The memory of beta

Article

Accepted Version

Becker, J., Hollstein, F., Prokopczuk, M. and Sibbertsen, P. (2021) The memory of beta. Journal of Banking & Finance, 124. 106026. ISSN 0378-4266 doi: 10.1016/j.jbankfin.2020.106026 Available at https://centaur.reading.ac.uk/95508/

To link to this article DOI: http://dx.doi.org/10.1016/j.jbankfin.2020.106026

Publisher: Elsevier

The Memory of Beta

October 29, 2020

Abstract

Researchers and practitioners employ a variety of time-series processes to forecast betas, either using short-memory models or implicitly imposing infinite memory. We find that both approaches are inadequate: beta factors show consistent long-memory properties. For the vast majority of stocks, we reject both the short-memory and difference-stationary (random walk) alternatives. A pure long-memory model reliably provides superior beta forecasts compared to all alternatives. Accounting for long memory in beta also pays off economically for portfolio formation. We widely document the robustness of these results.

JEL classification: G12, C58, G11

Keywords: Long memory, beta, persistence, forecasting, predictability

1 Introduction

In factor pricing models like the Capital Asset Pricing Model (CAPM) (Sharpe, 1964; Lintner, 1965; Mossin, 1966) or the arbitrage pricing theory (APT) (Ross, 1976) the drivers of expected returns are the stock's sensitivities to risk factors, i.e., beta factors. For many applications such as asset pricing, portfolio choice, capital budgeting, or risk management, the market beta is still the single most important factor. Indeed, Graham & Harvey (2001) document that chief financial officers of large U.S. companies primarily rely on one-factor market model cost-of-capital forecasts. In addition, Barber et al. (2016) and Berk & Van Binsbergen (2016) also show that investors mainly use the market model for capital allocation decisions. However, since beta factors are not directly observable, one needs to estimate them. For this purpose, researchers and practitioners alike typically use past information, i.e., employ time-series models.

The degree of memory is an important determinant of the characteristics of a time series. In an $I(0)$, or short-memory, process (e.g., AR(p) or ARMA(p, q)), the impact of shocks is short-lived and dies out quickly. On the other hand, for an $I(1)$, or difference-stationary, process like, for example, the random walk (RW), shocks persist infinitely. Thus, any change in a variable will have an impact on all future realizations. For an $I(d)$ process with $0 < d < 1$, shocks neither die out quickly nor persist infinitely, but have a hyperbolically decaying impact. In this case, the current value of a variable depends on past shocks, but the less so the further these shocks are past.

Researchers and practitioners estimate betas in several different ways. One approach is to use constant beta coefficients for the full sample (e.g., Fama & French, 1992). This relates to the most extreme $I(0)$ case possible. However, there is a strong consensus in the literature that betas vary over time. The usual approach to account for such time-variation is the use of rolling windows, where the most current estimate is taken as a forecast for the next month (e.g., Fama & MacBeth, 1973; Frazzini & Pedersen, 2014). This approach inherently imposes infinite memory and resembles a random walk, i.e., presuming that the best forecast for the future beta is today's estimate.[1]

Numerous other studies employ explicit or implicit short-memory processes for modeling beta dynamics. These include, among others, AR(1) processes in Ang & Chen (2007) and Levi & Welch (2017), an AR(1) process with further latent and exogenous variables in Adrian & Franzoni (2009), and an ARMA(1,1) process in Pagan (1980). Blume (1971) imposes a joint AR(1) process for the entire beta cross-section. The implications of these differing approaches for the modeling of betas, though, vary substantially.

Andersen et al. (2006) first tackle the issue of long memory in betas and conclude that betas do not exhibit such properties.[2] However, this conclusion is mainly based on a relatively small sample of daily data and only considering tests on the autocorrelation functions. In this study, we use a large dataset of high-frequency data to comprehensively reexamine whether betas are best described by either (i) short-memory processes, (ii) difference-stationary processes, or (iii) whether beta time series instead show long-memory properties.

First, we use 30-minute high-frequency data to estimate each month the realized betas for all stocks for which high-frequency data are available. Next, we estimate the memory of realized beta using the two-step exact local Whittle (2ELW) estimator by Shimotsu & Phillips (2005) and Shimotsu (2010).[3] We find that betas show consistent long-memory properties. The average estimate for the long-memory parameter d is 0.56. Adjusting for potential structural breaks in the beta series decreases the average d only modestly, to 0.51. For the vast majority of stocks, the statistical tests clearly reject both the short-memory ($d = 0$) and difference-stationary ($d = 1$) alternatives. Thus, most previous studies substantially misspecify the properties of the beta time series.

Our findings indicate that market betas have substantially longer memory than docu-

[1] Black et al. (1992), for example, explicitly model beta dynamics with a random walk.

[2] The literature on volatility modeling, though, documents that volatility has clear long-memory properties (Baillie et al., 1996; Bollerslev & Mikkelsen, 1996; Ding & Granger, 1996).

[3] In simulations, we show that, as opposed to the 2ELW estimator, the alternative, theoretically noise or structural break robust, estimators of Hurvich et al. (2005), Iacone (2010), Frederiksen et al. (2012), and Hou & Perron (2014) suffer from material biases in finite samples. Therefore, for our main analysis, we use the 2ELW estimator.

mented in Andersen et al. (2006). There are several potential reasons for this difference. First, our study has a substantially broader focus: we consider more than 1,700 stocks. Second, we use high-frequency data to estimate beta factors.[4] This enables us to obtain more precise and less noisy beta estimates (see also Bollerslev et al., 2016; Bollerslev et al., 2020; Hollstein et al., 2020). Noise in the beta series of Andersen et al. (2006) could potentially lead to a downward bias in memory estimates, as found by Deo & Hurvich (2001) and Arteche (2004). In contrast, we find that changing the bandwidth in the 2ELW estimator leads to similar estimates of the memory parameter. This suggests that the noise in our beta series is small. Third, and most importantly, using simulations, we show that for small samples, tests based on autocorrelation functions (as are used by Andersen et al., 2006), as opposed to direct estimates with the 2ELW estimator (which we rely on), have little power to detect *true* long memory.

Having documented that betas exhibit distinct long-memory properties, we next examine the implications of this result for forecasting. Beta forecasts are of paramount importance for many applications in finance. For example, capital allocation decisions, portfolio risk management (Daniel et al., 2020), as well as firms' cost of capital estimates (Levi & Welch, 2017) strongly depend on precise forecasts of betas. For our main analysis, we thus examine the out-of-sample forecast performance of the different models for 50 beta-sorted portfolios over a six-month horizon. We find that a FI model, which uses only the long-memory properties for beta forecasting, yields the lowest root mean squared error (RMSE). The FI model significantly outperforms both the short-memory (AR(p), ARMA(p, q)) and difference-stationary (RW as well as a modified RWW random walk estimator based on Welch, 2019) alternatives for a substantial fraction of the portfolios. A full-fledged FIARMA(d, p, q) alternative performs slightly worse than the pure FI model, but better than the AR, ARMA, RW, and RWW models. Thus, accounting for the long-memory property is very important for

[4] In a robustness analysis, Andersen et al. (2006) also use betas based on high-frequency data for the sample period 1993–1999, likewise concluding that betas do not show long-memory properties. The reason for this likely relates to the third point discussed below: their tests based on autocorrelation functions have very limited power to detect true long memory.

obtaining good beta forecasts.

In a next step, we examine the economic value of the forecasts by the different models in portfolio formation. Sorting the stocks into portfolios based on their beta forecasts, we detect the largest ex-post beta spread for the FIARMA and FI models. We run a battery of tests to document the robustness of these results. First, we alternatively estimate the short-memory and difference-stationary models in a state–space framework. In addition, we consider the HAR model augmented by jump beta components as well as a FI model, for which we set the long-memory parameter d to 0.5 instead of estimating it (FI(0.5)). We find that all alternative models underperform the FI model. Interestingly, though, the FI(0.5) model performs almost as well as the standard FI model. Second, we also document the long-memory properties of betas for the entire Center for Research in Security Prices (CRSP) sample. For this substantially larger sample and a much longer time period, we find that the FI model also outperforms all the alternatives.

In the Online Appendix, we present the results of further robustness tests. First, we document that the FI model is the only one that is consistently successful in generating ex-post market-neutral portfolios for all anomaly variables. All alternatives fail for at least one of the six anomalies examined. Second, we document substantial industry effects: stocks in the Energy, Utilities, and Manufacturing industries have comparably high memory in beta, while stocks in the Healthcare, HiTec Equipment, and Telephone industries tend to have relatively low memory in beta. The latter industries are and have been particularly prone to disruptions and creative destruction. The somewhat shorter memory of the betas of these stocks is thus consistent with what one might intuitively expect. One should note, however, that these still exhibit long memory: past shocks also have a long-lasting impact on their betas. Third, we document that for small, illiquid, and high-momentum stocks, using a RW model instead of the FI model yields particularly high errors. On the other hand, for liquid, low-leverage, and young stocks, it is most harmful to use an ARMA(p, q) model instead of the FI model.

Fourth, we show that the FI and FIARMA models also outperform their competitors when using hedging errors instead of the RMSE to evaluate the forecasts. Fifth, we study alternative forecast horizons between one month and one year, reaching similar conclusions. Finally, we perform the analysis at the individual stock level, consider the alternative estimator for the d parameter of Geweke & Porter-Hudak (1983), alternative intra-day sampling frequencies, alternative rolling estimation windows, bandwidths, and a correction for asynchronous trading. Our conclusions remain unchanged.

Our advice to the academic and professional community, thus, is to rely on long-memory beta forecasts whenever possible. Naturally, to consider long memory, one needs to be able to observe a sufficiently long sample of historical returns. Therefore, for newly listed or very young firms it is very difficult to accurately estimate the d parameters and make forecasts based on long-memory models. In this case, Welch's (2019) simple RWW estimator, which also works quite well for small stocks, emerges as a viable alternative.

Our paper contributes to the literature on beta estimation. Hollstein & Prokopczuk (2016) consider both $I(0)$ and $I(1)$ beta forecasts, but do not take into account models that account for long memory. Further contributions that deal with beta estimation include Buss & Vilkov (2012), Levi & Welch (2017), Hollstein et al. (2019), and Welch (2019). We complement these studies by explicitly considering long-memory processes to make beta forecasts. To the best of our knowledge, we are the first to show that forecasting beta with long-memory models yields superior forecasts compared to both $I(0)$ and $I(1)$ models.[5]

The literature has proposed several variations of models with heterogeneous agents (e.g., Müller et al., 1993; LeBaron, 2001; LeBaron, 2006; Alfarano & Lux, 2007; Corsi, 2009). In these models, agents incur heterogeneous planning and investment horizons. The interaction of different agents creates long memory both in total volatility and in systematic risk factors (Kamara et al., 2016; Brennan & Zhang, 2020). Heterogeneous agents' interactions, thus, are one (although not the only) possible mechanism generating the long memory in beta.

[5] Consistent with our main results, the complementary, concurrent study of Hollstein (2020) shows that the FI model also performs well for beta forecasting in an international setting.

We organize the remainder of this paper as follows. Section 2 introduces the data and shows summary statistics. We present results about long memory in betas in Section 3. In Section 4, we examine the impact of our findings for the forecasting of betas. Section 5 contains the economic implications of our findings and presents the results of several further analyses. In Section 6, we draw conclusions. The Online Appendix provides a further extensive set of robustness checks.

2 Data and Methodology

2.1 Data

Our main dataset covers U.S. stocks for the sample period from January 1996 to December 2014.[6] Following Hollstein et al. (2020), for our main analysis we restrict our attention to stocks for which we have high-frequency data. We collect high-frequency price data from the Thomson Reuters Tick History (TRTH) database. On average, the stocks for which high-frequency data are available represent 84 percent of the entire market capitalization of ordinary common U.S. stocks.

In order to process the final high-frequency dataset, we follow the data-cleaning steps outlined in Barndorff-Nielsen et al. (2009). First, we use only data with a time stamp during the exchange trading hours, i.e., between 9:30AM and 4:00PM Eastern Standard Time. Second, we remove recording errors in prices. To be more specific, we filter out prices that differ by more than 10 mean absolute deviations from a rolling centered median of 50 observations. Afterwards, we assign prices to every 30-minute interval using the most recent entry recorded that occurred at most one day before. Finally, we follow Bollerslev et al. (2016) and supplement the TRTH data with data on stock splits and distributions from CRSP to adjust the TRTH overnight returns.

[6] For a robustness check we also use the entire CRSP dataset and a time period starting from 1926. We present the results in Section 5.3. These are qualitatively similar to those of our main analysis.

2.2 Beta Estimation

Following Andersen et al. (2006), we use the realized beta estimator to obtain betas. We utilize intra-day high-frequency log-returns, sampled at intervals of 30 minutes to estimate

$$\beta_{i,t} = \frac{\sum_{\tau=1}^{O} r_{i,\tau} r_{M,\tau}}{\sum_{\tau=1}^{O} r_{M,\tau}^2},$$

where O is the number of high-frequency return observations during the time period under investigation.[7] $\beta_{i,t}$ is the beta estimate for asset i using data until the end of month t. $r_{i,\tau}$ and $r_{M,\tau}$ refer to the return of asset i and the market return at time τ, respectively. For the main analysis, we consider the time series of nonoverlapping monthly realized beta estimates.

The choice of sampling frequency underlies a delicate trade-off (Patton & Verardo, 2012). On the one hand, using low-frequency data could result in noisy estimates of beta (Andersen et al., 2005). On the other, pushing the analysis to a very high frequency introduces a number of microstructure issues (Scholes & Williams, 1977; Epps, 1979). To balance these effects, we focus our main analysis on a sampling frequency of 30 minutes. In Section A.2.9 of the Online Appendix, we show that our main results are robust to the choice of sampling frequency.

2.3 Long-Memory Estimation

Our estimation of the order of integration d of a beta time series relies on the 2ELW estimator as introduced in Shimotsu & Phillips (2005) and Shimotsu (2010). Given a time series y_t we can obtain this estimator as follows. We first consider the tapered local Whittle

[7] Note that this formula resembles the expanded formula for the covariance, while neglecting both the drift term and the risk-free rate. Andersen et al. (2006) note that the effect of the drift term vanishes as the sampling frequency increases, which effectively "annihilates" the mean. Empirically, for example, the average 30-minute return of the S&P 500 index amounts to 0.0017 percent. The average daily riskless interest rate during our sample period amounts to 0.01 percent, which is equivalent to an average risk-free rate as low as 0.0007 percent over 30-minute intervals. Thus, at this sampling frequency, both the drift and the risk-free rate can indeed be neglected.

estimator by Velasco (1999), which is obtained by

$$\hat{d}_{Vel} = argmin_{d\in(-0.5,2)} \left[log \left(\frac{3}{m} \sum_{j}^{m} \lambda_j^{2d} I_y^*(\lambda_j) \right) - 2d \frac{3}{m} \sum_{j}^{m} log\lambda_j \right].$$

Here, $I_y^*(\lambda_j)$ is the cosine-bell tapered periodogram of the series at frequency λ_j with $j = 3, 6, \ldots, m$. Furthermore, m is the bandwidth parameter, which determines the number of frequencies used for estimation. Larger m imply less variance of the estimates but then the estimator will be biased in case the underlying process exhibits short-run dependencies. We follow Shimotsu (2010) and consider $m = T^{0.7}$ in the following and report qualitatively similar results for alternative bandwidths of $m = T^{0.65}$ and $m = T^{0.75}$ as a robustness check in Section A.2.10 of the Online Appendix.

