The Memory of Beta Factors*

Janis Becker[†], Fabian Hollstein[†], Marcel Prokopczuk[†,‡],

and Philipp Sibbertsen[†]

September 24, 2019

Abstract

Researchers and practitioners employ a variety of time-series processes to forecast betas, using either 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. Finally, we document the relation of firm characteristics with the forecast error differentials that result from inadequately imposing short-memory or random walk instead of long-memory processes.

JEL classification: C58, G15, G12, G11

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

*Contact: becker@statistik.uni-hannover.de (J. Becker), hollstein@fmt.uni-hannover.de (F. Hollstein), prokopczuk@fmt.uni-hannover.de (M. Prokopczuk), and sibbertsen@statistik.uni-hannover.de (P. Sibbertsen).

[†]School of Economics and Management, Leibniz University Hannover, Koenigsworther Platz 1, 30167 Hannover, Germany.

[‡]ICMA Centre, Henley Business School, University of Reading, Reading, RG6 6BA, UK.

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 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]

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

Numerous other studies employ explicit or implicit short-memory processes for modeling beta dynamics. These include, among others, an AR(1) process in Ang & Chen (2007), 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.

However, the literature on volatility modeling documents that volatility has clear long-memory properties (Baillie et al., 1996; Bollerslev & Mikkelsen, 1996; Ding & Granger, 1996). It is thus natural to ask whether this is also true for beta. Andersen et al. (2006) tackle this issue and conclude that betas do not exhibit long memory. However, this conclusion is 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 each stock included in the S&P 500 during the 1996–2015 sample period. Next, we estimate the memory of realized beta using the two-step exact local Whittle (2ELW) estimator by Shimotsu & Phillips (2005) and Shimotsu (2010). 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 0.52. For virtually all stocks, the statistical tests clearly reject both the short-memory ($d = 0$) and difference-stationary ($d = 1$) alternatives. Thus, the vast majority of previous studies substantially misspecifies the properties of the beta time series.

Our findings differ considerably from those of Andersen et al. (2006). There are several causes for this difference. First, our study has a substantially broader focus: we consider more than 800 stocks. Second, we use high-frequency data to estimate beta factors. This enables us to obtain more precise and less noisy estimates of beta (see also

Hollstein et al., 2019a). 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). Third, using simulations, we show that for small samples, tests based on autocorrelation functions, as opposed to direct estimates with the 2ELW estimator, 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., 2018), as well as firms' cost of capital (Levi & Welch, 2017) strongly hinge on precise forecasts of betas. 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) alternatives for a substantial fraction of the stocks. A full-fledged FIARMA(d,p,q) alternative performs somewhat worse than the pure FI model, but better than the AR, ARMA, and RW models. We further show that the outperformance of the FI model over alternatives gets stronger for longer-horizon beta forecasts up to 1 year. Thus, incorporating the long-memory property is highly important for obtaining good beta forecasts.

In a next step, we examine which firm characteristics are associated with different degrees of memory in betas. We find that higher memory in beta is to some extent linked with higher levels of a stock's beta, book-to-market ratio, and leverage. In addition, stocks with high memory typically have lower market capitalization, turnover, idiosyncratic volatility, and short interest. Furthermore, we find substantial industry effects: stocks in the Energy and Manufacturing industries have comparably high memory in beta while stocks in the Healthcare, HiTec Equipment, Telephone, and Wholesale 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.

Finally, we document that for high-momentum stocks and those with high short interest, using a RW model instead of the FI model yields particularly high errors. On the other hand, for high-beta stocks, illiquid stocks, and those with high idiosyncratic volatility, it is most harmful to use an ARMA(p,q) model instead of the FI model.

We run a battery of tests to document the robustness of these results. First, we show that the FI model also outperforms its competitors when using hedging errors instead of the RMSE to evaluate the forecasts. Second, we also document 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 alternatives. Third, we estimate the short-memory and difference-stationary models in a state-space framework. In addition, we consider the Vasicek (1973) and Levi & Welch (2017) estimators, a heterogeneous AR (HAR) model, as well as a FI model, for which we set the long-memory parameter d to 0.5 instead of estimating it. We find that all alternative models underperform the FI model. Instead, the FI(0.5) model performs even somewhat better than the standard FI model. Fourth, we use the alternative estimator of the d parameter of Geweke & Porter-Hudak (1983) and obtain very similar results. Finally, we consider alternative intra-day sampling frequencies, alternative rolling estimation windows, and bandwidths. Our conclusions remain unchanged.

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), and Hollstein et al. (2019b). 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.

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

2 Data and Methodology

2.1 Data

Our dataset covers U.S. stocks for the sample period from January 1996 to December 2015. Following Bollerslev et al. (2016), for our main analysis we restrict our attention to stocks that are part of the S&P 500 index at least once during our sample period. 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 79 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.

In the Online Appendix, we also present results for the entire CRSP dataset and a time period starting from 1926. These results 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[2]

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

where O is the number of high-frequency return observations during the time period under investigation. $\beta_{i,t}$ 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 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 hand, 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 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 calculate the tapered local Whittle estimator by Velasco (1999) which is obtained by

$$\hat{d}_{Vel} = argmin_{d\in(-1/2,2)} \left[log\left(\frac{3}{m} \sum_{j}^{m} \lambda_j^{2d} I_y^*(\lambda_j) \right) - 2d\frac{3}{m} \sum_{j}^{m} log\lambda_j \right].$$

[2]Note that this formula resembles the expanded formula for the variance, 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.

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 the Online Appendix.

Under some mild assumptions this estimator is consistent and asymptotically normal for $d \in (-1/2, 2)$. However, as the estimator considers only every third frequency of the periodogram its variance exceeds that of the 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 $d > 1/2$, Shimotsu (2010) suggests using $\mu(d) = \bar{y}$ if $d < 1/2$, $\mu(d) = y_1$ if $d > 3/4$, and $\mu(d) = \omega(d)\bar{y} + (1-\omega(d))y_1$ with $\omega(d) = 1/2[1 + cos(4\pi d)]$ if $d \in [1/2, 3/4]$. This two-step estimator then has the same limiting variance as the standard local Whittle estimator while being consistent and asymptotically normally distributed for $d \in (-1/2, 2)$. Consequently, the 2ELW estimator can be used to distinguish short-memory series ($d = 0$), stationary long-memory series ($0 < d < 1/2$), nonstationary long-memory series ($1/2 < d < 1$), and difference-stationary series ($d = 1$) such as the random walk. This is an advantage over the standard local Whittle estimator, which can only be used for inference for $-1/2 < d < 3/4$ as it has a nonnormal limit distribution otherwise.

	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	0.561	0.112	0.998	0.999	0.523	0.136	0.993	0.999

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 $d = 0$ and $d = 1$, respectively, are rejected at the ten 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).

