Received: 3 April 2019 | Revised: 20 April 2020 | Accepted: 16 June 2020

DOI: 10.1002/for.2713

RESEARCH ARTICLE

WILEY

Estimating the volatility of asset pricing factors

Janis Becker | **Christian Leschinski**

Leibniz University Hannover, Germany

Correspondence
Janis Becker, Koenigsworther Platz 1, 30167 Hannover, Germany.
Email: janisbecker92@gmail.com

Abstract

Models based on factors such as size or value are ubiquitous in asset pricing. Therefore, portfolio allocation and risk management require estimates of the volatility of these factors. While realized volatility has become a standard tool for liquid assets, this measure is difficult to obtain for asset pricing factors such as size and value that include smaller illiquid stocks that are not traded at a high frequency. Here, we provide a simple approach to estimate the volatility of these factors. The efficacy of this approach is demonstrated using Monte Carlo simulations and forecasts of the market volatility.

KEYWORDS

asset pricing, factor models, realized volatility, volatility forecasting

JEL CLASSIFICATION

G11; G12; G32

1 | INTRODUCTION

Volatility permeates finance. It is central for everything from risk management to asset allocation. The fact that volatility is unobserved therefore poses a special challenge to practitioners that has been alleviated by the increased availability of high-frequency data and the advent of realized volatility, which led to major improvements of volatility estimates relative to GARCH models. For economy wide risk factors, such as the size and value factors used in asset pricing, however, obtaining high-frequency data proves to be difficult, so that one still relies on models based on squared returns (cf. e.g. He, Zhu, & Zhu, 2015; Moreira & Muir, 2017).

In this paper we propose a methodology to overcome this issue and estimate factor volatility with a precision comparable to that of realized volatility estimates. This is achieved by constructing approximate high-frequency returns of the respective risk factors.

There is a wide consensus that the cross section of asset returns is best described by factor models that proxy for economy wide risk factors. In addition to the established market, size, and value factors of Fama and French (1993), and the momentum factor of Carhart (1997), a plethora of anomalies has been uncovered in the literature that largely failed to attain the status of additional factors (cf. Stambaugh & Yuan, 2016). Recently, Fama and French (2015), Hou, Xue, and Zhang (2015), and Stambaugh and Yuan (2016) suggest investment, profitability, and mispricing factors that subsume a large proportion of these anomalies.

In a simplified form these factors are constructed as follows. First, all stocks in the asset universe are sorted according to some firm characteristic. Then, two value weighted portfolios are formed from those stocks whose firm characteristics fall into the highest and lowest $x\%$-quantile. The factor return is then obtained as the return from buying one of these portfolios and selling the other.

For risk management and portfolio formation purposes it is, however, not only the return but also the volatility of these factors that is of interest. Return volatility is a key variable for the pricing of options, speaks directly to the risk-return trade-off central to portfolio allocation, and even finds its way into government regulations.

For liquid individual assets the unobservability of volatility has been alleviated through the increased availability of high-frequency data and the advent of realized volatility. Given that returns of the asset can be observed frictionless in arbitrarily small time intervals, realized volatility provides a consistent estimate of the quadratic variation of the stock return. For a review of these concepts cf. Andersen and Benzoni (2009).

While this approach is straightforward for individual assets, the calculation of realized volatilities for empirical asset pricing factors is challenging. This is because the COMPUSTAT and CRSP data bases that are typically used to construct the factor returns do not provide high-frequency data. To calculate realized factor volatilities, it would therefore be necessary to match the stocks in these data bases with those from a high-frequency data provider.

This is the approach considered by Aït-Sahalia, Kalnina, and Xiu (2019). It is, however, not straightforward. High-frequency data is typically only available for the most liquid stocks that are traded regularly in short time intervals. The CRSP portfolios that are used to construct empirical factor models, on the other hand, contain much more illiquid stocks that are simply not traded often enough to calculate realized volatilities. Furthermore, high-frequency data bases are not necessarily free of survivorship bias, and finally — even if these hindrances would not exist — the matching of data bases typically constitutes a large effort and there tend to be non-negligible matching errors.

Practitioners or researchers that need to estimate factor volatilities therefore either use squared returns as a volatility measure as for example in Moreira and Muir (2017), or estimate the underlying volatility process through a GARCH model as for example in He et al. (2015). Both approaches have major drawbacks. Squared returns provide an unbiased but inconsistent estimate of the true variance and were the standard measure considered in the GARCH literature prior to the emergence of realized volatility. It is, however, well known that squared returns are extremely noisy. Andersen and Bollerslev (1998) show that, despite the high degree of persistence in stock return volatility, even the true model is only able to explain five to ten percent of the daily fluctuation in squared returns. Volatility estimates based on GARCH models, on the other hand, have a lower variance, but they are biased and inconsistent if the model is misspecified.

The main contribution of this paper is therefore to propose an estimation method for factor volatility that is close in precision to realized volatility. Our approach is applicable whenever the researcher has access to daily factor return series and some high-frequency data base. The idea is to approximate the factor return using a linear combination of the returns in the data base. In the first step, an appropriate linear combination is estimated using ridge regression. In the second step, the bias of the approximate factor is corrected, before the realized volatility of this approximate factor is calculated and used as an estimate for the volatility of the actual factor.

