

Risk Everywhere: Modeling and Managing Volatility

Tim Bollerslev

Duke University, NBER, and CREATES

Benjamin Hood

AQR Capital Management

John Huss

AQR Capital Management

Lasse Heje Pedersen

AQR Capital Management, Copenhagen Business School, and CEPR

Based on high-frequency data for more than fifty commodities, currencies, equity indices, and fixed-income instruments spanning more than two decades, we document strong similarities in realized volatility patterns within and across asset classes. Exploiting these similarities through panel-based estimation of new realized volatility models results in superior out-of-sample risk forecasts, compared to forecasts from existing models and conventional procedures that do not incorporate the similarities in volatilities. We develop a utility-based framework for evaluating risk models that shows significant economic gains from our new risk model. Lastly, we evaluate the effects of transaction costs and trading speed in implementing different risk models. (*JEL* C22, C51, C53, C58, G11)

Received March 7, 2016; editorial decision February 3, 2018 by Editor Andrew Karolyi. Authors have furnished an Internet Appendix, which is available on the Oxford University Press Web site next to the link to the final published paper online.

Measuring, forecasting, and controlling risk are at the heart of financial economic theory and practice. Investors and asset managers generally seek to maximize return while limiting risk. Many traders also have specific risk

We thank participants at the 2016 Volatility Institute Conference at NYU, Stern and Aaron Brown, Eric Ghysels, Antti Ilmanen, Ari Levine, Yao Hua Ooi, Scott Richardson, and Liuren Wu for useful comments and discussions and especially Andrew Karolyi (the Editor) and two anonymous referees for their detailed comments that greatly improved the paper. Pedersen gratefully acknowledges support from the European Research Council (ERC grant no. 312417) and the FRIC Center for Financial Frictions (grant no. DNRFF102). Hood, Huss, and Pedersen are affiliated with AQR Capital Management, and Bollerslev has a consulting relationship with AQR. AQR is a global investment management firm that may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR. Supplementary data can be found on *The Review of Financial Studies* Web site. Send correspondence to Lasse Heje Pedersen, Copenhagen Business School, Solbjerg Plads 3, 2000 Frederiksberg, Denmark. E-mail: lhp.fi@cbs.dk.

© The Author(s) 2018. Published by Oxford University Press on behalf of The Society for Financial Studies. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com doi:10.1093/rfs/hhy041

Advance Access publication May 22, 2018

limits or risk targets. Traders in a bank, for example, often face risk limits, while hedge funds often tell their investors that a particular fund is expected to realize a certain risk level. So-called “risk parity strategies” are explicitly designed to equate the risk stemming from the different investments included in a portfolio, rather than the capital allocated to the different components. Such investors therefore continually monitor global risk as it evolves and change their portfolios in response, reducing notional exposures as risk rises and increasing positions as risk fades.

This paper presents a framework for measuring, modeling, and forecasting risk across global assets and asset classes. Our main result is that exploiting commonality in risk everywhere has a statistically and economically significant impact. Our results are based on a broad data set of realized volatilities constructed from high-frequency data for more than fifty instruments across four asset classes. We find that our new risk models, new panel-based estimation techniques, and our global volatility factor—all designed to exploit the strong commonalities observed in the volatilities across assets and asset classes—result in statistically significant out-of-sample forecast improvements and nontrivial utility gains compared to more conventional individually estimated asset-specific risk models.

We start by constructing a comprehensive database comprising high-frequency intraday data from different global markets spanning more than two decades and covering 20 commodities, 21 equity indices, 8 fixed-income futures, and 9 currencies. We compute the realized volatilities (RV) for each day and asset in our sample. When qualitatively comparing the estimated RV s, the differences in risk levels across assets and asset classes immediately stand out. However, when we normalize each asset’s daily realized volatilities by their respective sample averages, striking similarities emerge. Indeed, these “normalized risk measures” have almost identical unconditional distributions and similar highly persistent autocorrelation structures when comparing across assets and asset classes. Hence, the volatilities of different assets—equities, bonds, commodities, or currencies—appear to behave almost the same over time. Going one step further, we document strong volatility spillover effects both within and across different markets and geographical regions. The existence of spillover effects and commonalities in the dynamic dependencies is, of course, well known from the already existing volatility literature and the estimation results obtained with traditional GARCH and stochastic volatility models (see, e.g., Taylor 2005; Andersen et al. 2006; and the references therein).

Next, we build risk forecasting models explicitly designed to exploit these strong similarities in the distributions of the volatilities across and within asset classes. The formulation of our models are motivated by the heterogeneous autoregressive (HAR) model of Corsi (2009) and draws on insights from the mixed data sampling (MIDAS) approach of Ghysels et al. (2006) and Ghysels, Sinko, and Valkanov (2007). First, we show how to simultaneously estimate

risk models across *many* assets using panel regressions that add power by exploiting the similarities in the cross-asset risk characteristics.¹ An important step needed to allow such a panel-based estimation is to “center” the models and eliminate the asset-specific intercept terms to ensure that all parameters are “scale-free” in the sense that they do not depend on the level of risk. Second, we introduce new “smooth” realized volatility models, in which the forecasted future volatilities depends on the past volatilities in a way that is continuous and decreasing in the lag lengths, thereby eliminating nonmonotonicities arising from estimation noise and predictable jumps in the risk forecast as time passes. Our preferred specification, which we denote the heterogeneous exponential realized volatility model (HExp for short), in particular, is based on a simple mixture of exponentially weighted moving average (EWMA) factors. Third, to account for the volatility spillover effects and strong commonalities observed not just across different assets but also across different geographical regions, we augment the asset-specific HExp model with a lagged “global” risk factor.

Looking at the in-sample results, we find that all of the *RV* models that we consider perform well compared to models that “only” use daily returns. By construction, when looking in-sample, the models that are tailored to each asset separately have larger predictive power in terms of R^2 than models that enforce a common risk model across assets. However, when looking at out-of-sample predictability, the models that impose common parameters generally perform better. In particular, enforcing common parameters across models within each asset class produces higher average out-of-sample R^2 s than individually estimated models. Even more surprisingly, enforcing common parameters not just within but across *all* asset classes, the properly “centered” risk models result in even higher average out-of-sample R^2 s. The basic HExp model and the HExp model with the “global” risk factor (termed “HExpGI”) result in the highest average out-of-sample predictability among all of the models, suggesting that the commonality and “smoothness” embedded in the HExp formulations ensure a robustness beyond that of the standard existing risk models.

Last, but not least, we present a simple framework for quantifying the utility benefits of risk modeling. Our approach is linked to the literature that seeks to assess the utility benefits of return predictability in the presence of empirically realistic transaction costs and other practical implementation issues (see, e.g., Balduzzi and Lynch 1999; Sangvinatsos and Wachter 2005; Lynch and Tan 2010). It is also related to the work of Fleming, Kirby, and Ostdiek (2001, 2003), and the idea of using a quadratic utility function to evaluate the benefits of

¹ The use of a panel-based estimation technique to enhance the efficiency of the individual forecasts parallels the use of Bayesian estimation procedures more generally, as exemplified by Karolyi (1993), who rely on Bayes-Stein shrinkage for improving the forecasts of individual stock return volatilities based on the cross-sectional dispersion.

volatility timing.² In contrast to all of these approaches that explicitly depend on forecasts of—and/or realizations of—both the future returns and volatilities, our method exclusively focuses on volatility forecasting. Specifically, we consider the expected utility of an investor with mean-variance preferences that trades an asset with a constant Sharpe ratio. The investor's optimal portfolio adjusts the position size to keep a constant volatility (and this "risk target" naturally depends on the investor's risk aversion). Correspondingly, the investor's utility is directly related to the volatility: the investor achieves the maximum utility by successfully targeting a constant risk level, while the utility decreases with the volatility-of-volatility. Hence, risk models that help the investor achieve more accurate volatility forecasts are associated with higher levels of utility.

We show that, in this situation, under realistic assumptions about the Sharpe ratio and the investor's risk target, using the HExp risk model augmented with our global risk factor (HExpGI) is worth about 48 basis points (bps) per year relative to using the best possible static risk model. Put differently, the assumed investor would in principle pay 48 bps of her/his wealth each year to have access to the HExpGI risk model developed here rather than using a static risk model. The utility benefit of HExpGI is also significant, but smaller, when compared to other sophisticated risk models. For example, the utility gain of the HExpGI model relative to a simple risk model based on daily data is 19 bps per year and the utility gain over a 21-day rolling average over realized volatilities is 8 bps per year. These utility gains are of the same magnitude as institutional asset management fees. While such fees are often thought as compensation for higher expected returns ("alpha"), our results show that the quality of the risk model can be equally important. Importantly, these benefits remain when we take realistic transaction costs into account since the new "smooth" risk models not only produce more accurate risk forecasts but also more stable forecasts resulting in less spurious trading than risk models based on daily data.

In summary, we contribute to the literature by exploring commonality in volatility across a broad set of asset classes, introducing a new class of risk models, showing how risk models can be centered to allow panel-based estimation, developing a utility-based framework to evaluate the economic importance of risk models, and, finally, combining all these to empirically showing that our HExp risk model estimated using our panel method produces statistical and economically significant gains relative to standard risk models.

² See also the benefits of volatility-timed time-series momentum factors (Moskowitz et al. 2012), value and momentum factors (Asness et al. 2013; Barroso and Santa-Clara 2015; Daniel and Moskowitz 2016), and other volatility-managed factors (Moreira and Muir 2017). Our utility framework can isolate the benefit coming from stabilizing volatility, as separate from predicting the return (or Sharpe ratio).

1. Realized Volatilities: Data Sources and Construction

There is a long history in finance of heuristically quantifying the ex post volatility based on the sum of intra-period squared returns.³ This approach may be formally justified by the theory of quadratic variation and the notion of ever-finer sampled returns over fixed time intervals, or so-called “in-fill asymptotic arguments” (see, e.g., the discussion and references in Andersen et al. 2013).

1.1 Realized volatilities and quadratic variation

To formally lay out the basic idea underlying the realized volatility concept, we let the unit time interval correspond to a day. The realized variation defined by the summation of high-frequency intraday squared returns,

$$RV_t \equiv \sum_{i=1}^{1/\Delta} [\log(P_{t-1+i\Delta}) - \log(P_{t-1+(i-1)\Delta})]^2, \quad (1)$$

then consistently estimates the quadratic variation, that is, the true variation, on day t as the number of intraday observations increases ($1/\Delta \rightarrow \infty$), or equivalently the length of the intraday return interval decreases ($\Delta \rightarrow 0$). This effectively renders the daily variation directly observable on an ex post basis.⁴ The volatility over longer, say weekly or monthly, horizons may similarly be estimated by summing the intraday squared returns over a week or a month or, equivalently, by summing the daily realized volatilities RV_t over the relevant longer multiday horizons.