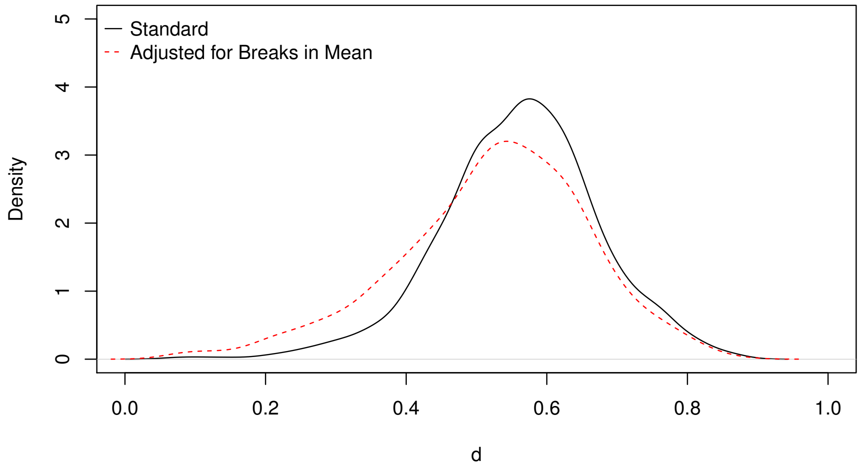

Figure 1: Density plot showing the distribution of the estimated beta memory parameters across stocks. For estimation we consider the Gaussian kernel and choose the bandwidth according to Silverman (1986).

3 Long Memory in Beta

3.1 Estimation Results

The left panel of Table 1 shows the average estimated d across the realized beta series of all stocks with more than 100 monthly observations (for $N = 823$ 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 ten percent level. To illustrate the variation in d across stocks, Figure 1 additionally plots the corresponding density of the estimates.

Table 1 reveals that the average d is approximately 0.56 and Figure 1 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 99 percent of the stocks it holds that $0 < d < 1$ at the ten percent level. At the one percent level this is still true for more than 98 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. $d < 1/2$, 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.[3]

The results are shown in the right panel of Table 1 and are visualized by the dashed line in Figure 1. We find that the average $\hat{d}$ decreases slightly to 0.523, implying that some stocks do indeed exhibit structural breaks in their betas time series. However, the reduction is small and for more than 99 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 analysis on daily data which leads to noisy

[3]Bai & Perron (1998) and Bai & Perron (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 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 break point selection, as for long-memory time series the standard procedure indicates too many break points.

estimates of beta, as also acknowledged by the authors themselves. Deo & Hurvich (2001) 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.2 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 $T = 100, 148, 240, 1000$. 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 shows 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.[4]

We therefore conclude that realized betas are highly persistent and are best described by either pure long-memory processes or a combination of break and long-memory process.

[4]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 in Section A.2 of the Online Appendix as a robustness check. This estimator is also unbiased, but exhibits a larger variance than the 2ELW estimator.

3.2 Beta Decomposition

Since beta is actually a combination of different components it seems 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, $\rho_{i,M,t}$ their realized correlation, and $\sigma_{i,t}$ is the realized volatility. Consequently, Equation (1) shows that the realized beta series evolves 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 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 70 percent of the stocks it holds that $d > 1/2$, 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.[5]

The results are shown in Table 2.[6] 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 0.56, while the d of realized volatility is even slightly higher on average, with 0.59. 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

[5] 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 $\rho_{i,M,t} = \frac{\sigma_{i,M,t}}{\sigma_{i,t}\sigma_{M,t}}$.

[6] 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. $d_i = 1$
$\rho_{i,M}$	0.559	0.096	1.000	0.999	0.557	0.099	1.000	1.000
σ_i	0.594	0.142	0.996	0.977	0.594	0.142	0.996	0.977
σ_M^{-1}	0.562	-	1.000	1.000	0.561	-	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 ($N = 823$), 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 $d = 0$ and $d = 1$, respectively, are rejected at the ten 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).

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, a natural next question to ask is: 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? Is accounting for long memory more important for long-term beta forecasts? 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 the first T betas of stock i, we first compute the $\hat{d}$th difference

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

where $\hat{d}_i$ is the estimate of the 2ELW estimator with a bandwidth of $m = T^{0.7}$. Again, we report qualitatively similar results for $m = T^{0.65}$ and $m = T^{0.75}$ in the Online Appendix.

We then set out to calculate the mean μ_i 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 μ_i. For this purpose, we perform the following regression

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

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

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

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 $\hat{Z}_{i,t} = \Delta^{-\hat{d}_i}\hat{\varepsilon}_{i,t}$ for $t = 1, \ldots, T+h$, respectively $t = 2, \ldots, T+h$. Forecasts of the original sequence then evolve as

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

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 $d < 1/2$. 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 $12[(T/100)^{0.25}]$.[7]

[7] To ease the presentation we focus on these five models. In the Online Appendix we additionally consider state-space AR, ARMA, and FIARMA models, the HAR-model of Corsi (2009), and the beta forecast approaches by Vasicek (1977) and Levi & Welch (2017). 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).[8] We test for significance in RMSE differences using the modified Diebold–Mariano (DM) test proposed by Harvey et al. (1997).

4.2 Forecast Results

The results of the various beta forecasts can be found in Table 3. We use a forecast window of 1 month and a rolling estimation window of 100 observations. Qualtitatively similar results for window sizes of 75 and 125 can be found in the Online Appendix. Table 3 presents the average RMSE across all stocks and the number of times the model yields the lowest RMSE when forecasting the realized beta of a stock. The remainder of the table indicates the number of stocks for which the column-model is significantly better than the row-model at the ten percent level. To allow for valid inference we only consider stocks for which we have at least 50 forecasts ($N = 689$ stocks fulfill this criterion).

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 more than 54 percent of the stocks. Second best is the FIARMA model, which is the best model for 24 percent of the stocks. 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 a combined 22 percent of the stocks. The outperformance of the long-memory models is

[8]The results when using the mean absolute error criterion instead of the RMSE are qualitatively similar.

	RW	AR	ARMA	FI	FIARMA
RMSE	0.3149	0.2942	0.2875	0.2792	0.2800
Best	4	24	123	374	164
vs RW	0	272	344	560	518
vs AR	3	0	177	305	307
vs ARMA	1	8	0	186	179
vs FI	0	3	14	0	7
vs FIARMA	0	2	13	24	0
N	689	689	689	689	689

Table 3: This table illustrates the forecast performance of the models for one-month beta forecasts from a rolling estimation window of 100 observations. The first row shows 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. 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 at the 10 percent level. Finally, N is the number of investigated stocks. To allow for reliable inference, we exclude all stocks for which we have less than 50 forecasts.

often also statistically significant. Compared to the RW forecasts, the FI forecasts are significantly better for 81 percent of the stocks; compared to AR and ARMA forecasts this number is 44 and 27 percent, respectively. On the other hand, the forecasts by the RW, AR, and ARMA models are almost never significantly better than those of the FI model. Consequently, we can conclude that accounting for the long-run dependence substantially improves 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. (2019a) 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.[9]