Under some mild assumptions, this estimator is consistent and asymptotically normal for $d \in (-0.5, 2)$. However, as the estimator considers only every third frequency of the periodogram its variance exceeds that of the standard local Whittle estimator by Robinson (1995). To account for this, the estimate is adjusted in the second step using

$$\hat{d}_{2ELW} = \hat{d}_{Vel} - \frac{L'(\hat{d}_{Vel})}{L''(\hat{d}_{Vel})}, \quad \text{with}$$
$$L(d) = log \left(\frac{1}{m} \sum_{j=1}^{m} I_{\Delta^d y - \mu(d)}(\lambda_j) \right) - 2d \frac{1}{m} \sum_{j=1}^{m} log\lambda_j.$$

Here, $I_{\Delta^d y-\mu(d)}(\lambda_j)$ is the periodogram of the demeaned series. Since the arithmetic mean $\bar{y}$ is inconsistent for $d > 0.5$, Shimotsu (2010) suggests using $\mu(d) = \bar{y}$ if $d < 0.5$, $\mu(d) = y_1$ if $d > 0.75$, and $\mu(d) = \omega(d)\bar{y} + (1 - \omega(d))y_1$ with $\omega(d) = 0.5[1 + cos(4\pi d)]$ if $d \in [0.5, 0.75]$. This two-step estimator then has the same limiting variance as the standard local Whittle estimator while being consistent and asymptotically normally distributed for $d \in (-0.5, 2)$. Consequently, the 2ELW estimator can be used to distinguish short-memory series ($d = 0$), stationary long-memory series ($0 < d < 0.5$), nonstationary long-memory series ($0.5 < d < 1$), and difference-stationary series ($d = 1$) such as the random walk. This is an

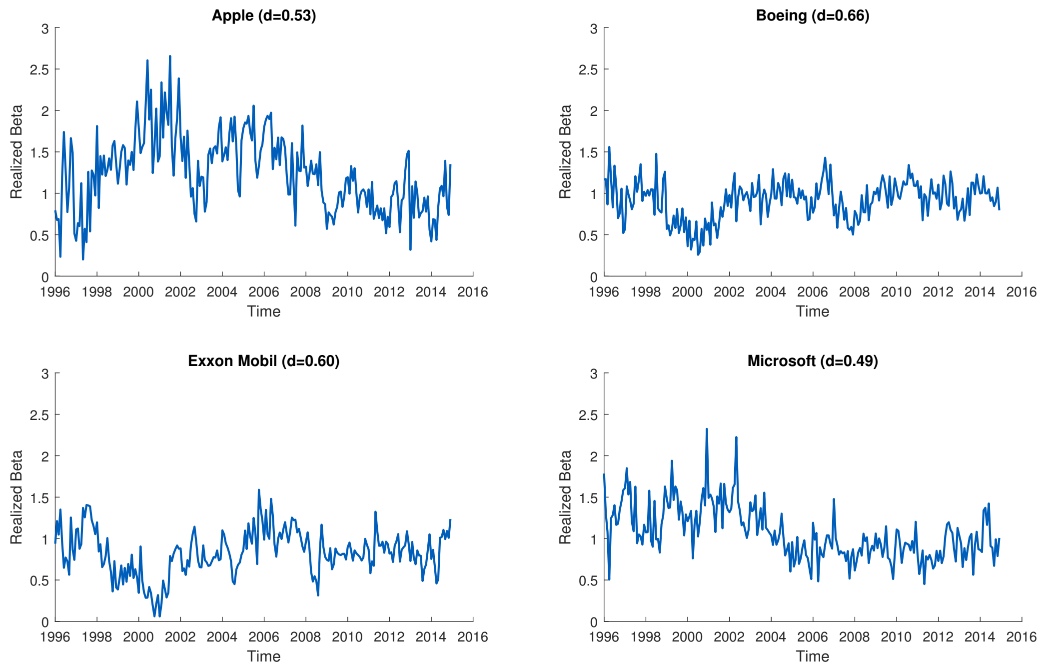

Figure 1: In this figure, we present the time series of monthly nonoverlapping realized beta estimates for Apple, Boeing, Exxon Mobil, and Microsoft. In parentheses, we present the companies' ds based on the 2ELW estimator.

advantage over the standard local Whittle estimator, which can only be used for inference for $-0.5 < d < 0.75$, as it has a nonnormal limit distribution otherwise.

3 Long Memory in Beta

3.1 Estimation Results

In Figure 1, we first present the realized beta time series of four exemplary companies: Apple, Boeing, Exxon Mobil, and Microsoft. For all companies, the beta time series contain several local trends and cycles, which are typical for long-memory processes. Thus, based on this first visual inspection, one is tempted to suspect that there could be long memory in the beta time series. Indeed, the d estimates for all four companies exceed 0.4.

Analyzing the memory of beta in our sample more systematically, the left panel of Table 1 shows the average estimated d across the realized beta series of all stocks with more than

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
β_i	0.555	0.158	0.971	0.996	0.509	0.172	0.963	0.997

Table 1: This table presents average estimates of the memory parameter of realized beta across all stocks ($\bar{\hat{d}}_i$) using the 2ELW estimator of Shimotsu & Phillips (2005) and Shimotsu (2010). Additionally, $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d_i = 0$ and $d_i = 1$, respectively, are rejected at the 10 percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle & Moulines (2000).

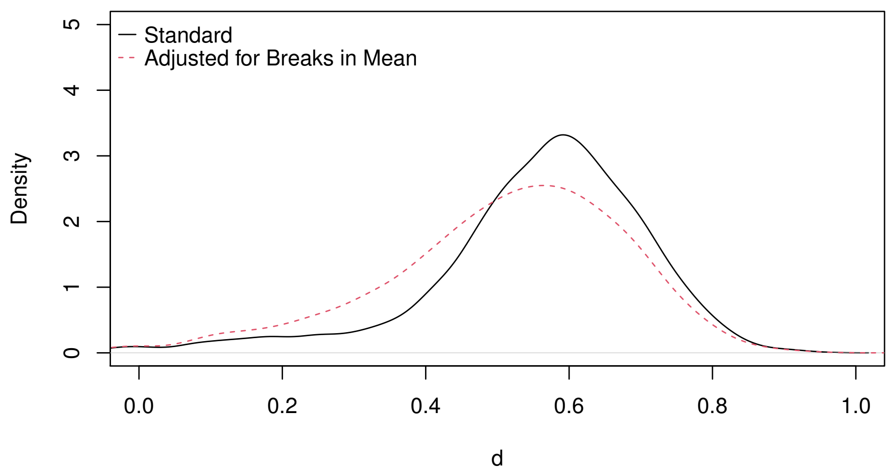

Figure 2: Density plot showing the distribution of the estimated beta memory parameters across stocks. For estimation we consider the Gaussian kernel and choose the bandwidth according to Silverman (1986).

100 monthly observations (for $N = 1{,}752$ stocks we have sufficient data) using the 2ELW estimator. Additionally, we present the standard deviation of the estimates across stocks and the relative frequency with which the d estimates of different stocks are significantly different from 0 and 1, respectively, at the 10 percent level. To illustrate the variation in d across stocks, Figure 2 additionally plots the corresponding density of the estimates.

Table 1 reveals that the average d is approximately 0.555 and Figure 2 shows that while there is some variation across stocks, most of them have a d between 0.4 and 0.8. A formal statistical test also confirms that for more than 96 percent of the stocks it holds that $0 < d < 1$ at the 10 percent level. At the 1 percent level this is still true for more than 92 percent of the stocks.

As a firm's business may change over time, some of the considered companies could exhibit a structural break in the realized beta series. When the underlying process is stationary, i.e., $d < 0.5$, but exhibits structural breaks in mean, then the local Whittle estimator and therefore also the 2ELW estimator is positively biased (e.g., Diebold & Inoue, 2001; Granger & Hyung, 2004). One way to account for this would be to use the estimators by Iacone (2010) or Hou & Perron (2014), as these remain consistent when structural breaks are present. However, as we also show in simulations in the Online Appendix, these are negatively biased for sample sizes smaller than 500, making them unsuitable for our application. To examine the robustness of our results, we therefore use an alternative two-step procedure. We first estimate the points at which the series exhibit structural breaks in mean using the procedure by Lavielle & Moulines (2000) and then apply the 2ELW estimator estimator for the cleaned series.[8]

The results are shown in the right panel of Table 1 and are visualized by the dashed line in Figure 2. We find that the average $\hat{d}$ decreases slightly to 0.509, implying that some stocks do indeed exhibit structural breaks in their beta time series. However, the reduction is small and for more than 96 percent of the stocks the null that $d = 0$ can still be rejected.

Our results stand in contrast to those by Andersen et al. (2006), who argue that betas are integrated of a much smaller order, often even $I(0)$. There are two main reasons for this difference in results.

First, Andersen et al. (2006) base their main analysis on daily data, which leads to noisy estimates of beta, as also acknowledged by the authors themselves. Deo & Hurvich (2001)

[8]Bai & Perron (1998, 2003) suggest estimating breaks in mean by minimizing the residual sum of squares (RSS) of $\beta_t = \mu_s + e_t$, where μ_s is the mean in segment s with $s = 1, \ldots, S$ and S being determined by means of the BIC. Lavielle & Moulines (2000) extend this approach by adding a penalty term to the BIC criterion, which is then $BIC = RSS(S) + 4S \log(T) T^{2d-1}$. This leads to a more parsimonious breakpoint selection, as for long-memory time series the standard procedure indicates too many breakpoints.

and Arteche (2004) show that for perturbed series any inference on the order of integration is biased such that the series appear to be less integrated. Our beta estimates based on intra-day observations, on the other hand, are less noisy, implying that the true order of integration can be better detected. To further illustrate this, one might think of comparing the 2ELW estimates to estimates made by noise robust estimators such as those of Sun & Phillips (2003) or Frederiksen et al. (2012). However, these are positively biased when the sample size is smaller than 500, making them inappropriate for our setup. As an alternative we show in Table A.15 of the Online Appendix that changing the bandwidth m in the 2ELW estimation leads to similar estimates of d. As demonstrated by Hurvich et al. (2005), this would not be the case if the series were seriously perturbed.

Second, Andersen et al. (2006) rely on graphical investigation of the first 36 autocorrelations instead of consistent estimation of the memory parameter. Particularly in small samples (Andersen et al., 2006 consider $T = 148$) this type of inference may lead to false conclusions. We illustrate this by means of a small simulation study for which we report the results in Table A.1 of the Online Appendix. We simulate fractionally integrated noise, i.e., $(1 - B)^d y_t = \epsilon_t$, with B being the backshift operator, for memory parameters of $d = 0.2, 0.4, 0.6$ and sample sizes of $T = 100, 148, 228, 1000$. The table reveals that on average only 24 percent of the first 36 autocorrelations of an $I(0.4)$ process with $T = 148$ are significantly larger than zero. From this result one might falsely infer that the series exhibit short memory. In contrast, the simulation results show that the 2ELW estimator is also unbiased in small samples, implying that the correct order of integration can be detected. For further details on the simulation setup and results we refer to Section A.1 of the Online Appendix.[9]

We therefore conclude that realized betas are highly persistent and are best described by

[9]Table A.1 also presents results for the estimators by Sun & Phillips (2003), Iacone (2010), Frederiksen et al. (2012), and Hou & Perron (2014) to validate our claim that these are biased in small samples. Additionally, the table presents results for the log periodogram estimator, which we consider for a robustness check (also presented in the Online Appendix). This estimator is also unbiased, but exhibits a larger variance than the 2ELW estimator.

either pure long-memory processes or a combination of break and long-memory process.

3.2 Beta Decomposition

Since beta is actually a combination of different components, it might be interesting to investigate which of these drives the persistence. For that purpose, consider the following decomposition

$$\beta_{i,t} = \sigma_{i,M,t}\sigma_{M,t}^{-2} = \rho_{i,M,t}\sigma_{i,t}\sigma_{M,t}\sigma_{M,t}^{-2} = \rho_{i,M,t}\sigma_{i,t}\sigma_{M,t}^{-1}, \tag{1}$$

where $\sigma_{i,M,t}$ is the realized covariance of asset i and the market M at time t, $\rho_{i,M,t}$ is their realized correlation, and $\sigma_{i,t}$ is the realized volatility. Consequently, Equation (1) shows that the realized beta series evolve as the product of realized correlation, realized volatility, and the inverse of realized market volatility. Leschinski (2017) shows theoretically that the products of stationary long-memory series (i.e., $0 < d < 0.5$) with nonzero mean are integrated with the maximum memory of the series. This would mean that one of the components needs to exhibit the same degree of memory as realized beta while the others could exhibit a smaller d, even $d = 0$. However, for approximately 73 percent of the stocks it holds that $d > 0.5$, meaning that the beta series exhibit nonstationary long memory. In these cases, it is theoretically unclear how products of such series behave. We therefore also estimate the order of integration of realized correlation, realized volatility, and the inverse of realized market volatility using the 2ELW estimator.[10]

The results are shown in Table 2.[11] Again, we consider the possibility of structural breaks and also report results when adjusting for these. The realized correlation and the inverse of realized market volatility on average exhibit a d of approximately 0.57 and 0.59, respectively,

[10]We obtain the realized volatility for stock i and the market ($i = M$) as $\sigma_{i,t} = \sqrt{\sum_{\tau=1}^{O} r_{i,\tau}^2}$, the realized covariance as $\sigma_{i,M,t} = \sum_{\tau=1}^{O} r_{i,\tau} r_{M,\tau}$, and the realized correlation as $\rho_{i,M,t} = \frac{\sigma_{i,M,t}}{\sigma_{i,t}\sigma_{M,t}}$.

[11]We Fisher-transform the realized correlation series to guarantee that there is no bias due to the restricted character of the variable. If we use the original series, the results are similar.

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
$\rho_{i,M}$	0.565	0.139	0.976	0.996	0.559	0.142	0.975	0.997
σ_i	0.598	0.154	0.991	0.967	0.598	0.154	0.991	0.967
σ_M^{-1}	0.588	-	1.000	1.000	0.585	-	1.000	1.000

Table 2: This table presents average estimates of the memory parameter of realized correlation (Fisher-transformed), and volatility across all stocks ($N = 1752$), as well as that of the inverse of the market volatility, using the 2ELW estimator of Shimotsu & Phillips (2005) and Shimotsu (2010). $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d_i = 0$ and $d_i = 1$, respectively, are rejected at the 10 percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle & Moulines (2000).

while the d of realized volatility is even slightly higher on average, with 0.60. Again, tests indicate that for almost all stocks the order of integration is different from 0 and 1 for all three components.

When adjusting for structural breaks, the d of the realized correlation decreases slightly, while the d of realized volatility does not. Consequently, it is rather breaks in realized correlation than breaks in volatility that drive the breaks observed in the realized betas. When comparing the actual estimate of d to the estimate of the memory of the realized beta series, it can be seen that all three components exhibit a slightly higher degree of persistence. Thus, it seems that no single component, but rather all of them, drives the persistence in realized betas.

4 Forecasting

Having shown that betas have consistent long-memory properties, the natural next questions to ask are: Can we leverage the long-memory properties in betas to make better forecasts? How big are the errors when inaccurately imposing $I(0)$ or $I(1)$ dynamics for forecasting betas? In this section, we set out to answer these questions. For this purpose, we compare pseudo out-of-sample forecasts for the realized beta series of models accounting for the long-memory characteristics with those for short-memory and difference-stationary processes.

4.1 Forecasting Methodology

For forecasting using long-memory models we follow the approach proposed by Hassler & Pohle (2019). Given the estimated order of integration of a series, we first remove the persistence by filtering. Then we calculate the mean of the series. In a next step, we forecast the filtered data accounting for potential short-run dependencies. Finally, we reintegrate the series to obtain a forecast.

In more detail, given a historical window of T betas of stock i, we first compute the $\hat{d}$th difference

$$\Delta^{\hat{d}_{i,T}}\beta_{i,t} = (1-L)^{\hat{d}_{i,T}}\beta_{i,t} = \sum_{j=0}^{t-1}\binom{\hat{d}_{i,T}}{j}(-1)^j\beta_{i,t-j}, \text{ with } t = 1,\ldots,T,$$

where $\hat{d}_{i,T}$ is the estimate of the 2ELW estimator with a bandwidth of $m = T^{0.7}$ for stock i from an estimation window ending at T. Again, we report qualitatively similar results for $m = T^{0.65}$ and $m = T^{0.75}$ in Section A.2.10 of the Online Appendix.