1.2 Data sources and “cleaning”

Our data covers a total of 58 different assets across commodities (20), equities (21), fixed income (8), and foreign exchange (9). The asset universe comprises global equity index futures (both developed and emerging markets), global developed fixed-income futures, commodity futures, and spot market foreign exchange rates. Our specific choice of assets is dictated by liquidity concerns and correspondingly the availability of reliable high-frequency intraday prices.

Our primary source of data for equities, fixed income, and commodities is the Thomson Reuters Tick History (TRTH) database. To extend the history for some of the assets, most notably fixed income and commodities, we use data from TickData.com (TDC). For the foreign exchange data, we exclusively rely on Olsen Data (OD). The data for all of the assets run through September 2014. The start of the sample period differs across assets, with some starting as

³ French, Schwert, and Stambaugh (1987) and Schwert (1989), for instance, rely on the sum of daily squared returns in their construction and modeling of monthly U.S. equity volatilities, while Hsieh (1991) estimates models for daily volatilities constructed from 15-minute S&P 500 returns.

⁴ Consistent with the common use of the two terms in the extant literature, we will use RV_t as the daily realized variation or the daily realized volatility interchangeably and also sometimes explicitly use $RV_t^{1/2}$ as the daily realized volatility.

early as October 1992. In general, the data for commodities are available the earliest, followed by equities and fixed income, with foreign exchange having the shortest time span. Table A1 in the appendix summarizes the exact start dates for all of the assets and the relevant data sources.⁵ Multiple futures contracts for the same underlying asset, but with different expiration, trade at the same time, and new contracts are opened as others expire. We focus on the most liquid contract, “rolling” from one contract to another at a regular schedule as further discussed in the appendix. We organize the resultant single time series of returns for each asset into minute bars based on the last observation prior to the end of each minute. For the TRTH and OD databases, we use the mid-quote price (average between the bid and ask price). For the TDC data we use the observed trade price (TDC does not provide quote-level data prior to 2010). To avoid “polluting” the high-frequency data with quote changes that occur during illiquid periods, we only use the minute bars for which there are at least one valid trade within that minute. Lastly, having organized all of the data into minute bars, we apply a series of “sanity filters” to clean out any obvious data errors. These filters are further discussed in the Appendix.

1.3 Overnight returns and intraday sampling

It is well established that volatility tend to be higher during exchange trading hours than during nontrading hours (see, e.g., French and Roll 1986). The theory underlying the consistency of the realized volatility measure portrays prices as evolving continuously through time. In actuality, of course, most markets close on weekends and certain holidays, change their trading hours, and sometimes experience “ghost” hours, where liquidity is very poor despite the markets technically being open. Accordingly, we only retain the trading hours for which the liquidity is sufficiently high to ensure a reasonable quality of the high-frequency data. We use the Financial Calendars (FinCal) database for market open and close times, together with so-called “liquidity plots,” to delineate the periods of actively operating markets; the appendix provides further details. Having determined the period for which reliable high-frequency data are available, we simply add the corresponding “overnight” squared returns to the daily realized volatilities constructed from the “intraday” squared returns to obtain an RV measure for the whole day.⁶

We rely on a common 5-minute sampling frequency for calculating the intraday RV s for all of the assets. This choice directly mirrors the sampling frequency used in much of the existing realized volatility literature. It may be

⁵ The staggered start dates are easily accommodated by our panel-based estimation procedures discussed below. We do not restrict the sample sizes to be the same for all assets.

⁶ Following Hansen and Lunde (2005b), we also experimented with “optimally” combining the squared overnight returns and the intraday RV , proportionally scaling the intraday RV , and separately modeling the overnight squared returns. None of these more complicated procedures clearly dominated the approach of simply adding the overnight squared returns to the intraday RV s.

justified by the volatility signature plots (Andersen et al., 2000) discussed in the appendix. To further enhance the efficiency of the RV estimates, we average the five different daily RV s obtained by starting the day at the first five unique 1-minute marks. A number of other consistent realized volatility estimators, requiring the choice of additional tuning and/or nuisance parameters, have been proposed in the literature.⁷ However, the theoretical comparisons in Andersen, Bollerslev, and Meddahi (2011) and Ghysels and Sinko (2011) show that from a theoretical forecasting perspective, the simple subsampled 5-minute RV estimator that we rely on here performs on par with or better than all of these more complicated estimators. The empirical study by Liu, Patton, and Sheppard (2015), comparing more than 400 different RV estimators across multiple assets, similarly concludes that “it is difficult to significantly beat 5-minute RV .”

2. Risk Characteristics Everywhere

To help guide the specification of empirically realistic risk models, we want to understand the distributional characteristics of the risks both within each of the four asset classes, equities, bonds, commodities, and currencies, as well as the similarities and differences across asset classes.

2.1 Unconditional distributions and dynamic dependencies

To begin, Figure 1 shows the time series of annualized realized volatilities for four representative assets, one from each asset class: S&P 500, 10-year Treasury bonds, Crude Oil, and Dollar/Euro. Even though the four volatilities obviously exhibit their own distinct behaviors, there is a clear commonality in the dynamic patterns observed across the four assets, with most of the peaks readily associated with specific economic events.

In spite of the similarities in the general patterns, the overall levels of the volatilities clearly differ across the four different assets. This is further evidenced by Figure 2, which plots the unconditional distribution of the daily realized volatilities for the same four representative assets. As the figure shows, Crude Oil is the most volatile on average, followed by the S&P 500, and the Dollar/Euro exchange rate. The volatility of 10-year Treasury bonds is by far the lowest.

This same ranking carries over to the four asset classes more generally. In particular, looking at the summary statistics for the daily realized volatilities averaged across each of the assets within each of the four asset classes reported in Table 1, the average annualized volatility for commodities and equities equal 25.4% and 20.6%, respectively, compared to 10.3% for foreign exchange, and just 3.1% for fixed income.

⁷ These include the two-scale RV of Zhang, Mykland, and Aït-Sahalia (2005), the kernel-based RV of Barndorff-Nielsen et al. (2008), and the pre-averaged RV of Jacod et al. (2009).

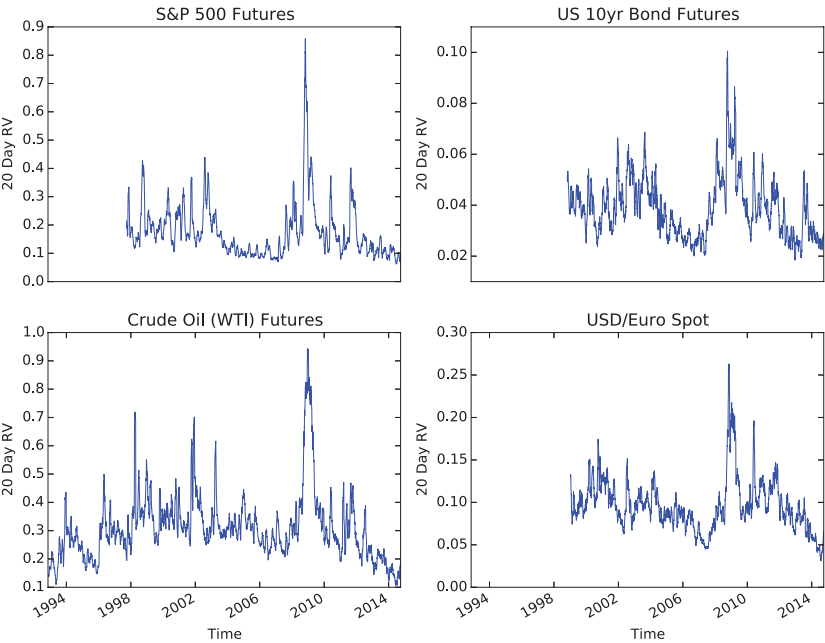


Figure 1
Monthly realized volatilities
This figure shows the time series of 20-day average realized volatilities (annualized) for four representative assets: S&P 500 Futures, U.S. 10-year Bond Futures, Crude Oil (WTI) Futures, and USD/Euro Spot.

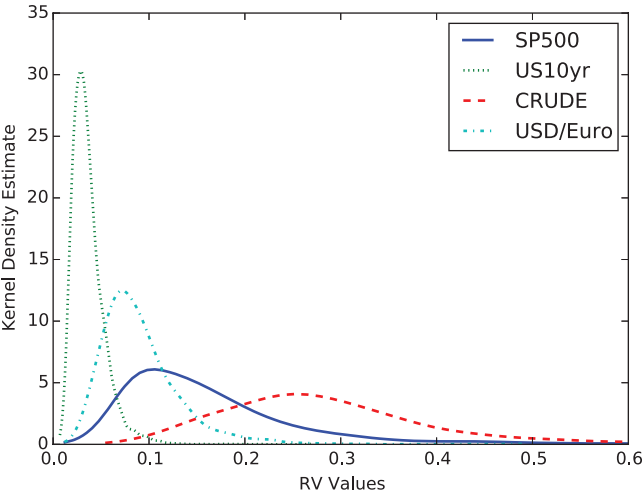


Figure 2
Unconditional daily RV distributions
This figure shows kernel density estimates of the unconditional daily realized volatility (annualized) for four representative assets: S&P 500 Futures, U.S. 10-year Bond Futures, Crude Oil (WTI) Futures, and USD/Euro Spot.

Table 1
RV summary statistics

	Commodities	Equities	Fixed income	Foreign exchange
Mean	25.4	20.6	3.1	10.3
SD	12.6	13.7	1.5	5.7
Skewness	2.6	3.4	2.3	3.1
Excess kurtosis	16.9	22.9	11.6	18.5
Maximum	185.6	186.6	19.4	74.1
95th percentile	47.8	44.8	5.8	20.4
50th percentile	22.7	17.0	2.8	9.0
5th percentile	11.6	8.2	1.5	4.6
Minimum	4.9	3.0	0.6	1.2
1-day autocorr.	0.516	0.707	0.481	0.517
20-day autocorr.	0.362	0.480	0.347	0.415
100-day autocorr.	0.195	0.228	0.197	0.221
250-day autocorr.	0.115	0.105	0.073	0.104
Number of assets	20	21	8	9
Avg. number of obs.	5,407	3,812	4,042	3,351
Earliest start date	Oct. 22, 1992	Jan. 3, 1996	Jan. 3, 1996	Jan. 1, 1999
Latest start date	Jan. 3, 1996	Dec. 13, 2005	Sept. 26, 2000	Jan. 1, 2004

The table presents summary statistics for daily realized volatilities averaged across all assets within a given asset class. All of the volatility numbers are reported in annualized percentage units.

