(t;\lambda_0) = 1,\ f_2(t;\lambda_0) = \frac{1-e^{-\lambda_{0,1}t}}{\lambda_{0,1}t}, f_3(t;\lambda_0) = \frac{1-e^{-\lambda_{0,1}t}}{\lambda_{0,1}t} - e^{-\lambda_{0,1}t}.$$

The fourth function, which adds a second 'hump', is the same as the curvature function in the Nelson–Siegel model but uses $\lambda_{0,1} < \lambda_{0,2}$, that is,

$$f_4(t;\lambda_0) = \frac{1-e^{-\lambda_{0,2}t}}{\lambda_{0,2}t} - e^{-\lambda_{0,2}t}.$$

Filipović (1999) points out that the Nelson–Siegel model is not arbitrage free. Diebold and Rudebusch (2013) argue that the four-factor model in Example 1.1 is not arbitrage free either. An arbitrage-free model requires that each curvature factor is paired with a slope factor with the same rate of mean reversion, which is impossible with a single slope factor used in Example 1.1. Example 1.2 gives an arbitrage-free model.

Example 1.2. (Christensen *et al.* 2009)**.** The five factors (with $\lambda_{0,1} < \lambda_{0,2}$)

$$f_1(t;\lambda_0) = 1,\ f_2(t;\lambda_0) = \frac{1-e^{-\lambda_{0,1}t}}{\lambda_{0,1}t}, f_3(t;\lambda_0) = \frac{1-e^{-\lambda_{0,2}t}}{\lambda_{0,2}t},$$
$$f_4(t;\lambda_0) = \frac{1-e^{-\lambda_{0,1}t}}{\lambda_{0,1}t} - e^{-\lambda_{0,1}t} \text{ and } f_5(t;\lambda_0) = \frac{1-e^{-\lambda_{0,2}t}}{\lambda_{0,2}t} - e^{-\lambda_{0,2}t},$$

are used to get an arbitrage-free model.

Example 1.3. (Chambers *et al.* 1984)**.** The polynomials of

$$f_1(t) = 1,\ f_2(t) = t,\ \dots,\ f_K(t) = t^K,$$

were used by Chambers *et al.* (1984) to model zero-coupon yield curves. Polynomials are able to approximate smooth functions if K is sufficiently large. This model does not contain the parameter λ.

Example 1.4. (Yallup 2012)**.** The exponential model

$$f_1(t;\lambda_0) = 1,\ f_2(t;\lambda_0) = e^{-\lambda_0 t},\ f_3(t;\lambda_0) = e^{-2\lambda_0 t},\ \dots,\ f_K(t;\lambda_0) = e^{-K\lambda_0 t},$$

is an alternative to the polynomial model of Example 1.3. The Bank of Canada uses this model with $K = 9$ to estimate yield curves on Canadian bond data (cf. Bank for International Settlements, 2005). Similar to Example 1.3, the exponential functions form a basis as well, so a linear combination of the f_i should capture EX_i if K is large. However, by choosing λ_0 suitably, a relatively small K can be used.

The models outlined in the above examples are often used in economic and finance research. Our objective is to develop inference for them. There are alternative non-parametric approaches that may be advantageous in some respects, see for example, Hays *et al.* (2012).

Our methodology and theory can accommodate temporally dependent and heteroscedastic functional errors ε_i in (1.1) and multi-variate errors $e_{i,\ell}$ in (1.2), as specified in Assumptions 2.5 and 2.6. The vector of coefficients cannot only follow an unknown time series model, but this model can change within the time period over which observations are available. This approach thus substantially extends homoskedastic parametric models such as the AR(1) in Diebold *et al.* (2006) and VAR(1) in Diebold and Li (2006). We emphasize that our theory does not use any specific model for the coefficients, but rather uses a general non-parametric assumption on the temporal dependence, broadly similar to a mixing condition. In different contexts, heteroskedastic errors were considered by Cavaliere *et al.* (2011), Xu (2015), and Górecki *et al.* (2018), among others.

DOI: 10.1111/jtsa.12642 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

There are several novel mathematical techniques used in the proofs of the results stated in Section 2. The broadest innovation is the treatment of functional factor models under the assumption of a changing stochastic structure of the errors. We developed several approaches, which may be useful in similar contexts. To prove the almost sure convergence in Theorem 2.1, we must study the behaviour of the variances $\xi_i = \int_{\mathcal{T}} [X_i(t) - EX_i(t)]^2 dm(t)$. We use a blocking technique by considering $2^k \leq N < 2^{k+1}$. This allows us to obtain delicate bounds in probability via Menshov's inequality, and then apply the Borel–Cantelli lemma. The proof of Theorem 2.2 is built up of very complex manipulations revealing the asymptotic distribution, which has a non-trivial form. The proofs of the remaining results use this form to justify the goodness-of-fit test. It would not have been possible to derive this test without knowing the precise limit. Numerical implementation of our method is not trivial either. The structure of the yields data must be taken into account. We provide precise algorithms.

The remainder of the article is organized as follows. Section 2 is dedicated to the development of the asymptotic theory for the models discussed above. This theory is conditional on the availability of suitable long-run covariance estimators, which remain consistent under heteroskedastic errors. Such estimators are studied in Section 3. Section 4 illustrates the statistical methodology and theory derived in this article by applying it to US and UK yield curves. Finite-sample properties of our methods are explored in Section 5. Online Supporting Information contains the proofs of the results stated in Sections 2 and 3, details of the numerical implementation of our methods, as well as additional results and figures.

2. ASSUMPTIONS AND LARGE-SAMPLE RESULTS

We work within the Hilbert space $L^2 = L^2(m)$ of square integrable functions defined on the compact interval $\mathcal{T} \subset [0, \infty)$. The integration over $\mathcal{T}$ is denoted simply by $\int$. To be able to handle various forms of data, we consider an abstract positive measure m on $\mathcal{T}$. This can be a discrete counting measure, which is convenient in many applications. We assume that the functions $f_\ell(t; \lambda)$ are elements of L^2 for any admissible value of the parameter vector λ. The functional observations X_i and the errors ε_i are random elements of L^2.

Our approach consists in minimizing the least-squares loss function

$$U_N(\mathbf{a}) = \frac{1}{N} \sum_{i=1}^{N} \int \left(X_i(t) - \sum_{\ell=1}^{K} c_\ell f_\ell(t; \lambda) \right)^2 dm(t), \tag{2.1}$$

where $\mathbf{a}$ is given by (1.3). The estimator $\hat{\mathbf{a}}_N$ is thus defined by

$$\hat{\mathbf{a}}_N = \arg\max_{\mathbf{a} \in \mathbf{A}} U_N(\mathbf{a}). \tag{2.2}$$

We begin by listing the assumptions on the measure m, the parameter space $\mathbf{A}$, and other objects in (1.1) and (1.2).

Assumption 2.1. The measure m satisfies one of the following assumptions:

(i) $m((-\infty, t]) = m(t)$, a function increasing by jumps at finitely many points in $\mathcal{T}$,
or
(ii) $m(B) = \int_B m^*(t)dt$ for all Borel subsets B of $\mathcal{T}$ with a function $m^*(t)$, bounded on $\mathcal{T}$.

Assumption 2.1(ii) means that the measure m is absolutely continuous with respect to the Lebesgue measure on $\mathcal{T}$ and its Radon–Nikodym derivative is bounded on $\mathcal{T}$.

Assumption 2.2. The parameter space $\mathbf{A} \subset R^{K+d}$ is a compact set and $\mathbf{a}_0$ is an interior point of $\mathbf{A}$.

Let $\mathbf{C} \subset R^K$ and $\mathbf{\Lambda} \subset R^d$ such that $\mathbf{A} \subset \mathbf{C} \times \mathbf{\Lambda}$.

We assume that the functions are smooth functions of λ.

© 2022 The Authors.
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

DOI: 10.1111/jtsa.12642

Assumption 2.3. Each function $f_\ell(t;\lambda)$ is continuous in $t \in \mathcal{T}$ and $|f_\ell(t;\lambda) - f_\ell(t;\lambda')| \le C\|\lambda - \lambda'\|$ with some $C > 0$ for all $1 \le \ell \le K$ and $\lambda, \lambda' \in \mathbf{\Lambda}$.

The next condition requires that the model is uniquely defined.

Assumption 2.4. For every $\mathbf{a} \in \mathbf{A}$, such that $\mathbf{a} \neq \mathbf{a}_0$, the set

$$\mathcal{T}(\mathbf{a}) = \left\{ t : \sum_{\ell=1}^{K} c_\ell f_\ell(t;\lambda) \neq \sum_{\ell=1}^{K} c_{\ell,0} f_\ell(t;\lambda_0) \right\}$$

has positive measure, $m(\mathcal{T}(\mathbf{a})) > 0$.

The above assumption means that the true mean curve corresponds only to the true parameter $\mathbf{a}_0$.

We model heteroscedasticity by assuming that the errors $\mathbf{e}_i = (e_{i,1}, e_{i,2}, \dots, e_{i,K}, \varepsilon_i(t))^\top$ are piecewise Bernoulli shifts. In particular, between the break points k_j in Assumption 2.5, these errors are strictly stationary. Their dependence is quantified in Assumption 2.6.

Assumption 2.5. The changes occur at times $k_1 = \lfloor N\tau_1 \rfloor < k_2 = \lfloor N\tau_1 \rfloor < \cdots < k_M = \lfloor N\tau_M \rfloor$ with $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_M < \tau_{M+1} = 1$.

We use the convention of $k_0 = 0$ and $k_{M+1} = N$. In contrast to Bardsley *et al.* (2017), we do not require that $\tau_1, \tau_2, \dots, \tau_M$ are known.