We then set out to calculate the conditional mean $\mu_{i,T}$ of the series, which is complicated by the long-memory characteristics. As discussed above, the arithmetic mean cannot be considered for nonstationary long-memory series as it does not exhibit a finite variance. We therefore consider the approach by Robinson (1994) to estimate $\mu_{i,T}$. For this purpose, we perform the following regression

$$\Delta^{\hat{d}_{i,T}}\beta_{i,t} = \psi_{i,t}\mu_{i,T} + \eta_{i,t}, \text{ with } \psi_{i,t} = \sum_{j=0}^{t-1}\binom{\hat{d}_{i,T}}{j}(-1)^j,$$

where $\eta_{i,t}$ is the error term that contains possible short-run dynamics. This allows us to calculate the residuals

$$\varepsilon_{i,t} = \Delta^{\hat{d}_{i,T}} \beta_{i,t} - \psi_{i,t} \hat{\mu}_{i,T},$$

which are not fractionally integrated any longer, but might exhibit short-run dependencies. We can optionally account for these using an ARMA(p, q) model

$$\varepsilon_{i,t} = \phi_{i,1}\varepsilon_{i,t-1} + \ldots + \phi_{i,p}\varepsilon_{i,t-p} + \theta_{i,1}\zeta_{i,t-1} + \ldots + \theta_{i,q}\zeta_{i,t-q} + \zeta_{i,t}, \qquad \text{with } t = 2, \ldots, T,$$

where $\zeta_{i,t}$ is the mean zero error term and p and q are determined by means of the BIC with a maximum lag length of $12[(T/100)^{0.25}]$. This allows us to forecast the residuals h steps ahead

$$\hat{\varepsilon}_{i,T+h} = \hat{\phi}_{i,1}\hat{\varepsilon}_{i,T+h-1} + \ldots + \hat{\phi}_{i,p}\hat{\varepsilon}_{i,T+h-p} + \hat{\theta}_{i,1}\hat{\zeta}_{i,T+h-1} + \ldots + \hat{\theta}_{i,q}\hat{\zeta}_{i,T+h-q}.$$

For $\hat{\varepsilon}_{i,T+h}$, the hat indicates that it is a forecast and h denotes the forecast window in months. In a case without short-run dependencies we simply set $\hat{\varepsilon}_{i,T+h} = 0$. We then reintegrate the series to account for the long-memory characteristics by calculating $\hat{Z}_{i,t} = \Delta^{-\hat{d}_{i,T}}\hat{\varepsilon}_{i,t}$ for $t = 1, \ldots, T+h$, respectively $t = 2, \ldots, T+h$. Forecasts of the original sequence then evolve as

$$\hat{\beta}_{i,T+h} = \mu_{i,T} + \hat{Z}_{i,T+h}.$$

The resulting forecasts for h-month betas can be computed as the averages of the point forecasts during the next h months.

This approach allows us to forecast stationary as well as nonstationary series while also accounting for potential short-run dynamics. We denote the model with short-run components by FIARMA in the following to emphasize that there is a difference from the standard ARFIMA models, as introduced by Granger & Joyeux (1980) and Hosking (1981), which only allow modeling and forecasting stationary series with $d < 0.5$. We refer to the model without short-run dependencies simply as FI.

As difference-stationary and short-memory competitor models, we consider the random walk model, for which $\hat{\beta}_{T+h} = \beta_T$, as well as AR($p$) and ARMA($p, q$) models, respectively. We estimate the latter models based on

$$\beta_{i,t} = a_i + \phi_{i,1}\beta_{i,t-1} + \ldots + \phi_{i,p}\beta_{i,t-p} + \theta_{i,1}e_{i,t-1} + \ldots + \theta_{i,q}e_{i,t-q} + e_{i,t}, \quad \text{with } t = 2, \ldots, T.$$

For the AR model we set $\theta_{i,1} = \ldots = \theta_{i,q} = 0$. Again, we choose p and q according to the BIC with a maximum lag length of $12\left[(T/100)^{0.25}\right]$.

Another popular way to model and forecast long-memory time series is to use the HAR model of Corsi (2009). For the realized beta series, it evolves as

$$\beta_{i,t} = a_i + \phi_{1,i}\beta_{i,t-1} + \frac{\phi_{2,i}}{5}\sum_{j=1}^{5}\beta_{i,t-j} + \frac{\phi_{3,i}}{22}\sum_{j=1}^{22}\beta_{i,t-j} + e_{i,t}, \tag{2}$$

where $e_{i,t}$ is a mean zero error term. While the HAR model does not formally belong to the class of long-memory models, when applied to return volatility time series, this model has been shown to be able to reproduce long-memory patterns.[12] We therefore also consider forecasts made by this model in the following. Finally, we also consider a random walk model based on the weighted slope-winsorized estimator of Welch (2019) (RWW). For this estimator, one first has to winsorize the individual stock returns between −2 and +4 times the contemporaneous market return. The second step of the Welch (2019) estimator consists of computing the beta estimates with exponentially decaying weights (with a decay parameter equal to 2/(252*14)).[13,14]

[12] Baillie et al. (2019), though, show that HAR models generally cannot capture the full scale of long memory. It is thus possible that the model performs less well than the FI and FIARMA alternatives.

[13] Note that we adjust the decay parameter of Welch (2019) to be consistent with the 30-minute high-frequency data, for which we have 14 observations per day. We consider an alternative version of the Welch (2019) estimator, which exactly follows the author's implementation, in Section 5.

[14] To keep the presentation manageable, we focus on these seven models. In Section 5.2, we consider several alternatives. The results are qualitatively similar to those presented here.

To examine the out-of-sample forecast accuracy of the different approaches, we perform the analysis using the root mean squared error (RMSE), a loss function commonly applied in the literature

$$\text{RMSE}_{i,h} = \sqrt{\frac{1}{\Upsilon}\sum_{T=1}^{\Upsilon}(\beta_{i,T+h} - \hat{\beta}_{i,T+h})^2},$$

where Υ is the number of out-of-sample observations of realized and predicted betas of one stock. $\beta_{i,T+h}$ is the realized beta and $\hat{\beta}_{i,T+h}$ denotes a beta forecast. The RMSE criterion is suitable since it is robust to the presence of (mean zero) noise in the evaluation proxy, while other commonly employed loss functions are not (Patton, 2011).[15] We test for significance in RMSE differences univariately using the modified Diebold–Mariano (DM) test proposed by Harvey et al. (1997) and additionally with the model confidence set (MCS) approach of Hansen et al. (2011). The MCS approach is designed such that it contains the best model based on a certain level of confidence.

To further mitigate the impact of errors-in-variables in the forecast evaluation proxy, we build portfolios. To do so, we follow Fama & MacBeth (1973) and Hollstein & Prokopczuk (2016). That is, each month we use the (realized) beta estimate for each firm obtained during the second-to-last nonoverlapping beta estimation window.[16] In doing so, we use a common sorting variable for all forecast approaches. That is, we employ one method to sort and then forecast betas for the same portfolios with different methodologies. We build 50 value-weighted portfolios. Figure 3 depicts the average market capitalization of the stocks

[15] The results when using the mean absolute error (MAE) criterion instead of the RMSE are qualitatively similar.

[16] We do so to ensure that (i) there is a spread in the market betas of the different portfolios. (ii) Using a sorting variable that is independent of the predictor variables is important to avoid discriminating against any of the predictors. Otherwise, the stocks with the highest positive and negative noise for the estimator upon which the betas are sorted are likely to end up in the extreme portfolios and that noise cannot be fully diversified. For stocks with missing estimates for the sorting variable, we set it to 1.

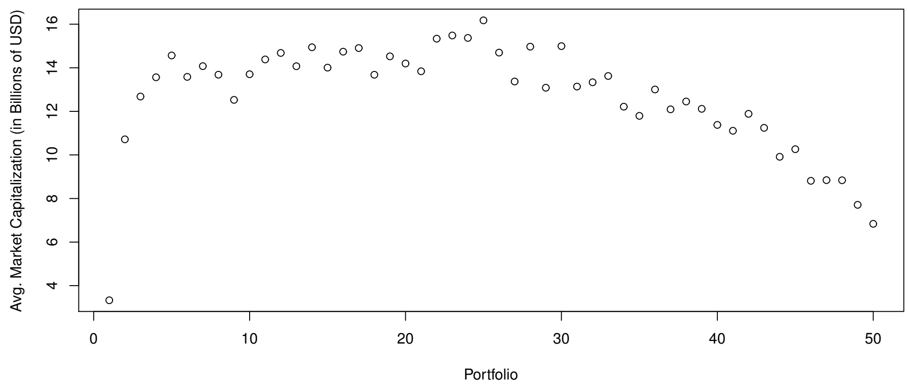

Figure 3: This figure depicts the average market capitalizations (in Billions of USD) of the stocks sorted into the 50 portfolios. The stocks are sorted each month (in ascending order) based on the (realized) beta estimate for each firm obtained during the second-to-last nonoverlapping beta estimation window. We sort them into 50 portfolios based on breakpoints derived from stocks traded at the New York Stock Exchange (NYSE), which have a sorting beta between 0.2 and 2. The minimum number of stocks in any portfolio at any time is 11. On average, each portfolio contains 22 stocks.

in the different portfolios.[17] All portfolio forecasts are based on the beta time series of the stocks newly allocated to these portfolios.

4.2 Forecast Results

The results of the various beta forecasts can be found in Table 3. We use a rolling estimation window of 100 observations along with a (monthly overlapping) forecast window of 6 months (Chang et al., 2012; Hollstein & Prokopczuk, 2016).[18] Table 3 presents the average RMSE across all stocks, the number of times the model yields the lowest RMSE

[17] As a further precaution against ending up with many portfolios populated only by small and illiquid stocks, we derive the sorting breakpoints only from stocks traded at the New York Stock Exchange (NYSE), which have a sorting beta between 0.2 and 2 (3.5% of (mostly small) NYSE stocks have betas below 0.2 and 2.1% betas above 2). The portfolios are then allocated based on these breakpoints, using all stocks. When simply using name breakpoints and no restrictions on beta, the market capitalization distribution is more uneven in that the extreme portfolios contain even smaller stocks on average. The (untabulated) results, though, are qualitatively similar. In addition, we present qualitatively similar results for individual stocks in Section A.2.6 of the Online Appendix. Finally, untabulated results for alternative numbers of portfolios are also qualitatively similar.

[18] Qualitatively similar results for forecast horizons of 1, 3, and 12 months and estimation windows of 75 and 125 months can be found in Sections A.2.5 and A.2.10 of the Online Appendix, respectively.

	RW	RWW	AR	ARMA	HAR	FIARMA	FI
RMSE	0.1484	0.1277	0.1405	0.1290	0.1300	0.1232	0.1185
Best	0	6	0	0	2	25	17
In MCS	21	46	27	43	49	50	50
vs. RW	0	17	5	11	12	25	27
vs. RWW	0	0	0	1	3	10	9
vs. AR	1	4	0	27	13	24	24
vs. ARMA	0	0	0	0	2	12	16
vs. HAR	0	1	0	0	0	10	9
vs. FIARMA	0	0	0	0	0	0	2
vs. FI	0	0	0	0	0	2	0
N	50	50	50	50	50	50	50

Table 3: This table illustrates the forecast performance of the models for six-month beta forecasts from a rolling estimation window of 100 observations. The first row shows the average RMSEs of different models across all 50 portfolios. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain portfolio. "In MCS" denotes the number of portfolios for which an estimator is in the Hansen et al. (2011) model confidence set. Furthermore, the rows denoted by "vs. X" correspond to modified DM-tests (Harvey et al., 1997), providing the number of times the column-model yields a significantly lower RMSE than the row-model. We examine statistical significance toward the 10 percent level. Finally, N is the number of investigated portfolios.

when forecasting the realized beta of one of the portfolios, and the number of portfolios for which a model makes it into the Hansen et al. (2011) model confidence set. The remainder of the table indicates the number of stocks for which the column-model is significantly better than the row-model. We examine statistical significance toward the 10 percent level.

Table 3 reveals that the FI model performs best across all considered models. It has the lowest RMSE on average and is the model with the lowest RMSE for 17 of the 50 portfolios. Second best is the FIARMA model, which even turns out best for 25 of the portfolios. The HAR model, on the other hand, performs substantially worse, indicating that it does not work well in capturing the long-memory characteristics of the stocks' betas. The models that do not account for the long-memory characteristics of the beta time series, on the other hand, are only the most accurate for 6 of the portfolios (for all of which the Welch, 2019 estimator performs best).[19] The differences in RMSE are economically large. When moving from the RW model to using the FI model, the RMSE decreases by 20.1 percent. For RWW, AR, and ARMA, the improvements are 7.2 percent, 15.7 percent, and 8.2 percent, respectively.

The outperformance of the long-memory models is often also statistically significant. The FI and FIARMA are the only models which are in the model confidence set for every single portfolio. Furthermore, in single comparisons, the FI forecasts are significantly better for 27 of the 50 portfolios when compared to RW (9 when compared to RWW); compared to AR and ARMA forecasts these numbers are 24 and 16, respectively. On the other hand, the forecasts by the RW, RWW, AR, and ARMA models are never significantly better than those of the FI and FIARMA models. Consequently, we can conclude that accounting for long-run dependence substantially improves the forecasts for realized betas.

Our finding that the FI model yields a significantly lower RMSE than the RW model for almost all stocks has broad implications. Hollstein et al. (2020) show that a RW model outperforms other predictors based on daily data as well as the Buss & Vilkov (2012) option-implied beta. Thus, the FI forecasts appear to be preferable not only to other time-series models but also to a broader set of potential estimators.[20]

To further investigate the causes of the differential forecast performance of the models, we follow Mincer & Zarnowitz (1969) and decompose the time-series mean squared error (MSE) in the following fashion

$$\text{MSE}_i = \underbrace{(\bar{\beta}_i - \bar{\hat{\beta}}_i)^2}_{\text{bias}} + \underbrace{(1-b_i)^2\sigma^2(\hat{\beta}_i)}_{\text{inefficiency}} + \underbrace{(1-\rho_i^2)\sigma^2(\beta_i)}_{\text{random error}}. \quad (3)$$

b_i is the slope coefficient of the regression $\beta_i = a_i + b_i\hat{\beta}_i + e_i$ and ρ_i^2 is the coefficient of determination of this regression. A bias indicates that the model is misspecified and the prediction is, on average, different from the realization. Inefficiency represents a tendency

[19]Further untabulated analyses reveal that the FI and FIARMA models perform even better when based on the time series of the noise-reduced Welch (2019) estimator instead of that of the plain realized beta.

[20]In untabulated results, we confirm this also empirically: the FI model outperforms estimators based on daily return data as well as option-implied estimators.

	RW	RWW	AR	ARMA	HAR	FIARMA	FI
Bias	0.0020	0.0004	0.0010	0.0004	0.0016	0.0007	0.0006
Inefficiency	0.0112	0.0042	0.0050	0.0035	0.0040	0.0054	0.0018
Random Error	0.0135	0.0128	0.0157	0.0148	0.0138	0.0132	0.0132

Table 4: This table shows the Mincer & Zarnowitz (1969) decomposition of the MSE as of Equation (3). The MSE is based on six-month forecasts of the realized beta series performed with a rolling estimation window of 100 observations. All numbers represent the average across 50 beta-sorted portfolios.

of an estimator to systematically yield positive forecast errors for low values and negative forecast errors for high values or vice versa. The remaining *random* forecast errors are unrelated to the predictions and realizations.