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

[9]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	AR	ARMA	FI	FIARMA
Bias	0.0000	0.0035	0.0023	0.0009	0.0009
Inefficiency	0.0290	0.0083	0.0062	0.0047	0.0048
Random Error	0.0910	0.0947	0.0922	0.0879	0.0887

Table 4: MSE Decomposition: This table shows the Mincer & Zarnowitz (1969) decomposition of the MSE as of Equation (2). The MSE is based on one-month forecasts of the realized beta series performed with a rolling estimation window of 100 observations. All numbers represent the average across all stocks for which at least 50 forecasts exist.

the following fashion

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

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 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 considered stocks. We find that the RW model is on average unbiased, but highly inefficient. Thus, particularly for high- and low-beta stocks, the RW approach generates sizable measurement errors. For the AR and ARMA models, the bias component is moderately larger than that of the RW model. Thus, these models appear to be somewhat misspecified. On the other hand, the inefficiency is dramatically smaller compared to the RW model. The random error component, which is the largest component for all models, is slightly higher for the AR and ARMA models than for the RW model.

The models that account for long memory are approximately unbiased and yield a low inefficiency on average. In particular 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 FI model also yields the lowest random error. Both inefficiency and random error are slightly higher for the FIARMA model. Thus, accounting for short-run dynamics in addition to long memory on average rather adds noise than helping to capture important parts of the variation in betas.

4.3 Longer Forecast Horizons

For many applications, such as capital budgeting decisions, managers typically plan over longer periods. Thus, they do not only need 1-month beta forecasts, but also forecasts over several months. Therefore, in this section, we also consider forecasts for three-month, six-month, and twelve-month horizons.

Table 5 presents the results for these forecast horizons, the table shows that the outperformance of the FI model forecasts persists and gets even stronger for horizons longer than one month. For all considered horizons the forecasts by the FI model have the lowest average RMSE and are the best for more than half of the stocks. It can further be seen that the absolute difference in RMSE between FI forecasts and RW, AR, and ARMA forecasts increases in the forecast horizon. Consequently, it is even more beneficial to consider long-memory models when forecasting for horizons longer than one month.

Not only is the magnitude of the forecast error loss differentials larger, but also is this differential statistically significant more often for longer horizons. For the three-month forecast horizon, the FI model is significantly better than the RW, AR, and ARMA models for 84, 52, and 30 percent of the stocks, respectively. These numbers are only slightly smaller for the twelve-month horizons with 73, 39, and 27 percent, respectively. In addition, the forecasts of the FI model are still barely ever outperformed by forecasts of models that do not account for the long-run dependencies. This is the case for less than 3 percent of the stocks, independently of the forecasts horizon.

To summarize, using models that account for long-run dependencies, instead of short-memory or difference-stationary alternatives, does not only improve one-month forecasts but also forecasts for longer horizons up to one year.

	RW	AR	ARMA	FI	FIARMA
Panel A: Three-Month Forecast Horizon					
RMSE	0.2918	0.2753	0.2589	0.2377	0.2417
Best	2	27	102	439	115
vs RW	0	184	303	572	617
vs AR	12	0	255	359	260
vs ARMA	4	7	0	210	130
vs FI	0	6	11	0	2
vs FIARMA	0	9	27	149	0
N	685	685	685	685	685
Panel B: Six-Month Forecast Horizon					
RMSE	0.2955	0.2882	0.2616	0.2286	0.2367
Best	3	28	81	451	115
vs RW	0	146	274	554	614
vs AR	29	0	240	334	200
vs ARMA	3	2	0	193	77
vs FI	0	3	12	0	0
vs FIARMA	0	16	41	208	0
N	678	678	678	678	678
Panel C: Twelve-Month Forecast Horizon					
RMSE	0.3075	0.3093	0.2766	0.2331	0.2451
Best	2	39	56	419	138
vs RW	0	139	236	478	588
vs AR	27	0	161	256	133
vs ARMA	11	4	0	177	61
vs FI	1	5	12	0	3
vs FIARMA	0	22	45	212	0
N	654	654	654	654	654

Table 5: In analogy to Table 3, this table illustrates the forecast performance of the models for three-, six-, and twelve-month beta forecasts from a rolling estimation window of 100 observations. The first row shows 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. 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 at the 10 percent level. Finally, N is the number of investigated stocks. To allow for reliable inference, we exclude all stocks for which we have less than 50 forecasts.

5 Economic Implications

5.1 The Memory in Beta and Stock Characteristics

We continue the empirical analysis by examining to what extent the memory in beta factors relates to different firm characteristics. There are various candidate variables that might explain part of the difference in a stock's beta-memory. It is, for example, possible that the beta estimates of small and illiquid stocks contain more random noise, which has zero autocorrelation. Furthermore, it is possible that growth stocks, firms that invest more, or those that are most profitable change more frequently, which might make past shocks to their systematic risk die out more quickly. On the other hand, it is possible that current loser stocks or firms whose stocks experience high shorting activity are more prone to change their business models, which likely changes their systematic risk. Finally, there may be industry effects: for some industries, the business models, and with that the constituent firms' systematic risk, may be more persistent, while others experience more frequent changes.

For this analysis, we sort the stocks into five portfolios (P1 up to P5), based on their estimates for $\boldsymbol{d}$. We do this at the end of each month using $\boldsymbol{d}$-estimates based on a 100-month rolling window. For each portfolio we record the average of several firm characteristics at the end of that month. Subsequently, we examine whether there are systematic differences in the average firm characteristics of the different $\boldsymbol{d}$-sorted portfolios. The variable definitions are in Section A.3 of the Online Appendix.

We present the results in Table 6. The quintile portfolio of the stocks with the lowest $\boldsymbol{d}$s (P1) on average has a memory parameter of 0.35 while that of the stocks with the highest $\boldsymbol{d}$s has a $\boldsymbol{d}$ of 0.73 on average. These averages are far away from both 0 and 1. This result is consistent with our previous finding that the betas of virtually all stocks have long-memory properties. Naturally, the difference between the memory parameters of portfolios 5 and 1 is highly statistically significant.