The details of this procedure are discussed in Section 2. We demonstrate the validity of our approach in a Monte Carlo study in Section 3. The empirical validity and usefulness of this approach for the estimation and prediction of volatility is demonstrated in Section 4. First, we analyze the relationship between our estimate and the squared returns for the factors considered by Fama and French (2015) and Carhart (1997) and show that both are estimates of the same underlying volatility process. Second, we consider the example of the market factor where we can use the realized volatility of the S&P 500 to evaluate the accuracy of volatility forecasts. Here, we find that using our measure improves forecasts of the factor volatility considerably compared to squared returns and GARCH-type models. Conclusions are discussed in Section 5.

2 | ESTIMATING FACTOR VOLATILITY

If asset returns are driven by a given factor model, then it holds true that the return of each asset is a linear combination of the returns of these factors and an idiosyncratic error term. Since our procedure is based on daily and high-frequency data, we assume that expected stock and factor returns are zero, so that they do not contain risk premia.

According to this model, the return of asset i at time t, $r_{it} = \frac{P_{it}}{P_{it-1}} - 1$ with P_{it} being the price of the asset at t, is given by

$$r_{it} = \sum_{k=1}^{K} \Lambda_{ik} f_{kt} + \varepsilon_{it}. \tag{1}$$

Here, f_{kt} is the return of factor $k = 1, \dots, K$ at time t, Λ_{ik} is the loading of the ith asset on the kth factor, $\varepsilon_{it} \sim (0, \sigma_\varepsilon^2)$, and $i = 1, \dots, N$. It is assumed that the ε_{it} have limited cross-sectional and serial dependence and that they are independent of all the Λ_{ik} and f_{kt}.

Conversely, it follows that the return of each factor can be approximated by a linear combination of the asset returns. For suitable β_{ik}, we therefore have

$$f_{kt} = \sum_{i=1}^{N} \beta_{ik} r_{it} + \nu_{kt}, \tag{2}$$

where ν_{kt} represents the approximation error which can be expected to be small for large N, since the idiosyncratic errors ε_{it} in (1) average out.

The rationale behind this approach becomes clear if we rewrite model (1) for a vector of N assets. With $R_t = (r_{1t}, \dots r_{Nt})'$, $F_t = (f_{1t}, \dots, f_{Kt})'$, $\Lambda_i = (\Lambda_{i1}, \dots, \Lambda_{iK})'$, $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$ and $\Lambda = (\Lambda_1, \dots, \Lambda_N)'$, we obtain $R_t = \Lambda F_t + \varepsilon_t$. If Λ was known (and $\Lambda'\Lambda$ invertible), we could estimate F_t, by $(\Lambda'\Lambda)^{-1}\Lambda' R_t = F_t + (\Lambda'\Lambda)^{-1}\Lambda'\varepsilon_t = F_t + \varepsilon_t^*$.

Since Λ is $N \times K$, and ε_t is $N \times 1$, ε_t^* is $K \times 1$. Therefore, every element of ε_t^* is a weighted average of the innovation terms $\varepsilon_{1t}, \dots, \varepsilon_{Nt}$ and the vector ε_t^* converges to zero by a suitable law of large numbers (cf. Stock & Watson, 2011 for a related discussion of cross-sectional averaging and statistical factor models).

The coefficient vector $\beta_k = (\beta_{1k}, \dots, \beta_{Nk})'$ in (2) corresponds to the kth row of the matrix $(\Lambda'\Lambda)^{-1}\Lambda'$. Since the returns f_{kt} of the observed factors are readily available, the problem in estimating β_k is that it is N dimensional and therefore potentially very variable if the number of observed days T is not large enough. In fact, it is likely that $N > T$ in empirical applications, so that standard estimation methods cannot be applied.

Our objective is not to recover which stocks are part of the portfolios that are used to derive the factor returns. Instead, we want to obtain a good approximation of the factor returns in terms of mean squared error (MSE). We therefore resort to regularization and estimate β_k using ridge regression. The estimator is given by

$$\hat{\beta}_k = \underset{\beta_{1k}, \dots, \beta_{Nk}}{\arg\min} \left\{ \sum_{t=1}^{T} \left(f_{kt} - \sum_{i=1}^{N} \beta_{ik} r_{it} \right)^2 + \gamma \sum_{i=1}^{N} \beta_{ik}^2 \right\}, \quad (3)$$

with $\gamma > 0$. This is a least squares estimator with an additional penalty term that shrinks the coefficients towards zero. The size of the penalty term depends on the parameter γ that can be selected using cross validation. Here, we select γ so that the out-of-sample mean squared error between the observed and estimated factor is minimized in 10-fold cross-validation. While the introduction of the penalty term introduces some bias, the rationale behind ridge regression is that for suitable γ, the reduction in variance outweighs the size of the bias, so that $\hat{\beta}_k$ is more accurate than the OLS estimator in terms of the mean squared error. Moreover, γ lowers the effective degrees of freedom, so that $N > T$ is permitted if γ is sufficiently large.