Table 4 presents the results of the MSE decomposition. Again, the numbers represent the averages across all portfolios. We find that the RW model has the highest bias and highest inefficiency components. Thus, particularly for high- and low-beta stocks, the RW approach generates sizable measurement errors. Welch's (2019) RWW approach improves upon the performance of RW in both dimensions. For the AR and ARMA models, the bias component is small. The inefficiency is also dramatically smaller compared to the RW model. However, the random error component, which is the largest component for all models, is highest for the AR and ARMA models.

The FI and FIARMA models are approximately unbiased. The FI model yields the lowest overall inefficiency component, which indicates that the model does well in particular for stocks with the most extreme betas. Finally, the two models also yield low random errors. Accounting for short-run dynamics in addition to long memory in betas appears to only affect the inefficiency component. On the other hand, the HAR model is not among the group of best models in any of the dimensions.

5 Additional Analyses

5.1 Economic Implications: Beta Portfolios

An economic criterion on which to assess different beta estimation approaches that is important is a model's performance in portfolio formation. Therefore, we also evaluate the estimators based on their ability to create ex-post spreads in realized betas (Daniel & Titman, 1997).[21]

For each forecast approach, we separately sort the stocks into 10 portfolios based on their respective current beta forecasts at the end of each month. We then calculate the ex-post realized beta of each portfolio. A good predictor should generate a monotonically increasing pattern in ex-post betas. Furthermore, the spread between the realized betas of high and low beta forecast portfolios should be large.

We present the results in Table 5. Consistent with the results in the previous section, we find that the long-memory FIARMA and FI models yield the lowest ex-post realized beta for their respective low-beta portfolio 1. Both generate a realized beta of 0.43. The difference-stationary random walk model, on the other hand, generates an ex-post realized beta of 0.49. The ARMA model yields an ex-post beta of 0.46. Considering the high-beta portfolio (10), the results are similar. The models that account for long memory yield the highest ex-post realized betas, while those of short-memory and difference-stationary models are somewhat lower. Consequently, the FIARMA and FI models also yield the highest 10–1 portfolio spreads, although, admittedly, the realized beta spreads of most other approaches are generally also within a two-standard-error range.

5.2 Alternative Models

Due to its great importance, there are numerous approaches and models to forecast beta. For ease of presentation in our main analysis, we compare the performance of the long-

[21] We thank an anonymous referee for suggesting this test design to us.

	1	2	3	4	5	6	7	8	9	10	10 − 1
RW	0.489	0.623	0.744	0.838	0.921	0.990	1.078	1.177	1.307	1.550	1.061
	(0.014)	(0.015)	(0.010)	(0.011)	(0.021)	(0.029)	(0.038)	(0.046)	(0.046)	(0.052)	(0.047)
RWW	0.475	0.622	0.747	0.840	0.922	0.983	1.075	1.168	1.282	1.533	1.058
	(0.014)	(0.013)	(0.010)	(0.010)	(0.020)	(0.030)	(0.037)	(0.045)	(0.046)	(0.054)	(0.047)
AR	0.479	0.578	0.700	0.798	0.903	0.980	1.051	1.142	1.251	1.479	1.000
	(0.021)	(0.017)	(0.012)	(0.011)	(0.016)	(0.028)	(0.035)	(0.044)	(0.047)	(0.047)	(0.041)
ARMA	0.462	0.584	0.709	0.810	0.908	0.984	1.077	1.164	1.275	1.522	1.061
	(0.021)	(0.016)	(0.010)	(0.009)	(0.021)	(0.032)	(0.039)	(0.039)	(0.046)	(0.049)	(0.041)
HAR	0.460	0.605	0.735	0.845	0.935	1.009	1.116	1.208	1.318	1.560	1.100
	(0.021)	(0.015)	(0.009)	(0.011)	(0.024)	(0.029)	(0.044)	(0.041)	(0.044)	(0.030)	(0.040)
FIARMA	0.432	0.588	0.708	0.816	0.904	0.993	1.081	1.179	1.302	1.560	1.128
	(0.019)	(0.014)	(0.011)	(0.010)	(0.020)	(0.039)	(0.046)	(0.052)	(0.051)	(0.051)	(0.038)
FI	0.432	0.588	0.705	0.813	0.899	0.987	1.076	1.173	1.299	1.556	1.124
	(0.019)	(0.014)	(0.011)	(0.010)	(0.018)	(0.036)	(0.042)	(0.049)	(0.051)	(0.052)	(0.040)

Table 5: This table presents the ex-post realized betas of portfolios sorted by the different beta forecasts. That is, at the end of each month we sort the stocks into 10 portfolios based on the beta forecasts. We do so separately for each forecast approach. Subsequently, we calculate the ex-post realized beta of each portfolio, as well as that of the high-minus-low (10–1) portfolio. In parentheses, we present the robust Andrews (1991) standard errors, using a quadratic spectral density and data-driven bandwidth selection. We omit stars to indicate significance, because all ex-post realized betas are statistically significant at the 1 percent level.

memory models only to the performance of the most popular competitors, RW, RWW, AR, and ARMA. In this section we now consider other approaches that have been proposed in the literature. First, for the analysis in this section we exactly follow Welch's (2019) suggestion and base the RWW model on one year of daily return data and a decay parameter equal to 2/252.[22]

Furthermore, Andersen et al. (2005) consider an AR(1) process to model beta in a state–space framework. Hollstein & Prokopczuk (2016) investigate the forecast performance of RW, AR(1), and ARMA(1,1) models in a state–space framework and find that the RW model performs somewhat better than the AR(1) and ARMA(1,1) models. Thus, in this section we also consider the forecasts from RW, AR(1), and ARMA(1,1) models when estimated as

[22] In untabulated analyses we also tried the shrinkage estimators of Vasicek (1973) and Levi & Welch (2017). Both underperform the simpler RWW model and, consequently, also the FI and FIARMA models.

a state–space system. The measurement equation for all three models is

$$\beta_{i,t} = \tilde{\beta}_{i,t} + \xi_{i,t},$$

where $\tilde{\beta}_{i,t}$ is the unobserved *true* beta. It evolves according to one of the following transition equations for the different models

$$\begin{aligned}
\tilde{\beta}_{i,t}^{RW} &= \tilde{\beta}_{i,t-1} + v_{i,t}, \\
\tilde{\beta}_{i,t}^{AR} &= \gamma_i + \phi_i \tilde{\beta}_{i,t-1} + v_{i,t}, \text{ and} \\
\tilde{\beta}_{i,t}^{ARMA} &= \gamma_i + \phi_i \tilde{\beta}_{i,t-1} + \theta_i v_{i,t-1} + v_{i,t}.
\end{aligned}$$

We estimate these models using the Kalman filter (Pagan, 1980; Black et al., 1992) and then perform forecasts as for the standard models.

To the best of our knowledge, long-memory models in a state–space framework have only been proposed for the stationary $d < 0.5$ case (Chan & Palma, 1998; Dissanayake et al., 2016). Since we mostly deal with nonstationary time series here, these models are likely inappropriate. Therefore, we do not consider a state–space variant of these models. As an alternative, we examine a variation of the FI model, in which we fix the parameter at 0.5 (Hassler & Pohle, 2019), referring to the model as FI(0.5). 0.5 approximates the average d obtained for our sample (the cross-sectional mean after adjusting for breaks is 0.509). The estimation of the memory parameter is nontrivial and point estimates often still have large confidence intervals. By fixing the d at a certain level, one essentially trades off the estimation noise for slight inaccuracy in the resulting persistence parameter. It is an empirical question whether this approach works better or worse than that involving actual estimation of the d parameter.

Finally, we also consider a HAR model augmented by jump betas, as in Andersen et al. (2007).[23] The main motivation for examining these models is that discontinuous processes

[23] We thank an anonymous referee for suggesting that we should analyze this model.

	RW	RWW	AR	ARMA	HAR	HARJump	FIARMA	FI	FI(0.5)
RMSE	0.1264	0.1426	0.1348	0.1358	0.1301	0.1610	0.1233	0.1185	0.1206
Best	12	1	0	4	1	0	13	5	14
In MCS	48	34	40	39	45	28	50	50	50
vs. RW	0	0	0	0	1	1	4	3	2
vs. RWW	17	0	9	9	11	1	15	17	8
vs. AR	21	1	0	0	6	2	13	14	9
vs. ARMA	20	1	13	0	6	2	14	14	11
vs. HAR	4	0	4	4	0	1	10	8	6
vs. HARJump	14	0	7	9	15	0	19	19	17
vs. FIARMA	1	0	0	0	0	0	0	2	5
vs. FI	0	0	0	0	0	0	2	0	3
vs. FI(0.5)	2	0	1	0	0	0	3	3	0
N	50	50	50	50	50	50	50	50	50

Table 6: In analogy to Table 3, this table illustrates the forecast performance of different additional models for six-month beta forecasts from a rolling estimation window of 100 observations. RW, AR, and ARMA are estimated in a state–space framework. RWW relies on a 12-month daily estimation window, FI(0.5) uses a FI model with d fixed at 0.5, and HARJump denotes a HAR model augmented by jump betas. The first row shows the average RMSEs of different models across all 50 portfolios. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain portfolio. "In MCS" denotes the number of portfolios for which an estimator is in the Hansen et al. (2011) model confidence set. Furthermore, the rows denoted by "vs. X" correspond to modified DM-tests (Harvey et al., 1997), providing the number of times the column-model yields a significantly lower RMSE than the row-model. We examine statistical significance toward the 10 percent level. Finally, N is the number of investigated portfolios.

can point to spurious long memory. It is thus possible that short-memory processes augmented with discontinuous components perform well for forecasting. We refer to this model as HARJump. To estimate the jump betas for each stock, we follow the approach of Bollerslev et al. (2016). We additionally include 3 lags of the jump betas over different historical horizons, similar to the approach for the HAR model of Equation (2).[24]

Table 6 shows the forecast results for these models and for comparison again the results of the FI and FIARMA models considered earlier. In line with the results of Hollstein & Prokopczuk (2016), we find that the performance of the RW model improves when estimated

[24]Following the advice of an anonymous referee, we have also examined the Corsi & Renò (2012) HAR model with jumps and leverage. Our untabulated analysis reveals that this model performs worse than the simpler HARJump model. This is likely because, as opposed to volatility, we are not aware of evidence pointing toward a leverage effect for beta.

within a state–space framework, as it on average now produces more accurate forecasts than the AR and ARMA models. However, the models that account for long-range dependencies still perform substantially better. The "original" RWW model also does not perform better than the FI and FIARMA models.

The HARJump model performs worse rather than better than the plain HAR model. The FI and FIARMA models yield substantially lower average RMSEs. Thus, it seems to be important to use true long-memory processes. Short-memory processes augmented by jump components do not suffice. Finally, it is noteworthy that the results for the FI(0.5) model are only slightly worse than those for the FI model considered earlier. Thus, fixing d at 0.5 instead of using estimates appears to be a practical and well-performing approach for beta forecasting.

5.3 Entire CRSP Dataset

In our main analysis, based on the need to have high-frequency data for liquid instruments, we restrict our dataset to firms with available high-frequency data and start in 1996. In this section, we examine whether the results found for this sample can also be generalized to a broader sample of stocks and for a longer sample period. We extend our dataset to consider the entire CRSP dataset starting from 1926. As intra-day observations are only available from 1996 onward, we calculate betas from daily returns. Since monthly beta estimates based on daily returns are too noisy, we follow Andersen et al. (2006) and consider quarterly estimates instead.[25]

Table 7 shows the estimated order of integration of the series averaged across all stocks for which more than 100 observations are available (N = 3,153). Again we present results when investigating the original series as well as when adjusting for structural breaks.

We find that the average d estimate decreases from 0.56 to 0.38 when considering the

[25]Since the zero-approximation to the risk-free rate becomes less reliable for daily returns, we deviate from the description in Equation (1) by using excess returns to estimate realized betas based on daily data.

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
β_i	0.382	0.157	0.916	0.999	0.331	0.181	0.841	0.999

Table 7: In analogy to Table 1, this table presents average estimates of the memory parameter of realized beta across all stocks ($\bar{\hat{d}}_i$) using the 2ELW estimator of Shimotsu & Phillips (2005) and Shimotsu (2010). The results are for the entire CRSP sample (3,153 stocks) and quarterly betas calculated from daily data. $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d_i = 0$ and $d_i = 1$, respectively, are rejected at the 10 percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle & Moulines (2000).

expanded sample of daily returns. This also holds when only considering the same stocks as in our main analysis, for which the average d estimate is now 0.46, and even when considering the same stocks and same time period as for our main analysis, where the average d is 0.50. Consequently, the observed reduction in d is in part due to the expanded sample with partly less liquid stocks and a longer time period. However, the change in the recording frequency also plays a role. As already discussed in Section 3.1, decreasing the recording frequency increases the level of noise in the realized beta time series. This then leads to a negative bias of the 2ELW estimator, which explains the reduction of the memory estimate.[26] Even though the d estimates appear to be slightly negatively biased, more than 84 percent of the stocks still have a d that is significantly greater than zero. For 99 percent, we can reject the null hypothesis of difference-stationarity.

In Table 8 we present the forecast results. As in the previous subsection RWW is based on one year of daily returns and Welch's (2019) original weighting scheme. In addition, based on the result that d is on average close to 0.4, we also consider a FI(0.4) estimator where we fix the parameter to 0.4 for all stocks instead of estimating it.

The forecast results echo our finding of long memory of betas in the CRSP dataset. We find that the FI(0.4) model performs overall best, closely followed by the FI and FIARMA

[26] In Section A.2.9 of the Online Appendix, we explore this issue further by considering alternative intra-day sampling frequencies of 15 minutes and 75 minutes. There, we already find that increased noise in realized betas derived from 75-minute data biases the d estimates negatively.

	RW	RWW	AR	ARMA	HAR	FIARMA	FI	FI(0.4)
RMSE	0.2168	0.1787	0.1890	0.1829	0.1801	0.1738	0.1738	0.1722
Best	0	8	0	0	0	11	2	29
In MCS	2	48	12	30	39	49	48	50
vs. RW	0	34	25	29	31	35	36	35
vs. RWW	0	0	0	0	1	4	4	7
vs. AR	0	3	0	15	8	30	27	30
vs. ARMA	0	2	0	0	5	22	14	21
vs. HAR	0	2	0	0	0	12	11	12
vs. FIARMA	0	0	0	0	0	0	0	3
vs. FI	0	0	0	0	0	0	0	4
vs. FI(0.4)	0	0	0	0	0	1	0	0
N	50	50	50	50	50	50	50	50

Table 8: In analogy to Table 3, this table illustrates the forecast performance of the models for six-month beta forecasts from a rolling estimation window of 100 observations. The first row shows the average RMSEs of different models across all 50 portfolios. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain portfolio. "In MCS" denotes the number of portfolios for which an estimator is in the Hansen et al. (2011) model confidence set. Furthermore, the rows denoted by "vs. X" correspond to modified DM-tests (Harvey et al., 1997), providing the number of times the column-model yields a significantly lower RMSE than the row-model. We examine statistical significance toward the 10 percent level. Finally, N is the number of investigated portfolios.

models. For 29 out of the 50 portfolios the FI(0.4) estimator yields the lowest average RMSE.[27] The FI(0.4) model is in the model confidence set in every single instance, while the difference-stationary RW model is only for 2 portfolios and the AR and ARMA models only for 12 and 30 of the 50 portfolios, respectively. The HAR model performs reasonably well, being in the model confidence set for 39 of the 50 portfolios.