The second variable of interest is beta itself. We find that the average beta of high-beta-memory stocks is significantly higher than that of low-beta-memory stocks. The

	P1	P2	P3	P4	P5	t−stat
$\boldsymbol{d}$	0.3494	0.4789	0.5491	0.6173	0.7291	22.9
$\boldsymbol{\beta}$	0.9326	0.9747	1.0219	1.0411	1.1182	3.02
log(Market Cap)	16.140	16.154	16.064	16.073	15.983	−2.40
BtM	0.4906	0.4972	0.4552	0.5334	0.5916	3.43
Investment	0.1073	0.0868	0.0880	0.1031	0.1033	−0.46
Profitability	0.2512	0.3398	−3.9065	−2.1864	−0.3117	−0.89
Momentum	0.1444	0.1252	0.1357	0.1368	0.1641	0.82
BAS	0.0008	0.0008	0.0008	0.0008	0.0009	1.52
Turnover	0.2339	0.2377	0.2444	0.2380	0.2503	0.90
iVol	0.0133	0.0134	0.0136	0.0134	0.0141	1.13
iSkew	0.1077	0.1169	0.1099	0.1048	0.1183	1.00
Short Interest	0.0397	0.0386	0.0402	0.0379	0.0374	−0.57
Leverage	0.5803	0.5775	0.5831	0.5759	0.6084	1.91
Age	34.330	36.799	36.839	36.302	35.992	0.75

Table 6: Portfolio sorts: This table presents portfolio sorts by the estimated d. At the end of each month, we sort the stocks in our sample based on the d-parameters estimated with the 2ELW estimator using a rolling window of 100 observations. Sorting the stocks into quintile portfolios, we save each portfolio's average of the firm characteristics at the end of the respective months. The main body of the table shows the average of the different firm characteristics over time. t-stat denotes the t-statistic of a test whether the firm characteristics of portfolio P5 and P1 are equal, with the standard errors calculated with the heteroscedasticity and autocorrelation robust approach by Andrews (1991), using a quadratic spectral density and data-driven bandwidth selection. Characteristics for which this difference is statistically significant at 10 percent are printed in **bold**.

relation appears to be monotonic, but overall economically not too strong. For the natural logarithm of a stock's market capitalization, we make an opposite observation. The stocks with the longest memory in beta appear to be somewhat smaller than those with the shortest memory in beta.

The average BtM ratio of the stocks in P1 is significantly smaller than that of P5. As the firms grow, the past shocks to their beta factors are essentially to those of different firms and their impact seems to die out more quickly. On the other hand, we detect no relation between the beta-memory and firms' investment, profitability, momentum, bid-ask spread, turnover, idiosyncratic volatility, idiosyncratic skewness, and short interest. Low-d stocks on average exhibit lower leverage than high-d stocks. Age appears to be unrelated to the memory in betas.

Finally, we turn the focus on the stocks' industries. We present the results in Table 7.

	$\bar{\hat{d}}$	t-stat
Durables	0.5188	-1.70
Energy	0.6013	3.64
Healthcare	0.5529	-0.58
HiTec Equipment	0.5243	-3.85
Manufacturing	0.6079	4.77
NonDurables	0.5642	0.20
Other	0.5915	3.19
Telephone	0.5209	-1.40
Utilities	0.5489	-0.82
Wholesale	0.5036	-5.12

Table 7: 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 heteroscedasticity and autocorrelation robust approach by Andrews (1991), using a quadratic spectral density and data-driven bandwidth selection. Industries for which the average d is significantly higher or lower than this value at the 10 percent level are printed in **bold**.

Stocks in the Energy, Manufacturing, and Other 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 Durables, HiTec Equipment, Telephone, and Wholesale industries, the opposite holds true. These industries have in part been particularly prone to disruptions and creative destruction during the recent 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.

In Table A.3 of the Online Appendix, we also present the results of portfolios sorted on beta. We confirm that the relation of beta and beta-memory is on average positive, but weakly so. There is very little difference in the d-parameters of the first three beta quintiles. Only for the two quintiles of the highest betas is the d estimate somewhat larger.

5.2 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 processes or those that impose infinite memory. To that end, we regress the difference in absolute forecast errors on different firm characteristics. In more detail, we perform the following regressions

$$abs(\hat{\beta}_{i,t}^{RW} - \beta_{i,t}) - abs(\hat{\beta}_{i,t}^{FI} - \beta_{i,t}) = a + bx_{i,t-1} + e_{i,t},$$
$$abs(\hat{\beta}_{i,t}^{ARMA} - \beta_{i,t}) - abs(\hat{\beta}_{i,t}^{FI} - \beta_{i,t}) = a + bx_{i,t-1} + e_{i,t}.$$

Here, $\hat{\beta}_{i,t}$ are the forecasts made by the RW, ARMA, and FI models as presented in Section 4.2 and $x_{i,t-1}$ contains the set of explanatory variables lagged by one period.

We present the result for the forecast error differential between the RW model and the FI model in Table 8 and that between the ARMA(p,q) model and the FI model in Table 9.

Starting with the errors made when inadequately imposing a difference-stationary RW model in Table 8, we first obtain an economically large and statistically highly significant intercept term. This echoes our previous findings 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 tenth.

The level of the idiosyncratic volatility has a positive effect on the forecast error differential. This effect is economically large: for an idiosyncratic volatility two-standard-deviation below the average, all else being equal, the forecast error of RW and FI processes are approximately the same. Thus, for high volatility stocks a random walk assumption appears to be less suitable.

We further observe that a one-standard-deviation increase in momentum, short interest, and leverage increases the forecast error differential by on average 0.21, 0.20, and 0.13 percentage points, respectively. It is well known that betas of stocks with extreme momentum are highly time-varying (Grundy & Martin, 2001). Similarly, firms whose

	coef	se	$t-$stat	$p-$value
Intercept	0.0253	0.0014	18.26	0.0000
$\boldsymbol{d}$	−0.0034	0.0007	−5.28	0.0000
β	0.0013	0.0014	0.94	0.3200
log(Market Cap)	0.0006	0.0008	0.73	0.4730
BtM	0.0013	0.0037	0.34	0.7010
Investment	0.0000	0.0006	0.03	0.9790
Profitability	0.0000	0.0001	0.29	0.7200
Momentum	0.0021	0.0009	2.36	0.0170
BAS	−0.0071	0.0038	−1.84	0.0630
Turnover	−0.0039	0.0013	−2.93	0.0050
iVol	0.0127	0.0016	7.93	0.0000
iSkew	−0.0003	0.0006	−0.47	0.6420
Short Interest	0.0020	0.0012	1.71	0.0870
Leverage	0.0013	0.0008	1.71	0.0870
Age	−0.0005	0.0005	−1.01	0.3120
Durables	0.0008	0.0047	0.17	0.8620
Energy	−0.0030	0.0023	−1.33	0.1710
Healthcare	0.0010	0.0018	0.55	0.5600
HiTec Equipment	−0.0002	0.0021	−0.09	0.9160
Manufacturing	−0.0015	0.0017	−0.90	0.3540
NonDurables	0.0009	0.0023	0.40	0.6680
Telephone	−0.0069	0.0030	−2.28	0.0230
Utilities	0.0030	0.0232	0.13	0.6500
Wholesale	0.0007	0.0018	0.39	0.6590

Table 8: In this table, we run regressions of the difference in absolute forecast errors from the RW and FI models on different firm characteristics variables. Firm characteristics (except for the dummy variables) are 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. Characteristics which yield a statistically significant regression coefficient (coef) at 10 percent are printed in **bold**.

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 bid-ask spread and the turnover have a negative impact on the loss differential. The beta of highly liquid stocks should therefore be predicted with long memory models rather than the random walk.