Holding the weights $\hat{\beta}_k$ in the linear combination constant then allows to obtain approximate high-frequency factor returns. Denote the m-th of M intraday returns of stock i on day t by $r_{it}^{(m)}$, then the m-th intraday return of factor k on day t is given by

$$\hat{f}_{kt}^{(m)} = \sum_{i=1}^{N} \hat{\beta}_{ik} r_{it}^{(m)}. \quad (4)$$

This allows for a realized-volatility-type estimation of the daily factor volatility V_{kt}.

The approach is subject to two sources of bias. On the one hand, regularization shrinks the coefficients towards zero so that the volatility is underestimated. On the other hand, the variance of the coefficient estimates can be translated to the volatility estimate, which causes a positive bias. To correct for these biases, we include an auxiliary regression step. We calculate the predicted daily values of the factors $\hat{f}_{kt} = \hat{\beta}_k' R_t$ based on (2) and then use ordinary least squares to estimate $f_{kt} = \delta \hat{f}_{kt} + \eta_{kt}$, where η_{kt} is assumed to be a mean-zero martingale difference sequence. Since $V_{kt} = \delta^2 Var(\hat{f}_{kt}) + \sigma_{\eta_k}^2$, we can use the estimated coefficient $\hat{\delta}$ and the residual variance estimate $\hat{\sigma}_{\eta_k}^2$ to correct for the bias.

Consequently, an unbiased estimator for V_{kt} analogous to realized volatility (RV) is given by

$$\hat{V}_{kt} = \hat{\delta}^2 \sum_{m=1}^{M} \left(\log \hat{f}_{kt}^{(m)} + 1 \right)^2 + \hat{\sigma}_{\eta_k}^2. \quad (5)$$

We refer to $\hat{V}_{kt}$ as the Ridge-RV estimator.[1]

To summarize, our method proceeds as follows:

1. Regress the daily factor return f_{kt} on the daily returns of the N stocks in the data base to obtain the coefficient vector $\hat{\beta}_k$ from (3).
2. Estimate the auxiliary regression model $f_{kt} = \delta \widetilde{f}_{kt} + \eta_{kt}$.
3. Obtain estimates $\hat{f}_{kt}^{(m)}$ of the intraday returns of the factors using (4).
4. Estimate the volatility of the factor from the estimated intraday returns $\hat{f}_{kt}^{(m)}$ and the estimated coefficients $\hat{\delta}$ and $\hat{\sigma}_{\eta_k}^2$ using the Ridge-RV estimator in (5).

It should be noted that it is not necessary to have high-frequency returns of all stocks that are part of the original portfolios used to construct the asset pricing factors. As long as the assumed empirical asset pricing model is a linear factor model and it is a good approximation of the true underlying process, a large number of stocks should have non-zero loadings on the factor. For example, the return of the size factor can be estimated from large stocks that have negative loadings on the size factor. High-frequency observations of small illiquid stocks are not required.

[1]Note that when speaking of volatility, some researchers refer to the variance and others to the standard deviation of asset returns. By defining $\hat{V}_{kt}$ as in (5), we implicitly follow Andersen and Benzoni (2009) and Aït-Sahalia, Mykland, and Zhang (2011) and refer to volatility as the variance of asset returns. Performing the analyses in Section 3 and 4 for $\sqrt{\hat{V}_{kt}}$ leads to qualitatively similar results.

3 | MONTE CARLO SIMULATION

To demonstrate the validity of the Ridge-RV estimator, we conduct a simulation study that is tailored to resemble the setup in the empirical applications in Section 4.

It is well known that stock volatilities tend to have long memory and are well described by fractionally integrated processes (cf. Andersen, Bollerslev, Diebold, & Labys, 2001). A fractionally integrated process X_t is given by

$$(1-B)^d X_t = \nu_t, \tag{6}$$

where B defined by $BX_t = X_{t-1}$ is the lag operator, ν_t is a short memory process, and $-1/2 < d \leq 1$. The fractional difference operator $(1-B)^d$ is defined in terms of generalized binomial coefficients. For details confer the original contributions of Granger and Joyeux (1980) or Hosking (1981). A process that fulfills (6) — such as the well known ARFIMA model — is referred to as $I(d)$. Standard short memory processes are included for $d = 0$ and unit root processes are obtained for $d = 1$.

To resemble these long-memory patterns in the daily volatilities V_{kt} of the K factors, we use the long-memory stochastic volatility framework of Breidt, Crato, and De Lima (1998) and simulate T daily observations (with 250 burn-in observations) for each factor using

$$V_{kt} = \exp(X_{kt}), \text{ with } X_{kt} \sim ARFIMA(0, d, 0).$$

The log-volatilities therefore follow a fractionally integrated model. Applying the exponential function guarantees that all volatilities are positive. The V_{kt} obtained this way are used as the true daily volatilities.

Based on these, we subsequently draw M intraday factor returns $f_{kt}^{(m)} \overset{iid}{\sim} N(0, V_{kt}/M)$ for each day and factor. The daily factor returns are obtained as $\sum_{m=1}^{M} f_{kt}^{(m)}$, so that they have volatility V_{kt}. Using these intraday factor returns, we can simulate intraday returns of N stocks. In analogy to Equation (1), the mth return of stock i at day t evolves as

$$r_{it}^{(m)} = \sum_{k=1}^{K} \Lambda_{ik} f_{kt}^{(m)} + \varepsilon_{it}^{(m)},$$

with $\varepsilon_{it}^{(m)} \overset{iid}{\sim} N\left(0, \sigma_\varepsilon^2/M\right)$ being a noise component. As for the daily factor returns, daily stock returns are obtained as the sum over the M intraday returns so that $r_{it} = \sum_{m=1}^{M} r_{it}^{(m)}$.