The piecewise Bernoulli shifts are defined in the following assumption:

Assumption 2.6. Let $\mathbf{e}_i = (e_{i,1}, e_{i,2}, \dots, e_{i,K}, \varepsilon_i(t))^\top = \mathbf{g}_j(\eta_i, \eta_{i-1}, \eta_{i-2}, \dots)$, $k_{j-1} < i \le k_j$, $1 \le j \le M+1$, where $\mathbf{g}_j, 1 \le j \le M+1$ are non-random functionals defined on S^∞ with values in $R^K \times L^2$, and S is a measurable space. Also, $\eta_i = \eta_i(t,\omega)$ is jointly measurable in $(t,\omega), -\infty < i < \infty$. The sequence $\mathbf{e}_i, -\infty < i < \infty$ can be approximated with ν–dependent sequences $\mathbf{e}_{i,\nu} = (e_{i,1,\nu}, e_{i,2,\nu}, \dots, e_{i,K,\nu}, \varepsilon_{i,\nu}(t))^\top$ in the sense that

$$\sum_{\nu=1}^{\infty} (E(e_{i,\ell} - e_{i,\ell,\nu})^2)^{1/2} < \infty \ \text{ for all } \ 1 \le \ell \le K, \tag{2.3}$$

and

$$\sum_{\nu=1}^{\infty} \left(E \int (\varepsilon_i(t) - \varepsilon_{i,\nu}(t))^2(t) dm(t) \right)^{1/2} < \infty, \tag{2.4}$$

where $\mathbf{e}_{i,\nu} = \mathbf{g}_j(\eta_i, \eta_{i-1}, \eta_{i-2}, \dots, \eta_{i-\nu+1}, \boldsymbol{\eta}^*_{i,\nu})$, $k_{j-1} < i \le k_j$, $1 \le j \le M+1$, $\boldsymbol{\eta}^*_{i,\nu} = (\eta^*_{i,\nu,i-\nu}, \eta^*_{i,\nu,i-\nu-1}, \eta^*_{i,\nu,i-\nu-2}, \dots)$, and the $\eta^*_{i,\nu,r}$'s are independent copies of η_0, independent of $\{\eta_i, -\infty < i < \infty\}$.

The idea of approximation of a stationary sequence with random variables which exhibit finite dependence first appeared in Ibragimov (1962). Examples of models that satisfy Assumption 2.6 (and similar assumptions that follow) are discussed in Chapter 16 of Horváth and Kokoszka (2012). Basically, all stationary time series models in practical use can be represented as Bernoulli shifts, see Wu (2005), Shao and Wu (2007), Aue *et al.* (2009), Hörmann and Kokoszka (2010), Horváth *et al.* (2013), Kokoszka *et al.* (2015), and Zhang (2016), among many other contributions. The nonlinear moving average representation admits heteroskedastic models for the errors, like those studied in Hörmann *et al.* (2013).

Theorem 2.1. If (1.1), (1.2), and Assumptions 2.1–2.6 are satisfied, then $\hat{\mathbf{a}}_N \to \mathbf{a}_0$ a.s.

To derive the limit distribution of $N^{1/2}(\hat{\mathbf{a}}_N - \mathbf{a}_0)$, we need to introduce further notation and additional assumptions.

Assumption 2.7. The functions $f_\ell(t;\lambda), 1 \le \ell \le K$, are twice continuously differentiable in a neighbourhood of λ_0.

Let $\mathbf{f}(t,\lambda) = ((f_1(t;\lambda), f_2(t;\lambda), \ldots, f_K(t;\lambda))^\top$ and

$$\mathfrak{f}(t;\lambda) = \left(\mathbf{f}(t,\lambda)^\top, \sum_{\ell=1}^K c_{\ell,0} g_{\ell,1}(t;\lambda), \sum_{\ell=1}^K c_{\ell,0} g_{\ell,2}(t;\lambda), \ldots, \sum_{\ell=1}^K c_{\ell,0} g_{\ell,d}(t;\lambda)\right)^\top, \tag{2.5}$$

where

$$g_{\ell j}(t;\lambda) = \frac{\partial f_\ell(t;\lambda)}{\partial \lambda_j}, \quad 1 \le \ell \le K, 1 \le j \le d,$$

$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_d)^\top$. Also, with $\mathbf{e}_j = (e_{j,1}, e_{j,2}, \ldots, e_{j,K})^\top$ we define for $1 \le i \le M+1$

$$\lim_{N\to\infty} \frac{1}{k_i - k_{i-1}} E\left(\sum_{j=k_{i-1}+1}^{k_i} \mathbf{e}_j \mathbf{e}_j^T\right) = \mathbf{D}_i, \tag{2.6}$$

$$\lim_{N\to\infty} \frac{1}{k_i - k_{i-1}} E\left(\sum_{j=k_{i-1}+1}^{k_i} \mathbf{e}_j \varepsilon_j(t)\right) = \mathbf{d}_i(t), \tag{2.7}$$

and

$$\lim_{N\to\infty} \frac{1}{k_i - k_{i-1}} E\left(\sum_{j=k_{i-1}+1}^{k_i} \varepsilon_j(t)\varepsilon_j(s)\right) = d_i(s,t). \tag{2.8}$$

We note that the existence of $\mathbf{D}_i, \mathbf{d}_i(t)$, and $d_i(s,t)$ follows from Assumption 2.6. Let $\Gamma(t)$ be a Gaussian process with $E\Gamma(t) = 0$ and

$$\begin{aligned} C(t,s) &= E\Gamma(t)\Gamma(s) \\ &= \sum_{i=1}^{M+1} (\tau_i - \tau_{i-1}) \left(\mathbf{f}^\top(t)\mathbf{D}_i\mathbf{f}(s) + \mathbf{f}^\top(t)\mathbf{d}_i(s) + \mathbf{f}^\top(s)\mathbf{d}_i(t) + d_i(t,s)\right). \end{aligned}$$

To derive the limit distribution of $N^{1/2}(\hat{\mathbf{a}}_N - \mathbf{a}_0)$, we need to replace (2.3) and (2.4) with mildly stronger conditions.

Assumption 2.8. There are $\kappa_1 > 2$, $\kappa_2 > 2$, and $C > 0$ such that

$$(E(e_{i,\ell} - e_{i,\ell,v})^{\kappa_1})^{1/\kappa_1} < Cv^{-\kappa_2} \text{ for all } 1 \le \ell \le K, \tag{2.10}$$

and

$$\left(E\int (\varepsilon_i(t) - \varepsilon_{i,v}(t))^{\kappa_1}(t)dm(t)\right)^{1/\kappa_1} < Cv^{-\kappa_2}, \tag{2.11}$$

where $e_{i,\ell,v}$ and $\varepsilon_{i,v}(t), 1 \le i \le N, 1 \le \ell \le K$ are defined in Assumption 2.6.

The next assumption is needed to have a non-degenerate limit.

© 2022 The Authors.
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

DOI: 10.1111/jtsa.12642

Assumption 2.9. The matrix $\mathfrak{B}$ defined by

$$\mathfrak{B} = \int \mathfrak{f}(t;\lambda_0)\mathfrak{f}^\top(t;\lambda_0)dm(t) \tag{2.12}$$

is non-singular. (The vector $\mathfrak{f}(t;\lambda)$ is defined in (2.5).)

We note that Assumption 2.9 means that the coordinates of $\mathfrak{f}(t;\lambda_0)$ are linearly independent, that is, $f_1, f_2, \ldots, f_K$ and their derivatives with respect to λ are linearly independent at λ_0 as a function of t in L^2.

Theorem 2.2. If (1.1), (1.2), and Assumptions 2.1–2.9 are satisfied, then we have that

$$N^{1/2}(\hat{\mathbf{a}}_N - \mathbf{a}_0) \mathcal{D}\to \ \boldsymbol{\xi} = \mathfrak{B}^{-1}\left(\int \Gamma(t)\mathfrak{f}(t,\lambda_0)dm(t)\right).$$

It is clear that $\boldsymbol{\xi}$ is a $(K+d)$-dimensional normal random variable with $E\boldsymbol{\xi} = \mathbf{0}$, and the covariance matrix

$$\mathfrak{C} = E\boldsymbol{\xi}\boldsymbol{\xi}^\top = \mathfrak{B}^{-1}\left(\iint \mathfrak{f}(t;\lambda_0)C(t,s)\mathfrak{f}^\top(s;\lambda_0)dm(t)dm(s)\right)\mathfrak{B}^{-1}.$$

We discuss the estimation of $\mathfrak{C}$ in Section 3.

We now proceed to results that will allow us to investigate whether a specific model defined by (1.1) and (1.2) fits the data, that is, the goodness-of-fit testing. Recall that the estimator for $\mathbf{a}_0$ has the form

$$\hat{\mathbf{a}}_N = \left(\hat{c}_{1,N}, \ldots, \hat{c}_{K,N}, \hat{\boldsymbol{\lambda}}_N^\top\right)^\top \in R^{K+d}.$$

Our testing procedure is based on a functional of the process

$$Z_N(t) = N^{-1/2}\sum_{i=1}^{N}\left(X_i(t) - \sum_{\ell=1}^{K}\hat{c}_{\ell,N}f_\ell(t;\hat{\boldsymbol{\lambda}}_N)\right), \quad t \in \mathcal{T}.$$

If the null limit distribution of the process $\{Z_N(t), t \in \mathcal{T}\}$ can be established, in a suitable function space, then any continuous functional of this process can serve as a test statistic. In this article, we focus on the L^2 norm with respect to another abstract measure, n, which can be taken to be m.

Assumption 2.10. The measure n satisfies one of the following assumptions:

(i) $n((-\infty, t]) = n(t)$, a function increasing only at the jump points on m. (Not all jump points of m need to be used.)
or
(ii) $n(B) = \int_B n^*(t)dm(t)$ for all Borel subsets B of $\mathcal{T}$ with a function $n^*(t)$, bounded on $\mathcal{T}$.

The testing procedure needs a simple modification of Assumption 2.8. We require higher moments and v-approximability in L^2 with respect to the measure n on $\mathcal{T}$.