In Table 9, we analyze the determinants of the ARMA and FI model error differentials. Consistent with our previous results, we also detect a strongly statistically significant

	coef	se	t−stat	p−value
Intercept	0.0103	0.0009	11.13	0.0000
d	−0.0005	0.0005	−0.98	0.3260
$\boldsymbol{\beta}$	0.0059	0.0011	5.41	0.0000
log(Market Cap)	−0.0004	0.0005	−0.65	0.5250
BtM	0.0003	0.0033	0.10	0.8960
Investment	0.0007	0.0007	1.00	0.3020
Profitability	−0.0002	0.0000	−4.90	0.0000
Momentum	−0.0005	0.0005	−1.03	0.3030
BAS	0.0040	0.0018	2.21	0.0310
Turnover	0.0016	0.0011	1.46	0.1490
iVol	−0.0004	0.0009	−0.41	0.6770
iSkew	−0.0002	0.0003	−0.59	0.5540
Short Interest	−0.0015	0.0007	−2.23	0.0260
Leverage	−0.0006	0.0006	−0.89	0.3590
Age	−0.0014	0.0004	−3.28	0.0010
Durables	−0.0025	0.0021	−1.21	0.2210
Energy	0.0017	0.0022	0.78	0.3870
Healthcare	0.0005	0.0017	0.27	0.7730
HiTec Equipment	0.0003	0.0014	0.23	0.8220
Manufacturing	−0.0010	0.0015	−0.65	0.5310
NonDurables	0.0014	0.0020	0.69	0.5210
Telephone	−0.0036	0.0043	−0.83	0.4780
Utilities	−0.0056	0.0059	−0.95	0.3760
Wholesale	0.0006	0.0013	0.47	0.6380

Table 9: In this table, we run regressions of the difference in absolute forecast errors from the ARMA and FI models on different firm characteristics variables. Firm characteristics (except for the dummy variables) are 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. Characteristics which yield a statistically significant regression coefficient (coef) at 10 percent are printed in **bold**.

intercept term of 0.0103. This intercept term is substantially smaller than that for the RW–FI forecast error differential.

The forecast error differential increases with beta. The impact of beta on these forecast error differentials is economically large: for betas two-standard-deviation below the average, all else being equal, the forecast error of ARMA and FI processes are approximately the same.

The profitability, short interest, and age all have a significant negative impact on the forecast error differential. Smaller firms and firms with higher short interest might be more prone to short-run changes in betas. Thus, the short-memory models perform a little less badly for these. The impact of each of these variables, however, is economically substantially smaller than that of the level of beta. The bid-ask spread has a positive impact on the forecast error differential. Thus, for the rather illiquid stocks the betas might contain more noise. The short-memory models might pick up too much of this noise to generate reliable forecasts.

6 Conclusion

In this paper, we analyze the memory of beta factors. 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 or difference-stationary models. For longer forecast horizons, the errors made by falsely imposing structures that do not account for long memory increase further.

Failing to account for the long-memory properties of betas can lead to very high errors, in particular for high-momentum stocks, those with strong short-selling pressure, high-beta stocks, illiquid stocks, and those with high idiosyncratic volatility. For the former two, imposing a random walk is most hurtful while for the latter short-memory processes are particularly inadequate.

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.

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.

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*, 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.

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.

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

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.

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.

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

Daniel, K., Mota, L., Rottke, S., & Santos, T. (2018). The cross-section of risk and return. *NBER Working Paper*.

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.

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

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., & 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.

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

Hollstein, F., Prokopczuk, M., & Wese Simen, C. (2019a). The conditional Capital Asset Pricing Model revisited: Evidence from high-frequency betas. *Management Science, forthcoming*.

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

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.

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.

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.

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.

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.

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*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

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

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, *5*(2), 177–188.

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.

The Memory of Beta

Online Appendix

JEL classification: G15, G12, G11, C58

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, 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 $T = 100, 148, 240, 1000$ 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. 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 240. The noise robust estimators, on the other hand, are positively biased for sample sizes of 100, 148, and 240.

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

	$d = 0.2$				$d = 0.4$				$d = 0.6$			
	$T = 100$	$T = 148$	$T = 240$	$T = 1000$	$T = 100$	$T = 148$	$T = 240$	$T = 1000$	$T = 100$	$T = 148$	$T = 240$	$T = 1000$
$\bar{\hat{d}}_{2ELW}$	0.22	0.21	0.22	0.20	0.43	0.42	0.41	0.40	0.63	0.62	0.62	0.61
$sd(\hat{d}_{2ELW})$	0.13	0.11	0.09	0.05	0.13	0.11	0.09	0.05	0.13	0.11	0.09	0.05
$\bar{\hat{d}}_{GPH}$	0.20	0.20	0.21	0.20	0.42	0.41	0.41	0.40	0.62	0.62	0.62	0.61
$sd(\hat{d}_{GPH})$	0.16	0.13	0.11	0.06	0.16	0.14	0.12	0.06	0.17	0.14	0.11	0.06
$\bar{\hat{d}}_{HP}$	0.11	0.12	0.16	0.19	0.27	0.29	0.33	0.38	0.38	0.43	0.49	0.57
$sd(\hat{d}_{HP})$	0.19	0.16	0.12	0.06	0.23	0.21	0.15	0.06	0.36	0.30	0.22	0.08
$\bar{\hat{d}}_{trLW}$	0.30	0.19	0.14	0.17	0.47	0.38	0.32	0.35	0.66	0.55	0.53	0.55
$sd(\hat{d}_{trLW})$	0.42	0.34	0.24	0.11	0.44	0.34	0.24	0.11	0.45	0.32	0.25	0.12
$\bar{\hat{d}}_{LWN}$	0.36	0.36	0.32	0.25	0.52	0.49	0.47	0.43	0.68	0.67	0.65	0.63
$sd(\hat{d}_{LWN})$	0.29	0.26	0.20	0.08	0.21	0.17	0.13	0.06	0.16	0.14	0.11	0.06
$\bar{\hat{d}}_{LPWN}$	0.40	0.39	0.37	0.28	0.55	0.53	0.50	0.45	0.69	0.69	0.68	0.65
$sd(\hat{d}_{LPWN})$	0.36	0.33	0.29	0.16	0.29	0.26	0.20	0.09	0.24	0.19	0.15	0.08
Sign. ac (%)	4.30	6.49	10.53	30.29	13.56	22.70	37.79	88.41	28.21	45.74	67.28	99.55
Ljung–Box	0.006	0.008	0.006	0.000	0.004	0.003	0.001	0.000	0.008	0.001	0.002	0.000

Table A.1: Simulation results: 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 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 percent 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.

observed order of integration is decreased, such that the 2ELW estimator yields a d of 0.4 on average. They then fractionally differenced 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 to detect 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 Additional Analyses and Robustness

A.2.1 Tables Referenced in the Main Manuscript

	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	0.127	0.996	0.989	0.532	0.156	0.987	0.989
$\rho_{i,M}$	0.554	0.102	1.000	0.998	0.561	0.109	1.000	0.996
σ_i	0.586	0.146	0.991	0.968	0.586	0.146	0.991	0.968
σ_M^{-1}	0.544	-	1.000	1.000	0.544	-	1.000	1.000
Panel B: Bandwidth $m = T^{0.75}$								
β_i	0.549	0.103	0.999	0.998	0.517	0.122	0.996	0.998
$\rho_{i,M}$	0.546	0.088	0.995	1.000	0.543	0.090	1.000	1.000
σ_i	0.592	0.137	1.000	0.989	0.592	0.137	1.000	0.989
σ_M^{-1}	0.591	-	1.000	1.000	0.590	-	1.000	1.000

Table A.2: 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 $d = 0$ and $d = 1$, respectively, are rejected at the ten 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).