All parameters are chosen such that the situation in our empirical application in Section 4 is replicated as closely as possible. This means we consider $K = 6$ factors whose correlation matrix matches the correlation matrix of the market, size, value, profitability, investment, and momentum factors considered there, we chose the memory parameter d to be 0.6 for all factors as the literature suggests the memory parameter of volatility to be in this region (cf. Wenger, Leschinski, & Sibbertsen, 2018), we simulate $M = 78$ intraday returns which corresponds to five minute stock data, the factor loadings Λ_{ik} used for the simulation of stock returns are given by regression estimates of the factor loadings of $N = 500$ randomly chosen stocks that were in the S&P 500 at some point in the last 20 years, and σ_ε^2 evolves as the residual variance of this regression. Moreover, we set $T = 750$.

Based on this simulated data we then apply the procedure described in Section 2 based on Equations (3) to (5). Using the intraday factor returns $f_{kt}^{(m)}$, we can also compute the actual realized volatility. As a comparison, we further fit a GARCH(1,1) and a FIGARCH(1,d,1) model and we consider the squared daily factor returns as an estimate of V_{kt}, too.

The results from 1,000 Monte Carlo repetitions can be found in the upper panel of Table 1 that shows the bias compared to the true volatility V_{kt} and the root mean squared error (RMSE) of all the procedures considered. The results are qualitatively similar for all factors and indicate RV and Ridge-RV to be the best estimators of the true volatility process. They are both unbiased and exhibit a similar degree of variance resulting in comparable RMSEs.

The squared returns are unbiased, but their large variance leads to an RMSE that is several times larger than that of the Ridge-RV estimator. The GARCH model cannot remedy the noise problem, and is biased since it does not allow for long memory but the data generating process is $I(d)$. The FIGARCH model achieves an improvement since it allows for long memory, but it is still too noisy resulting in a RMSE six times that of the Ridge-RV estimator.

As a robustness check we repeat the same simulation but with a stochastic volatility ARMA(1,1) process, since it is still often assumed that short-memory GARCH-type models allow for an accurate description of the volatility process. The parameter values used for the simulation are obtained via estimation of an ARMA(1,1) for the respective Ridge-RV series.

The results are shown in the lower panel of Table 1. It can be seen, that Ridge-RV still performs comparable to the infeasible RV estimate and it is considerably better than the competitors. In situations where the intraday returns of a portfolio cannot be observed, the Ridge-RV estimator is therefore the best available choice.

4 | VOLATILITY ESTIMATION AND FORECASTING

In the following, we consider the market (MKT), size (SMB), and value (HML) factors included in the 3-factor model of Fama and French (1993), the profitability (RMW) and investment (CMA) factors added in the 5-factor model of Fama and French (2015), and the momentum factor (MOM) included by Carhart (1997). These factors are com-

TABLE 1 Simulation results: RMSE × 1000 and Bias × 1000 for different volatility estimation approaches

		F1	F2	F3	F4	F5	F6
ARFIMA(0,d,0)							
RV	RMSE	1.042	1.195	1.088	1.278	1.612	1.238
	Bias	0.002	-0.005	0.007	0.000	-0.011	-0.006
Ridge-RV	RMSE	1.046	1.202	1.082	1.322	1.611	1.235
	Bias	0.002	-0.011	0.001	0.016	-0.016	-0.014
Squared Return	RMSE	8.383	9.787	8.708	12.188	12.486	10.163
	Bias	0.000	0.014	-0.041	0.073	0.103	-0.010
GARCH(1,1)	RMSE	8.337	10.401	8.717	13.266	14.237	10.738
	Bias	0.998	1.234	1.042	1.440	1.708	1.234
FIGARCH(1,d,1)	RMSE	6.257	7.424	6.460	8.207	9.079	7.811
	Bias	0.020	0.019	-0.039	0.053	0.113	0.002
ARMA(1,1)							
RV	RMSE	0.440	0.440	0.435	0.440	0.439	0.439
	Bias	0.001	-0.000	-0.001	0.000	0.001	0.001
Ridge-RV	RMSE	0.435	0.463	0.440	0.477	0.452	0.450
	Bias	0.005	-0.022	-0.021	-0.013	0.021	-0.020
Squared Return	RMSE	3.792	3.838	3.735	3.760	3.709	3.752
	Bias	0.002	0.012	-0.005	0.002	-0.008	-0.002
GARCH(1,1)	RMSE	2.215	2.212	2.201	2.225	2.212	2.216
	Bias	0.007	0.016	-0.001	0.006	-0.005	0.000
FIGARCH(1,d,1)	RMSE	2.221	2.224	2.207	2.232	2.217	2.220
	Bias	0.010	0.021	0.002	0.011	-0.002	0.004

Note. The true volatility processes of the six factors (F1, F2,..) evolve as $V_{kt} = \exp(X_{kt})$, with $X_{kt} \sim ARFIMA(0, d, 0)$ respectively $X_{kt} \sim ARMA(1, 1)$. Moreover, the correlation matrix of the simulated processes matches the correlation matrix of the six factors considered in the empirical application.

monly used in the asset pricing literature and their validity is widely accepted. Daily returns of these factors are freely available on the homepage of Kenneth R. French.