These differences in the mean levels of the volatilities across the different assets and asset classes are, of course, well known. More interesting features arise when we consider the volatilities normalized by their sample mean, $RV_t/Mean(RV_t)$, corresponding to the risk of a leveraged (or deleveraged) position. For example, if $Mean(RV_t)=0.5$, then the normalized volatility measure corresponds to the risk of a position that is always leveraged 2-to-1. Hence, normalizing each individual contract by its own average volatility is equivalent to measuring risk on a *common scale*, in the sense that each position is leveraged to the same common average risk level.

Interestingly, the unconditional distributions of these daily normalized realized volatilities are remarkably similar, both across assets and asset classes. Indeed, Figure 3 shows that the sampling distributions of the four representative assets are obviously very close.⁸ The unconditional distributions for the normalized volatility of each of the individual assets within each of the four asset classes are also very similar, as reported in the Online Appendix.

It is important to note that the “width” of these distributions are not similar by construction, as would be the case if we normalized by subtracting the mean and dividing by the standard deviation, $[RV_t - Mean(RV_t)]/std(RV_t)$, which would match both the mean and standard deviation by construction (or, said differently, such a normalization implies a loss of two degrees of freedom, rather than just one for our normalization). Hence, the fact that the $RV_t/Mean(RV_t)$ normalized distributions are so similar is not hard-wired,

⁸ Although seemingly very close, pairwise Kolmogorov-Smirnov tests easily reject that the four sampling distributions in Figure 3 are identical, with a p -value of .0087 for commodities and fixed income, and p -values below .0001 for all of the other pairwise comparisons.

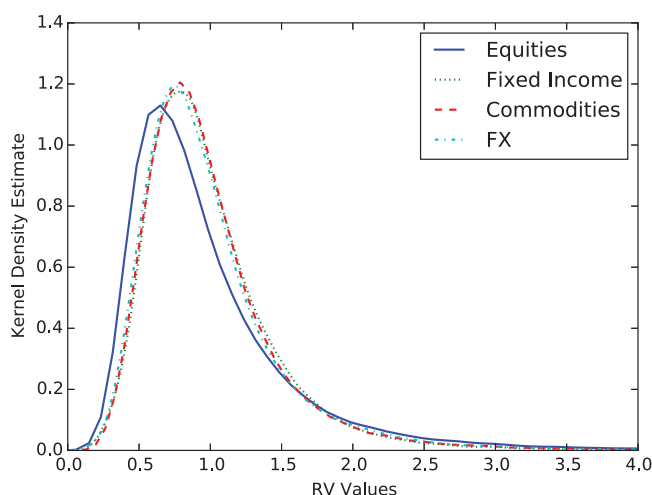


Figure 3
Normalized unconditional daily RV distributions

This figure shows kernel density estimates of the normalized daily realized volatility (annualized) of all assets within each of the four asset classes. In particular, we show the density of each asset's daily RV divided by its average daily RV to account for level differences in individual asset volatilities.

but rather direct evidence that risks do indeed behave similarly across asset classes. Importantly, these similarities also imply that risk parity strategies designed to match the average volatility of different assets and/or asset classes will not only equate the average volatility levels but also effectively the entire unconditional distributions of the ex post realized volatilities for the leveraged positions.

These commonalities in the unconditional distributions carry over to the general dynamic dependencies. The autocorrelations for the daily realized volatilities averaged across the different assets within each of the four asset classes shown in Table 1 and Figure 4 make clear that the general patterns and decay rates are very similar. This is, again, in line with the equities perhaps being slightly different for the short-term lags.⁹ Similar dynamic dependencies, of course, have been extensively documented in the burgeoning volatility literature.¹⁰

⁹ Because of the varying degrees of measurement errors in the realized volatilities for the assets within the four asset classes, the levels of the autocorrelations for the shortest lags are not directly comparable (see also the discussion in Hansen and Lunde 2014).

¹⁰ The book by Taylor (1986), in particular, provides some of the earliest empirical evidence explicitly highlighting the similarities in the autocorrelations of absolute and squared daily returns across assets and asset classes. GARCH and stochastic volatility models also typically result in very similar dynamic parameter estimates for different assets. For instance, applying a GARCH(1,1) model to daily returns, the estimates for α and β in the usual notation of the model are typically around 0.03–0.08 and 0.92–0.97, respectively, with the sum very close to unity (see, e.g., Hansen and Lunde 2005a).

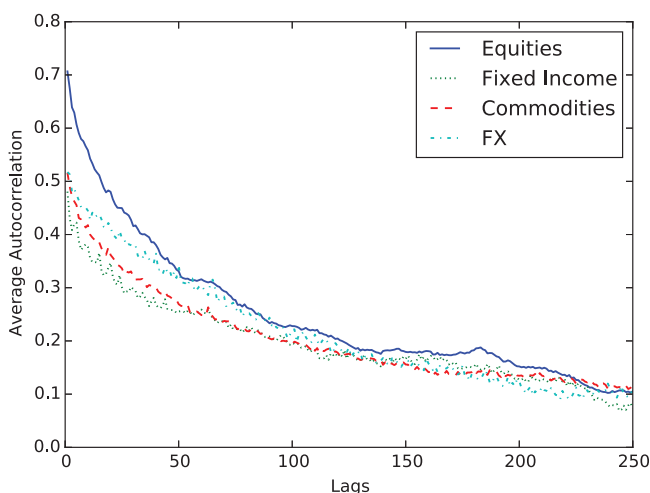


Figure 4

Daily realized volatility autocorrelations

This figure shows the average autocorrelation function of the daily realized volatilities averaged across all of the assets within each of the four asset classes.

2.2 Cross-asset dependencies, spillovers, and global volatility

The averages of the standard sample correlations for the realized volatilities reported in the top panel in Table 2 are all positive. This comovement of risk across assets and asset classes is consistent with the visual impression from the time-series plots for the four representative assets previously discussed in Figure 1. Looking at the actual numbers, commodity volatilities are generally the least correlated, both within the asset class and across other asset classes. That is, the risk of different commodities comove less with each other than do equity risks that comove with each other, and so on, and commodity risks also comove less with the risk of equity, fixed income, and currencies than they comove with each other. In fact, commodity risk comoves about as much across asset classes as within the asset class.

In addition to the within and across asset class correlations, the last row reports the average correlations with a “global risk factor” that we denote *GIRV*. The global risk factor is defined as the average normalized *RV*s across all assets. Since we will also use *GIRV* for forecasting, we construct this factor on an asset-specific basis to prevent any look-ahead biases due to time-zone effects. In particular, for any specific asset, we construct the corresponding *GIRV* so that today’s *GIRV* does not use any data that overlap with tomorrow’s trading hours for the specific asset, lagging by 1 day any asset that would otherwise create such an overlap.¹¹ Based on this lagging convention, on

¹¹ For instance, the day $t - 1$ market hours for U.S. equities overlap with the overnight portion of day t for Australian equities. Hence, in the construction of the global factor used for Australian equities, we shift the *RV* for U.S.

Table 2
RV contemporaneous and partial correlations

	Commodities	Equities	Fixed income	Foreign exchange
<i>A. Contemporaneous correlations</i>				
Commodities	0.277	0.298	0.217	0.355
Equities	–	0.668	0.407	0.554
Fixed income	–	–	0.470	0.433
Foreign exchange	–	–	–	0.710
Global	0.410	0.721	0.547	0.633
<i>B. Partial lead-lag correlations and ΔR^2</i>				
Commodities	0.059	0.063	0.074	0.132
Equities	0.081	0.173	0.122	0.198
Fixed income	0.072	0.078	0.093	0.151
Foreign exchange	0.078	0.108	0.110	0.144
Global	0.171	0.328	0.267	0.456
Commodities	0.005	0.004	0.007	0.017
Equities	0.011	0.021	0.024	0.041
Fixed income	0.006	0.005	0.011	0.016
Foreign exchange	0.007	0.010	0.015	0.014
Global	0.012	0.022	0.033	0.061

The table presents correlations of daily realized volatilities averaged within and across asset classes. Panel A reports the standard contemporaneous correlations. Panel B reports the average partial correlation coefficients obtained from regressions of the daily RVs for the assets within the asset classes indicated in the columns on their own daily lags and the lagged values of the RVs for the assets in the asset classes indicated in the rows (top panel), together with the average absolute percentage increases in the R^2 s compared to regressions of the assets indicated in the columns on their own daily lags only (bottom panel). The “Global” volatility factor is constructed as a weighted average of the daily RVs, as further discussed in the main text.

each day and for each asset i , we compute the asset-specific $GIRV$ as the average normalized RV scaled back to the asset’s own level of volatility; that is, $\left(\frac{1}{J} \sum_{j=1, \dots, J} \frac{RV_{i,j}}{RV_j}\right) \overline{RV}_i$.

This global volatility factor captures well the overall volatility dynamic across asset classes as seen in the last row of the top panel in Table 2. Indeed, as a sign of the strong commonalities in the realized volatilities, the average correlations with this new global risk factor systematically exceed the across asset class correlations. With the exception of foreign exchange, the global risk factor correlations are also all higher than the within-asset class correlations.

The literature is rife with studies seeking to model these strong dependencies and possible volatility spillover effects using multivariate GARCH and related procedures (see, e.g., Engle et al. 1990; Karolyi 1995 for some of the earliest evidence). In contrast to these earlier studies, which rely on parametric volatility models for inferring the dependencies, our use of realized volatilities allow us to directly quantify the strengths of any spillover effects. To this end, the middle panel in Table 2 reports the average partial lead-lag correlation coefficients

equities back by 1 day. The RV for Australian equities, on the other hand, need not be shifted in the construction of the global factor for U.S. equities. To also account for changes in market hours over the sample period, we perform this shifting on a day-by-day basis for each asset.