Assumption 2.11. There are $\kappa_1 > 4$, $\kappa_2 > 4$, and $C > 0$ such that

$$(E(e_{i,\ell} - e_{i,\ell,v})^{\kappa_1})^{1/\kappa_1} < Cv^{-\kappa_2} \ \text{ for all } \ 1 \le \ell \le K, \tag{2.13}$$

and

$$\left(E\int(\varepsilon_i(t) - \varepsilon_{i,v}(t))^{\kappa_1}(t)dn(t)\right)^{1/\kappa_1} < Cm^{-\kappa_2}, \tag{2.14}$$

where $e_{i,\ell,v}$ and $\varepsilon_{i,v}(t), 1 \le i \le N, 1 \le \ell \le K$ are defined in Assumption 2.6.

To formulate the null limit distribution of our goodness-of-fit statistic, we introduce Gaussian processes $\{\Gamma_N^0(t), t \in \mathcal{T}\}$ defined by

$$\Gamma_N^0(t) = \Gamma_N(t) - \left(\int \Gamma_N(u) \mathfrak{f}^T(u; \lambda_0) dm(u) \right) \mathfrak{B}^{-1} \mathfrak{f}(t; \lambda_0), \quad t \in \mathcal{T},$$

where the processes Γ_N have the same distribution as the Gaussian process Γ with covariances (2.9). The existence of the sequence $\{\Gamma_N\}$ is part of the claim of the following theorem. Each process Γ_N^0 is clearly Gaussian and its distribution does not depend on N.

Theorem 2.3. If (1.1), (1.2), and Assumptions 2.2–2.11 are satisfied, then

$$\int (Z_N(t) - \Gamma_N^0(t))^2 dn(t) = o_P(1).$$

Let $\{\Gamma^0(t), 0 \le t \le 1\} \mathcal{D} = \{\Gamma_N^0(t), 0 \le t \le 1\}$. We note that $E\Gamma^0(t) = 0$ and

$$\begin{aligned} \mathcal{D}(t,s) = C(t,s) &- \left(\int C(s,u) \mathfrak{f}^\top(u; \lambda_0) dm(u) \right) \mathfrak{B}^{-1} \mathfrak{f}(t; \lambda_0) \\ &- \left(\int C(t,v) \mathfrak{f}^\top(v; \lambda_0) dm(v) \right) \mathfrak{B}^{-1} \mathfrak{f}(s; \lambda_0) \\ &+ \mathfrak{f}^\top(t; \lambda_0) \mathfrak{B}^{-1} \left\{ \iint \mathfrak{f}(u; \lambda_0) C(u,v) \mathfrak{f}^\top(v; \lambda_0) dm(u) dm(v) \right\} \mathfrak{B}^{-1} \mathfrak{f}(s; \lambda_0). \end{aligned}$$

Next, we discuss the consistency of the testing procedure. Under the alternative, model (1.1)–(1.2) does not fit the data, which is formalized in the following assumption:

Assumption 2.12. The sequence $\{X_i\}$ is strictly stationary in $L_2(m)$ and

$$\inf_{\mathbf{a} \in \mathbf{A}} \int \left(EX_1(t) - \sum_{\ell=1}^{K} c_\ell f_\ell(t; \lambda) \right)^2 dn(t) > 0,$$

where $\mathbf{a} = (c_1, c_2, \ldots c_K, \lambda^\top)^\top$, and $\mathbf{A}$ is defined in Assumption 2.2.

Theorem 2.4. If Assumptions 2.2,2.3,2.5,2.8, and 2.12 are satisfied, then we have that

$$\frac{1}{N} \int Z_N^2(t) dn(t) \; P\to \; c_0$$

with some $c_0 > 0$.

The distributions of functionals of Γ^0 depend on several unknown parameters, so it is not immediately obvious how to use Theorem 2.3 to test the validity of Nelson–Siegel-type models. We now explain how critical values can be obtained.

Theorem 2.3 implies that

$$\widehat{T}_N = \int Z_N^2(t) dn(t) \; \xrightarrow{\mathcal{D}} \; \int (\Gamma^0(t))^2 dn(t). \tag{2.15}$$

By the spectral theorem, there are non-negative eigenvalues $\zeta_1 \ge \zeta_2 \ge \zeta_3 \ge \cdots$ and corresponding orthonormal eigenfunctions $\phi_1(t), \phi_2(t), \phi_3(t), \ldots$ satisfying

$$\zeta_i \phi_i(t) = \int \mathcal{D}(t,s) \phi_i(s) dn(s), \quad i = 1, 2, \ldots \tag{2.16}$$

© 2022 The Authors.
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

DOI: 10.1111/jtsa.12642

According to the Karhunen–Loéve expansion

$$\int (\Gamma^0(t))^2 dn(t) = \sum_{j=1}^{\infty} \zeta_j \mathcal{N}_j^2, \tag{2.17}$$

where $\mathcal{N}_1, \mathcal{N}_2, \ldots$ are independent standard normal random variables. The eigenvalues in (2.17) are unknown, so we need to estimate them from the observations. Assume that $\widehat{\mathcal{D}}_N$ is an L^2 estimator of $\mathcal{D}$, that is,

$$\iint (\widehat{\mathcal{D}}_N(t,s) - \mathcal{D}(t,s))^2 dn(t) dn(s) = o_P(1). \tag{2.18}$$

A construction leading to such an estimator is described in Section 3. Using again the spectral theorem, there are non-negative empirical eigenvalues $\hat{\zeta}_{1,N} \geq \hat{\zeta}_{2,N} \geq \hat{\zeta}_{3,N} \geq \ldots$ and corresponding orthonormal eigenfunctions $\hat{\phi}_{1,N}(t), \hat{\phi}_{2,N}(t), \hat{\phi}_{3,N}(t), \ldots$ such that

$$\hat{\zeta}_{i,N} \hat{\phi}_{i,N}(t) = \int \widehat{\mathcal{D}}_N(t,s) \hat{\phi}_{i,N}(s) dn(s). \tag{2.19}$$

Putting together (2.16), (2.18), and (2.19), we obtain automatically (cf. Horváth and Kokoszka, 2012, p. 34)

$$|\hat{\zeta}_{i,N} - \zeta_i| = o_P(1)$$

for every fixed i. Hence

$$\chi_{L,N}^2 = \sum_{j=1}^{L} \hat{\zeta}_{j,N} \mathcal{N}_j^2 \tag{2.20}$$

is a good approximation for the distribution of the integral in (2.17) if L and N are suitably large. Thus, if we compute the estimates $\hat{\zeta}_{i,N}$, we can approximate the null distribution of the test statistics $\widehat{T}_N$ in (2.15). These estimates can be obtained by numerically solving (2.19). If the critical value is defined by

$$P\{\chi_{L,N}^2 > c_{L,N}(\alpha)\} = \alpha, \tag{2.21}$$

then under the null hypothesis we have that

$$\lim_{L\to\infty} \lim_{N\to\infty} P\left\{ \int Z_N^2(t) dn(t) > c_{L,N}(\alpha) \right\} = \alpha, \tag{2.22}$$

and under the alternative

$$\lim_{L\to\infty} \lim_{N\to\infty} P\left\{ \int Z_N^2(t) dn(t) > c_{L,N}(\alpha) \right\} = 1. \tag{2.23}$$

Thus, our procedure has asymptotic size α and it is consistent. Conditions (2.22) and (2.23) are verified in the online Supporting Information.

3. ESTIMATION OF $\mathfrak{C}$, $C(t,s)$ AND $\mathcal{D}$

The estimation of the matrix $\mathfrak{B}$ is conceptually easy. Recall that

$$\mathfrak{B}(\lambda) = \int \mathfrak{f}(t;\lambda) \mathfrak{f}^{\top}(t;\lambda) dm(t).$$

The plug-in estimator for $\mathfrak{B}$ is

$$\widehat{\mathfrak{B}}_N = \mathfrak{B}(\hat{\lambda}_N). \tag{3.1}$$

The continuity of $\mathfrak{B}(\lambda)$ in a neighbourhood of λ_0 and Theorem 2.1 immediately imply that

$$\widehat{\mathfrak{B}}_N \xrightarrow{P} \mathfrak{B}(\lambda_0) = \mathfrak{B}. \tag{3.2}$$

Thus we need to consider only

$$\mathfrak{C}_1 = \iint \mathfrak{f}(t;\lambda_0)C(t,s)\mathfrak{f}^\top(s;\lambda_0)dm(t)dm(s)$$

to construct an estimator for $\mathfrak{C}$.

Let

$$u_j(t) = \sum_{\ell=1}^{K} e_{j,\ell} f_\ell(t;\lambda_0) + \varepsilon_j(t), \quad t \in \mathcal{T}. \tag{3.3}$$

The observations, $X_i(t)$, $1 \le i \le N$, as well as the errors form heterogenous sequences. Thus, we need to verify if the standard methods can be used to estimate $\mathfrak{C}_1$ as well as $C(t,s)$. We write

$$\sum_{i=1}^{N} u_i(t) = \sum_{\ell=1}^{M+1} \sum_{j=k_{\ell-1}+1}^{k_\ell} u_j(t)$$

and observe that

$$E\left\{\left(\sum_{i=1}^{N} u_i(t)\right)\left(\sum_{i'=1}^{N} u_{i'}(s)\right)\right\} = \sum_{\ell,\ell'=1}^{M+1} E\left\{\left(\sum_{j=k_{\ell-1}+1}^{k_\ell} u_j(t)\right)\left(\sum_{j'=k_{\ell'-1}+1}^{k_{\ell'}} u_{j'}(s)\right)\right\}.$$

Using Assumption 2.11, we get that

$$\lim_{N\to\infty} \frac{1}{N} E\left\{\left(\sum_{j=k_{\ell-1}+1}^{k_\ell} u_j(t)\right)\left(\sum_{j'=k_{\ell'-1}+1}^{k_{\ell'}} u_{j'}(s)\right)\right\} = 0,$$

if $\ell \neq \ell'$ and

$$\lim_{N\to\infty} \frac{1}{k_\ell - k_{\ell-1}} E\left\{\left(\sum_{j=k_{\ell-1}+1}^{k_\ell} u_j(t)\right)\left(\sum_{j'=k_{\ell-1}+1}^{k_\ell} u_{j'}(s)\right)\right\} = \mathbf{f}^\top(t)\mathbf{D}_\ell\mathbf{f}(s) + \mathbf{f}^\top(t)\mathbf{d}_\ell(s) + \mathbf{f}^\top(s)\mathbf{d}_\ell(t) + d_\ell(t,s).$$