In addition to the daily factor returns we require daily stock returns r_{it} and high-frequency returns $r_{it}^{(m)}$ for the estimation of (2) and the calculation of approximate 5-minute factor returns $f_{kt}^{(m)}$ from (4).

Since it is common to calculate realized volatilities from 5-minute returns, we extract five-minute prices of all stocks that were part of the S&P 500 at some point between 1996 and 2017 from the *Thomson Reuters Tick History* data base. This results in a total amount of 1,367 stocks that are considered. Since high-frequency data is often subject to minor recording mistakes, it is common practice to apply some form of data cleaning. Here, we adopt the approach of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009), which comprises, among other things, the removal of observations with negative stock prices and abnormal high or low entries in comparison to other observations on the same day.

Due to the long time span, it cannot be expected that the coefficients β_{ik} stay constant over time. The loading of individual stocks on factors can change as competitors are acquired that have a different exposure to market risk, small firms grow into large firms, and growth stocks turn into value stocks as companies mature. We therefore conduct the estimation of the coefficient vector $\hat{\beta}_k$ according to (3) in a rolling window of size W. For the factors MKT, SMB, HML, and RMW which are based on firm characteristics that are relatively stable over time we set $W = 750$. The factors MOM and CMA that are based on more dynamic features are estimated in a window of size $W = 125$. Results for other values of W are qualitatively similar and available upon request.

To demonstrate the empirical validity of our factor volatility estimates, the next section shows a number of model diagnostics. Afterwards, Section 4.2 demonstrates that volatility forecasts can be improved by using our measure.

4.1 | In-sample volatility estimates and model diagnostics

When trying to evaluate the performance of the Ridge-RV estimator, we face the problem that the true volatility process is unobserved and realized volatilities are not available for the factors. Only squared returns can be observed. We therefore consider a number of model diagnostics that demonstrate the satisfactory performance of our procedure, before turning to the application in Section 4.2.

Figure 1 plots the logarithms of the squared returns and our volatility estimate over time. Two main observations can be made. First, our measure is comoving with the squared factor returns, which is a first indication that both measures estimate the same underlying volatility process.

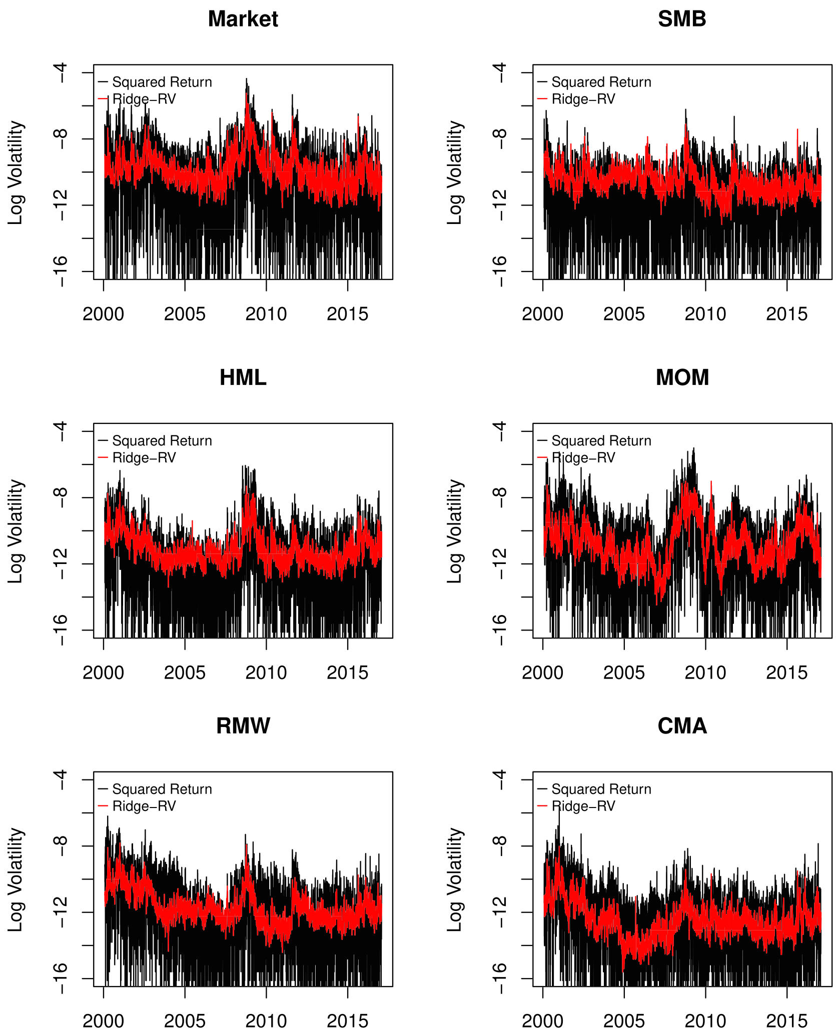

FIGURE 1 Time series plots of the logarithms of Ridge-RV and squared returns for the six factors [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Ridge regression results: coefficient of determination R^2 in percent for the ridge regression as considered in (2)

	MKT	SMB	HML	MOM	RMW	CWA
R^2	98.63	88.68	88.78	90.10	88.42	89.71

Larger values of the squared factor returns are associated with larger values of the Ridge-RV and vice versa. This holds for all factors and all time periods. Second, the Ridge-RV appears to be far less perturbed than the squared returns.