Table 3
Global volatility correlations

	Sentiment	Unusual sentiment	VRP	News
Global Volatility	−0.123*	0.197**	−0.454**	0.183**
Sentiment	—	0.568**	−0.097	0.171**
Unusual sentiment	—	—	0.051	0.136*
VRP	—	—	—	−0.087

The table reports the monthly (end-of-month) contemporaneous correlations between the global volatility factor *ExpGIRV* formally defined in Section 3.6 and four other variables. *Sentiment* refers to the “orthogonalized” investor sentiment measure of Baker and Wurgler (2006). *Unusual sentiment* is defined as the absolute value of the Sentiment measure minus its sample mean. The variance risk premium (*VRP*) is equal to the difference between the VIX and the realized U.S. equity volatility over the past month. The *News* surprise variable is constructed as the average value of the standardized absolute news announcement surprises observed over the month. Statistical significance at the 5% and 1% levels are indicated by * and **, respectively.

obtained from regressions of the daily *RV*s for each of the assets within the asset classes indicated in the columns on their own daily lagged value and the lagged values of the *RV*s for the assets in the asset classes indicated in the rows. To allow for a scale-invariant interpretation, we further normalize the realized volatility RV_j of any asset j by its sample mean \bar{RV}_j and regress the normalized *RV* for asset j on a constant, its own lag, and the normalized lagged *RV* for asset i , where the latter regression coefficient is the partial correlation of interest. Specifically, the first-order partial autocorrelation between asset j and asset i is the estimated $b_{2,ij}$ coefficient from the regressions $RV_{t,j}/\bar{RV}_j = b_{0,ij} + b_{1,ij}RV_{t-1,j}/\bar{RV}_j + b_{2,ij}RV_{t-1,i}/\bar{RV}_i + u_{t,ij}$, with Table 2 reporting the average of these within-asset classes. The bottom panel shows the resultant average increases in the R^2 s compared to simple first order autoregressions that only control for the own lagged dynamic dependencies; that is, $RV_{t,j}/\bar{RV}_j = b_{0,j} + b_{1,j}RV_{t-1,j}/\bar{RV}_j + u_{t,j}$.

Consistent with the presence of strong cross-market linkages and spillover effects, all of the average partial correlations are positive.¹² Comparing the results across the different asset classes, equity volatilities as a whole tend to exert the largest impact on the other asset class volatilities. Meanwhile, the magnitude of the equity partial correlations and the resultant increases in the R^2 s are all dominated by those of the global risk factor. This naturally raises a few questions: What is behind these linkages? And, in particular, what might explain the dynamic variation in the global risk factor?

The economic forces behind volatility clustering per se remain poorly understood, and a full-fledged analysis of that question is beyond the scope of the present paper. However, in an effort to shed some light on the mechanisms at work, Table 3 reports the monthly (end-of-month) correlations between an exponentially smoothed version of the global volatility factor *ExpGIRV*, as used in our preferred HExpGI model formally defined in Section 3.6, and four

¹² As reported in the Online Appendix, the vast majority of the individually estimated coefficients are also statistically significant.

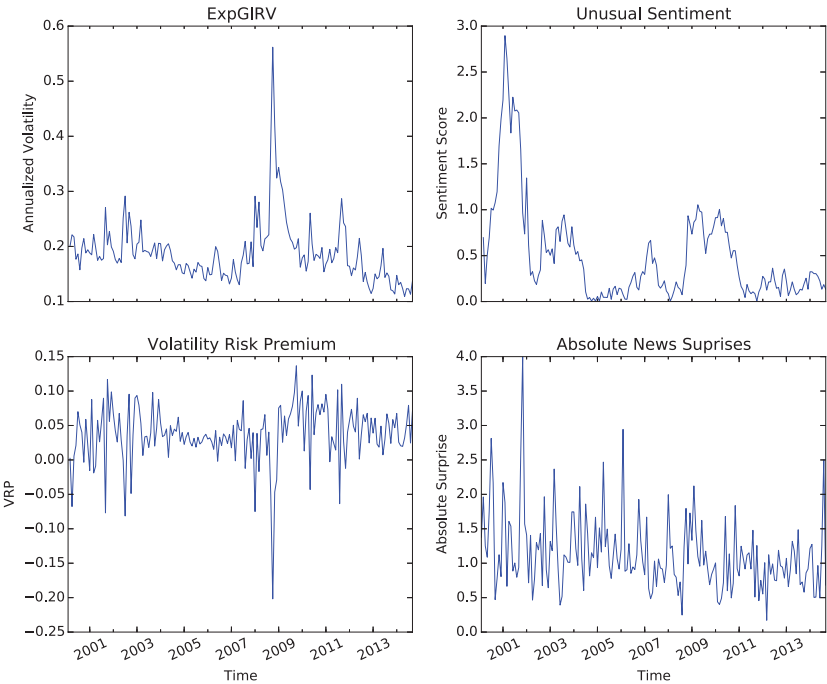


Figure 5
Global volatility, sentiment, VRP, and news
This figure shows the global volatility factor (ExpGIRV), unusual monthly sentiment, the volatility risk premium (VRP), and the absolute news surprise variable, as formally defined in the main text. All of the variables are plotted at a monthly frequency.

other variables naturally related to volatility. Figure 5 plots the global volatility factor, together with three of the four variables.

The first entry in Table 3, in particular, shows that the global volatility factor is negatively correlated with the U.S. investor sentiment index of Baker and Wurgler (2006).¹³ Baker, Wurgler, and Yuan (2012) have previously argued that U.S. investor sentiment is contagious, and that international capital flows may in part account for this contagion. The well-established strong link between trading volume and return volatility (see, e.g., for a survey of some of the earliest empirical evidence Karpoff 1987) may help explain the connection between investor sentiment and global volatility. Further along these lines, Karolyi, Lee, and van Dijk (2012) have previously documented strong commonalities in equity trading volumes across countries, arguably driven by correlated trading activity among institutional investors and common investor sentiment.

¹³ We rely on the “orthogonalized” sentiment measure available on Wurgler’s Web site: <http://people.stern.nyu.edu/jwurgler/>.

The previous correlation portrays a monotonic relationship between volatility and sentiment, possibly driven by correlated trading. However, if “noise” traders acting on sentiment affect prices, unusually high or low levels of sentiment should both be associated with high levels of volatility (see, e.g., Brown 1999). Consistent with this idea, the monthly correlation between the global volatility factor and “unusual” sentiment, defined as the absolute value of the same sentiment measure minus its sample mean, equals 0.197.¹⁴ Thus, whereas the general level of U.S. investor sentiment is negatively correlated with global market volatility, unusual U.S. investor sentiment is positively correlated with global market volatility.

The variance risk premium, formally defined as the difference between the actual and risk-neutral expectation of the future return variation, is naturally interpreted as a measure of aggregate risk aversion (see, e.g., Bakshi and Madan 2006). Supporting the idea that risk aversion, and in turn risk bearing capacity, influence volatility, the global volatility factor is strongly negatively correlated with the U.S. equity variance risk premium.¹⁵

Numerous studies have sought to relate low-frequency variation in return volatility to directly observable macroeconomic variables or indicators (see, e.g., Schwert 1989; Engle and Rangel 2008). The reported relations, however, are weak at best. At the other end of the spectrum, a number of studies have documented a sharp, but short lived, increase in intraday volatility following macroeconomic and other public news announcements (see, e.g., Ederington and Lee 1993; Andersen et al. 2003). The last entry in Table 3 reports the correlation between the global volatility factor and a news surprise variable, constructed as the average of the standardized absolute surprises for five of the most important U.S. macroeconomic news announcements released over the month.¹⁶ Corroborating the idea that the volatility in global financial markets may in part be attributed to “news” about the U.S. economy, the correlation between this news surprise variable and the global volatility factor equals 0.183.

Taken as a whole, the results in Table 3 support the idea that commonalities in trading behavior possibly induced by changes in investor sentiment and/or notions of risk aversion, together with unanticipated news, all serve important roles in accounting for the strong commonalities in the dynamic dependencies observed in the volatilities within and across asset classes. We will not pursue this any further here. Instead, we turn to a discussion of the practical risk models that we use for modeling and forecasting these dynamic dependencies.

¹⁴ Restricting the sentiment measure to just two states, Yu and Yuan (2011) also find that especially high sentiment tends to be associated with high volatility.

¹⁵ We follow Bollerslev, Tauchen, and Zhou (2009) in quantifying the variance risk premium as the difference between the VIX and the realized U.S. equity volatility over the past month.

¹⁶ We follow Andersen et al. (2003) in identifying the five most important regularly scheduled U.S. macroeconomic news announcement as nonfarm payroll, durable goods orders, retail sales, housing starts, and the Philadelphia Fed’s business outlook survey.

3. Risk Modeling

Our new risk models are explicitly designed to incorporate the strong similarities observed in risk characteristics across assets and asset classes.

3.1 Omnibus RV-based risk models: $AR(\infty)$

To facilitate the discussion, it is instructive to consider the omnibus $AR(\infty)$ risk model,

$$RV_{t+1} = b_0 + b_1 RV_t + b_2 RV_{t-1} + \dots + \epsilon_t = b_0 + b(L)RV_t + \epsilon_t, \quad (2)$$

in which the realized volatility on day $t+1$ is determined by a distributed lag of past realized volatilities.¹⁷ The estimation of an infinite number of b_i coefficients implicit in this representation is, of course, not practically feasible, and the different risk models in effect represent alternative ways of restricting the $b(L)$ lag polynomial to allow for its meaningful estimation, as exemplified by the RV -based ARIMA model originally proposed by Andersen et al. (2003), and the MIDAS model advocated by Ghysels, Santa-Clara, and Valkanov (2006) in which $b(L)$ is parameterized in terms of beta functions.

The notion of multiple volatility components, or factors, is commonly used for parsimony, representing the $b(L)$ lag polynomial. The HAR model of Corsi (2009), for example, is based on a weighted sum of a daily, weekly and monthly volatility factors.¹⁸ In this situation, what ultimately matters from a practical forecasting perspective is the factors' ability to capture the influence of the lagged RV 's. Ideally, we want a set of factors that "span" the $b(L)$ lag space well, while still enforcing some commonality and "smoothness," thereby enabling a unified set of factors to be used for all assets and asset classes by simply altering the weights of the different factors.

Regardless of the way in which the $b(L)$ lag polynomial is parameterized, the b_i coefficients are usually estimated on an asset-by-asset basis. This ignores the cross-asset similarities in the dynamic dependencies discussed in the previous section. We therefore also explore the use of panel regression techniques that force the coefficients to be the same within and across different asset classes. As demonstrated below, doing so imbues the resultant risk models with a built-in robustness and statistically significant superior out-of-sample forecast performance.