	P1	P2	P3	P4	P5	t−stat
$\boldsymbol{\beta}$	0.4830	0.7726	0.9769	1.2075	1.6887	22.1
d	0.5333	0.5213	0.5394	0.5550	0.5822	3.57
log(Market Cap)	16.175	16.215	16.177	16.016	15.673	−4.55
BtM	0.5380	0.5060	0.5016	0.5299	0.5628	0.49
Investment	0.0938	0.1067	0.1077	0.1191	0.1177	1.59
Profitability	−0.5928	−3.4268	−0.5836	0.2827	0.1098	1.07
Momentum	0.1276	0.1329	0.1359	0.1485	0.2029	1.39
BAS	0.0009	0.0007	0.0007	0.0008	0.0010	0.98
Turnover	0.1846	0.1893	0.2114	0.2540	0.3882	6.40
iVol	0.0115	0.0115	0.0125	0.0143	0.0193	5.10
iSkew	0.0834	0.1044	0.1038	0.1269	0.1500	5.06
Short Interest	0.0302	0.0337	0.0369	0.0419	0.0558	11.2
Leverage	0.5920	0.5889	0.5774	0.5745	0.5922	0.01
Age	37.611	33.049	33.317	32.864	29.838	−11.0
Durables	0.0048	0.0107	0.0230	0.0384	0.0309	10.3
Energy	0.0309	0.0319	0.0441	0.0783	0.1137	3.42
Healthcare	0.1299	0.1096	0.0664	0.0448	0.0378	−2.79
HiTec Equipment	0.0812	0.1401	0.1790	0.1776	0.2001	2.41
Manufacturing	0.0781	0.1153	0.1655	0.1721	0.1661	9.19
NonDurables	0.1624	0.0819	0.0518	0.0364	0.0309	−18.0
Telephone	0.0397	0.0452	0.0340	0.0227	0.0132	−8.33
Utilities	0.1721	0.0806	0.0341	0.0230	0.0108	−10.0
Wholesale	0.1185	0.1374	0.1259	0.1128	0.0775	−2.97

Table A.3: This table presents portfolio sorts by beta. At the end of each month, we sort the stocks in our sample based on the realized beta during the past month. Sorting the stocks into quintile portfolios, we save each portfolio's average of the firm characteristics and dummy variables at the end of the respective months. The main body of the table shows the average of the different firm characteristics over time ($T = 141$ months). t-stat denotes the t-statistic of a test whether the firm characteristics of portfolio P5 and P1 are equal with the standard errors being calculated using the heteroscedacity and autocorrelation robust approach by Andrews (1991). Characteristics, for which this difference is statistically significant at 10 percent are printed in **bold**.

A.2.2 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 our main

	RW	*AR*	*ARMA*	*FI*	*FIARMA*
Mean	4.3890	4.3765	4.3699	4.3582	4.3640
ΔRW	0.0000	−0.0124**	−0.0191***	−0.0308***	−0.0250***
		(−2.0180)	(−2.9422)	(−5.1588)	(−3.4989)
ΔARMA	0.0191***	0.0067	0.0000	−0.0117***	−0.0059
	(2.9422)	(1.5468)		(−3.5983)	(−0.9233)

Table A.4: Hedging errors: 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 $(r_{i,T+1} - r_{f,T+1}) - \hat{\beta}_{i,T+1}(r_{M,T+1} - r_{f,T+1})$. 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.

approaches. We compute the hedging error for each stock as

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

$r_{i,T+1}$ 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. We use 1-month returns. $\hat{\beta}_{i,T+1}$ 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 $\frac{\text{var}(H_{i,T+1})}{\text{var}(r_{M,T+1} - r_{f,T+1})}$ is approximately equal to the mean squared error relative to the *true* realized beta plus a term that is constant for all beta forecasts. We follow Liu et al. (2018) and estimate the variance ratios using rolling five-year windows to account for 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.4. These are consistent with our previous results relying on the RMSE. The average hedging error of the FI-model forecasts is lowest. In particular, the average hedging error is significantly lower than both that of the difference-stationary RW and the short-memory ARMA models.

	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	0.330	0.181	0.841	0.999

Table A.5: 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 $d = 0$ and $d = 1$, respectively, are rejected at the ten 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.2.3 Entire CRSP Dataset

In our main analysis, based on the need to have high-frequency data for liquid instruments, we restrict our dataset to the S&P 500 firms 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.[10]

Table A.5 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 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 0.36, and even when considering the same stocks and same time period as for our main analysis, where the average d is 0.42. Consequently, the observed reduction in d is mainly due to the change of the recording frequency and not to the expanded set of stocks and time period. As

[10]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.

	RW	AR	ARMA	FI	FIARMA
RMSE	0.5654	0.4981	0.4881	0.4720	0.4724
Best	27	82	157	821	282
vs RW	0	721	802	1104	1052
vs AR	29	0	345	656	701
vs ARMA	15	38	0	458	480
vs FI	0	9	27	0	28
vs FIARMA	1	7	25	27	0
N	1369	1369	1369	1369	1369

Table A.6: In analogy to Table 3, this table illustrates the forecast performance of the models for quarterly beta forecasts, based on daily data, from a rolling estimation window of 100 observations. The first row shows 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. 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 at the 10 percent level. Finally, N is the number of investigated stocks. To allow for reliable inference, we exclude all stocks for which we have less than 50 forecasts.

already discussed in Section 3, 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.[11]

Even though the d estimates are negatively biased, more than 84 percent of the stocks still have a d that is significantly greater than zero. The forecast results displayed in Table A.6 also echo this finding. It can be seen that the FI model still significantly outperforms all models that do not account for the long-range dependencies. Its forecasts obtain the lowest average RMSE and are the most accurate for almost half of the stocks. Forecasts by RW, AR, or ARMA models, on the other hand, are only the most accurate for a combined 19 percent of the stocks. These models significantly outperform FI forecasts for less than three percent of the stocks in total. We should further note that as the FI model relies on a biased estimate of d, its performance would likely be even better if we accounted for this bias by adding a constant to each d estimate or even fixed d at a certain level for all stocks.

[11] In Section A.2.6, 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.