The Ridge-RV estimate is based on the approximation of the factor of interest by a linear combination of stock returns. If this approximation in (2) is sufficiently accurate, so are those in (4) and (5). A first indication of the quality of the estimate can therefore be obtained from the coefficients of determination R^2 in a regression of f_{kt} on $\widetilde{f}_{kt}$. Table 2 shows that the measure is above 88% for all of the six considered factors indicating a high precision of the estimates.

Since squared returns and Ridge-RV are both estimates of the same unobserved volatility process, they can both be understood as differently perturbed versions of it. An approach to test the validity of the Ridge-RV estimator in this empirical setup is therefore to test for fractional cointegration between the squared returns and $\hat{V}_{kt}$. Fractional cointegration is a natural generalization of cointegration to fractionally integrated series. Two time series X_t and Y_t are said to be fractionally cointegrated, if both are $I(d)$ and there exists a linear combination $X_t - \alpha - \beta Y_t = u_t$, so that u_t is $I(d-b)$ for some $0 < b \leq d$. As in standard cointegration, both series must be highly persistent and they are (fractionally) cointegrated if a linear combination of them has reduced persistence. The extension lies in the fact that the reduction of persistence does not have to be from $I(1)$ to $I(0)$, but can be from $I(d)$ to $I(d-b)$.

When modeling volatility time series it is common practice to work with the log of the volatility series since it is better approximated by the normal distribution (cf. Andersen et al. 2001). If $\log V_{kt}$ denotes the true volatility process, then $\log f_{kt}^2 = \log V_{kt} + \omega_{kt}$ and $\log \hat{V}_{kt} = \log V_{kt} + \eta_{kt}$, where ω_{kt} and η_{kt} are the respective estimation errors. Therefore,

TABLE 3 Fractional cointegration test results: test statistics and critical values for the tests by Chen and Hurvich (2006) (CH) and Souza et al. (2018) (SRF)

	MKT	SMB	HML	MOM	RMW	CMA	
CH	4.438	2.542	3.673	4.923	1.897	4.780	(1.697)
SRF	3.807	1.381	3.020	3.263	2.751	3.770	(1.960)

Note. Here, the null of no fractional cointegration between log-squared returns and log-Ridge-RVs is tested against the alternative of fractional cointegration. The values in brackets are critical values at the five percent level,

if $\log V_{kt}$ is $I(d)$, then $\hat{V}_{kt}$ can only be a reasonable estimator of $\log V_{kt}$, if it is fractionally cointegrated with $\log f_{kt}^2$, so that $\log \hat{V}_{kt} - \log f_{kt}^2 = \eta_{kt} - \omega_{kt}$ is $I(d-b)$.

To formally test the hypothesis of fractional cointegration between both volatility measures, we apply the tests of Chen and Hurvich (2006) and Souza, Reisen, Franco, and Bondon (2018) for the null hypothesis of no fractional cointegration. Under the alternative a fractional cointegration relationship exists.

Table 3 reports the results of the tests. As expected from Figure 1, the test by Chen and Hurvich (2006) rejects the null of no fractional cointegration for all factors and the test by Souza et al. (2018) rejects the null for all factors, except for the size factor. Therefore, we can conclude that squared returns and Ridge-RV are fractionally cointegrated.

All of the statistics presented so far show that our Ridge-RV estimator works well. However, as discussed above, the evidence provided is indirect, since the actual volatility process is unobserved. For the market factor, we can, however, conduct one experiment that provides insight into the actual accuracy of the Ridge-RV estimate. Even though we do not have realized volatilities for the market factor, it is well known that the value weighted CRSP return, which is generally regarded as the best available market proxy, is highly correlated with the return of the S&P 500. The correlation coefficient is about 99 percent, meaning that the direction of the variation and its scaling over time is essentially the same. For the S&P 500 it is possible to obtain intraday prices, meaning that we can calculate realized volatilities. Consequently, we can compare our estimate of the market volatility with the realized volatility of the S&P 500. As Andersen and Benzoni (2009) stress, the realized volatility is the natural ex post measure of the underlying volatility process to consider. Figure 2 shows that the two measures are close to identical. In fact, they have a correlation of 91.4 percent, are fractionally cointegrated, and regressing our volatility estimate on the realized volatility yields a slope of 0.99, even though it is significantly different from 1.

We therefore conclude that our estimate is appropriate for describing the volatility of the market factor. Even though the results in Tables 2 and 3 indicate that the procedure works slightly better for the market factor than for the other factors, the degree of precision obtained for the market implies that the Ridge-RV should still be a good estimate for the volatility of the other factors.