¹⁷ In addition to the risk models nested in this $AR(\infty)$ representation, we also explored models that decompose the daily RV 's into continuous and jump components (Andersen, Bollerslev, and Diebold, 2007), and models that differentiate between up and down realized semivariances (Patton and Sheppard 2015). For the monthly forecast horizon primarily analyzed below, none of these alternative RV measures and models resulted in superior out-of-sample forecasts. Chen and Ghysels (2011) report largely symmetric "news impact curves" for U.S. aggregate equity indexes at the 1-month horizon.

¹⁸ Related representations based on the combination of multiple distributed lag polynomials also have been proposed in the MIDAS literature and elsewhere (see, e.g., Ghysels et al. 2007).

3.2 “Centering”: Eliminating the level parameter in a robust fashion

Even though one might naturally restrict the dynamic $b(L)$ lag coefficients to be the same across assets to exploit the commonalities in the dynamic dependencies and distribution of the standardized volatilities, $RV_t / \text{Mean}(RV_t)$, the very different volatility levels for different asset classes means that it is unreasonable to force the b_0 intercepts to be the same. To circumvent this and allow for meaningful cross-asset estimation, we “center” the risk models by replacing the intercept with a long-run volatility factor RV_t^{LR} , equal to the expanding sample mean of the daily RV ’s from the start of the sample up until day t . Forcing all of the b_i coefficients to sum to one, including the coefficient for the RV_t^{LR} factor, ensures that the iterated long-run forecasts from the model constructed on day t converges to this day t estimate of the “unconditional” volatility.¹⁹

Although seemingly complicated to implement, this “centering” of the risk models is easily enforced by subtracting the long-run volatility factor from all of the RV s, including the left-hand-side RV forecast target:

$$RV_{t+1} - RV_t^{LR} = b_1(RV_t - RV_t^{LR}) + b_2(RV_{t-1} - RV_t^{LR}) + \dots + \epsilon_t. \quad (3)$$

When the regression is run in this way, the coefficients are free (i.e., need not sum to one), but, if we collect terms for RV_t^{LR} on the right-hand side, then RV_t^{LR} has an implied coefficient of $1 - b_1 - b_2 - \dots = 1 - b(1)$ such that all the implied coefficients do sum to one. By eliminating the level of the volatility, this alternate representation allows for the meaningful estimation of common dynamic b_i coefficients by panel regression techniques. More complicated Bayesian shrinkage-type procedures, in which the dynamic coefficients are allowed to differ across assets, of course, could be applied. However, we purposely restrict the coefficients to be the same within-asset classes or across all assets, to allow for a direct comparison with the individually estimated risk models.²⁰

3.3 Multiperiod and other volatility forecasts

The $AR(\infty)$ model in (2) and the centered version thereof in (3) are directly geared to forecasting the one-day-ahead variance. Longer-run forecasts, say over weekly or monthly horizons, may be obtained by recursively substituting the forecasts for the future daily RV ’s into the right-hand side of the model, subsequently adding up the resultant one, two, three, etc., days-ahead forecasts to achieve the multiperiod RV forecast over the requisite horizon. Instead,

¹⁹ This mirrors the idea of variance targeting in GARCH models first proposed by Engle and Mezrich (1996), in which the intercept in the conditional variance equation is replaced by a (scaled) estimate of the long-run “variance target” to which the forecasts converge. It also resembles the Spline-GARCH model of Engle and Rangel (2008) and the use of a low-frequency volatility component to scale the forecasts.

²⁰ We also experimented with the use of common long-run asset class RV_t^{LR} s, but we did not find that these models resulted in obviously superior forecasts.

a much simpler approach for constructing multi-day-ahead forecasts is to estimate the risk model as such from the start. That is, by replacing the daily variance RV_{t+1} on the left-hand side of the risk model, with the realized variance over the forecast horizon h of interest, say $RV_t^h \equiv \frac{1}{h} \sum_{i=1}^h RV_{t-h+i}$. In the forecasting literature, this approach is commonly referred to as direct as opposed to iterated forecasts.²¹

In particular, for the “monthly,” or 20-day, forecast that we focus on below, and the generic risk model in (2), we have

$$RV_{t+h}^h = b_0^h + b^h(L)RV_t + \epsilon_t^h, \quad (4)$$

with $h=20$. The b_i^h coefficients, which dictate the “speed” of the model, will obviously depend on the forecast horizon, as indicated by the superscripts h . For notational simplicity, however, we will drop this superscript in the sequel. Also, even though ϵ_t^h generally will be serially correlated up to the order of $h-1$, we will simply denote the residuals in all of the models discussed below as ϵ_t for short.²² In theory, if the model for the 1-day-ahead RV_{t+1} in (2) is correctly specified, the iterated forecasts from that model would be the most efficient. However, ample empirical evidence shows that even minor model misspecifications tend to be amplified in iterated volatility forecasts, and, as a result, the direct forecasts produced from a model, such as (4), are often superior in practice (see, e.g., Andersen et al. 2003; Ghysels et al. 2009; Sizova 2011).²³

This same basic idea also may be used for forecasting other functions of the variance, by simply replacing RV_{t+h}^h on the left-hand side in Equation (4) with the volatility object of interest. For instance, the future volatility as opposed to the variance, or the inverse of the variance, are often of primary import. Unless the volatility is constant or perfectly predictable, simply transforming the forecast for the variance will result in a systematically biased forecast.²⁴ We turn next to a brief discussion of the specific risk models, old and new, that we actually rely on.

3.4 HAR models

The original HAR model of Corsi (2009) has proven very successful. It has emerged as somewhat of a benchmark in the financial econometrics

²¹ A similar distinction has been made in the context of option pricing and the use of exactly matched-by-horizon versus cumulated volatility estimates (see, e.g., Karolyi 1993). This issue also has been extensively studied in the MIDAS literature (see, e.g., Ghysels et al. 2009).

²² The use of overlapping daily data in the estimation of the models means that the conventional standard errors for the coefficient estimates will have to be adjusted to account for the overlap (see, e.g., Hansen and Hodrick 1980; Newey and West 1987).

²³ Ghysels, Rubia, and Valkanov (2009) explicitly refer to direct forecasts as MIDAS.

²⁴ By a standard first-order Taylor series expansion, $E(\sqrt{X}) \cong E(X)^{1/2} - \frac{1}{8} \text{Var}(X)$. Hence, $[E_t(RV_{t+h}^h)]^{1/2}$ is an upward-biased forecast of future volatility. Similarly, $E(1/X) \cong 1/E(X) + \text{Var}(X)/E(X)^3$, so that one over $E_t(RV_{t+h}^h)$ will result in a downward-biased forecast for the relevant ratio. The appendix provides some additional discussion and empirical results along these lines.

literature for judging other RV -based forecasting procedures. The model may be succinctly expressed in variance form as

$$RV_{t+h}^h = \beta_0 + \beta_D RV_t + \beta_W RV_t^W + \beta_M RV_t^M + \epsilon_t, \quad (5)$$

where RV_t^W and RV_t^M denote the 5-day (weekly) and 20-day (monthly) realized volatilities, respectively, thus implying a step function for the b_i coefficients in the omnibus $AR(\infty)$ representation in (2).²⁵ In addition to the results from individual asset-by-asset estimation of this “uncentered” HAR model, we report the results from a fixed effects panel-based estimation of a “centered” version of the model, in which we restrict the β_D , β_W and β_M coefficients to be the same, but allow the β_0 coefficients to differ across different assets.

The stepwise nature of the volatility factors employed in the HAR models, imply that the forecasts from the models are subject to potentially abrupt changes as an unusually large/small daily lagged RV drops out of the sums for the longer-horizon lagged volatility factors.²⁶ Our remaining risk models rely on alternative $b(L)$ polynomials for “smoothing” out these problems.

3.5 MIDAS models

The original HAR model may be interpreted as a special case of MIDAS regressions with step functions (see, e.g., the discussion in Andersen et al. 2007, Ghysels et al. 2007, Corsi 2009), while the HAR-free model discussed in the footnote above is closely related to the so-called “U-MIDAS model” (Foroni et al. 2015). In contrast to HAR models, however, one of the main objectives of the MIDAS approach is the specification of “smooth” distributed lag polynomials for representing the dynamic dependencies. Another main theme of the MIDAS literature, of course, relates to the use of data sampled at different frequencies, and the choice of sampling frequency for the regressor variables. Addressing both of those issues, Ghysels, Santa-Clara, and Valkanov (2006) conclude that the direct modeling of high-frequency data does not result in systematically better volatility forecasts compared to the forecasts from a model of the form in (2) based on the daily RV s only.²⁷ They also propose a specific parameterization for the $b(L)$ lag polynomial based on beta functions.

²⁵ Analogous multifactor formulations previously have been used as way to approximate long-memory dynamic dependencies in the context of parametric stochastic volatility models (see, e.g., Gallant et al. 1999). The use of monthly lagged volatility in the daily ARCH model estimated by French, Schwert, and Stambaugh (1987) provides an early precedent to this formulation.

²⁶ In addition to the basic HAR model in (5), we also experimented with a model in which we freely estimated the coefficients for the first six daily lagged RV s, together with an annual volatility factor RV_t^A , defined as the (normalized) sum of the daily RV 's over the past 261 days. The Online Appendix reports the results for this HAR-free model.

²⁷ Echoing this same theme, Clements, Galvao, and Kim (2008) also find that the parameterization implicit in the original HAR model generally results in superior out-of-sample forecasts compared to a freely parameterized MIDAS model for the high-frequency intraday squared returns.

This representation has now emerged as somewhat of a standard in the MIDAS literature.

Thus, directly following Ghysels, Santa-Clara, and Valkanov (2006), we implement a MIDAS model of the form,

$$RV_{t+h}^h = \beta_0 + \beta_1[a(1)^{-1}a(L)]RV_t + \epsilon_t, \quad (6)$$

in which the nonzero coefficients in the $a(L)$ lag polynomial are defined by scaled beta functions:

$$a_i = \left(\frac{i}{k}\right)^{\theta_1-1} \left(1 - \frac{i}{k}\right)^{\theta_2-1} \Gamma(\theta_1 + \theta_2) \Gamma(\theta_1)^{-1} \Gamma(\theta_2)^{-1}, \quad i = 1, \dots, k, \quad (7)$$

where $\Gamma(\cdot)$ denotes the Gamma function. The normalization by $a(1) \equiv a_1 + \dots + a_k$ in Equation (6) ensures that the coefficients in the $[a(1)^{-1}a(L)]$ lag polynomial sum to unity, so that the β_1 coefficient is uniquely identified. Implementation of the model still requires a choice of the cutoff k , and the two tuning parameters θ_1 and θ_2 . Again, directly mirroring Ghysels, Santa-Clara, and Valkanov (2006), we fix the cutoff at $k=50$, and set the tuning parameter $\theta_1=1$. This choice of θ_1 is now also commonly employed in the MIDAS literature more generally. Lastly, following the approach of Ghysels and Qian (2016), we determine the remaining θ_2 tuning parameter by a grid search, in which we profile the predictive R^2 from the model as a function of θ_2 , together with the freely estimated β_0 and β_1 parameters, and choose the value of θ_2 that maximizes the predictability over the full sample.²⁸ As a result, our subsequent predictive MIDAS analysis is not truly out-of-sample. However, the computational burden does not allow us to perform a rolling grid search for the θ_2 parameter.