The RWW model yields an average RMSE that is 3.6 percent higher than that of the FI(0.4) model and is in the model confidence set for 48 of the 50 portfolios. However, the simple Welch (2019) RWW estimator appears to work well for small stocks. An untabulated analysis reveals that for equally weighted portfolios the RWW estimator performs almost as

[27]Indeed, we find that a FI(0.5) estimator performs slightly worse than the FI(0.4) estimator. Thus, even though the d estimates appear to be slightly downward biased, the average underlying parameter seems to be closer to 0.4 than to 0.5.

well as the FI and FI(0.4) models. In addition, for considering long memory, we need to be able to observe a sufficiently long sample of historical returns. Therefore, for newly listed or very young firms it is very difficult to accurately estimate the d parameters and make forecasts based on long-memory models. Thus, for small stocks and those with only a short return history, the RWW estimator appears to be a viable alternative.

5.4 Further Analyses and Robustness Tests

In Section A.2 of the Online Appendix, we present the results of further analyses and robustness tests. We study market-neutral anomaly portfolios in Section A.2.1, the relation between the memory of beta and industries in Section A.2.2, and the determinants of forecast errors in Section A.2.3. An analysis of hedging errors is in Section A.2.4. We consider alternative forecast horizons in Section A.2.5 and an analysis based on individual stocks in Section A.2.6. The results of a cross-sectional analysis are in Section A.2.7. Additionally, we use an alternative long-memory estimator (Section A.2.8), alternative sampling frequencies (Section A.2.9), alternative estimation windows and bandwidths in the long-memory estimation (Section A.2.10), and impose a correction for asynchronous trading (Section A.2.11). None of these robustness tests changes our main conclusions. The results are all qualitatively similar to those presented in the main part of the paper.

6 Conclusion

In this paper, we analyze the memory of beta. We first document that the betas of virtually all stocks exhibit long-memory properties. We further show that accounting for these long-memory properties is very important for forecasting. A pure long-memory FI model outperforms all other short-memory and difference-stationary models. Accounting long memory is also important economically: the FI and FIARMA models perform best in portfolio formation exercises.

We also document that setting the memory parameter to $d = 0.5$ on average yields results that are almost as well as (or even better) when first estimating the d parameters. In Section A.3 of the Online Appendix, we provide a simple R-algorithm for making forecasts with this model.

References

Adrian, T., & Franzoni, F. (2009). Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM. *Journal of Empirical Finance*, *16*(4), 537–556.

Alfarano, S., & Lux, T. (2007). A noise trader model as a generator of apparent financial power laws and long memory. *Macroeconomic Dynamics*, *11*(S1), 80–101.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, *89*(4), 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Wu, J. G. (2005). A framework for exploring the macroeconomic determinants of systematic risk. *American Economic Review: Papers and Proceedings*, *95*(2), 398–404.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Wu, J. G. (2006). Realized beta: Persistence and predictability. In T. B. Fomby, & D. Terrell (Eds.) *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series*, (pp. 1–40). New York: Elsevier.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariant matrix estimation. *Econometrica*, *59*(3), 817–858.

Ang, A., & Chen, J. (2007). CAPM over the long run: 1926–2001. *Journal of Empirical Finance*, *14*(1), 1–40.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, *61*(1), 259–299.

Arteche, J. (2004). Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. *Journal of Econometrics*, *119*(1), 131–154.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, *66*(1), 47–78.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*(1), 1–22.

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*(1), 3–30.

Baillie, R. T., Calonaci, F., Cho, D., & Rho, S. (2019). Long memory, realized volatility and heterogeneous autoregressive models. *Journal of Time Series Analysis*, *40*(4), 609–628.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3–18.

Barber, B. M., Huang, X., & Odean, T. (2016). Which factors matter to investors? Evidence from mutual fund flows. *Review of Financial Studies*, *29*(10), 2600–2642.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal*, *12*(3), 1–32.

Berk, J. B., & Van Binsbergen, J. H. (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics*, *119*(1), 1–23.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance*, *43*(2), 507–528.

Black, A., Fraser, P., & Power, D. (1992). UK unit trust performance 1980–1989: A passive time-varying approach. *Journal of Banking & Finance*, *16*(5), 1015–1033.

Blume, M. E. (1971). On the assessment of risk. *Journal of Finance*, *26*(1), 1–10.

Boehme, R. D., Danielsen, B. R., & Sorescu, S. M. (2006). Short-sale constraints, differences of opinion, and overvaluation. *Journal of Financial and Quantitative Analysis*, *41*(2), 455–487.

Bollerslev, T., Li, S. Z., & Todorov, V. (2016). Roughing up beta: Continuous versus discontinuous betas and the cross-section of expected stock returns. *Journal of Financial Economics*, *120*(3), 464–490.

Bollerslev, T., & Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, *73*(1), 151–184.

Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2020). Realized semibetas: Signs of things to come. *Duke University Working Paper*.

Boyer, B., Mitton, T., & Vorkink, K. (2009). Expected idiosyncratic skewness. *Review of Financial Studies*, *23*(1), 169–202.

Brennan, M. J., & Zhang, Y. (2020). Capital asset pricing with a stochastic horizon. *Journal of Financial and Quantitative Analysis*, *55*(3), 783–827.

Buss, A., & Vilkov, G. (2012). Measuring equity risk with option-implied correlations. *Review of Financial Studies*, *25*(10), 3113–3140.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, *90*(3), 414–427.

Chan, N. H., & Palma, W. (1998). State space modeling of long-memory processes. *Annals of Statistics*, *26*(2), 719–740.

Chang, B.-Y., Christoffersen, P., Jacobs, K., & Vainberg, G. (2012). Option-implied measures of equity risk. *Review of Finance*, *16*(2), 385–428.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Corsi, F., & Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, *30*(3), 368–380.

Daniel, K., Mota, L., Rottke, S., & Santos, T. (2020). The cross-section of risk and return. *Review of Financial Studies*, *33*(5), 1927–1979.

Daniel, K., & Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance*, *52*(1), 1–33.

Deo, R. S., & Hurvich, C. M. (2001). On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory*, *17*(4), 686–710.

Diebold, F. X., & Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, *105*(1), 131–159.

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, *7*(2), 197–226.

Ding, Z., & Granger, C. W. (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, *73*(1), 185–215.

Dissanayake, G., Peiris, M., & Proietti, T. (2016). State space modeling of Gegenbauer processes with long memory. *Computational Statistics & Data Analysis*, *100*, 115–130.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, *74*(366a), 291–298.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607–636.

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25.

Frederiksen, P., Nielsen, F. S., & Nielsen, M. Ø. (2012). Local polynomial Whittle estimation of perturbed fractional processes. *Journal of Econometrics*, *167*(2), 426–447.

Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, *4*(4), 221–238.

Graham, J. R., & Harvey, C. R. (2001). The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics*, *60*(2), 187–243.

Granger, C. W., & Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, *11*(3), 399–421.

Granger, C. W., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, *1*(1), 15–29.

Grundy, B. D., & Martin, J. S. M. (2001). Understanding the nature of the risks and the source of the rewards to momentum investing. *Review of Financial Studies*, *14*(1), 29–78.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281–291.

Hassler, U., & Pohle, M.-O. (2019). Forecasting under long memory and nonstationarity. *Goethe University Working Paper*.

Hollstein, F. (2020). Estimating beta: The international evidence. *Journal of Banking and Finance*, *121*, 105968.

Hollstein, F., & Prokopczuk, M. (2016). Estimating beta. *Journal of Financial and Quantitative Analysis*, *51*(4), 1437–1466.

Hollstein, F., Prokopczuk, M., & Wese Simen, C. (2019). Estimating beta: Forecast adjustments and the impact of stock characteristics for a broad cross-section. *Journal of Financial Markets*, *44*, 91–118.

Hollstein, F., Prokopczuk, M., & Wese Simen, C. (2020). The conditional Capital Asset Pricing Model revisited: Evidence from high-frequency betas. *Management Science*, *66*(6), 2474–2494.

Hosking, J. (1981). Fractional differencing. *Biometrika*, *68*(1), 165–176.

Hou, J., & Perron, P. (2014). Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations. *Journal of Econometrics*, *182*(2), 309–328.

Hurvich, C. M., Moulines, E., & Soulier, P. (2005). Estimating long memory in volatility. *Econometrica*, *73*(4), 1283–1328.

Iacone, F. (2010). Local Whittle estimation of the memory parameter in presence of deterministic components. *Journal of Time Series Analysis*, *31*(1), 37–49.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, *48*(1), 65–91.

Kamara, A., Korajczyk, R. A., Lou, X., & Sadka, R. (2016). Horizon pricing. *Journal of Financial and Quantitative Analysis*, *51*(6), 1769–1793.

Lavielle, M., & Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, *21*(1), 33–59.

LeBaron, B. (2001). Stochastic volatility as a simple generator of apparent financial power laws and long memory. *Quantitative Finance*, *1*(6), 621–631.

LeBaron, B. (2006). Agent-based financial markets: Matching stylized facts with style. In D. Colander (Ed.) *Post Walrasian Macroeconomics: Beyond the DSGE Model*, (pp. 221–235). Cambridge: Cambridge University Press.

Leschinski, C. (2017). On the memory of products of long range dependent time series. *Economics Letters*, *153*, 72–76.

Levi, Y., & Welch, I. (2017). Best practice for cost-of-capital estimates. *Journal of Financial and Quantitative Analysis*, *52*(2), 427–463.

Lewellen, J., & Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, *82*(2), 289–314.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, *47*(1), 13–37.

Liu, J., Stambaugh, R. F., & Yuan, Y. (2018). Absolving beta of volatility's effects. *Journal of Financial Economics*, *128*(1), 1–15.

Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In J. Mincer (Ed.) *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, (pp. 1–46). Cambridge, MA: Elsevier.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, *34*(4), 768–783.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Pictet, O. V., Olsen, R. B., & Ward, J. R. (1993). Fractals and intrinsic time: A challenge to econometricians. *39th International AEA Conference on Real Time Econometrics, Luxembourg*.

Pagan, A. (1980). Some identification and estimation results for regression models with stochastically varying coefficients. *Journal of Econometrics*, *13*(3), 341–363.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Patton, A. J., & Verardo, M. (2012). Does beta move with news? Firm-specific information flows and learning about profitability. *Review of Financial Studies*, *25*(9), 2789–2839.

Robinson, P. M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association*, *89*(428), 1420–1437.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, *23*(5), 1630–1661.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, *13*(3), 341–360.

Scholes, M., & Williams, J. (1977). Estimating betas from nonsynchronous data. *Journal of Financial Economics*, *5*(3), 309–327.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(3), 425–442.

Shimotsu, K. (2010). Exact local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory*, *26*(2), 501–540.

Shimotsu, K., & Phillips, P. C. (2005). Exact local Whittle estimation of fractional integration. *Annals of Statistics*, *33*(4), 1890–1933.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, vol. 26 of *Monographs on Statistics and Applied Probability*. London: Chapman & Hall.

Sun, Y., & Phillips, P. C. (2003). Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics*, *115*(2), 355–389.

Vasicek, O. A. (1973). A note on using cross-sectional information in Bayesian estimation of security betas. *Journal of Finance*, *28*(5), 1233–1239.

Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, *20*(1), 87–127.

Welch, I. (2019). Simpler better market betas. *NBER Working Paper*.

Zhang, X. (2006). Information uncertainty and stock returns. *Journal of Finance*, *61*(1), 105–137.

The Memory of Beta

Online Appendix

JEL classification: G12, C58, G11

Keywords: Long memory, beta, persistence, forecasting, predictability

A.1 Simulation Study

To investigate the performance of different approaches for estimating the memory parameter d in small samples, we perform a small simulation study. For this purpose we simulate data according to

$$(1 - B)^d y_t = \epsilon_t,$$

where $\epsilon \sim N(0, 1)$. To account for the high persistence in the series we consider a burn-in period of 250 observations.

We then infer on the order of integration of the series using various approaches. These include the two-step exact local Whittle estimator by Shimotsu (2010) (2ELW) as considered in this paper, the log periodogram estimator by Geweke & Porter-Hudak (1983) (GPH) as considered in Section A.2.8, the structural break robust estimators by Iacone (2010) (trLW) and Hou & Perron (2014) (HP), and the noise robust estimators by Hurvich et al. (2005) (LWN) and Frederiksen et al. (2012) (LPWN). Additionally, we consider the approach by Andersen et al. (2006) to infer on the order of integration. They investigate the autocorrelation function of the beta series and perform Ljung–Box tests on the residuals when estimating an AR(p) model to the realized beta series where p is determined by means of the AIC.

Table A.1 reports results for $d = 0.2, 0.4, 0.6$ and $T = 100, 148, 228, 1000$ averaged across 1000 repetitions. The table reveals that the 2ELW and GPH estimators are almost unbiased, also for a small sample of size $T = 100$. We further find that the variance of the 2ELW estimator is smaller than that of the GPH estimator, which is in line with the results presented in Section A.2.8. Concerning the break robust estimators, it can be seen that both the HP estimator is negatively biased and the trLW estimator is positively biased in sample sizes of 100, 148, and 228. The noise robust estimators, on the other hand, are positively biased for sample sizes of 100, 148, and 228.

Andersen et al. (2006) investigate quarterly betas for which, due to the noise, the observed

	$d = 0.2$				$d = 0.4$				$d = 0.6$			
	$T = 100$	$T = 148$	$T = 228$	$T = 1000$	$T = 100$	$T = 148$	$T = 228$	$T = 1000$	$T = 100$	$T = 148$	$T = 228$	$T = 1000$
$\bar{\hat{d}}_{2ELW}$	0.22	0.21	0.21	0.20	0.43	0.42	0.41	0.40	0.62	0.63	0.62	0.61
$sd(\hat{d}_{2ELW})$	0.13	0.11	0.09	0.05	0.14	0.11	0.09	0.05	0.13	0.11	0.09	0.05
$\bar{\hat{d}}_{GPH}$	0.20	0.20	0.20	0.20	0.41	0.41	0.41	0.40	0.61	0.63	0.62	0.61
$sd(\hat{d}_{GPH})$	0.16	0.13	0.11	0.06	0.17	0.13	0.12	0.06	0.16	0.13	0.11	0.07
$\bar{\hat{d}}_{HP}$	0.11	0.12	0.16	0.19	0.27	0.30	0.34	0.38	0.38	0.45	0.48	0.57
$sd(\hat{d}_{HP})$	0.19	0.16	0.12	0.06	0.24	0.20	0.15	0.06	0.33	0.28	0.24	0.08
$\bar{\hat{d}}_{trLW}$	0.30	0.19	0.14	0.17	0.50	0.37	0.31	0.35	0.66	0.56	0.53	0.55
$sd(\hat{d}_{trLW})$	0.42	0.34	0.27	0.11	0.42	0.34	0.26	0.11	0.43	0.33	0.27	0.11
$\bar{\hat{d}}_{LWN}$	0.36	0.36	0.32	0.25	0.50	0.49	0.46	0.43	0.67	0.67	0.65	0.63
$sd(\hat{d}_{LWN})$	0.29	0.26	0.21	0.09	0.21	0.17	0.13	0.06	0.17	0.14	0.11	0.06
$\bar{\hat{d}}_{LPWN}$	0.40	0.39	0.36	0.28	0.53	0.52	0.50	0.45	0.68	0.69	0.68	0.65
$sd(\hat{d}_{LPWN})$	0.36	0.33	0.30	0.16	0.30	0.26	0.21	0.09	0.24	0.19	0.15	0.08
Sign. ac (%)	4.08	6.36	9.97	31.01	14.19	23.24	36.71	88.59	27.32	45.35	65.65	99.43
Ljung–Box	0.006	0.008	0.003	0.000	0.001	0.005	0.000	0.000	0.003	0.007	0.002	0.000

Table A.1: We simulate T observations of fractional white noise that is integrated of order $I(d)$ and then compare different approaches to infer on the memory parameter d. This table reports the average d estimate and standard deviation ($sd()$) for the estimators by Shimotsu (2010) (2ELW), Geweke & Porter-Hudak (1983) (GPH), Hou & Perron (2014) (HP), Iacone (2010) (trLW), Hurvich et al. (2005) (LWN), and Frederiksen et al. (2012) (LPWN). Additionally, the table shows the average percentage of the first 36 autocorrelations that are indicated to be significantly larger than zero by 95 percent Bartlett confidence intervals. This is the technique Andersen et al. (2006) use to decide on the order of integration of the series. They further consider Ljung–Box tests on the residuals of AR(p) processes, where p is selected by the AIC. In case there is significant autocorrelation in the residuals, the null is rejected, indicating that there is long memory in the series. The last row reports the power of this approach for the simulated series, i.e., the relative number of times the null hypothesis is rejected. All results are the averages over 1000 repetitions.

order of integration is decreased, such that the 2ELW estimator yields a d of 0.4 on average. They then fractionally difference the series by 0.2, such that the resulting series should be approximately I(0.2). For such a series, the simulations indicate that only 6 percent of the first 36 autocorrelations are significantly greater than zero according to 95 percent Bartlett confidence intervals. It is understandable that, based on such autocorrelation functions, the authors conclude that realized betas exhibit a d of 0.2 or smaller. The simulations further reveal that Ljung–Box tests on the residuals of an $AR(p)$ with p selected by the AIC are not particularly useful for detecting long-memory time series. The order p is simply chosen to be high, such that the long-memory characteristics can be captured by the AR model.