It should be noted, however, that the procedure is based on the assumption that the factors under consideration are actually relevant for the cross section of stock returns. This may be an issue if one wishes to apply the procedure to any of the many weak factors discussed in the literature.

4.2 | Out-of-sample forecasts of market volatility

For portfolio allocation and risk management purposes, accurate forecasts are needed in addition to ex post and on-line estimates of the factor volatility. In this section we therefore compare the performance of forecasts using squared returns and GARCH-type models with those using Ridge-RV.

When trying to evaluate these forecasts, we again face the problem that the true factor volatility is unobserved. As shown by Andersen and Bollerslev (1998), considering squared returns as a proxy for the true factor volatility when evaluating volatility forecasts is not suitable since the tremendous amount of noise in the return generating process inevitably causes a poor performance of the forecasting models. On the other hand, it seems tautological to show superior performance of our Ridge-RV measure when considering it as the true factor volatility. We therefore proceed as in the previous section and conduct a forecast comparison for the volatility of the market factor, where we can use realized volatilities of the S&P 500 to proxy for the true factor volatility. This makes for a fair comparison, since both types of models (Ridge-RV and models based on squared returns) do not use the realized volatilities of the S&P 500 in any way.

The Ridge-RV is predicted using the HAR model of Corsi (2009). We refer to this forecast as the HAR-Ridge-RV model. As a benchmark, we also consider the standard HAR-RV model, which is possible for the market but not for the other factors. It can thus be interpreted as the "infeasible" model that we try to approximate when predicting factors such as SMB, HML, or others. As feasible benchmark models we include a GARCH(1,1) and due to the long range dependence in factor volatility we also use a FIGARCH(1,d,1) model, as proposed by Baillie, Bollerslev, and Mikkelsen (1996), fitted to the squared

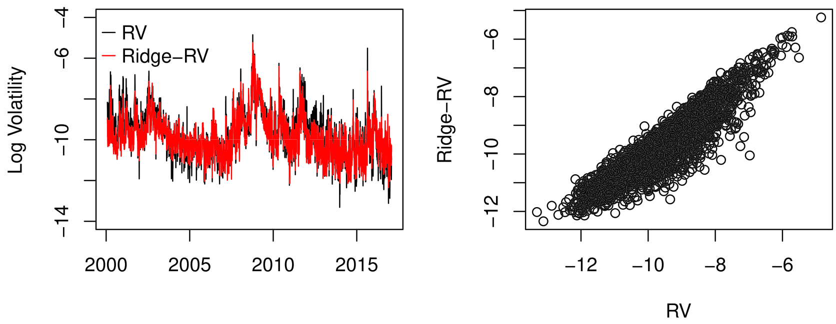

FIGURE 2 Both plots display the Ridge-RV estimate of market factor volatility and the true volatility of the market factor approximated by the realized volatility of the S&P 500. While the left plot shows the two measures over time, the right plot displays a scatter plot [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 Forecast results: RMSE $\times 10^3$, QLIKE, and R^2 from Mincer-Zarnowitz regressions for the competing models and different forecast horizons

	1-Step			5-Step			22-Step		
	RMSE	QLIKE	R^2	RMSE	QLIKE	R^2	RMSE	QLIKE	R^2
GARCH(1,1)	0.211	0.298	0.492	0.171	0.238	0.596	0.181	0.259	0.513
FIGARCH(1,d,1)	0.222	0.299	0.444	0.168	0.230	0.606	0.175	0.250	0.535
HAR-RV	0.175	0.215	0.603	0.120	0.195	0.745	0.113	0.248	0.710
HAR-Ridge-RV	0.175	0.220	0.608	0.123	0.196	0.731	0.120	0.240	0.664

Note. GARCH and FIGARCH use squared returns to forecast the market factor volatility, HAR-Ridge-RV uses the Ridge-RV estimate, and HAR-RV uses the true volatility given by the realized volatility of the S&P 500.

returns. All estimations are carried out in a rolling window of 750 observations.

For the evaluation of the forecasts, we consider the RMSE and the QLIKE loss function, since Patton (2011) shows that these are the only commonly used measures that preserve the true ordering of the forecasts if they are evaluated on a perturbed volatility proxy. Furthermore, we report the coefficient of determination R^2 from Mincer-Zarnowitz Mincer and Zarnowitz (1969) regressions given by

$$\frac{1}{h}\sum_{j=1}^{h} RV_{t+j} = b_0 + b_1 \frac{1}{h}\sum_{j=1}^{h} \hat{V}_{t+j} + u_{kt}.$$

Here, RV_{t+j} is the observed volatility approximated by the realized volatility of the S&P 500, $\hat{V}_{t+j}$ is the predicted volatility based on all information available in t, h is the forecast horizon, and u_{kt} is an error term. Consequently, larger values of R^2 imply that the forecasts are performing better in predicting the true volatility.

Table 4 shows the results of this forecasting exercise for 1-step, 5-step, and 22-step forecasts. It can be seen that for all forecasting horizons and for all evaluation measures, the HAR-Ridge-RV model performs better than all of the models based on squared daily returns. For 1-step forecasts, for example, the RMSE $\times 10^3$ of the HAR-Ridge-RV model is 0.175, QLIKE is 0.220, and the R^2 is 0.608, while for the GARCH(1,1) model, which is the best model using squared returns, the RMSE $\times 10^3$ is 0.211, QLIKE is 0.298, and the R^2 is 0.492. Due to the averaging, the forecasting performance of the models becomes slightly better on longer horizons. The ranking of the models, however, stays the same.