3.6 HExp models

The “smooth” beta polynomial employed in the MIDAS model avoids the stepwise changes inherent in the forecast from the HAR component-type structure. Further extending this idea, our last set of risk models rely on a mixture of “smooth” exponentially weighted moving averages (EWMA) of the past realized volatilities. Simple EWMA filters with a pre-specified center of mass are often used in practice. Instead, we explicitly estimate the relative importance of different EWMA factors constructed from the past daily RV ’s:

$$ExpRV_t^{CoM(\lambda)} \equiv \sum_{i=1}^{500} \frac{e^{-i\lambda}}{e^{-\lambda} + e^{-2\lambda} + \dots + e^{-500\lambda}} RV_{t+1-i}, \quad (8)$$

²⁸ In addition to this benchmark MIDAS model, we also experimented with a series of more elaborate specifications based on the mixture of multiple beta polynomials. The Online Appendix provides further details concerning these additional models.

where λ defines the decay rate of the weights, and $CoM(\lambda)$ denotes the corresponding center-of-mass, $CoM(\lambda) = e^{-\lambda} / (1 - e^{-\lambda})$.²⁹ The center-of-mass of each exponential RV measure effectively summarizes the “average” horizon of the lagged realized volatilities that it uses. Conversely, for each center-of-mass, we can compute the corresponding rate of decay as $\lambda = \log(1 + 1/CoM)$. Hence, we can think of $\lambda = \log(1 + 1/125) = 0.008$ as an annual $ExpRV$ risk measure because the corresponding center of mass is 125 trading days, that is, about half a year, just like an annual equal-weighted average realized volatility, RV_t^A .

Focussing on similar horizons to the ones used in the HAR model augmented with an annual volatility factor, we rely on the exponential RV 's to “span” the universe of past RV 's in a way that is both parsimonious and “smooth,” mixing four $ExpRV_t^{CoM(\lambda)}$ factors with λ chosen to equate the center-of-mass to 1, 5, 25, and 125 days, respectively. Further “centering” the model around the expanding long-run volatility factor, we obtain the following new risk model:

$$RV_{t+h}^h - RV_t^{LR} = \sum_{j=1,5,25,125} \beta_j (ExpRV_t^j - RV_t^{LR}) + \epsilon_t. \quad (9)$$

We will refer to this specification as the Heterogeneous Exponential, or HExp, model. This model, of course, is still nested in the omnibus distributed-lag model in (3). In contrast to the MIDAS specification above, this HExp model uses pre-specified volatility factors for characterizing the volatility dynamics that do not depend on any unknown tuning parameters, which implies that the model is straightforward to estimate by standard OLS on an asset-by-asset basis or panel regression procedures that restrict the beta coefficients to be the same across groups of assets.³⁰

Motivated by the cross-asset and cross-market volatility spillover effects discussed in Section 2, our final risk model augments the asset-specific HExp model in (9) with a global risk factor. Specifically, $ExpGIRV_t^5$ is defined as the 5-day center-of-mass EWMA of the realizations of the $GIRV_t$ global risk

²⁹ The center of mass is formally defined as the weighted-average time period for the lags used,

$$CoM(\lambda) \equiv \frac{\sum_{t=0}^{\infty} e^{-\lambda t} t}{\sum_{t=0}^{\infty} e^{-\lambda t}} = \frac{e^{-\lambda}}{1 - e^{-\lambda}},$$

where we ignore that the sum in (8) only uses the first 500 lags to achieve a simple formula (the influence of the remaining lags is numerically immaterial).

³⁰ We also experimented with related specifications based on linearly and hyperbolically decaying volatility factors, mixtures of multiple beta polynomials as the one employed in the MIDAS model, and mixtures of both exponentially and hyperbolically decaying factors. For the monthly forecast horizon primarily analyzed below, none of these alternative models resulted in systematically superior forecasts compared to the HExp model in (9). The Online Appendix provides further details concerning some of these additional results.

factor defined in Section 2.2, yielding the following model:

$$RV_{t+h}^h - RV_t^{LR} = \sum_{j=1,5,25,125} \beta_j (ExpRV_t^j - RV_t^{LR}) + \beta_5^{GI} (ExpGIRV_t^5 - RV_t^{LR}) + \epsilon_t. \quad (10)$$

As discussed in Section 2.2, the global risk factor is purposely defined on an asset-specific basis to avoid any overlap between RV_t and $GIRV_{t-1}$, and further normalized to have the same time t expanding sample mean as the specific asset. The inclusion of the $ExpGIRV_t^5$ risk factor thus naturally enforces a degree of commonality over-and-above that afforded by restricting the beta coefficients to be the same. We will refer to this specification as the HExpGI risk model.

4. Risk Inference: Model Estimation and Forecasting

We will focus our discussion on a 1-month (i.e., 20-day) forecast horizon. We report both in-sample results, in which we rely on all the available data, as well as out-of-sample forecasts, in which we estimate the models based on an expanding window of the data up to that point in time.³¹ We rely on the explained sum of squares divided by the total sum of squares within and across asset classes as way to succinctly summarize the performance of the different models. To allow for meaningful comparisons between the in- and out-of-sample results, we use the expanding long-run sample mean, or the RV_t^{LR} factor, in the calculations of the out-of-sample R^2 's.³² All of the models are estimated by OLS on an individual asset-by-asset basis, by panel regressions that restrict the coefficients to be the same for all of the assets within an asset class, and by panel regressions that restrict the coefficients to be the same across *all* assets.³³ We refer to these alternative estimation schemes as "Individual Asset," "Panel," and "Mega" panel, respectively.

4.1 Basic estimation and in-sample forecasting results

We begin our discussion by considering the in-sample estimation results, and the different risk models' ability to forecast the future variance, as measured by

³¹ We require that an asset has at least 1 full calendar year of RV 's before it is included in the estimation.

³² The out-of-sample R^2 is formally defined as $R^2 = 1 - \frac{\sum_{t=1}^T (RV_{t+20}^M - \widehat{RV}_{t+20}^M)^2}{\sum_{t=1}^T (RV_{t+20}^M - RV_t^{LR})^2}$, where \widehat{RV}_{t+20}^M refer to the predictions from one of the risk models. This mirrors the out-of-sample R^2 's commonly used in evaluating return predictability (see, e.g., Campbell and Thompson 2008). The HAR and MIDAS models in a few instances produce very large out-of-sample forecasts. To avoid deflating the corresponding R^2 's for these models and to allow for empirically more meaningful comparisons, we follow Swanson and White (1997) in applying an "insanity filter," in which we replace any forecast that exceeds the maximum RV_t^M observed up to that point with RV_t^{LR} ; that is, "insanity" is replaced by "ignorance."

³³ As noted above, the additional θ_2 parameter needed for the MIDAS model is estimated by profiling. Since the R^2 's are fairly "flat" over a wide range of θ_2 's, we rely on the full in-sample estimates for the out-of-sample analysis as well, implicitly ignoring any look-ahead biases. The single MIDAS model that restricts the parameters to be the same across *all* assets results in a common $\hat{\theta}_2 = 6.50$. The Online Appendix provides further details concerning the θ_2 MIDAS.

the in-sample predictive R^2 's. It is instructive to think about the results in terms of different types, or "generations," of risk models: *static* (based on the idea that risk is constant, as exemplified by the expanding sample average of the realized RV s); *first-generation dynamic* (based on daily data, as exemplified by the 21-day rolling sums of the daily squared returns); *second-generation dynamic* (based on the use of high-frequency intraday data, as exemplified by the 21-day rolling RV s, corresponding to a simple random-walk-type forecast); and *state-of-the-art dynamic* (based on directly modeling the RV s, as exemplified by the HAR, MIDAS, and HExp risk models).

Table 4 shows that, not surprisingly, having a static risk model performs the worst. The use of the 21-day rolling sample variance constructed from the daily squared returns is obviously better, and for foreign exchange, in particular, by quite a wide margin. Still, the simple 21-day RV performs substantially better, while the more sophisticated dynamic RV -based models perform the best. At the same time, the differences across the sophisticated risk models are modest, with the HExp models generally performing the best overall.

In addition to the results for the individually estimated risk models, the table also shows the results from our panel-based estimation techniques that restricts the parameters in the dynamic risk models to be the same for all assets within a given asset class (panel) and across all assets (mega). By construction, of course, the individually estimated risk models always result in larger in-sample R^2 's compared to any of the panel-based versions of the identical models. Interestingly, however, the differences in R^2 's across the individual versus panel-based models are fairly small. Given the robustness afforded by the panel estimation, this therefore also suggests that the results may look different out-of-sample.

To more formally assess the statistical significance of the differences in in-sample predictability, the bottom panel of Table 4 reports the results of Diebold and Mariano (1995) (DM) tests. Specifically, taking the individually estimated HExp model as the benchmark, we test the null hypotheses of equal predictive ability by calculating robust t -statistics for the sample means of the time series comprising the average standardized (by the mean of the realized variation) squared error losses for each of the different models, estimation procedures, and asset classes.³⁴ With the exception of the predictions pertaining to the foreign exchange market and the predictions from the HExpGI model, the majority of the t -statistics are significant at the usual 5% level. Looking across the different estimation procedures, the by-construction higher in-sample R^2 s for the individually estimated models also generally translate into statistically

³⁴ The results in Patton (2011) formally justify the use of the DM tests for judging the statistical significance in this situation. The standardization of the squared error loss allows for a more meaningful comparison of the losses across the different assets. It also reflects the empirical observation that the normalized distributions of the realized volatilities are similar across the different assets and more directly corresponds to the scale-invariant R^2 s reported in the top panel. However, similar average DM-test statistics are obtained for nonstandardized loss differentials; the Online Appendix reports further details.