A.2.4 Alternative Models

Due to its great importance there are numerous approaches and models to forecast beta. For the ease of presentation in our main analysis, we compare the performance of the long-memory models only to the performance of the most popular competitors, RW, AR, and ARMA. In this section we now consider other approaches that have been proposed in the literature.

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 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

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

We estimate those 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 investigated for the stationary $d < 0.5$ case (Chan & Palma, 1998; Dissanayake et al., 2016). As we investigate mostly nonstationary time series here, these models are likely inappropriate. As an alternative, we consider a FIARMA model as before, but with the short-run dynamics now estimated with an ARMA(1,1) model in a state-space framework.

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

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

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. We therefore also consider forecasts made by this model in the following.

Hassler & Pohle (2019) argue that although local Whittle-based approaches yield better results than other estimators, they still have a large variance. Moreover, as discussed above, the estimators are negatively biased when the degree of noise in the series becomes large. These considerations lead the authors to believe that it might be beneficial for forecasting to fix d at a certain value instead of estimating it. This eliminates estimation uncertainty, while the model is still able to capture the long-memory characteristics of the series. Based on the results of Section 3 we fix d to 0.5. We refer to this model as FI(0.5) in the following.

Finally, we also consider two popular shrinkage approaches. First, we apply the Vasicek (1973) estimator as modification to the RW forecast. We obtain a posterior beta by combining the RW forecast with a prior ($b_{j,t}$) in the following way

$$\beta_{i,t}^{\text{RWV}} = \frac{s_{b_{i,t}}^2}{\sigma_{\beta_{i,t}}^2 + s_{b_{i,t}}^2}\beta_{i,t} + \frac{\sigma_{\beta_{i,t}}^2}{\sigma_{\beta_{i,t}}^2 + s_{b_{i,t}}^2}b_{i,t}.$$

$\sigma_{\beta_{i,t}}^2$ and $s_{b_{i,t}}^2$ are the squared standard errors of the beta estimate and the prior, respectively. Hence, the degree of shrinkage depends on the relative precision of the historical estimate and the prior. As prior, we use the cross-sectional average beta, as suggested by Vasicek (1973).

Levi & Welch (2017) argue that a simple Vasicek (1973) shrinkage is not sufficient to

	RW	RWV	RWLW	AR	ARMA	FIARMA	HAR	FI(0.5)	FI
RMSE	0.2820	0.3124	0.3862	0.2854	0.2850	0.2808	0.3176	0.2775	0.2792
Best	92	39	4	11	19	84	4	313	123
vs RW	0	2	1	2	2	53	0	85	80
vs RWV	279	0	0	248	243	315	78	394	335
vs RWLW	569	522	0	560	560	572	432	592	584
vs AR	171	6	1	0	41	81	1	107	109
vs ARMA	139	5	1	40	0	81	0	103	101
vs FIARMA	15	1	1	9	9	0	1	68	49
vs HAR	489	81	2	425	419	495	0	440	400
vs FI(0.5)	9	1	0	3	4	8	0	0	11
vs FI	24	2	0	12	11	22	0	82	0
N	689	689	689	689	689	689	689	689	689

Table A.7: In analogy to Table 3, this table illustrates the forecast performance of different additional models for one-month beta forecasts from a rolling estimation window of 100 observations. RW, AR, ARMA, and FIARMA are estimated in a state-space framework. RWV and RWLW correspond to the forecasts from the approaches of Vasicek (1973) and Levi & Welch (2017), respectively. Finally, HAR corresponds to the model by Corsi (2009) and FI(0.5) uses a FI model with d fixed at 0.5. The first row shows 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. 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 at the 10 percent level. Finally, N is the number of investigated stocks. To allow for reliable inference, we exclude all stocks for which we have less than 50 forecasts.

create good forecasts for beta. They suggest further shrinkage using

$$\beta_{i,t}^{\text{RWLW}} = 0.75\beta_{i,t}^{\text{RWV}} + 0.25\beta_i^{\text{target}},$$

where β_i^{target} is set to 0.5 for the smallest market capitalization tercile, to 0.7 for the middle tercile, and to 0.9 for the highest market capitalization tercile. One has to bear in mind, though, that Levi & Welch (2017) optimize this double-shrinkage for betas based on daily return data. Since we rely on a highly liquid subset of stocks and use more precise estimates based on high-frequency data it is likely that this approach does not work too well.

Table A.7 shows the forecast results for these models and for comparison again the results by the FI model considered before. In line with the results by Hollstein & Prokopczuk (2016), we find that the performance of the RW model improves when esti-

	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	0.133	0.994	0.953	0.518	0.158	0.967	0.965
$\rho_{i,M}$	0.572	0.122	0.995	0.966	0.572	0.128	0.991	0.962
σ_i	0.593	0.164	0.978	0.930	0.593	0.164	0.978	0.930
σ_M^{-1}	0.591	-	1.000	1.000	0.590	-	1.000	1.000

Table A.8: 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 $d = 0$ and $d = 1$, respectively, are rejected at the ten 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).

mated with a state-space framework as it on average now produces more accurate forecasts than AR and ARMA models. However, the models that account for long-range dependencies still perform substantially better and are more accurate for almost 80 percent of the stocks. The RMV model performs somewhat better than the simple RW model. The performance of the RVLW model, on the other hand, is very poor, as was expected.

Finally, it is noteworthy that the results for the FI(0.5) model are even slightly better than those by the FI model considered before. Thus, fixing d at 0.5 instead of using estimates appears to be a practical and well-performing approach for beta forecasting.

A.2.5 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.8 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 95 percent of the stocks exhibit significant long

	RW	AR	ARMA	FI	FIARMA
RMSE	0.3149	0.2942	0.2878	0.2812	0.2814
Best	7	30	128	332	192
vs RW	0	271	343	507	455
vs AR	3	0	155	228	259
vs ARMA	1	5	0	140	135
vs FI	1	5	22	0	21
vs FIARMA	1	4	11	21	0
N	689	689	689	689	689

Table A.9: In analogy to Table 3, this table illustrates the forecast performance of the models for one-month beta forecasts from a rolling estimation window of 100 observations. FI and FIARMA model are now calculated using d estimates by the log periodogram estimator instead of the 2ELW estimator. The first row shows 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. 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 at the 10 percent level. Finally, N is the number of investigated stocks. To allow for reliable inference, we exclude all stocks for which we have less than 50 forecasts.

memory in beta. We can therefore conclude that with the log periodogram estimator realized betas are also highly persistent.

Table A.9 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, AR, and ARMA models. It can be seen that compared to the results using the 2ELW estimate, the performance of the FI and FIARMA model slightly decreases, which is probably due to the higher variance of the estimates. However, the forecasts by the FI model still clearly outperform all forecasts by models that do not account for the long-memory characteristics.

A.2.6 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.