Thus we get

$$\lim_{N\to\infty} E\left\{\left(\sum_{i=1}^{N} u_i(t)\right)\left(\sum_{i'=1}^{N} u_{i'}(s)\right)\right\} = C(t,s).$$

The estimation of $\mathfrak{C}_1$ is based on the projection vectors

$$\mathbf{v}_i = \int u_i(t)\mathfrak{f}(t;\lambda_0)dm(t).$$

© 2022 The Authors.
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

DOI: 10.1111/jtsa.12642

The residuals of the model are

$$q_i(t) = X_i(t) - \sum_{\ell=1}^{K} \hat{c}_{\ell,N} f_\ell(t; \hat{\boldsymbol{\lambda}}_N), \quad t \in \mathcal{T},$$

so we estimate $\mathbf{v}_i$ with

$$\hat{\mathbf{v}}_i = \int q_i(t) \mathfrak{f}(t; \hat{\boldsymbol{\lambda}}_N) dm(t).$$

The data are heteroscedastic, but we show that the long-run covariance matrix estimator still can be used. As usual, we start with the estimation of the covariances of lag ℓ:

$$\widehat{\mathfrak{B}}_\ell = \frac{1}{N-\ell} \sum_{i=1}^{N-\ell} \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i^\top.$$

The estimator for $\mathfrak{C}_1$ is

$$\hat{\mathfrak{C}}_1 = \widehat{\mathfrak{B}}_0 + \sum_{\ell=1}^{N-1} J\left(\frac{\ell}{h}\right) (\widehat{\mathfrak{B}}_\ell + \widehat{\mathfrak{B}}_\ell^\top), \tag{3.4}$$

where J is a kernel function. In a similar manner, we can define an estimator for $C(t,s)$ which satisfies (2.18). The estimator is based on the sample covariances defined for $\ell \geq 0$ by

$$\hat{\gamma}_{\ell,N}(t,s) = \frac{1}{N-h} \sum_{i=1}^{N-h} q_i(t) q_{i+\ell}(s).$$

We note that $Eq_i(t)q_{i+h}(s)$ depends on i because of the heteroscedasticity of the data. However, we show that the classical kernel estimator for the long-run variance of sums of dependent variance can still be used. The proof of (2.18) requires new arguments under our more general assumptions. The estimator is defined by

$$\widehat{C}_N(t,s) = \hat{\gamma}_{0,N}(t,s) + \sum_{\ell=1}^{N-1} J\left(\frac{\ell}{h}\right) (\hat{\gamma}_{\ell,N}(t,s) + \hat{\gamma}_{\ell,N}(s,t)). \tag{3.5}$$

The following conditions on the kernel J and the window (smoothing parameter) h are standard.

Assumption 3.1. (*i*) $J(0) = 1$ (*ii*) $J(u) = 0$, if $|u| > c_*$ with some $c_* > 0$ (*iii*) J is Lipschitz continuous on the real line.

Assumption 3.2. $h = h(N) \to \infty$ and $h/N \to 0$, as $N \to \infty$.

Theorem 3.1. If (1.1), (1.2), and Assumptions 2.2–3.2 are satisfied, then

$$\iint \left(\widehat{C}_N(t,s) - C(t,s)\right)^2 dn(t) dn(s) = o_P(1).$$

Since

$$\widehat{\mathfrak{C}}_1 = \iint \mathfrak{f}(t; \hat{\boldsymbol{\lambda}}_N) \widehat{C}_N(t,s) \mathfrak{f}^\top(t; \hat{\boldsymbol{\lambda}}_N) dm(t) dm(s),$$

we have the following immediate consequence of Theorem 3.1.

Corollary 3.1. If (1.1), (1.2), and Assumptions 2.2–3.2 are satisfied, then

$$\widehat{\mathfrak{C}}_1 \xrightarrow{P} \mathfrak{C}_1.$$

Now it is easy to find an estimator for $\mathcal{D}(t,s)$

$$\begin{aligned}\widehat{\mathcal{D}}_N(t,s) = \widehat{\mathcal{C}}_N(t,s) &- \left(\int \hat{\mathcal{C}}_N(s,u)\mathfrak{f}^\top(u;\hat{\boldsymbol{\lambda}}_N)dm(u)\right)\widehat{\mathfrak{B}}_N^{-1}\mathfrak{f}(t;\hat{\boldsymbol{\lambda}}_N)\\ &- \left(\int \widehat{\mathcal{C}}_N(t,v)\mathfrak{f}^\top(v;\hat{\boldsymbol{\lambda}}_N)dm(v)\right)\widehat{\mathfrak{B}}_N^{-1}\mathfrak{f}(s;\hat{\boldsymbol{\lambda}}_N)\\ &+ \mathfrak{f}^\top(t;\hat{\boldsymbol{\lambda}}_N)\widehat{\mathfrak{B}}_N^{-1}\left\{\iint \mathfrak{f}(u;\hat{\boldsymbol{\lambda}}_N)\widehat{\mathcal{C}}_N(u,v)\mathfrak{f}^\top(v;\hat{\boldsymbol{\lambda}}_N)dm(u)dm(v)\right\}\widehat{\mathfrak{B}}_N^{-1}\mathfrak{f}(s;\hat{\boldsymbol{\lambda}}_N),\end{aligned}$$

where $\widehat{\mathfrak{B}}_N$ is defined in (3.1). Observing that $\mathfrak{f}(s;\boldsymbol{\lambda})$ is continuous in $\boldsymbol{\lambda}$, Theorem 2.1, (3.2), and Theorem 3.1 imply that (2.18) holds.

4. APPLICATION TO US AND UK YIELD CURVES

This section demonstrates our estimation method and the goodness-of-fit test by applying them to the yield curves in the United States and the United Kingdom. Our framework is general and can accommodate both sparsely and densely observed data. Researchers have the freedom to choose the measures m and n. They can be the counting measure, potentially with some weights, the Lebesgue measure, or any absolutely continuous measure with a specific density of $\mathcal{T}$. For the sake of conserving space, the method of parameter estimation is demonstrated by the traditional Nelson–Siegel model, and other models can be estimated in the same manner.[1] The goodness-of-fit test is illustrated by a range of models.[2]

4.1. US Yield Curves

The website of the Board of Governors of the Federal Reserve System provides daily nominal interest rates for selected non-inflation-indexed US Treasury securities.[3] The market yields are calculated from composites of quotations on actively traded treasury securities in the over-the-counter market. Such a calculation method provides yields at fixed maturities even if there is no outstanding security with one of those exact maturities. The maturities are at fixed at 1, 3, 6, 12, 24, 36, 60, 84, 120, 240, and 360 months. Since the data are sparsely observed at these maturities, we set m and n as the counting measure for the US yield curves. This corresponds to the established way of fitting yield curves to such data. Additionally, we produce yield curves in the form of functional data from the discrete observations using the linear interpolation.[4] The data we used in this study are daily yield curves from 1 January 1980 to 31 August 2020. For the purpose of demonstration, we investigate the Nelson–Siegel model over

[1] For the empirical application, we impose a common sense restriction that the hump in the Nelson–Siegel model cannot occur before the first available maturity or after the last available maturity. The numerical method to minimize the least-squares loss function in (2.1) is the 'interior-point' algorithm provided by the 'fmincon' function in MATLAB. This algorithm is a local minimizer, rather than the global minimizer. We repeat the numerical method 100 times with different random initial values and choose the final solution as the one giving minimum value in the objective function among 100 trials. In this way, the global minimum can generally be reached. However, there is no guarantee and, in principle, the global minimum need not be unique.

[2] In the models of Svensson (1994) and Christensen *et al.* (2009), we have the assumption of $\lambda_{0,1} < \lambda_{0,2}$. Combined with Assumption 2.2 (compactness), this means that we actually assume $\lambda_{0,2} - \lambda_{0,1} > c$, for some positive c. Practically, we restrict $\lambda_{0,2} - \lambda_{0,1} > 0.05$ in our empirical applications.

[3] https://www.federalreserve.gov/releases/h15/default.htm. Accessed on 12 September 2020.

[4] If the data are available at densely observed maturities and may also suffer measurement errors, it is more appropriate to use a smoothing procedure with a basis expansion as a means to construct functional data (see, Ramsay and Silverman, 2005; Kokoszka and Reimherr, 2017).

© 2022 The Authors.
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

DOI: 10.1111/jtsa.12642

Table I. Parameter estimates of the Nelson–Siegel model on US yield curves.

Start	Jan 1980	Jul 1981	Jul 1990	Mar 2001	Dec 2007	Feb 2020
End	Jul 1980	Nov 1982	Mar 1991	Nov 2001	Jun 2009	Aug 2020*
No. of observations	147	352	185	188	395	147
$\hat{c}_1$	10.999	13.446	8.637	4.837	4.735	1.908
	(0.181)	(0.167)	(0.068)	(0.198)	(0.070)	(0.039)
$\hat{c}_2$	1.352	−3.566	−1.566	−1.792	−3.679	−1.563
	(0.670)	(0.699)	(0.099)	(0.308)	(0.123)	(0.093)
$\hat{c}_3$	−1.373	5.908	−1.645	4.803	−2.099	−1.729
	(0.267)	(0.432)	(0.112)	(0.346)	(0.339)	(0.118)
$\hat{\lambda}$	0.070	0.124	0.096	0.010	0.036	0.020
	(0.010)	(0.010)	(0.005)	(0.001)	(0.002)	(0.001)

Note: * indicates that the last contraction period ends at the most recent available date. The numbers in parentheses are the standard deviation of the estimated parameters.

Table II. p-values of the goodness-of-fit test for different models on US yield curves.