Table 4
In-sample predictions

		Static	21-daily	21-day <i>RV</i>	HAR	MIDAS	HExp	HExpGI
		<i>R</i> ²						
Commodities	Indiv.	−2.3%	2.5%	27.1%	45.3%	46.1%	45.9%	47.1%
	Panel	—	—	—	43.4%	43.6%	43.8%	44.8%
	Mega	—	—	—	43.2%	43.5%	43.7%	44.4%
Equities	Indiv.	−3.2%	2.3%	20.9%	43.6%	43.4%	43.8%	46.9%
	Panel	—	—	—	42.1%	41.6%	42.2%	45.2%
	Mega	—	—	—	41.9%	41.5%	42.0%	44.1%
Fixed income	Indiv.	−4.8%	2.5%	27.6%	43.1%	44.1%	46.7%	47.4%
	Panel	—	—	—	42.5%	43.4%	45.9%	46.1%
	Mega	—	—	—	41.6%	42.2%	43.9%	42.3%
Foreign exchange	Indiv.	−1.8%	28.9%	39.9%	53.7%	54.3%	54.4%	61.1%
	Panel	—	—	—	53.0%	53.5%	53.5%	58.8%
	Mega	—	—	—	52.4%	52.7%	52.6%	56.2%
All assets	Indiv.	−2.8%	2.6%	24.2%	44.5%	44.8%	44.9%	47.1%
	Panel	—	—	—	42.8%	42.7%	43.1%	45.1%
	Mega	—	—	—	42.6%	42.6%	42.9%	44.3%
		DM <i>t</i> -tests						
Commodities	Indiv.	−3.86	−7.09	−4.60	−2.22	−0.46	NA	1.68
	Panel	—	—	—	−2.45	−2.22	−2.00	−1.27
	Mega	—	—	—	−2.55	−2.32	−2.11	−1.64
Equities	Indiv.	−2.86	−2.78	−2.40	−0.15	−0.65	NA	1.29
	Panel	—	—	—	−1.30	−1.32	−1.41	0.28
	Mega	—	—	—	−1.31	−1.36	−1.44	−0.36
Fixed income	Indiv.	−5.02	−3.63	−3.46	−2.10	−1.42	NA	1.44
	Panel	—	—	—	−2.30	−1.75	−2.65	−1.92
	Mega	—	—	—	−3.10	−2.84	−3.66	−2.07
Foreign exchange	Indiv.	−1.96	−2.31	−1.28	−0.60	−0.15	NA	1.08
	Panel	—	—	—	−1.20	−0.97	−1.79	0.66
	Mega	—	—	—	−2.01	−1.79	−1.54	0.98
All assets	Indiv.	−3.42	−4.64	−3.64	−0.86	−0.23	NA	1.39
	Panel	—	—	—	−2.20	−2.02	−2.20	0.17
	Mega	—	—	—	−2.24	−2.10	−2.21	−0.65

This table reports the in-sample results for predicting the 20-day future realized volatility using the different predictor variables and risk models. The top panel reports the average in-sample regression R^2 s by asset class and across all assets. The bottom panel reports Diebold-Mariano (DM) t -statistics for testing the significance of the average standardized loss differentials relative to the individually estimated HExp model. A positive (negative) DM-test statistic indicates that the model and the estimation procedure outperform (underperform) the individually estimated HExp model in-sample.

significant lower losses compared to the panel and mega models that restrict the coefficients to be the same across groups of assets. As a case in point, the t -statistic for the individually estimated HExp model versus the mega HExp model for all assets equals −2.21. Among the mega-based models, only the mega HExpGI model does not result in statistically inferior in-sample predictions compared to the individually estimated HExp model.

The different specifications of the models complicate any direct comparisons of the estimated regressions coefficients. However, all of the risk models, except for the HExpGI model, are nested in the univariate AR(∞) representation in (2). Hence, whereas the estimated β coefficients are not directly comparable, the dynamics of the different risk models may be meaningfully compared in

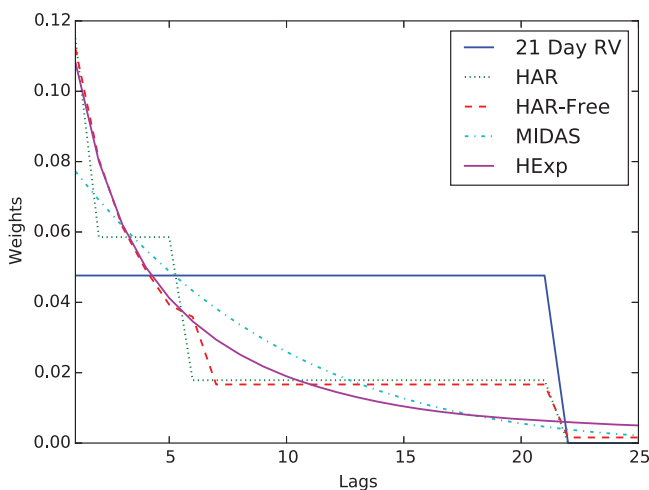


Figure 6

Implied lag coefficients for different risk models

This figure shows the lag coefficients implied by the regression coefficients of full-period models pooled across all assets for each of five different RV-based models: 21-day-RV, HAR, HAR-free, MIDAS, and HExp.

terms of the implied b_i coefficients in that representation. To this end, Figure 6 depicts the implied $b(L)$ polynomials out to a lag length of 25 days for the mega-panel-based estimated models.³⁵ For comparison purposes, we also include the weights for the 21-day RV , or equivalently a HAR model with $\beta_M=1$ and $\beta_0=\beta_D=\beta_W=0$, together with a HAR-free model in which we freely estimate the impact of the first six daily lagged RV s. As the figure shows, with the exception of the flat weights for the 21-day RV , the estimates are generally fairly close. Nonetheless, the HAR-free and HExp models both appear slightly “faster” than the MIDAS model, with a more rapid initial decay and less weight assigned to the intermediate lags ranging between five days and two weeks.

The figure also visualizes the stepwise nature of the implied b_i coefficients for both of the HAR models. As a result, forecasts constructed from these models are more susceptible to abrupt changes, and therefore potentially also more costly to implement, than the forecasts from the “smoother” risk models.³⁶ We will return to this issue and the “speed” of the models in our discussion of the utility-based comparisons in Section 5 below.

³⁵ The Online Appendix provides similar graphs for the individually estimated risk models for the four representative assets shown in Figures 1 and 2 and the panel-based estimation of the four asset classes.

³⁶ For instance, looking at the daily change in the in-sample forecasts averaged across all assets, the sample variance $Var(RV_{t+21|t+1}^M - RV_{t+20|t}^M)$ equals $2.11 \cdot 10^{-4}$ for the HAR model, compared to $1.37 \cdot 10^{-4}$ for the HExp model. Moreover, the first-order autocorrelation of the change in the daily forecasts equals -0.21 for the HAR model, compared to -0.01 for the HExp model.

Table 5
Out-of-sample predictions

		Static	21-daily	21-day <i>RV</i>	HAR	MIDAS	HExp	HExpGI
		<i>R</i> ²						
Commodities	Indiv.	1.2%	5.9%	29.7%	44.5%	45.0%	46.4%	47.6%
	Panel	—	—	—	45.5%	45.7%	46.9%	47.9%
	Mega	—	—	—	45.6%	45.8%	47.5%	48.1%
Equities	Indiv.	2.9%	8.1%	25.6%	41.2%	41.9%	48.3%	51.6%
	Panel	—	—	—	45.8%	46.4%	50.7%	53.8%
	Mega	—	—	—	47.0%	47.8%	51.0%	53.3%
Fixed income	Indiv.	−1.5%	5.5%	29.9%	42.6%	44.3%	47.2%	45.7%
	Panel	—	—	—	43.3%	44.7%	47.5%	46.1%
	Mega	—	—	—	43.2%	43.1%	47.0%	44.0%
Foreign exchange	Indiv.	1.4%	31.1%	41.8%	28.6%	30.4%	46.8%	54.6%
	Panel	—	—	—	27.8%	30.4%	46.5%	55.1%
	Mega	—	—	—	47.7%	49.2%	52.6%	56.8%
All assets	Indiv.	2.1%	7.1%	27.7%	42.8%	43.4%	47.3%	49.5%
	Panel	—	—	—	45.5%	45.9%	48.7%	50.7%
	Mega	—	—	—	46.2%	46.7%	49.2%	50.6%
		DM <i>t</i> -tests						
Commodities	Indiv.	−3.93	−7.01	−3.68	−3.18	−1.82	−0.76	0.29
	Panel	—	—	—	−3.34	−3.68	−2.31	0.85
	Mega	—	—	—	−4.01	−4.09	NA	1.31
Equities	Indiv.	−2.42	−2.46	−1.85	−1.69	−1.72	−1.40	−0.07
	Panel	—	—	—	−2.91	−3.71	−0.77	1.03
	Mega	—	—	—	−2.95	−3.62	NA	1.16
Fixed income	Indiv.	−4.70	−3.07	−2.50	−1.32	−0.96	−0.23	−0.78
	Panel	—	—	—	−1.78	−1.29	−0.52	−0.77
	Mega	—	—	—	−1.71	−1.78	NA	−0.86
Foreign exchange	Indiv.	−1.96	−2.01	−0.93	−1.21	−1.15	−0.93	0.67
	Panel	—	—	—	−1.20	−1.15	−0.90	0.68
	Mega	—	—	—	−1.30	−1.32	NA	1.30
All assets	Indiv.	−3.42	−3.70	−2.56	−2.01	−2.02	−1.93	−0.23
	Panel	—	—	—	−2.83	−3.43	−2.05	0.91
	Mega	—	—	—	−3.64	−4.80	NA	1.21

This table reports the out-of-sample results for predicting the 20-day future realized volatility using the different predictor variables and risk models. The top panel reports the average out-of-sample predictive R^2 s by asset class and across all assets. The bottom panel reports Diebold-Mariano (DM) t -statistics for testing the significance of the average standardized loss differentials relative to the mega HExp model that restricts the coefficients to be the same across all assets. A positive (negative) DM-test statistic indicates that the model and the estimation procedure outperform (underperform) the mega HExp model out-of-sample.

4.2 Out-of-sample forecasting results

Before this discussion, however, Table 5 reports on the out-of-sample performance of the identical set of risk models using the same R^2 metric as above.³⁷ Looking across the columns, we see a similar ranking as for the in-sample results: the static risk model (the expanding average RV) naturally performs the worst, the first-generation risk model based on daily data is second, followed by the simple 21-day RV , while the more sophisticated RV models perform the best, with the HExp models again preforming the very best overall.