A.2 Further Analyses and Robustness Tests

A.2.1 Market-Neutral Anomaly Portfolios

An alternative economic criterion on which to assess different beta estimation approaches that is important is a model's performance in portfolio formation. Therefore, we also evaluate the estimators based on their ability to create ex-post market-neutral anomaly portfolios (Hollstein et al., 2019).

We generate anomaly long–short portfolios by sorting the stocks each month based on NYSE breakpoints. We follow Lewellen & Nagel (2006) and build 25 independently sorted size/value portfolios. The SMB portfolio is the difference between the average returns of the 5 low-market-cap portfolios and those of the 5 high-market-cap portfolios. Similarly, the HML portfolio is the difference between the average returns of the 5 high-book-to-market portfolios and those of the 5 low-book-to-market portfolios. For momentum, we sort the stocks into 10 portfolios based on their return over the past 12 months while skipping the most recent month (Jegadeesh & Titman, 1993). WML denotes the difference between the winner (high-momentum) and loser (low-momentum) decile portfolio returns. For beta, idiosyncratic volatility, and leverage we sort the stocks into 5 portfolios and form high-minus-low (5–1) portfolios. As before, we sort beta portfolios based on the estimate during the second-to-last nonoverlapping beta estimation window. Detailed definitions of the other sorting variables are in Appendix B.

For each beta estimator, we use the forecasts to compute the long and short portfolio beta predictions. We set the weight $\upsilon_{i,T+h}$ so that it fulfills the condition $\hat{\beta}^{\text{long}}_{i,T+h} - \upsilon_{i,T+h}\hat{\beta}^{\text{short}}_{i,T+h} = 0$.[1] We thus create anomaly portfolios that are ex-ante market-neutral. Then we test whether the ex-post realized beta of the anomaly portfolios is indeed 0 on average.

We present the results in Table A.2. Our main finding is that accounting for the long-

[1] The results are qualitatively similar when keeping the weight of the short side at 1 and instead weighting the long side to make the portfolios market-neutral.

	RW	*RWW*	*AR*	*ARMA*	*HAR*	*FIARMA*	*FI*
SMB	−0.003	0.053***	0.069**	0.043	−0.036	0.042	0.048
(s.e.)	(0.021)	(0.019)	(0.031)	(0.027)	(0.033)	(0.031)	(0.029)
HML	0.002	−0.001	0.038	0.029	0.003	0.017	0.019
(s.e.)	(0.012)	(0.015)	(0.032)	(0.026)	(0.027)	(0.023)	(0.023)
WML	−0.017	0.025	0.105	0.087	0.072	0.072	0.071
(s.e.)	(0.046)	(0.037)	(0.071)	(0.067)	(0.057)	(0.052)	(0.051)
Beta	0.059***	0.023	−0.025	0.003	0.044**	0.033*	0.029
(s.e.)	(0.016)	(0.015)	(0.024)	(0.022)	(0.019)	(0.019)	(0.018)
iVol	0.012	−0.034**	−0.065**	−0.037	0.031	−0.011	−0.013
(s.e.)	(0.015)	(0.015)	(0.028)	(0.025)	(0.021)	(0.020)	(0.021)
Lev	0.000	0.020	0.048**	0.040*	0.011	0.023	0.026
(s.e.)	(0.014)	(0.013)	(0.023)	(0.024)	(0.030)	(0.025)	(0.024)

Table A.2: Market-Neutral Anomaly Portfolios: This table presents average ex-post 6-month realized betas of ex-ante beta-neutral anomaly portfolios based on the different beta forecasts. We generate anomaly long–short portfolios by sorting the stocks each month based on NYSE breakpoints. We build 25 independently sorted size–value portfolios. The SMB portfolio is the difference between the average returns of the 5 low-market-cap portfolios and those of the 5 high-market-cap portfolios. Similarly, the HML portfolio is the difference between the average returns of the 5 high-book-to-market portfolios and those of the 5 low-book-to-market portfolios. For momentum, we sort the stocks into 10 portfolios based on their return over the past 12 months while skipping the most recent month (Jegadeesh & Titman, 1993). WML denotes the difference between the winner (high-momentum) and loser (low-momentum) decile portfolio returns. For beta, idiosyncratic volatility, and leverage we sort the stocks into 5 portfolios and form high-minus-low (5–1) portfolios. Beta portfolios are sorted on the realized beta 2 months prior to the current month. We make the anomaly portfolios ex-ante beta-neutral by solving the condition $\hat{\beta}^{\text{long}}_{i,T+h} - \upsilon_{i,T+h}\hat{\beta}^{\text{short}}_{i,T+h} = 0$ and applying the resulting weight $\upsilon_{i,t}$ to the long side of the anomaly. In parentheses, we present the robust Andrews (1991) standard errors, using a quadratic spectral density and data-driven bandwidth selection. *, **, and *** indicate significance at the 10 percent, 5 percent, and 1 percent level, respectively.

memory property of beta is economically important in an anomaly portfolio setting. The FI model performs best. It yields ex-post market-neutral portfolios for all anomaly variables. The FIARMA model performs somewhat worse, leaving a significant exposure (at 10 percent) for beta-sorted portfolios. The RW model, too, fails for beta portfolios. The RWW model fails for size and idiosyncratic volatility portfolios. The short-memory models, on the other hand, collectively leave significant market exposures for portfolios sorted on size, idiosyncratic volatility, and leverage. Finally, the HAR model performs reasonably well, failing only for portfolios sorted by beta.

A.2.2 The Memory of Beta and Industries

We continue the empirical analysis by examining to what extent the memory in beta relates to different industries. For some industries, the business models, and with that the constituent firms' systematic risk, may be more persistent, while others experience more frequent changes.

	$\bar{\hat{d}}$	$t-$stat
Durables	0.5509	-0.19
Energy	0.5951	2.81
Healthcare	0.5042	-2.99
HiTec Equipment	0.5014	-4.59
Manufacturing	0.5945	4.31
NonDurables	0.5153	-2.67
Other	0.5697	2.02
Telephone	0.5047	-2.35
Utilities	0.5808	2.00
Wholesale	0.5223	-3.08

Table A.3: This table shows the average estimate of d in each industry. The t-stat corresponds to t-statistics testing the null that the average d of the industry equals the average across all industries. Standard errors are calculated with the heteroskedasticity and autocorrelation robust approach by Andrews (1991), using a quadratic spectral density and data-driven bandwidth selection. The names of the industries for which the average d is significantly higher or lower than this value at the 10 percent level are printed in **bold**.

We present the results in Table A.3. Stocks in the Energy and Manufacturing industries have on average the highest ds. Thus, these traditional industries tend to have higher persistence in their systematic risk than many others. For the Healthcare, HiTec Equipment, NonDurables, and Telephone industries, the opposite holds true. These industries have in part been particularly prone to disruptions and creative destruction during the past two decades. Thus, many of these firms and/or their market environment have experienced substantial changes, and past shocks to their systematic risk die out more quickly.

A.2.3 The Determinants of Forecast Errors

Having documented that accounting for long memory in betas substantially improves the forecasts, we next analyze for *which* stocks one makes the biggest mistakes when using short-memory or difference-stationary processes. To that end, we regress the difference in absolute forecast errors on different firm characteristics. In more detail, we perform the following regressions

$$abs(\hat{\beta}^{RW}_{i,T+h} - \beta_{i,T+h}) - abs(\hat{\beta}^{FI}_{i,T+h} - \beta_{i,T+h}) = a + bx_{i,T} + e_{i,T+h},$$
$$abs(\hat{\beta}^{ARMA}_{i,T+h} - \beta_{i,T+h}) - abs(\hat{\beta}^{FI}_{i,T+h} - \beta_{i,T+h}) = a + bx_{i,T} + e_{i,T+h}.$$

Here, $\hat{\beta}_{i,T+h}$ are the forecasts made by the RW, ARMA, and FI models as presented in Section 4.2 and $x_{i,T}$ contains the set of explanatory variables observed at time T.

We present the result for the forecast error differential between the RW model and the FI model in Table A.4 and that between the ARMA(p,q) model and the FI model in Table A.5.

In Table A.4, we start with the errors made when inadequately imposing a difference-stationary RW model. First, we observe an economically large and statistically highly significant intercept term. This echoes our previous finding that the FI model yields substantially lower forecast errors on average than the RW model. Second, consistent with what one would intuitively expect, the slope coefficient on d is highly significantly negative. Thus, the higher the memory in betas, the less inadequate becomes the RW assumption. However, a one-standard-deviation increase in d from its average, while keeping all else equal, reduces the average forecast error differential (implied by the intercept term) by only one third.

Momentum, the bid–ask spread, idiosyncratic volatility, short interest, and leverage have positive effects on the forecast error differential. It is well known that the betas of stocks with

	coef	se	t-stat	p-value
Intercept	0.0482	0.0014	35.6091	0.0000
$\boldsymbol{d}$	−0.0176	0.0011	−15.9259	0.0000
β	−0.0017	0.0011	−1.4521	0.1700
log(Market Cap)	−0.0049	0.0011	−4.6082	0.0000
BtM	0.0020	0.0014	1.3682	0.2690
Investment	−0.0027	0.0007	−3.8931	0.0000
Profitability	−0.0025	0.0008	−3.1070	0.0030
Momentum	0.0074	0.0009	8.3117	0.0000
BAS	0.0043	0.0013	3.3480	0.0010
Turnover	0.0023	0.0019	1.2402	0.3370
iVol	0.0032	0.0011	2.8463	0.0060
iSkew	0.0002	0.0005	0.4083	0.6560
Short Interest	0.0033	0.0012	2.6895	0.0020
Leverage	0.0021	0.0009	2.4059	0.0170
Age	0.0011	0.0007	1.5973	0.1060
Durables	0.0048	0.0039	1.2508	0.1910
Energy	0.0049	0.0033	1.4868	0.1360
Healthcare	0.0015	0.0032	0.4606	0.6560
HiTec Equipment	0.0016	0.0030	0.5405	0.5800
Manufacturing	−0.0018	0.0021	−0.8509	0.3880
NonDurables	−0.0036	0.0026	−1.3917	0.1960
Telephone	−0.0013	0.0045	−0.2821	0.8110
Utilities	−0.0087	0.0020	−4.3542	0.0000
Wholesale	0.0048	0.0024	1.9582	0.0510

Table A.4: In this table, we run regressions of the difference in absolute forecast errors from the RW and FI models on different firm characteristics. Firm characteristics (except for the dummy variables) are winsorized (at the 1 and 99 percent levels) and standardized to have zero mean and a volatility of one. The standard errors (se) are bootstrapped using the procedure of Cameron et al. (2008). t-stat and p-value denote the corresponding t-statistics and p-values, respectively. The names of the characteristics, which yield a statistically significant regression coefficient (coef) at 10 percent, are printed in **bold**.

extreme momentum are highly time-varying (Grundy & Martin, 2001). Similarly, firms whose stocks exhibit very high short interest are also prone to substantial changes in systematic risk. For these stocks, in particular, it is therefore advisable to rely on the long-range dependencies when making forecasts. On the other hand, the stocks' size, investment, and profitability have a negative impact on the loss differential. In particular the betas of the smaller and unprofitable stocks that invest little should therefore be predicted with long-memory models rather than the random walk.

In Table A.5, we analyze the determinants of the ARMA and FI forecast error differentials. Consistent with our previous results, we also detect a strongly statistically significant intercept term of 0.0117. This intercept term is substantially smaller than that for the RW–FI forecast error differential. The bid–ask spread, short interest, leverage, and age all have a significant negative impact on the forecast error differential. On the other hand, turnover has a positive impact. Illiquid firms that are not old and use high leverage might be more prone to short-run changes in betas. Thus, the short-memory models perform a little less poorly for these.

A.2.4 Hedging Errors

To account for the possibility that the ex-post realized betas are measured with error, we follow Liu et al. (2018) and examine the out-of-sample hedging errors of the main approaches.[2] We compute the hedging error for each stock as

$$H_{i,T+1} \quad = \quad (r_{i,T+1} - r_{f,T+1}) - \hat{\beta}_{i,T+h}(r_{M,T+1} - r_{f,T+1}).$$

$r_{i,T+1}$ is the return of stock i in month $T+1$. $r_{f,T+1}$ and $r_{M,T+1}$ are the risk-free rate and the return on the market portfolio over the same horizon. $\hat{\beta}_{i,T+1}$ is the forecast for beta using data up to month T. Liu et al. (2018) show that under certain assumptions the hedging error variance ratio $\frac{\mathrm{var}(H_{i,T+1})}{\mathrm{var}(r_{M,T+1}-r_{f,T+1})}$ is approximately equal to the mean squared error relative to

[2] Although set up as hedging exercise like that in Section A.2.1, the analysis is more closely related to the statistical analysis of Section 4.2 of the main paper. This is because the hedging error variance ratio presented below is a direct function of the MSE.

	coef	se	t-stat	p-value
Intercept	0.0117	0.0012	9.8194	0.0000
d	−0.0008	0.0006	−1.3609	0.1560
β	−0.0001	0.0008	−0.1206	0.9090
log(Market Cap)	0.0005	0.0009	0.6190	0.5130
BtM	−0.0001	0.0006	−0.1912	0.8730
Investment	0.0003	0.0006	0.5803	0.5530
Profitability	0.0003	0.0007	0.3855	0.7390
Momentum	−0.0006	0.0005	−1.1634	0.2270
BAS	−0.0021	0.0008	−2.5773	0.0080
Turnover	0.0051	0.0010	5.2143	0.0000
iVol	0.0010	0.0008	1.2100	0.2140
iSkew	0.0003	0.0002	1.1294	0.2620
Short Interest	−0.0018	0.0006	−2.9454	0.0040
Leverage	−0.0035	0.0007	−4.9695	0.0000
Age	−0.0013	0.0006	−2.1408	0.0360
Durables	−0.0012	0.0035	−0.3602	0.6990
Energy	0.0080	0.0033	2.4115	0.0160
Healthcare	−0.0055	0.0023	−2.4314	0.0200
HiTec Equipment	0.0045	0.0028	1.6271	0.0920
Manufacturing	−0.0023	0.0021	−1.0853	0.2680
NonDurables	−0.0049	0.0018	−2.7636	0.0080
Telephone	0.0015	0.0034	0.4405	0.6360
Utilities	−0.0040	0.0017	−2.2906	0.0290
Wholesale	−0.0029	0.0021	−1.3864	0.1770

Table A.5: In this table, we run regressions of the difference in absolute forecast errors from the ARMA and FI models on different firm characteristics. Firm characteristics (except for the dummy variables) are winsorized (at the 1 and 99 percent levels) and standardized to have zero mean and a volatility of one. The standard errors (se) are bootstrapped using the procedure of Cameron et al. (2008). t-stat and p-value denote the corresponding t-statistics and p-values, respectively. The names of the characteristics, which yield a statistically significant regression coefficient (coef) at 10 percent, are printed in **bold**.

the *true* realized beta plus a term that is constant for all beta forecasts.[3] We follow Liu et al. (2018) and estimate the variance ratios using rolling 5-year windows to account for

[3]Because of these constant terms, one can strictly not interpret the levels of the hedging errors, but only their differences across models.