Start	Jan 1980	Jul 1981	Jul 1990	Mar 2001	Dec 2007	Feb 2020
End	Jul 1980	Nov 1982	Mar 1991	Nov 2001	Jun 2009	Aug 2020*
Nelson and Siegel (1987)	12.1%	0.0%	0.0%	0.0%	0.6%	44.8%
Svensson (1994)	0.2%	0.0%	0.0%	0.0%	0.0%	0.3%
Christensen *et al.* (2009)	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%
Chambers *et al.* (1984)	18.3%	0.0%	0.1%	0.0%	23.3%	40.4%
Yallup (2012), $K = 7$	100.0%	100.0%	100.0%	0.0%	7.9%	7.5%
Yallup (2012), $K = 8$	100.0%	100.0%	100.0%	0.0%	100.0%	79.0%
Yallup (2012), $K = 9$	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Note: * indicates that the last contraction period ends at the most recent available date.

six most recent contraction periods. The dates of US contraction periods are obtained from the National Bureau of Economic Research.[5]

We apply our estimation method to the yield curves in each contraction period. Table I reports estimated parameters of the Nelson–Siegel Model. As can be observed, the level parameter $\hat{c}_1$ has a substantial reduction from 11.00% in 1980 to 1.91% in 2020. The slope parameter $\hat{c}_2$ is positive in the first contraction period (Jan 1980–Jul 1980), indicating that the yield curve is inverted. In the other five periods, the yield curves are generally upward-sloping with negative $\hat{c}_2$. The curvature parameter $\hat{c}_3$ is positive in the second (Jul 1981–Nov 1982) and fourth (Mar 2001–Nov 2001) contraction periods, implying the upside hump in yield curves. The negative $\hat{c}_3$ in the other four periods indicates the downside hump in the shape of yield curves. $\hat{\lambda}$ controls the exponential decay rate and the location of hump. We find that $\hat{\lambda}$ changes in different periods, unlike Diebold and Li (2006), who assume that λ is constant.

Next, we apply our goodness-of-fit test to a range of yield curve models, including Nelson and Siegel (1987), Svensson (1994), Christensen *et al.* (2009), Chambers *et al.* (1984), and Yallup (2012). We set the polynomial basis up to the fourth degree in the model of Chambers *et al.* (1984), as suggested in their original paper. We also consider different numbers of exponential basis in the model of Yallup (2012) in order to explore the effects of more basis functions. Table II shows the p-values of the goodness-of-fit test for those models on US yield curves. The general pattern is that more factor functions enable larger flexibility in the yield curve modelling, which results in fewer rejections by the goodness-of-fit test. The Nelson–Siegel model is the most parsimonious model among

[5] https://www.nber.org/cycles.html. Note that the most recent contraction period starts from February 2020 and does not terminate yet. We use the data up to August 2020, which is the most available period at the time of this research.

Table III. Brexit timeline.

Period label	Key events in the period	Period
P0	Before Brexit initiative	01/01/2014 – 16/12/2015
P1	Leading up to the EU referendum	17/12/2015 – 22/06/2016
P2	Referendum	23/06/2016 – 07/06/2017
P3	General election	08/06/2017 – 10/12/2017
P4	Close of phase 1	11/12/2017 – 25/06/2018
P5	EU (withdrawal) Act becomes law	26/06/2018 – 15/01/2019
P6	The meaningful vote	16/01/2019 – 23/07/2019
P7	Boris Johnson becomes PM	24/07/2019 – 30/01/2020
P8	UK left EU & COVID-19	31/01/2020 – 31/08/2020

the considered candidates, and it cannot be rejected in the first and last contraction periods, but can be rejected at 5% significance level in other periods. The model of Yallup (2012) with $K = 9$, used by the Bank of Canada, is the most flexible one among all models, and it cannot be rejected in all contraction periods. It is also interesting to see that the Nelson–Siegel model outperforms the models of Svensson (1994) and Christensen *et al.* (2009), based on the number of rejections. This can be possibly explained by the additional uncertainty of parameter estimates in the extra factor functions, in particular when the second hump is not present. We note that while the Nelson–Siegel model may not offer a good fit, it may be useful for other purposes.

4.2. UK Yield Curves

The Bank of England provides daily estimated nominal government yield curves for the United Kingdom on their website.[6] We choose to use the short-end curves with monthly resolution.[7] Unlike the sparsely observed US yield data, the short-end of the UK yield data is densely observed. It should be emphasized that the estimated yield data in the United Kingdom is derived from spline-based smoothing methods. Specifically, Bank of England used the 'variable roughness penalty' (VRP) method of Anderson and Sleath (1999) to fit a curve to the data, with constraints imposed to ensure that the overall curve is continuous and smooth. One can say that the Bank of England has already converted the discrete data to functional data by penalized smoothing. Since the UK yields are available at densely observed time points, m and n are set to be the Lebesgue measure because it can take into account more information contained in the curves.

For the UK yield curves, we focus on the recent years because of the frequent change in the uncertainty related to the Brexit, which has a substantial impact on the shape of yield curves that can reflect expectations on the future of the UK economy. According to Walker (2020), the timeline of Brexit can be summarized in Table III. We are interested in the evolution of the parameters of the Nelson–Siegel Model in different periods related to the Brexit. Our sample period is from 1 January 2014 to 31 August 2020, containing 1685 trading days in total. Figure 1 shows the UK yield curves in the whole sample period, and Figure S1 in the online Supporting Information presents the mean yield curves in different periods related to the Brexit.

Table IV presents the estimated parameters of the Nelson–Siegel Model. The first two parameters, $\hat{c}_1$ and $\hat{c}_2$, are associated with level and slope of the yield curves. There is a general pattern that $\hat{c}_1$ decreases and $\hat{c}_2$ increases as time moving on to the later periods related to the Brexit. This indicates that the yield curves gradually move to a lower level and become flatter, which is consistent with the observation in Figure 1. The curvature parameter $\hat{c}_3$

[6] https://www.bankofengland.co.uk/statistics/yield-curves. Accessed on 12 September 2020.
[7] The Bank of England provides two types of estimated yield curves with different maturity range. The short-end curves have the maturity at monthly resolution ranging from 1 month to 5 years. The long-end curves have half-year resolution ranging from 0.5 year to a very large maturity (25 or 40 years). The short-end curves in the United Kingdom is available at more maturities because they are based on the yields of generalized collateral (GC) repo agreements, in additional to gilts (UK government bond). The Bank of England use GC repo to conduct its Open Market Operations, and the GC repo contracts are actively traded for maturities up to 1 year.

© 2022 The Authors.
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

DOI: 10.1111/jtsa.12642

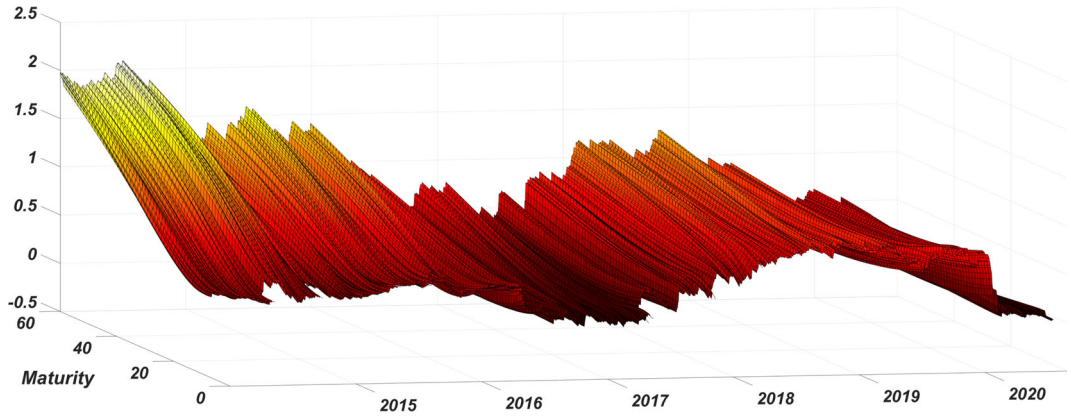

Figure 1. UK yield curves.

Table IV. Parameter estimates of the Nelson–Siegel model on UK yield curves.

Period label	P0	P1	P2	P3	P4	P5	P6	P7	P8
No. of observations	497	128	242	131	134	142	131	133	146
$\hat{c}_1$	2.765	1.928	2.479	2.508	2.082	1.952	1.574	0.512	0.415
	(0.076)	(0.044)	(0.150)	(0.078)	(0.153)	(0.071)	(0.109)	(0.033)	(0.151)
$\hat{c}_2$	−2.272	−1.352	−2.338	−2.126	−1.605	−1.198	−0.778	0.288	−0.222
	(0.087)	(0.063)	(0.160)	(0.088)	(0.166)	(0.093)	(0.109)	(0.037)	(0.195)
$\hat{c}_3$	−3.305	−2.843	−3.099	−3.023	−1.129	−1.636	−1.560	−0.676	-0.757
	(0.109)	(0.066)	(0.142)	(0.104)	(0.093)	(0.028)	(0.067)	(0.046)	(0.250)
$\hat{\lambda}$	0.070	0.062	0.030	0.030	0.030	0.038	0.032	0.059	0.031
	(0.001)	(0.002)	(0.002)	(0.001)	(0.005)	(0.003)	(0.002)	(0.003)	(0.008)

Note: The numbers in parentheses are the standard deviation of the estimated parameters.

is negative in all periods, indicating a downside hump in the shape of the yield curves. In terms of $\hat{\lambda}$, it starts from 0.070 in P0 and decreases to 0.030 in P2 and generally stays at that level afterwards, except for a bounce in P7.

We further perform the goodness-of-fit test for the same range of yield curve models. Table V shows the p-values of the goodness-of-fit test for different models on UK yield curves. A remarkably similar pattern as in Section 4.1 is observed. At the 5% significance level, the Nelson–Siegel model cannot be rejected in P5 and P6, while the model of Yallup (2012) with $K = 9$ cannot be rejected in all periods. Additionally, the Nelson–Siegel model has less number of rejections than the models of Svensson (1994) and Christensen *et al.* (2009). Again, this can be explained by the estimation uncertainty in the parameters related to the second hump, which may not exist.