Table A.10 shows that decreasing the frequency to 75-minute data decreases the es-

	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	0.593	0.112	0.999	0.996	0.553	0.139	0.994	0.998
$\rho_{i,M}$	0.583	0.099	1.000	0.996	0.585	0.101	1.000	0.996
σ_i	0.598	0.139	0.996	0.977	0.597	0.139	0.996	0.977
σ_M^{-1}	0.553	-	1.000	1.000	0.553	-	1.000	1.000
Panel B: 75-Minute Data								
β_i	0.498	0.123	0.994	0.998	0.457	0.148	0.977	0.998
$\rho_{i,M}$	0.505	0.096	0.999	0.998	0.496	0.099	1.000	1.000
σ_i	0.578	0.140	0.995	0.984	0.578	0.140	0.995	0.984
σ_M^{-1}	0.568	-	1.000	1.000	0.566	-	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 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 $d = 0$ and $d = 1$, respectively, are rejected at the ten 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).

timated memory in realized beta from 0.56 to 0.50. 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 97 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 0.56 to 0.51 and for 15-minute data there is a small increase to 0.58. 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-minute data ranging from 0.58 to 0.60 for stock volatility and 0.55 to 0.57 for the inverse of market volatility.

	RW	AR	ARMA	FI	FIARMA
Panel A: 15-Minute Data					
RMSE	0.2876	0.2713	0.2660	0.2586	0.2595
Best	5	27	124	362	171
vs RW	0	268	320	530	499
vs AR	2	0	139	277	293
vs ARMA	1	8	0	182	180
vs FI	1	4	15	0	13
vs FIARMA	1	4	18	19	0
N	689	689	689	689	689
Panel B: 75-Minute Data					
RMSE	0.3724	0.3421	0.3355	0.3241	0.3248
Best	2	25	90	432	140
vs RW	0	291	360	574	541
vs AR	4	0	156	333	353
vs ARMA	2	9	0	231	222
vs FI	0	1	4	0	12
vs FIARMA	0	2	11	21	0
N	689	689	689	689	689

Table A.11: In analogy to Table 3, this table illustrates the forecast performance of the models for one-month beta forecasts from a rolling estimation window of 100 observations. For the different panels, the realized beta series are now, however, based on 15-minute and 75-minute data. The first row shows 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. 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 at the 10 percent level. Finally, N is the number of investigated stocks. To allow for reliable inference, we exclude all stocks for which we have less than 50 forecasts.

Table A.11 presents the forecast performance of the different models. We find that the ranking of the models stays the same for all considered frequencies. The FI model is 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 changing from 30- to 75-minute data than when changing from 30- to 15-minute data.

Table A.11 further reveals that changing the recording frequency only leads to small changes when comparing the models against each other. For 15-minute data the FI model significantly outperforms the RW, AR, and ARMA models for 77, 40, and 26 percent of the stocks while for 75-minute data this holds for 83, 48, 34 percent of the stocks, respectively.

Overall, the main message of Section 4 remains unchanged: accounting for long-range dependencies significantly improves the forecasting performance for realized betas.

A.2.7 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.12 shows the results for window sizes of 75 and 125 observations.

While the smaller estimation window allows for more stocks to be included in the analysis, it can be seen that the results are qualitatively similar. The forecasts by the FI model perform the best and are outperformed by models that do not account for long-run dependencies only for a tiny number of stocks.

We also consider alternative bandwidths of $m = T^{0.65}$ and $m = T^{0.75}$ for forecasting as a final robustness check in Table A.13 . These results are qualitatively similar as for our main bandwidth choice of $m = T^{0.7}$.

	RW	AR	ARMA	FI	FIARMA
Panel A: Rolling Window of 75 Observations					
RMSE	0.3102	0.2914	0.2864	0.2764	0.2775
Best	4	33	78	446	208
vs RW	0	263	318	612	570
vs AR	3	0	168	367	384
vs ARMA	2	22	0	235	241
vs FI	0	0	3	0	11
vs FIARMA	0	2	9	28	0
N	769	769	769	769	769
Panel B: Rolling Window of 125 Observations					
RMSE	0.3017	0.2800	0.2746	0.2676	0.2683
Best	4	19	120	330	130
vs RW	0	255	302	479	433
vs AR	4	0	114	205	201
vs ARMA	3	15	0	116	113
vs FI	0	1	11	0	13
vs FIARMA	0	1	17	25	0
N	603	603	603	603	603

Table A.12: In analogy to Table 3, this table illustrates the forecast performance of the models for one-month beta forecasts from a rolling estimation window of 75 as well as 125 observations. The first row shows 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. 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 at the 10 percent level. Finally, N is the number of investigated stocks. To allow for reliable inference, we exclude all stocks for which we have less than 50 forecasts.

	RW	AR	ARMA	FI	FIARMA
Panel A: Bandwidth $m = T^{0.65}$					
RMSE	0.3149	0.2942	0.2878	0.2797	0.2806
Best	3	23	108	367	188
vs RW	0	271	343	558	506
vs AR	3	0	155	270	295
vs ARMA	1	5	0	166	172
vs FI	0	2	13	0	21
vs FIARMA	0	2	12	24	0
N	689	689	689	689	689
Panel B: Bandwidth $m = T^{0.75}$					
RMSE	0.3149	0.2942	0.2878	0.2794	0.2801
Best	6	30	100	399	154
vs RW	0	271	343	564	534
vs AR	3	0	155	310	313
vs ARMA	1	5	0	192	189
vs FI	0	4	7	0	10
vs FIARMA	0	4	9	20	0
N	689	689	689	689	689

Table A.13: In analogy to Table 3, this table illustrates the forecast performance of the models for one-month beta forecasts from a rolling estimation window of 100 observations. FI and FIARMA model now calculated using d estimates of the 2ELW estimator calculated with bandwidths of $m = T^{0.65}$ and $m = T^{0.75}$. The first row shows 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. 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 at the 10 percent level. Finally, N is the number of investigated stocks. To allow for reliable inference, we exclude all stocks for which we have less than 50 forecasts.

A.3 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 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.

- **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.

- **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 Non-Durables, "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.

- **Marked Cap** (Banz, 1981) is the current market capitalization of a firm. Market capitalization is computed as the product of the stock price and the number of shares outstanding. In regressions, we take the natural logarithm to remove the extreme iSkew in this variable.

- **Momentum** (Jegadeesh & Titman, 1993) is the cumulative stock return over the period from $t-12$ until $t-1$.

- **Profitability** (Fama & French, 2015) is a firm's operating profitability. Operating profitability is revenues minus cost of goods sold minus selling, general, and administrative expenses minus interest expense, all divided by current book equity. As for BtM, we assume that accounting data become available by the end of June of year t.

- **Short interest (RSI)** (Boehme et al., 2006) is the ratio of short interest of a firm, obtained from Compustat, over the number of shares outstanding. If available, we use the short interest as of the end of month t, otherwise we use the last observation recorded in that month.

References

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

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

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.

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.

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

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

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

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

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.

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.

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

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.

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.

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

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

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

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.

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

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