	RW	*RWW*	*AR*	*ARMA*	*HAR*	*FIARMA*	*FI*
Mean	4.8277	4.8153	4.8366	4.8168	4.7990	4.7983	4.7987
ΔRW		−0.0123	0.0090	−0.0108	−0.0287***	−0.0294***	−0.0290***
		(−1.5149)	(1.0002)	(−1.3140)	(−3.1339)	(−3.9139)	(−4.1694)
ΔRWW	0.0123		0.0213***	0.0015	−0.0164	−0.0170***	−0.0167***
	(1.5149)		(3.3871)	(0.2550)	(−1.5986)	(−3.0731)	(−3.0949)
ΔARMA	0.0108	−0.0015	0.0198***		−0.0179**	−0.0185***	−0.0182***
	(1.3140)	(−0.2550)	(5.9062)		(−2.1839)	(−5.7269)	(−5.0750)

Table A.6: This table presents the ratio of hedging error variances to the market variance for different approaches. For each stock, estimator, and month, we obtain the hedging error over the next month as $(r_{i,T+1} - r_{f,T+1}) - \hat{\beta}_{i,T+h}(r_{M,T+1} - r_{f,T+1})$. We estimate the hedging error and market variances using rolling 5-year windows and use the average ratio over time. The table presents the average ratio of the hedging error variance to the market variance across all stocks. Additionally, ΔRW and ΔARMA report the differences between the hedging errors of RW and ARMA, respectively, and the other models. In parentheses, we present the robust Andrews (1991) t-statistics, using a quadratic spectral density and data-driven bandwidth selection, of a test for equal average hedging errors. *, **, and *** indicate significance at the 10 percent, 5 percent, and 1 percent level, respectively.

the possibility that the variances in the numerator and denominator change over time. We report the average ratio over time.

We present the results in Table A.6. These are consistent with our previous findings relying on the RMSE. The average hedging errors of the FI and FIARMA model forecasts are lowest. The differences in hedging errors are highly statistically significant compared to the difference-stationary RW, RWW, and the short-memory ARMA models.[4]

A.2.5 Alternative Forecast Horizons

There are various different applications for which investors and company managers need beta estimates. While a six-month forecast horizon appears plausible for both investors and

[4] Assuming that the MSE of the ARMA approach was 0.09 (which is approximately equivalent to a RMSE of 0.3), the difference of −0.0182 between the hedging errors for ARMA and FI reported would correspond to a decrease of $1 - 0.0718/0.09 = 20.22\%$ when moving from ARMA to FI. Assuming that the MSE is 0.16 (RMSE$\approx$ 0.4), the difference would still be 11.39%. Thus, we believe that the results are of economic relevance.

	RW	RWW	AR	ARMA	HAR	FIARMA	FI
Panel A: One-Month Forecast Horizon							
RMSE	0.1527	0.1409	0.1458	0.1432	0.1440	0.1404	0.1363
Best	1	9	1	0	4	14	21
In MCS	25	40	41	45	48	50	50
vs. RW	0	5	9	11	9	16	16
vs. RWW	0	0	4	4	1	8	11
vs. AR	0	0	0	8	3	12	12
vs. ARMA	0	1	0	0	1	7	5
vs. HAR	0	0	2	3	0	6	8
vs. FIARMA	0	1	1	0	0	0	6
vs. FI	0	0	1	0	0	0	0
N	50	50	50	50	50	50	50
Panel B: Three-Month Forecast Horizon							
RMSE	0.1440	0.1283	0.1345	0.1282	0.1303	0.1241	0.1200
Best	1	3	0	0	3	24	19
In MCS	22	42	33	43	46	50	50
vs. RW	0	14	5	10	10	19	22
vs. RWW	0	0	2	3	1	9	9
vs. AR	0	1	0	22	10	21	24
vs. ARMA	0	0	0	0	1	4	8
vs. HAR	0	0	0	0	0	10	7
vs. FIARMA	0	0	0	0	0	0	2
vs. FI	0	0	0	0	0	0	0
N	50	50	50	50	50	50	50

to be continued on the next page

Table A.7: In analogy to Table 3, this table illustrates the forecast performance of the models for one-, three-, and twelve-month beta forecasts from a rolling estimation window of 100 observations. The first row shows the average RMSEs of different models across all 50 portfolios. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain portfolio. "In MCS" denotes the number of portfolios for which an estimator is in the Hansen et al. (2011) model confidence set. Furthermore, the rows denoted by "vs. X" correspond to modified DM-tests (Harvey et al., 1997), providing the number of times the column-model yields a significantly lower RMSE than the row-model. We examine statistical significance toward the 10 percent level. Finally, N is the number of investigated portfolios.

Table A.7 (continued):

	RW	RWW	AR	ARMA	HAR	FIARMA	FI
Panel C: Twelve-Month Forecast Horizon							
RMSE	0.1593	0.1355	0.1559	0.1386	0.1455	0.1300	0.1266
Best	0	7	0	0	1	27	15
In MCS	34	48	28	46	50	50	50
vs. RW	0	17	2	7	6	22	21
vs. RWW	0	0	0	1	0	9	6
vs. AR	2	2	0	25	15	25	27
vs. ARMA	1	1	0	0	4	22	24
vs. HAR	0	0	0	2	0	8	7
vs. FIARMA	0	0	0	0	0	0	1
vs. FI	0	0	0	0	0	3	0
N	50	50	50	50	50	50	50

company managers to use, for some applications they may plan over shorter or even longer periods. Thus, they also need forecasts over alternative horizons. Therefore, in this section, we also consider forecasts for one-month, three-month, and twelve-month horizons.

Table A.7 presents the results. We find that the FI and FIARMA models perform best independently of the forecast horizon. They yield the lowest average RMSEs, deliver the best forecasts for at least two thirds of the portfolios, and are in the model confidence set for every single portfolio for all forecast horizons. To summarize, using models that account for long-run dependencies, instead of short-memory or difference-stationary alternatives, does not only improve six-month forecasts but also those for shorter and longer horizons reaching from one month to one year.

A.2.6 Individual Stocks

Next, we also present the results when directly using individual stocks instead of building portfolios. We show these both for the six-month and the one-month horizon. The results are in Table A.8. Consistent with our main results, the FIARMA and FI models also perform

	RW	RWW	AR	ARMA	HAR	FIARMA	FI
Panel A: Six-Month Forecast Horizon (Stocks)							
RMSE	0.3002	0.2606	0.2631	0.2450	0.2569	0.2311	0.2314
Best	4	105	52	197	122	459	228
In MCS	450	773	987	1117	1131	1160	1154
vs. RW	0	510	417	627	533	776	771
vs. RWW	1	0	253	430	362	588	596
vs. AR	3	19	0	230	90	302	302
vs. ARMA	1	4	8	0	27	115	117
vs. HAR	1	10	10	45	0	80	60
vs. FIARMA	0	1	6	20	11	0	6
vs. FI	0	0	13	36	14	44	0
N	1167	1167	1167	1167	1167	1167	1167
Panel B: One-Month Forecast Horizon (Stocks)							
RMSE	0.3251	0.3026	0.3006	0.2944	0.2949	0.2883	0.2870
Best	1	112	42	195	132	412	324
In MCS	463	931	1020	1157	1193	1208	1201
vs. RW	0	346	320	393	409	591	702
vs. RWW	3	0	116	176	170	269	316
vs. AR	4	41	0	190	127	311	284
vs. ARMA	0	19	14	0	62	150	145
vs. HAR	6	11	14	53	0	85	77
vs. FIARMA	1	3	3	25	12	0	20
vs. FI	0	5	9	36	15	16	0
N	1218	1218	1218	1218	1218	1218	1218

Table A.8: In analogy to Table 3, this table illustrates the forecast performance of the models for individual stocks and six-, and one-month beta forecasts from a rolling estimation window of 100 observations. The first row shows the average RMSEs of different models across all stocks. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain stock. "In MCS" denotes the number of stocks for which an estimator is in the Hansen et al. (2011) model confidence set. Furthermore, the rows denoted by "vs. X" correspond to modified DM-tests (Harvey et al., 1997), providing the number of times the column-model yields a significantly lower RMSE than the row-model. We examine statistical significance toward the 10 percent level. Finally, N is the number of investigated stocks.

best on a disaggregated basis.

The average RMSEs are largest for individual stocks. For the FI model, we observe a reduction from 0.29 to 0.23 when moving from a six-month to a one-month window for individual stocks. When building 50 portfolios instead, the average RMSEs decrease to 0.14 for the one-month horizon and 0.12 for the six-month horizon. Thus, building portfolios appears to diversify measurement errors and reduce errors-in-variables. However, in any case they seem to be larger for the one-month than for the six-month horizon.

A.2.7 Cross-Sectional Results

We also repeat the analysis in the cross-section. That is, we examine whether an estimator significantly outperforms another in predicting the cross-section of future betas of individual stocks. We present the results in Table A.9. We find that the cross-sectional results are similar to those in the time series. The FI and ARFIMA models yield the lowest average RMSEs and are in the model confidence set for the vast majority of periods.

A.2.8 Alternative Long-Memory Estimator

We base our main analysis on the 2ELW estimator, as we believe it is the most suitable estimator in our setup. A popular alternative is the log periodogram estimator by Geweke & Porter-Hudak (1983). Although the variance of log periodogram-based approaches commonly exceeds that of local Whittle-based approaches, they are often considered due to their simplicity in application and calculation.

Table A.10 shows the average estimate of d when using the log periodogram estimator. While the average estimates of d are almost equal, the relative number of stocks for which d is significantly different from 0 and 1 decreases slightly due to the higher variance of the estimates. However, still more than 90 percent of the stocks exhibit significant long memory in beta. Thus, the results of the log periodogram estimator confirm that realized betas are highly persistent.

	RW	RWW	AR	ARMA	HAR	FIARMA	FI
RMSE	0.3197	0.2703	0.2840	0.2645	0.2859	0.2495	0.2487
Best	3	20	3	3	14	42	37
In MCS	14	56	20	37	40	106	105
vs. RW	0	67	43	58	32	79	80
vs. RWW	5	0	16	25	16	41	40
vs. AR	20	49	0	86	25	81	76
vs. ARMA	7	25	0	0	12	53	50
vs. HAR	8	33	17	29	0	50	47
vs. FIARMA	3	13	2	2	2	0	9
vs. FI	3	11	2	3	3	14	0
N	122	122	122	122	122	122	122

Table A.9: This table illustrates the cross-sectional forecast performance of the models for six-month beta forecasts from a rolling estimation window of 100 observations. The first row shows the average RMSEs of different models across all stocks over the different time periods. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain time period. "In MCS" denotes the number of times at which an estimator is in the Hansen et al. (2011) model confidence set. Furthermore, the rows denoted by "vs. X" correspond to modified DM-tests (Harvey et al., 1997), providing the number of times the column-model yields a significantly lower RMSE than the row-model. We examine statistical significance toward the 10 percent level. Finally, N is the number of investigated months.

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
β_i	0.557	0.181	0.942	0.940	0.507	0.197	0.915	0.951
$\rho_{i,M}$	0.550	0.160	0.955	0.958	0.550	0.165	0.953	0.953
σ_i	0.592	0.184	0.974	0.886	0.592	0.184	0.974	0.886
σ_M^{-1}	0.446	-	1.000	1.000	0.281	-	1.000	1.000

Table A.10: In analogy to Tables 1 and 2, this table presents average estimates of the memory parameter of realized betas, realized correlation (Fisher-transformed), and volatility across all stocks ($N = 823$), as well as that of the inverse of the market volatility, using the log periodogram estimator by Geweke & Porter-Hudak (1983). $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d_i = 0$ and $d_i = 1$, respectively, are rejected at the 10 percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle & Moulines (2000).

	RW	RWW	AR	ARMA	HAR	FIARMA	FI
RMSE	0.1467	0.1264	0.1395	0.1277	0.1282	0.1206	0.1187
Best	0	10	0	0	3	24	13
In MCS	31	48	30	45	50	50	50
vs. RW	0	20	3	7	12	19	19
vs. RWW	0	0	0	0	3	5	4
vs. AR	1	3	0	23	13	23	22
vs. ARMA	0	0	0	0	2	15	16
vs. HAR	0	1	0	0	0	4	1
vs. FIARMA	0	0	0	0	0	0	2
vs. FI	0	0	0	0	0	8	0
N	50	50	50	50	50	50	50

Table A.11: In analogy to Table 3, this table illustrates the forecast performance of the models for six-month beta forecasts from a rolling estimation window of 100 observations. FI and FIARMA use d estimates by the log periodogram estimator instead of the 2ELW estimator. The first row shows the average RMSEs of different models across all 50 portfolios. The row “Best” indicates the number of times a model achieves the lowest RMSE for a certain portfolio. “In MCS” denotes the number of portfolios for which an estimator is in the Hansen et al. (2011) model confidence set. Furthermore, the rows denoted by “vs. X” correspond to modified DM-tests (Harvey et al., 1997), providing the number of times the column-model yields a significantly lower RMSE than the row-model. We examine statistical significance toward the 10 percent level. Finally, N is the number of investigated portfolios.

Table A.11 repeats the analysis of Table 3 and shows the forecast performance of the FI and FIARMA model when estimating d using the log periodogram estimator. For comparison, we also present the results for the RW, RWW, AR, ARMA, and HAR models. It can be seen that compared to the results using the 2ELW estimator, the performance of the FI and FIARMA models is similar. The forecasts by the FI model clearly outperform all forecasts by models that do not account for long memory.

A.2.9 Alternative Sampling Frequencies

In our main analysis, our results are based on measures calculated with 30-minute data. Since the sampling frequency influences the bias as well as the variance of the estimates, we repeat our analysis for realized betas calculated from 15- and 75-minute data.

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
Panel A: 15-Minute Data								
β_i	0.586	0.166	0.973	0.989	0.542	0.178	0.967	0.991
$\rho_{i,M}$	0.593	0.151	0.977	0.989	0.588	0.155	0.977	0.988
σ_i	0.604	0.153	0.993	0.958	0.604	0.153	0.993	0.958
σ_M^{-1}	0.577	-	1.000	1.000	0.575	-	1.000	1.000
Panel B: 75-Minute Data								
β_i	0.509	0.154	0.971	0.998	0.464	0.176	0.945	0.998
$\rho_{i,M}$	0.518	0.129	0.980	0.999	0.503	0.132	0.979	0.999
σ_i	0.578	0.149	0.992	0.980	0.578	0.149	0.992	0.980
σ_M^{-1}	0.581	-	1.000	1.000	0.576	-	1.000	1.000

Table A.12: In analogy to Tables 1 and 2, this table presents average estimates of the memory parameter of realized betas, realized correlation (Fisher-transformed), and volatility across all stocks ($N = 823$), as well as that of the inverse of the market volatility, using the 2ELW estimator of Shimotsu & Phillips (2005) and Shimotsu (2010). The realized measures are now calculated from 15- and 75-minute data. $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d_i = 0$ and $d_i = 1$, respectively, are rejected at the 10 percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle & Moulines (2000).

Table A.12 shows that decreasing the frequency to 75-minute data decreases the estimated memory in realized beta from 0.56 to 0.51. This is again due to an increase of the noise level in the ex-post realized betas, which negatively biases the 2ELW estimator. When increasing the recording frequency from 30- to 15-minute data the estimated d increases only slightly to 0.59, implying that the amount of noise in the betas calculated from 30-minute data is already small. Despite these smaller changes, it still holds for at least 94 percent of the stocks that the order of integration of their betas is significantly different from 0 and 1.