When assessing goodness-of-fit, we assume that one model should be suitable for a given stretch of data. We addressed this issue by considering relatively short time periods. However, the partitions into these shorter periods may not be optimal. It might be useful to develop a change point detection and estimation procedure for these models, and use the resulting data-driven partitions. We hope that our work will motivate research in this direction.

5. FINITE-SAMPLE PERFORMANCE

We now examine the finite-sample performance of our estimation method and the goodness-of-fit test. We first describe the data-generating process (DGP) designed to be in line with the characteristics of the empirical yield curves in the United States and the United Kingdom analysed in Section 4. We simulate realizations of the

Table V. p-values of the goodness-of-fit test for different models on UK yield curves.

Period label	P0	P1	P2	P3	P4	P5	P6	P7	P8
Nelson and Siegel (1987)	1.90%	1.23%	3.64%	1.91%	0.00%	25.76%	16.98%	0.43%	4.51%
Svensson (1994)	0.00%	0.00%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Christensen *et al.* (2009)	90.54%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Chambers *et al.* (1984)	0.00%	4.85%	16.67%	12.79%	0.39%	10.44%	3.49%	0.47%	0.33%
Yallup (2012), $K=7$	47.41%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Yallup (2012), $K=8$	99.98%	10.17%	25.76%	98.25%	86.18%	0.00%	33.60%	0.11%	4.72%
Yallup (2012), $K=9$	100.00%	99.60%	100.00%	100.00%	99.85%	95.23%	99.91%	12.26%	58.49%

Nelson–Siegel model

$$X_i(t) = b_{i,1}f_1(t;\lambda) + b_{i,2}f_2(t;\lambda) + b_{i,3}f_3(t;\lambda) + \varepsilon_i(t), \quad t \in [0,360],\ 1 \le i \le N, \tag{5.1}$$

where the random coefficients follow

$$b_{i,\ell} = c_\ell + e_{i,\ell},\ \ 1 \le \ell \le 3,\ \ 1 \le i \le N, \tag{5.2}$$

and the factor functions are

$$f_1(t;\lambda) = 1,\ \ f_2(t;\lambda) = \frac{1-e^{-\lambda t}}{\lambda t} \ \text{ and } \ f_3(t;\lambda) = \frac{1-e^{-\lambda t}}{\lambda t} - e^{-\lambda t}, \quad t \in [0,360].$$

The true values of the model parameters are set to be $c_1 = 7.579$, $c_2 = -2.098$, $c_3 = -0.162$, and $\lambda = 0.0609$. The chosen values are taken from Diebold and Li (2006).

Recall that Assumptions 2.5 and 2.6 allow for heteroscedasticity in the errors. This enables us to consider both homoscedastic and heteroscedastic errors to check the robustness of the developed methods. Under homoscedasticity, the errors of the random coefficients in (5.2) are generated by

$$e_{i,\ell} \sim i.i.d.\ \ \mathcal{N}(0,\sigma_b^2),\ \ 1 \le \ell \le 3,\ \ 1 \le i \le N.$$

The error term $\varepsilon_i(t)$ in (5.1) is generated by

$$\varepsilon_i(t) = \sum_{r=1}^{R} \xi_{i,r}\psi_r(t), \quad 1 \le i \le N,$$

where $\psi_r(t)$ are Fourier basis functions, and $\xi_{i,r}$ follows an autoregressive process

$$\xi_{i,r} = \rho\xi_{i-1,r} + Z_{i,r}, \ \text{ and } \ Z_{i,r} \sim i.i.d.\ \mathcal{N}(0,\sigma_{Z_r}^2),\ \ 1 \le r \le R. \tag{5.3}$$

Under heteroscedasticity, the errors are generated by

$$\begin{cases} e_{i,\ell} \sim i.i.d.\ \mathcal{N}(0,\sigma_b^2), & 1 \le i < k^*, \\ e_{i,\ell} \sim i.i.d.\ \mathcal{N}(0,2\sigma_b^2), & k^* \le i \le N. \end{cases}$$

© 2022 The Authors.
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

DOI: 10.1111/jtsa.12642

The error term $\varepsilon_i(t)$ is generated by

$$\varepsilon_i(t) = \begin{cases} \sum_{r=1}^{R} \xi_{i,r}\psi_r(t), & 1 \le i < k^*, \\ \sum_{r=1}^{R} \sqrt{2}\xi_{i,r}\psi_r(t), & k^* \le i \le N, \end{cases}$$

and $\xi_{i,r}$ follows the same autoregressive process in (5.3). For the parameters to generate errors under both homoscedasticity and heteroscedasticity, we set $R = 5$, $\sigma_b = 0.2$, $\sigma_{Z_1} = 0.1$, $\sigma_{Z_2} = 0.05$, $\sigma_{Z_3} = 0.05$, $\sigma_{Z_4} = 0.025$, $\sigma_{Z_5} = 0.025$, and $\rho = 0.1$ or 0.5. For the heteroscedastic errors, we set $k^* = N/2$.

As shown in Section 4, the US yields data are sparsely observed, while the UK data come in the form of densely observed yield curves. In order to closely mimic the characteristics of the empirical data, we devise two specific versions of DGPs: *functional DGP* and *discrete DGP*. For the functional DGP, the data are simulated in the domain of $t \in [0,360]$, which is already in the form of functional data. For the discrete DGP, the data is first generated only at 11 fixed maturities, $t \in \{1, 3, 6, 12, 24, 36, 60, 84, 120, 240, 360\}$, and then linear interpolation is employed to convert discrete data into continuous curves as functional data. To compare the two DGPs, Figure S2 in the online Supporting Information shows one realization of simulated yield curves based on the functional DGP (upper panel) and the discrete DGP (lower panel) of sample size $N = 100$ with heteroscedastic errors.

Before presenting the simulation results, it is worthwhile to discuss two practical issues: the measures and the kernel estimators. Since our framework allows us to choose various measures m and n, we have explored many options and found that the Lebesgue measure is suitable for the functional DGP and the counting measure is appropriate for the discrete DGP. In the following, the reported simulation results for the functional DGP are based on the Lebesgue measure and those for the discrete DGP on the counting measure. The kernel estimator of $\hat{\mathfrak{C}}_1$ in (3.4) and $\widehat{C}_N(t,s)$ in (3.5) needs the specification of the kernel and the smoothing bandwidth, and we use the flat-top kernel

$$J(u) = \begin{cases} 1, & 0 \le |u| < 0.5, \\ 2 - 2|u|, & 0.5 \le |u| < 1, \\ 0, & |u| \ge 1, \end{cases}$$

with the bandwidth $h = \log(N)$. Our simulations show that this bandwidth produces stable and satisfactory results. In Section D of the online Supporting Information, we show that the test is not sensitive to the bandwidth choice. Data-driven methods of bandwidth selection might improve the results further. Politis (2003) studies adaptive bandwidth for scalar data and Rice and Shang (2017) propose a plug-in bandwidth selection for functional data.

5.1. Parameter Estimation

Based on the two DGPs, we investigate the consistency and the normality in finite samples of the estimator developed in Section 2. It is also of interest to compare the asymptotic standard deviation of the estimator in Theorem 2.2 with the sample standard deviation. In Section B of the online Supporting Information, we provide a practical step-by-step algorithm to implement our estimation approach and the details of using a quasi-Newton method to numerically minimize the least-squares loss function in (2.1), together with an improved algorithm with fast computation if the resolution (sampling frequency) of yield curves is monthly.

We generate 1000 independent realizations of the two DGPs, and record the estimated values of $\mathbf{a} = (c_1, c_2, c_3, \lambda)$. The summary statistics of estimated parameters are reported in Tables VI and VII for the functional DGP and the discrete DGP respectively. Generally, the consistency of the four parameters is well supported by the simulation results for the two DGPs because the bias and standard deviation become smaller with the increase in

Table VI. Summary statistics of the estimated parameters for simulated data based on the functional DGP.

		Homoskedastic errors ($\rho = 0.1$)				Heteroscedastic errors ($\rho = 0.1$)			
		$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\hat{\lambda}$	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\hat{\lambda}$
$N = 100$	Mean	7.579	−2.099	−0.127	0.060	7.579	−2.098	−0.130	0.060
	SD	0.020	0.020	0.101	0.003	0.026	0.025	0.111	0.003
	Skewness	0.136	−0.003	1.818	−1.759	−0.112	0.035	1.647	−1.531
	Kurtosis	2.639	3.139	6.901	6.699	3.380	3.143	6.503	6.130
	Median of asy. SD	0.019	0.019	0.061	0.002	0.023	0.023	0.073	0.002
$N = 200$	Mean	7.579	−2.098	−0.145	0.060	7.579	−2.098	−0.140	0.060
	SD	0.014	0.014	0.070	0.002	0.017	0.017	0.089	0.002
	Skewness	0.033	−0.045	2.536	−2.457	0.001	−0.011	2.181	−2.094
	Kurtosis	2.821	2.967	12.583	12.341	3.289	2.889	9.079	8.727
	Median of asy. SD	0.014	0.014	0.044	0.001	0.017	0.017	0.052	0.001
$N = 500$	Mean	7.579	−2.098	−0.156	0.061	7.579	−2.098	−0.148	0.061
	SD	0.009	0.009	0.043	0.001	0.011	0.011	0.063	0.002
	Skewness	−0.119	−0.072	4.027	−3.944	0.113	−0.176	3.475	−3.401
	Kurtosis	2.934	2.771	31.858	31.070	3.174	2.980	19.601	19.166
	Median of asy. SD	0.009	0.009	0.028	0.001	0.011	0.011	0.035	0.001
		Homoskedastic errors ($\rho = 0.5$)				Heteroscedastic errors ($\rho = 0.5$)			
		$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\hat{\lambda}$	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\hat{\lambda}$
$N = 100$	Mean	7.578	−2.097	−0.122	0.060	7.578	−2.097	−0.112	0.060
	SD	0.020	0.020	0.128	0.004	0.025	0.025	0.149	0.004
	Skewness	−0.105	−0.004	1.092	−0.933	0.001	−0.049	0.955	−0.789
	Kurtosis	2.985	3.227	4.378	3.963	2.898	3.144	3.797	3.448
	Median of asy. SD	0.019	0.019	0.100	0.003	0.023	0.023	0.115	0.003
$N = 200$	Mean	7.579	−2.098	−0.120	0.060	7.579	−2.098	−0.122	0.060
	SD	0.014	0.014	0.113	0.003	0.017	0.017	0.122	0.003
	Skewness	−0.040	0.028	1.477	−1.300	0.026	−0.230	1.216	−1.045
	Kurtosis	2.863	3.060	5.666	5.038	2.910	3.263	4.754	4.221
	Median of asy. SD	0.014	0.014	0.076	0.002	0.017	0.017	0.091	0.003
$N = 500$	Mean	7.579	−2.098	−0.141	0.060	7.579	−2.098	−0.135	0.060
	SD	0.009	0.009	0.077	0.002	0.011	0.011	0.093	0.003
	Skewness	0.145	0.063	2.156	−1.966	−0.004	0.000	1.941	−1.783
	Kurtosis	3.211	2.679	9.883	8.784	2.768	2.703	8.068	7.381
	Median of asy. SD	0.009	0.009	0.048	0.001	0.011	0.011	0.059	0.002