When comparing the forecasts based on our volatility estimate with the HAR-RV forecasts based on the realized volatility of the S&P 500, it can be seen that the two models deliver qualitatively similar results.

Consequently, forecasts based on Ridge-RV achieve their objective to approximate those that are obtained if realized volatilities are available and they strongly outperform forecasts of the market volatility compared to models using squared returns. For factors other than the market, where realized volatilities are not available, they can therefore be expected to provide results that are far better than standard approaches.

5 | CONCLUSION

Although the volatilities of economy wide risk factors such as the size and value factors of Fama and French (1993) are of importance for risk management and portfolio allocation purposes, the development of methods for their estimation has lagged behind that for liquid individual assets or indices, where intraday returns are available.

The Ridge-RV approach suggested in this paper circumvents the lack of high-frequency data for factor returns and provides a volatility measure that is closely related to realized volatility. This is achieved by approximating the daily factor returns by a linear combination of the returns of assets for which intraday returns are available. Holding the weights in the linear combination constant then allows to obtain approximate high-frequency factor returns that are the basis for the estimation of the factor volatility. Due to the large number of parameters in the linear combination that have to be estimated, it is necessary to apply a regularized estimation method such as ridge regression.

This approach to estimate the factor volatility is subject to two sources of bias. On the one hand, regularization shrinks the coefficients towards zero so that the volatility is underestimated. On the other hand, the variance of the coefficient estimates can be translated to the volatility estimate, which causes a positive bias. To correct for these biases, we include an auxiliary regression step.

The subsequent applications to the market, size, value, momentum, investment, and profitability factors demonstrate that the proposed measure performs well in practice and outperforms competing approaches such as GARCH-type models. We therefore find that adopting the proposed approach has the potential for significant improvements in asset allocation decisions and risk management.

ACKNOWLEDGEMENT

We want to thank Philipp Sibbertsen, Fabian Hollstein, Christoph Wegener and Gregor von Schweinitz for their helpful comments. Furthermore, we are grateful for financial support from the University of Hannover through the program “Wege in die Forschung”.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in figshare at https://doi.org/10.6084/m9.figshare.7946438, reference number 7946438

ORCID

Janis Becker https://orcid.org/0000-0001-5104-6330
Christian Leschinski https://orcid.org/0000-0002-1456-6211

REFERENCES

Aït-Sahalia, Y., Kalnina, I., & Xiu, D. (2019). High-frequency factor models and regressions. *Journal of Econometrics*, *216*(1), 86–105.

Aït-Sahalia, Y., Mykland, P., & Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, *160*(1), 160–175.

Andersen, T., & Benzoni, L. (2009). Realized volatility, *Handbook of Financial Time Series* (pp. 555–575). Berlin, Heidelberg: Springer.

Andersen, T., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*, 885–905.

Andersen, T., Bollerslev, T., Diebold, F., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, *96*(453), 42–55.

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

Barndorff-Nielsen, O., Hansen, P., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, *12*(3), C1–C32.

Breidt, J., Crato, N., & De Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, *83*(1-2), 325–348.

Carhart, M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57–82.

Chen, W., & Hurvich, C. (2006). Semiparametric estimation of fractional cointegrating subspaces. *The Annals of Statistics*, *34*(6), 2939–2979.

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

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

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

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

He, Z., Zhu, J., & Zhu, X. (2015). Multi-factor volatility and stock returns. *Journal of Banking & Finance*, *61*, 132–149.

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

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, *28*(3), 650–705.

Mincer, J., & Zarnowitz, V. (1969). The evaluation of economic forecasts, *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3–46). Cambridge. NBER.

Moreira, A., & Muir, T. (2017). Volatility-managed portfolios. *The Journal of Finance*, *72*(4), 1611–1644.

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

Souza, I., Reisen, V., Franco, G., & Bondon, P. (2018). The estimation and testing of the cointegration order based on the frequency domain. *Journal of Business & Economic Statistics*, *36*(4), 695–704.

Stambaugh, R., & Yuan, Y. (2016). Mispricing factors. *The Review of Financial Studies*, *30*(4), 1270–1315.

Stock, J., & Watson, M. (2011). Dynamic factor models, *Oxford Handbook on Economic Forecasting* (pp. 1–44). Oxford: Oxford University Press.

Wenger, K, Leschinski, C., & Sibbertsen, P. (2018). The memory of volatility. *Quantitative Finance and Economics*, *2*(1), 137–159.

AUTHOR BIOGRAPHIES

Janis Becker received his PhD in economics in 2020 from the University of Hannover. He now works as a data scientist for Hannover Re SE.

Christian Leschinski received his PhD in economics in 2016 from the University of Hannover. He now works as a data scientist for SPRING Axel Springer Digital News Media GmbH & Co. KG.

How to cite this article: Becker J, Leschinski C. Estimating the volatility of asset pricing factors. *Journal of Forecasting*. 2021;40:269–278. https://doi.org/10.1002/for.2713