Concerning the order of integration of the realized correlation series we observe a similar pattern. For 75-minute data the estimate decreases from the original value of 0.57 to 0.52 and for 15-minute data there is a small increase to 0.59. The ex-post estimates of stock and market volatility, on the other hand, seem to be less perturbed when decreasing the recording frequency. Here, the estimated memory is almost the same for 15-, 30-, and 75-

	RW	RWW	AR	ARMA	HAR	FIARMA	FI
Panel A: 15-Minute Data							
RMSE	0.1435	0.1231	0.1340	0.1244	0.1518	0.1159	0.1157
Best	0	6	0	0	4	25	15
In MCS	36	48	34	46	49	50	50
vs. RW	0	24	1	6	13	17	15
vs. RWW	0	0	0	1	2	4	5
vs. AR	2	2	0	23	11	19	22
vs. ARMA	0	1	0	0	3	15	16
vs. HAR	0	0	0	0	0	8	9
vs. FIARMA	0	0	0	0	0	0	1
vs. FI	0	0	0	0	0	5	0
N	50	50	50	50	50	50	50
Panel B: 75-Minute Data							
RMSE	0.1604	0.1352	0.1514	0.1366	0.1365	0.1253	0.1254
Best	0	3	0	1	3	19	24
In MCS	24	46	26	40	50	50	50
vs. RW	0	26	4	9	16	24	23
vs. RWW	0	0	0	1	3	6	8
vs. AR	2	6	0	21	9	26	26
vs. ARMA	0	1	0	0	3	13	15
vs. HAR	0	0	0	0	0	3	4
vs. FIARMA	0	0	0	0	0	0	3
vs. FI	0	0	0	0	0	3	0
N	50	50	50	50	50	50	50

Table A.13: In analogy to Table 3, this table illustrates the forecast performance of the models for six-month beta forecasts from a rolling estimation window of 100 observations. For the different panels, the realized beta series are based on 15-minute and 75-minute data. The first row shows the average RMSEs of different models across all 50 portfolios. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain portfolio. "In MCS" denotes the number of portfolios for which an estimator is in the Hansen et al. (2011) model confidence set. Furthermore, the rows denoted by "vs. X" correspond to modified DM-tests (Harvey et al., 1997), providing the number of times the column-model yields a significantly lower RMSE than the row-model. We examine statistical significance toward the 10 percent level. Finally, N is the number of investigated portfolios.

minute data ranging from 0.59 to 0.60 for stock volatility and 0.57 to 0.59 for the inverse of market volatility.

Table A.13 presents the forecast performance of the different models. We find that the ranking of the models stays the same for all considered frequencies. The FIARMA and FI models are the best independently of the sampling frequency. In addition, models that account for long-range dependencies perform substantially better than those that do not. Due to the difference in noise of the ex-post realized beta estimates, however, the average RMSE increases with decreasing sampling frequency. In line with the discussion above, this effect is more pronounced when switching from 30- to 75-minute data than when going from 30- to 15-minute data. Table A.13 further reveals that changing the recording frequency only leads to small changes when comparing the models against each other. Overall, the main message of Section 4 of the main paper remains unchanged: accounting for long-range dependencies significantly improves the forecasting performance for realized betas.

	RW	RWW	AR	ARMA	HAR	FIARMA	FI
Panel A: Rolling Window of 75 Observations							
RMSE	0.1395	0.1210	0.1339	0.1250	0.1263	0.1182	0.1133
Best	0	4	0	0	0	20	26
In MCS	36	50	35	45	50	50	50
vs. RW	0	26	3	8	8	20	21
vs. RWW	0	0	0	0	0	5	6
vs. AR	1	3	0	30	5	24	27
vs. ARMA	0	0	0	0	0	15	20
vs. HAR	0	0	0	0	0	13	14
vs. FIARMA	0	0	0	0	0	0	2
vs. FI	0	0	0	0	0	3	0
N	50	50	50	50	50	50	50
Panel B: Rolling Window of 125 Observations							
RMSE	0.1473	0.1303	0.1317	0.1214	0.1243	0.1138	0.1138
Best	0	1	0	2	1	19	27
In MCS	21	34	34	42	48	50	50
vs. RW	0	14	5	14	16	24	23
vs. RWW	0	0	0	9	7	14	15
vs. AR	1	1	0	18	8	19	20
vs. ARMA	0	1	0	0	2	11	11
vs. HAR	0	0	0	3	0	11	11
vs. FIARMA	0	0	0	0	0	0	2
vs. FI	0	0	0	0	0	1	0
N	50	50	50	50	50	50	50

Table A.14: In analogy to Table 3, this table illustrates the forecast performance of the models for six-month beta forecasts from rolling estimation windows of 75 and 125 observations. The first row shows the average RMSEs of different models across all 50 portfolios. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain portfolio. "In MCS" denotes the number of portfolios for which an estimator is in the Hansen et al. (2011) model confidence set. Furthermore, the rows denoted by "vs. X" correspond to modified DM-tests (Harvey et al., 1997), providing the number of times the column-model yields a significantly lower RMSE than the row-model. We examine statistical significance toward the 10 percent level. Finally, N is the number of investigated portfolios.

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
Panel A: Bandwidth $m = T^{0.65}$								
β_i	0.575	0.171	0.963	0.988	0.521	0.190	0.949	0.988
$\rho_{i,M}$	0.561	0.143	0.973	0.996	0.562	0.150	0.973	0.992
σ_i	0.592	0.162	0.986	0.941	0.592	0.162	0.986	0.941
σ_M^{-1}	0.567	-	1.000	1.000	0.559	-	1.000	1.000
Panel B: Bandwidth $m = T^{0.75}$								
β_i	0.539	0.146	0.977	1.000	0.499	0.157	0.972	1.000
$\rho_{i,M}$	0.551	0.129	0.983	1.000	0.544	0.131	0.983	0.999
σ_i	0.602	0.152	0.995	0.975	0.602	0.152	0.995	0.975
σ_M^{-1}	0.621	-	1.000	1.000	0.618	-	1.000	1.000

Table A.15: In analogy to Tables 1 and 2, this table presents average estimates of the memory parameter of realized betas, realized correlation (Fisher-transformed), and volatility across all stocks ($N = 823$), as well as that of the inverse of the market volatility, using the 2ELW estimator of Shimotsu & Phillips (2005) and Shimotsu (2010) with alternative bandwidths of $m = T^{0.65}$ and $m = T^{0.75}$. $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d_i = 0$ and $d_i = 1$, respectively, are rejected at the 10 percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle & Moulines (2000).

A.2.10 Alternative Estimation Windows and Bandwidths

Our main analysis regarding the forecast performance of the models uses a rolling estimation window of 100 observations. To show that the results are robust to other specifications of the estimation window, Table A.14 presents the results for window sizes of 75 and 125 observations.

While the shorter estimation window allows for more stocks to be included in the analysis, it can be seen that the results are qualitatively similar for both the shorter and longer estimation windows. The forecasts by the FI model perform the best and are never outperformed by models that do not account for long-run dependencies. A longer estimation window further improves the performance of the FI and FIARMA models.

	RW	RWW	AR	ARMA	HAR	FIARMA	FI
Panel A: Bandwidth $m = T^{0.65}$							
RMSE	0.1467	0.1264	0.1395	0.1277	0.1282	0.1217	0.1172
Best	0	5	0	0	3	27	15
In MCS	20	48	30	44	50	50	50
vs. RW	0	20	3	7	12	25	26
vs. RWW	0	0	0	0	3	8	7
vs. AR	1	3	0	23	13	22	21
vs. ARMA	0	0	0	0	2	18	19
vs. HAR	0	1	0	0	0	10	8
vs. FIARMA	0	0	0	0	0	0	3
vs. FI	0	0	0	0	0	4	0
N	50	50	50	50	50	50	50
Panel B: Bandwidth $m = T^{0.75}$							
RMSE	0.1467	0.1264	0.1395	0.1277	0.1282	0.1225	0.1197
Best	0	9	0	0	3	22	16
In MCS	22	47	27	46	50	50	50
vs. RW	0	20	3	7	12	23	22
vs. RWW	0	0	0	0	3	6	7
vs. AR	1	3	0	23	13	23	25
vs. ARMA	0	0	0	0	2	12	15
vs. HAR	0	1	0	0	0	6	6
vs. FIARMA	0	0	0	0	0	0	3
vs. FI	0	0	0	0	0	2	0
N	50	50	50	50	50	50	50

Table A.16: In analogy to Table 3, this table illustrates the forecast performance of the models for six-month beta forecasts from a rolling estimation window of 100 observations. The FI and FIARMA models are based on d estimates of the 2ELW estimator calculated with bandwidths of $m = T^{0.65}$ and $m = T^{0.75}$. The first row shows the average RMSEs of different models across all 50 portfolios. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain portfolio. "In MCS" denotes the number of portfolios for which an estimator is in the Hansen et al. (2011) model confidence set. Furthermore, the rows denoted by "vs. X" correspond to modified DM-tests (Harvey et al., 1997), providing the number of times the column-model yields a significantly lower RMSE than the row-model. We examine statistical significance toward the 10 percent level. Finally, N is the number of investigated portfolios.

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
β_i	0.467	0.144	0.973	0.999	0.423	0.167	0.942	0.999

Table A.17: This table presents average estimates of the memory parameter of realized beta, estimated with 2 lags, across all stocks ($\bar{\hat{d}}_i$) using the 2ELW estimator of Shimotsu & Phillips (2005) and Shimotsu (2010). Additionally, $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d_i = 0$ and $d_i = 1$, respectively, are rejected at the 10 percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle & Moulines (2000).

	RW	RWW	AR	ARMA	HAR	FIARMA	FI
RMSE	0.1483	0.1417	0.1404	0.1289	0.1300	0.1232	0.1185
Best	0	0	0	1	2	25	22
In MCS	30	41	22	42	49	50	50
vs. RW	0	8	5	9	14	25	25
vs. RWW	1	0	0	5	6	14	15
vs. AR	1	16	0	28	13	24	24
vs. ARMA	0	0	0	0	2	13	16
vs. HAR	0	1	0	0	0	10	9
vs. FIARMA	0	0	0	0	0	0	2
vs. FI	0	0	0	0	0	2	0
N	50	50	50	50	50	50	50

Table A.18: In analogy to Table 3, this table illustrates the forecast performance of the models for six-month beta forecasts from a rolling estimation window of 100 observations. The realized beta series are estimated with 2 lags. The first row shows the average RMSEs of different models across all 50 portfolios. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain portfolio. "In MCS" denotes the number of portfolios for which an estimator is in the Hansen et al. (2011) model confidence set. Furthermore, the rows denoted by "vs. X" correspond to modified DM-tests (Harvey et al., 1997), providing the number of times the column-model yields a significantly lower RMSE than the row-model. We examine statistical significance toward the 10 percent level. Finally, N is the number of investigated portfolios.

We also consider alternative bandwidths of $m = T^{0.65}$ and $m = T^{0.75}$ as a further robustness check in Tables A.15 and A.16. These results are qualitatively similar to those for our main bandwidth choice of $m = T^{0.7}$.

A.2.11 Correction for Asynchronous Trading

In this section, we test the robustness of our results to accounting for potentially asynchronous trading. That is, we repeat the analysis using a realized beta estimator with 2 lags.[5]

The results are in Tables A.17 and A.18. We find that the average d estimates are somewhat smaller. This is in line with the findings of Hollstein et al. (2019) that adding lags increases the noise in betas. The d estimates are nevertheless significantly different from 0 and 1 for more than 94% of the stocks. The forecast results are also qualitatively similar to those of our main analysis.

[5]Adding lags follows the idea of Dimson (1979) and consists of adding further regression coefficients toward market excess returns lagged by 1 and 2 periods. See Hollstein (2020) for details on the implementation of the realized beta estimator with leads and lags.

A.3 R-Algorithm for FI(0.5) Forecasts

The following algorithm can be used to forecast a time series using the FI(0.5) model as considered in Section 5.2 in R. The function requires the LongMemoryTS package and then generates h step ahead forecasts (as averages of the h point forecasts) of a time series contained in the object *series* as demonstrated in the example.

```
require(LongMemoryTS)
FI.Forecast<-function(series,h){
  T<-length(series)
  d<-0.5
  dif_series<-fdiff(series,d=d)
  pi<-fdiff(rep(1,T),d=d)
  mu<-coef(lm(dif_series~pi+0))
  epsilon<-dif_series-pi*mu
  epsilon<-c(epsilon,rep(0,h))
  dif_epsilon<-fdiff(epsilon,d=-d)
  point_forecasts<-(dif_epsilon+mu)[(T+1):length(dif_epsilon)]
  horizon_forecast<-mean(point_forecasts)
  return(horizon_forecast)
}
# Example for simulated series x
x<-FI.sim(T=200,q=1,rho=0,d=0.5)
FI.Forecast(x,h=1)
```

Appendix

B Firm Characteristics

- **Age** (Zhang, 2006) is the number of years up to time t since a firm first appeared in the CRSP database.
- **Beta** is the median beta estimate for a certain stock across all estimation approaches considered.
- **Bid–ask spread (BAS)** is the stock's average daily relative bid–ask spread over the previous month.
- **Book-to-market (BtM)** (Fama & French, 1992) is the most current observation for "book equity" divided by the market capitalization. Following the standard literature, we assume that the book equity of the previous year's balance sheet statement becomes available at the end of June and use the market capitalization at the end of the corresponding fiscal year. Book equity is defined as stockholders' equity, plus balance sheet deferred taxes and investment tax credit, plus post-retirement benefit liabilities, minus the book value of preferred stock.
- **Idiosyncratic skewness (iSkew)** (Boyer et al., 2009) is the iSkew of the residuals $\epsilon_{i,\tau}$ in the Fama & French (1993) 3-factor model $r_{i,\tau} - r_{f,\tau} = \alpha_{i,t} + \beta^M_{i,t}(r_{M,\tau} - r_{f,\tau}) + \beta^S_{i,t} SMB_\tau + \beta^H_{i,t} HML_\tau + \epsilon_{i,\tau}$, using daily returns over the previous month.
- **Idiosyncratic volatility (iVol)** (Ang et al., 2006) is the standard deviation of the residuals $\epsilon_{i,\tau}$ in the Fama & French (1993) 3-factor model $r_{i,\tau} - r_{f,\tau} = \alpha_{i,t} + \beta^M_{i,t}(r_{M,\tau} - r_{f,\tau}) + \beta^S_{i,t} SMB_\tau + \beta^H_{i,t} HML_\tau + \epsilon_{i,\tau}$, using daily returns over the previous month. SMB_τ and HML_τ denote the returns on the Fama & French (1993) factors.
- **Industry classifications** employ the definition for 10 industry portfolios applied by Kenneth French. "Durable" is Consumer Durables, "Energy" is the oil, gas, and coal extraction industry, "Healthcare" is Healthcare, Medical Equipment, and Drugs, "HiTec Equipment" is Business Equipment, "NonDurables" is Consumer NonDurables, "Telephone" is Telephone and Television Transmission, "Wholesale" is Wholesale, Retail, Services, and "Other" contains Mines, Construction, Construction Materials, Transport, Hotels, Bus Services, Entertainment, as well as Finance.
- **Investment** (Fama & French, 2015) is the change in total assets from the fiscal year ending in year $t-2$ to that ending in $t-1$, divided by the total assets of year $t-2$. As for BtM, we assume that accounting data become available by the end of June of year t.
- **Leverage** (Bhandari, 1988) is defined as one minus book equity (see "Book-to-market") divided by total assets (Compustat: AT). Book equity and total assets are updated every 12 months at the end of June.