Note: The measure m is the Lebesgue measure. The true values for the parameters are $c_{1,0} = 7.579$, $c_{2,0} = -2.098$, $c_{3,0} = -0.162$, and $\lambda_0 = 0.0609$.

sample size. Additionally, the median of the asymptotic standard deviation obtained by Theorem 2.2 is close to the standard deviation in the simulations.[8] Importantly, our estimation procedure is robust to the heteroscedastic errors in the data. Compared to those based on the functional DGP, the estimated parameters based on the discrete DGP follow a normal distribution more closely. This is also reflected in the QQ-plots. For example, Figure S3 in the online Supporting Information shows QQ-plots under the discrete DGP with heteroskedastic errors for sample size $N = 100$. For homoskedastic errors and larger sample sizes, the QQ-plots look even better.

[8] We report the median of the asymptotic standard deviation, rather than the mean, because of the outliers.

© 2022 The Authors.
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

DOI: 10.1111/jtsa.12642

Table VII. Summary statistics of the estimated parameters for simulated data based on the discrete DGP.

		Homoskedastic errors ($\rho = 0.1$)				Heteroscedastic errors ($\rho = 0.1$)			
		$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\hat{\lambda}$	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\hat{\lambda}$
$N = 100$	Mean	7.579	−2.098	−0.157	0.061	7.579	−2.098	−0.157	0.061
	SD	0.020	0.020	0.044	0.001	0.024	0.024	0.054	0.001
	Skewness	0.081	0.040	0.838	−0.826	0.015	−0.079	0.884	−0.817
	Kurtosis	2.718	2.906	4.132	4.343	2.798	2.773	3.813	3.863
	Median of asy. SD	0.019	0.019	0.037	0.001	0.023	0.023	0.045	0.001
$N = 200$	Mean	7.579	−2.098	−0.160	0.061	7.579	−2.098	−0.159	0.061
	SD	0.014	0.015	0.029	0.001	0.017	0.018	0.039	0.001
	Skewness	-0.036	−0.051	0.814	−0.684	−0.077	0.101	0.937	−0.776
	Kurtosis	2.944	3.081	5.270	5.189	2.787	2.822	4.783	4.623
	Median of asy. SD	0.014	0.014	0.027	0.001	0.017	0.017	0.033	0.001
$N = 500$	Mean	7.579	−2.098	−0.162	0.061	7.579	−2.098	−0.161	0.061
	SD	0.009	0.009	0.018	0.000	0.011	0.011	0.022	0.001
	Skewness	−0.054	−0.007	0.373	−0.260	0.097	0.052	0.492	−0.349
	Kurtosis	3.227	3.405	3.424	3.212	3.122	2.943	3.526	3.185
	Median of asy. SD	0.009	0.009	0.017	0.000	0.011	0.011	0.021	0.001
		Homoskedastic errors ($\rho = 0.5$)				Heteroscedastic errors ($\rho = 0.5$)			
		$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\hat{\lambda}$	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\hat{\lambda}$
$N = 100$	Mean	7.579	−2.098	−0.151	0.061	7.581	−2.097	−0.151	0.061
	SD	0.020	0.019	0.071	0.002	0.025	0.025	0.077	0.002
	Skewness	−0.050	−0.022	0.824	−0.756	0.126	0.036	0.537	−0.451
	Kurtosis	2.861	2.987	3.551	3.463	2.971	2.951	2.464	2.558
	Median of asy. SD	0.019	0.019	0.056	0.002	0.023	0.023	0.068	0.002
$N = 200$	Mean	7.579	−2.098	−0.158	0.061	7.579	−2.098	−0.154	0.061
	SD	0.014	0.014	0.051	0.001	0.017	0.017	0.062	0.002
	Skewness	0.118	0.117	0.922	−0.900	0.121	0.037	0.917	−0.781
	Kurtosis	2.867	3.445	4.070	3.999	3.084	2.903	3.425	3.351
	Median of asy. SD	0.014	0.014	0.042	0.001	0.017	0.017	0.050	0.001
$N = 500$	Mean	7.579	−2.098	−0.159	0.061	7.579	−2.097	−0.157	0.061
	SD	0.009	0.009	0.029	0.001	0.011	0.011	0.040	0.001
	Skewness	0.039	0.045	0.688	−0.618	0.022	0.035	0.992	−0.942
	Kurtosis	3.153	3.191	4.753	4.597	3.029	3.113	5.005	4.814
	Median of asy. SD	0.009	0.009	0.027	0.001	0.011	0.011	0.033	0.001

Note: The measure function m is the counting measure. The true values for the parameters are $c_{1,0} = 7.579$, $c_{2,0} = -2.098$, $c_{3,0} = -0.162$, and $\lambda_0 = 0.0609$.

5.2. The goodness-of-fit test

We study the empirical size and power of the goodness-of-fit test developed in Section 2. In Section C of the online Supporting Information, we provide the detailed steps for implementing our test. Under the null hypothesis, we simulate the data based on the functional DGP and the discrete DGP using 1000 independent replications, and calculate the test statistics $\widehat{T}_N$ in (2.15) and approximate its limit distribution by (2.20). For each replication, we compare the test statistics $\widehat{T}_N$ with the critical values obtained by (2.21) to decide the rejection of the null hypothesis.

Table VIII shows the empirical size of the test for the functional DGP in Panel A and for the discrete DGP in Panel B. Generally, the empirical size is close to the theoretical levels for the two DGPs, with a tendency to

Table VIII. Empirical size of the goodness-of-fit test.

Significance level	10%	5%	1%	10%	5%	1%
Panel A: the functional DGP						
	Homoscedastic errors ($\rho = 0.1$)			Heteroscedastic errors ($\rho = 0.1$)		
$N = 100$	12.3%	6.5%	2.0%	12.7%	6.9%	1.6%
$N = 200$	10.6%	4.8%	1.0%	12.5%	5.6%	1.4%
$N = 500$	10.0%	5.1%	1.5%	11.2%	5.8%	1.3%
	Homoscedastic errors ($\rho = 0.5$)			Heteroscedastic errors ($\rho = 0.5$)		
$N = 100$	13.7%	7.9%	1.8%	14.2%	8.8%	2.8%
$N = 200$	14.1%	7.4%	1.3%	11.5%	6.3%	1.5%
$N = 500$	10.3%	6.1%	1.2%	12.6%	7.5%	1.9%
Panel B: the discrete DGP						
	Homoscedastic errors ($\rho = 0.1$)			Heteroscedastic errors ($\rho = 0.1$)		
$N = 100$	13.0%	7.1%	2.1%	12.7%	7.5%	2.0%
$N = 200$	10.6%	5.1%	1.7%	11.6%	7.2%	2.1%
$N = 500$	11.3%	6.1%	0.8%	10.5%	5.0%	0.8%
	Homoscedastic errors ($\rho = 0.5$)			Heteroscedastic errors ($\rho = 0.5$)		
$N = 100$	16.2%	8.6%	3.5%	16.3%	8.7%	2.6%
$N = 200$	15.7%	7.8%	2.8%	13.5%	7.0%	2.0%
$N = 500$	12.7%	6.5%	1.6%	12.0%	6.9%	1.9%

over-reject, which becomes smaller as the sample size increases. The size under $\rho = 0.5$ is marginally higher than the size under $\rho = 0.1$. This is expected because the convergence rate is slower under stronger dependence. Our goodness-of-fit test is robust to heteroscedasticity in the data. We also employed the technique of circular block bootstrap (Politis and Romano, 1992) to improve the size of the test. Details are provided in Section E of the online Supporting Information.

Under the alternative hypothesis, we generate the data from the Svensson (1994) model (two humps) and test the goodness-of-fit of the Nelson–Siegel model (one hump). Explicitly, we generate realizations of the form

$$X_i(t) = \sum_{\ell=1}^{4} b_{i,\ell} f_\ell(t;\lambda) + \varepsilon_i(t), \quad t \in [0,360],\ 1 \le i \le N, \tag{5.4}$$

where the random coefficients follow (1.2) and the functions $f_\ell(t;\lambda)$ are as specified in Example 1.1.

Depending on the combination of parameter values in (5.4), we have the following four alternatives:

$$\begin{aligned}
&H_{A,1}\colon c_1 = 7.579,\ c_2 = -2.098,\ c_3 = -0.162,\ c_4 = -0.040, \text{ and } \lambda_1 = 0.0609,\ \lambda_2 = 0.15,\\
&H_{A,2}\colon c_1 = 7.579,\ c_2 = -2.098,\ c_3 = -0.162,\ c_4 = -0.040, \text{ and } \lambda_1 = 0.0609,\ \lambda_2 = 0.30,\\
&H_{A,3}\colon c_1 = 7.579,\ c_2 = -2.098,\ c_3 = -0.162,\ c_4 = -0.081, \text{ and } \lambda_1 = 0.0609,\ \lambda_2 = 0.15,\\
&H_{A,4}\colon c_1 = 7.579,\ c_2 = -2.098,\ c_3 = -0.162,\ c_4 = -0.081, \text{ and } \lambda_1 = 0.0609,\ \lambda_2 = 0.30.
\end{aligned}$$

We choose those specific parameter combinations in order to explore the empirical power performance under different scenarios. Essentially, there is an additional hump at the maturity of 12 months under $H_{A,1}$ and $H_{A,3}$, and at the maturity of 6 months under $H_{A,2}$ and $H_{A,4}$. The magnitude of the second hump is around a quarter of the first hump under $H_{A,1}$ and $H_{A,2}$, and half under $H_{A,3}$ and $H_{A,4}$. The other settings of the DGPs under the alternative

© 2022 The Authors.
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

DOI: 10.1111/jtsa.12642

Table IX. Empirical power of the goodness-of-fit test at the 5% significance level.

Panel A: the functional DGP								
	Homoscedastic errors ($\rho = 0.1$)				Heteroscedastic errors ($\rho = 0.1$)			
	$H_{A,1}$	$H_{A,2}$	$H_{A,3}$	$H_{A,4}$	$H_{A,1}$	$H_{A,2}$	$H_{A,3}$	$H_{A,4}$
$N = 50$	18.5%	23.7%	50.9%	66.2%	12.6%	18.4%	38.4%	51.2%
$N = 100$	26.6%	40.8%	81.9%	91.7%	20.5%	30.5%	63.8%	79.5%
$N = 200$	50.3%	67.7%	98.8%	100.0%	34.7%	51.9%	91.0%	97.8%
	Homoscedastic errors ($\rho = 0.5$)				Heteroscedastic errors ($\rho = 0.5$)			
$N = 50$	15.3%	17.2%	31.5%	43.2%	13.6%	13.3%	23.5%	34.8%
$N = 100$	19.3%	20.2%	41.4%	72.5%	14.7%	12.7%	31.6%	47.6%
$N = 200$	24.0%	34.3%	73.6%	97.2%	16.0%	23.0%	54.1%	82.5%
Panel B: the discrete DGP								
	Homoscedastic errors ($\rho = 0.1$)				Heteroscedastic errors ($\rho = 0.1$)			
	$H_{A,1}$	$H_{A,2}$	$H_{A,3}$	$H_{A,4}$	$H_{A,1}$	$H_{A,2}$	$H_{A,3}$	$H_{A,4}$
$N = 50$	39.3%	36.8%	84.6%	83.3%	28.0%	29.7%	70.4%	70.1%
$N = 100$	56.9%	56.9%	98.1%	97.5%	44.5%	42.7%	91.2%	92.3%
$N = 200$	79.9%	81.2%	100.0%	100.0%	63.3%	64.9%	99.4%	99.5%
	Homoscedastic errors ($\rho = 0.5$)				Heteroscedastic errors ($\rho = 0.5$)			
$N = 50$	31.2%	38.0%	73.2%	84.4%	23.9%	28.6%	60.6%	70.6%
$N = 100$	47.0%	53.3%	95.4%	97.6%	34.7%	42.7%	84.9%	89.8%
$N = 200$	73.5%	81.3%	99.9%	100.0%	54.9%	66.1%	98.6%	99.6%

hypothesis are the same as those of the null hypothesis. Based on 1000 replications, Table IX shows the empirical power of the test at 5% significance level for the functional DGP in Panel A and for the discrete DGP in the Panel B. The test can achieve high power even under small sample sizes, such as $N = 50$. The power is higher under the discrete DGP than under the functional DGP. With a higher value of ρ, the power deteriorates slightly under the functional DGP but not under the discrete DGP. Additionally, the heteroscedasticity marginally reduces the power of the test. Comparing the four alternatives, the empirical power is higher if (1) the two humps are further apart (under $H_{A,2}$ and $H_{A,4}$), and (2) the second hump has a larger magnitude (under $H_{A,3}$ and $H_{A,4}$).

ACKNOWLEDGEMENTS

The authors are grateful to the Editor-in-Chief, one anonymous Co-Editor, and two Referees for their helpful insights and remarks, which have improved the manuscript significantly. The research of Piotr Kokoszka was partially supported by NSF grants DMS–1923142, DMS–1914882, and DMS–2123761. The authors also thank the Reading Academic Computing Cluster at the University of Reading for providing the high-performance computer facility to run the computational program.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at the website of the Board of Governors of the Federal Reserve System (https://www.federalreserve.gov/releases/h15/default.htm) and the website of the Bank of England (https://www.bankofengland.co.uk/statistics/yield-curves).

SUPPORTING INFORMATION

Additional Supporting Information can be found online in the supporting information tab for this article.

REFERENCES

Anderson N, Sleath J. 1999. New estimates of the UK real and nominal yield curves. *Bank of England Quarterly Bulletin* **39**:384–392.

Annaert J, Claes AG, De Ceuster MJ, Zhang H. 2013. Estimating the spot rate curve using the Nelson–Siegel model: A ridge regression approach. *International Review of Economics & Finance* **27**:482–496.

Aue A, Hörmann S, Horváth L, Reimherr M. 2009. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* **37**:4046–4087.

Bank for International Settlements. 2005. Zero-Coupon Yield Curves: Technical Documentation.

Bank of England. 2020. Daily Estimated Yield Curves for the UK. https://www.bankofengland.co.uk/statistics/yield-curves

Bardsley P, Horváth L, Kokoszka P, Young G. 2017. Change point tests in functional factor models with application to yield curves. *The Econometrics Journal* **20**(1):86–117.

Bliss R, Fama E. 1987. The information in long-maturity forward rates. *The American Economic Review* **77**:680–692.

Board of Governors of the Federal Reserve System. 2020. Selected Interest Rates (Daily). https://www.federalreserve.gov/releases/h15/default.htm

Cavaliere G, Harvey D, Leybourne S, Taylor R. 2011. Testing for unit roots in the presence of a possible break in in trend and nonstationary volatility. *Econometric Theory* **27**:957–991.

Chambers DR, Carleton WT, Waldman DW. 1984. A new approach to estimation of the term structure of interest rates. *Journal of Financial and Quantitative Analysis* **19**(3):233–252.

Christensen JH, Diebold FX, Rudebusch GD. 2009. An arbitrage-free generalized Nelson–Siegel term structure model. *The Econometrics Journal* **12**(3):C33–C64.

Diebold FX, Li C. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* **130**(2): 337–364.

Diebold FX, Rudebusch GD. 2013. *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach* Princeton University Press, Princeton, NJ.

Diebold FX, Rudebusch GD, Aruoba SB. 2006. The macroeconomy and the yield curve: A dynamic latent factor approach. *Journal of Econometrics* **131**(1-2):309–338.

Filipović D. 1999. A note on the Nelson–Siegel family. *Mathematical Finance* **9**(4):349–359.

Górecki T, Horváth L, Kokoszka P. 2018. Change point detection in heteroscedastic time series. *Econometrics and Statistics* **7**:63–88.

Hays S, Shen H, Huang JZ. 2012. Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics* **6**:870–894.

Hörmann S, Horváth L, Reeder R. 2013. A functional version of the ARCH model. *Econometric Theory* **29**:267–288.

Hörmann S, Kokoszka P. 2010. Weakly dependent functional data. *The Annals of Statistics* **38**:1845–1884.

Horváth L, Kokoszka P. 2012. *Inference for Functional Data with Applications* Springer, New York, NY.

Horváth L, Kokoszka P, Reeder R. 2013. Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society (B)* **75**:103–122.

Ibragimov IA. 1962. Some limit theorems for stationary processes. *Theory of Probability & Its Applications* **7**(4):349–382.

Kokoszka P, Miao H, Zhang X. 2015. Functional dynamic factor model for intraday price curves. *Journal of Financial Econometrics* **13**:456–477.

Kokoszka P, Reimherr M. 2017. *Introduction to Functional Data Analysis* CRC Press, Boca Raton, FL.

Nelson CR, Siegel AF. 1987. Parsimonious modeling of yield curves. *Journal of Business* **60**(4):473–489.

Politis DN. 2003. Adaptive bandwidth choice. *Journal of Nonparametric Statistics* **15**(4-5):517–533.

Politis DN, Romano JP. 1992. *A circular block-resampling procedure for stationary data*. In *Exploring the Limits of Bootstrap*, Lepage R, Billard L (eds.). Wiley, New York, NY.

Ramsay J, Silverman BW. 2005. *Functional Data Analysis* Springer, New York, NY.

Rice G, Shang HL. 2017. A plug-in bandwidth selection procedure for long-run covariance estimation with stationary functional time series. *Journal of Time Series Analysis* **38**(4):591–609.

Shao X, Wu WB. 2007. Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics* **35**:1773–1801.

Svensson LE. 1994. Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994. Technical Report, National Bureau of Economic Research.

Vermani V. 2012. On estimability of parsimonious term structure models: An experiment with the Nelson-Siegel specification. *Applied Economics Letters* **19**:1703–1706.

Walker N. 2020. Brexit Timeline: Events Leading to the UK's Exit from the European Union. Technical Report, House of Commons Library.

© 2022 The Authors.
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

DOI: 10.1111/jtsa.12642

Wu W. 2005. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences* **102**(40):14150–14154.
Xu K-L. 2015. Testing for structural change under non-stationary variances. *The Econometrics Journal* **18**:274–305.
Yallup PJ. 2012. Models of the yield curve and the curvature of the implied forward rate function. *Journal of Banking & Finance* **36**(1):121–135.
Zhang X. 2016. White noise testing and model diagnostic checking for functional time series. *Journal of Econometrics* **194**:76–95.

DOI: 10.1111/jtsa.12642 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.