³⁷ We note again that the results for the MIDAS models are not truly out-of-sample, as the θ_2 tuning parameters have been chosen using the full sample.

Meanwhile, looking across the rows, we see that the systematic ordering of the individual versus panel-based estimated models is completely *reversed* relative to the in-sample results in Table 4. The mega-panel estimation that restricts the coefficients to be the same across *all* assets now typically results in the highest R^2 's. This is especially true for foreign exchange, where the individually estimated dynamic risk models perform rather poorly, clearly underscoring the importance of more robust forecasting procedures.

The DM-tests based on the average out-of-sample standardized squared error losses, reported in the bottom part of the table, again corroborate these conclusions.³⁸ Now taking the mega HExp model as the benchmark, the pairwise tests show that the forecasts from that model result in significantly lower losses than the forecasts from *all* of the other models, except for the different HExpGI models. Importantly, the forecasts from the mega HExp model also result in significantly lower losses than the forecasts from the individually estimated HExp models and panel HExp models that allow for different coefficients across asset classes. Intuitively, the mega estimation approach provides a built-in robustness against influential outliers, and in turn superior out-of-sample forecast.

To further appreciate this point, the expected squared forecast error loss may be expressed as the sum of the squared forecast bias, the variance of the forecasts, plus the variance of the “irreducible error” associated with the true (unknown) conditional expectation $RV_{t+h}^h - E_t(RV_{t+h}^h)$ (see, e.g., the discussion in Hastie et al. 2009). Looking at the bias-variance trade-off implicit in the squared forecast error losses thus help explain why the pooling and panel-based estimation that exploit the commonalities and reduce the parameter estimation error uncertainty generally works the best from an out-of-sample forecasting perspective and result in the highest predictive R^2 's. In particular, while the squared out-of-sample forecast biases are very small for all of the different models and estimation methods and effectively immaterial, the individually estimated risk models systematically result in the most variable forecasts by quite a wide margin. For the HExp model, for example, the average variance of the forecasts is reduced by almost 30% for the mega model that restrict the coefficients to be the same for all assets compared to the average forecast variance for the individually estimated HExp models; the Online Appendix provides more detailed results along these lines for all of the different models. Importantly, by the same reasoning, this does not necessarily imply that the best performing forecasting model is somehow closest to the “true” model, only that more parsimonious risk models tend to produce better out-of-sample forecasts.

³⁸ Note that the out-of-sample DM-tests are not formally justified as tests for the correct model specification, as the magnitude of the parameter estimation errors decrease with the size of the expanding estimation window, thereby rendering the losses nonstationary. Nonetheless, taking the model forecasts as the primitives, the tests still may be justified as tests for how competing risk forecasters using the different models would have fared (see also the discussion in Diebold 2015).

The practical uses of risk models, of course, face a host of other issues and tradeoffs related to the actual costs and benefits of implementing the forecasts from the models. To illustrate these issues, we turn next to a utility-based framework for evaluating the benefits of an investment strategy involving the notion of equal risk shares. To keep the results simple and directly comparable to the ones discussed in the previous sections, we purposely focus on the risk models estimated to forecast the variance.

5. Risk Models in Action: Quantifying the Utility Benefits

We consider a simple utility-based framework: an investor with mean-variance preferences investing in an asset with time-varying volatility and a constant Sharpe ratio. In contrast to the related approach of Fleming, Kirby, and Ostdiek (2001, 2003), which depends on forecasts for both returns and volatilities, our framework relies exclusively on volatility forecasts.

5.1 Expected utility and risk targeting

By standard arguments, the time- t expected utility may, up to a factor of proportionality, conveniently be approximated as (dropping constant terms that only depend on time- t variables):

$$E_t(u(W_{t+1})) = E_t(W_{t+1}) - \frac{1}{2} \gamma^A \text{Var}_t(W_{t+1}), \quad (11)$$

where $\gamma^A \equiv -u''/u'$ denotes the absolute risk aversion of the investor. We will assume that the investor allocates a fraction x_t of his current wealth to a risky asset with return r_{t+1} and the rest to a risk-free money market account earning r_t^f . Correspondingly, his future wealth becomes $W_{t+1} = W_t(1 + x_t r_{t+1} + (1 - x_t) r_t^f) = W_t(1 + r_t^f) + W_t x_t r_{t+1}^e$, where $r_{t+1}^e \equiv r_{t+1} - r_t^f$ denotes the excess return, resulting in an expected utility of (again dropping constant terms):

$$\begin{aligned} U(x_t) &= W_t \left(x_t E_t(r_{t+1}^e) - \frac{\gamma}{2} x_t^2 \text{Var}_t(r_{t+1}^e) \right) \\ &= W_t \left(x_t E_t(r_{t+1}^e) - \frac{\gamma}{2} x_t^2 E_t(RV_{t+1}) \right), \end{aligned} \quad (12)$$

where $\gamma \equiv \gamma^A W_t$ refers to the investor's relative risk aversion.

To focus on risk modeling, we assume that the conditional Sharpe ratio, defined as $SR \equiv E_t(r_{t+1}^e)/\sqrt{E_t(RV_{t+1})}$, is constant.³⁹ Under this assumption, the

³⁹ This assumption naturally corresponds to our modeling of the realized variation in the previous sections. Alternatively, one might assume that $E_t(r_{t+1}^e)/\sqrt{E_t(RV_{t+1})}$ is constant, resulting in a Jensen's inequality adjustment term $E_t(\sqrt{RV_{t+1}})/E_t(RV_{t+1})$ for the optimal investment positions and expected utilities derived below. This same general setup also generalize to time-varying conditional Sharpe ratios, $SR_t \equiv E_t(r_{t+1}^e)/\sqrt{E_t(RV_{t+1})}$, as long as SR_t is independent of the risk $E_t(RV_{t+1})$.

expected utility simply depends on the position x_t , together with the expected realized volatility $E_t(RV_{t+1})$:

$$U(x_t) = W_t \left(x_t SR \sqrt{E_t(RV_{t+1})} - \frac{\gamma}{2} x_t^2 E_t(RV_{t+1}) \right). \quad (13)$$

The optimal portfolio that maximizes this utility is obtained by investing the fraction of wealth $x_t^* = E_t(r_{t+1}^e) / (\gamma E_t(RV_{t+1}))$ in the risky asset, or, alternatively,

$$x_t^* = \frac{SR/\gamma}{\sqrt{E_t(RV_{t+1})}}. \quad (14)$$

In other words, the investor optimally targets a volatility of SR/γ , since the conditional standard deviation of the x_t^* portfolio equals $\sqrt{\text{Var}_t(x_t^* r_{t+1}^e)} = SR/\gamma$. When the predicted volatility $\sqrt{E_t(RV_{t+1})}$ is above the “risk target” SR/γ , the agent only invests part of his wealth in the risky asset ($x_t^* < 1$). Conversely, when the predicted risk is below the target, the investor applies leverage ($x_t^* > 1$) to reach his risk target. This “volatility-timing” behavior mimics in a simple way the actual trading behavior of many hedge funds with explicit volatility targets and so-called “risk parity investors.”

This in turn results in an expected utility of

$$U(x_t^*) = \frac{SR^2}{2\gamma} W_t = \underbrace{\frac{1}{2}}_{\substack{\text{fraction of expected} \\ \text{return not lost to} \\ \text{disutility of risk}}} \times \underbrace{\frac{SR}{\gamma}}_{\substack{\text{reward-to-risk} \\ \text{expected excess return}}} \times \underbrace{\frac{SR}{\gamma}}_{\text{risk target}} W_t. \quad (15)$$

We see that the expected utility (as a fraction of wealth) is half of the expected return of the optimal position size; the other half of the expected return is “lost” to disutility of risk.

For concreteness and in parallel to the forecasting results discussed in the previous section, we focus on a monthly forecast horizon. Guided by the typical results reported in the extant investment literature, we take the corresponding annualized Sharpe ratio and coefficient of risk aversion to be $SR=0.4$ and $\gamma=2$, respectively (see, e.g., Pedersen 2015 for empirical evidence pertaining to the same broad set of assets analyzed here).⁴⁰ Correspondingly, the investor optimally targets an annualized volatility of 20% (similar to the average volatility across all of the assets considered here):

$$x_t^* = \frac{20\%}{\sqrt{E_t(RV_{t+1})}}. \quad (16)$$

By substitution into (13) or (15), the associated utility for this optimally targeted portfolio equals

$$U(x_t^*) = 4\% W_t, \quad (17)$$

⁴⁰ However, we also consider the results for other values of SR and γ . As discussed further below, higher Sharpe ratios and lower risk aversion coefficients will generally result in greater utility benefits.

meaning that the investor would be willing to give up 4% of his wealth to have access to the x_t^* portfolio rather than simply investing in the risk-free asset. Put differently, since the utility from (13) of a risk-free position equals $U(0)=0$, the investor would receive the same utility by either (1) trading the risky asset optimally while paying a fee of 4% times his wealth or (2) putting all of his money in the risk-free asset.

To further appreciate this number, consider the expected return of the investor's strategy. Given a Sharpe ratio of 0.4 and a 20% risk target, the investor expects to make an excess return of 8% per year. However, by the decomposition in (15), at the optimally targeted position half of this return is "lost" due to the disutility of risk, so the investor is left with a benefit of only 4%.⁴¹

To explicitly quantify the utility gains from different risk models, let $E_t^\theta(\cdot)$ denote the expectations from model θ . Also, let $E_t(\cdot)$ denote the expectations from the true (unknown) risk model. Assuming that the investor uses model θ , to choose the position $x_t^\theta = 20\% / \sqrt{E_t^\theta(RV_{t+1})}$, the expected utility per unit of wealth, $UoW_t^\theta \equiv U_t(x_t^\theta) / W_t$, may be expressed as

$$UoW_t^\theta = 8\% \frac{\sqrt{E_t(RV_{t+1})}}{\sqrt{E_t^\theta(RV_{t+1})}} - 4\% \frac{E_t(RV_{t+1})}{E_t^\theta(RV_{t+1})}. \quad (18)$$

We evaluate this expected utility empirically by averaging the corresponding realized expressions over the same rolling out-of-sample risk model forecasts underlying the results discussed in Section 4.2:

$$UoW^\theta = \frac{1}{T} \sum_{t=1}^T \left(8\% \frac{\sqrt{RV_{t+1}}}{\sqrt{E_t^\theta(RV_{t+1})}} - 4\% \frac{RV_{t+1}}{E_t^\theta(RV_{t+1})} \right). \quad (19)$$

For short, we will simply refer to this as the "realized utility." In parallel to (15), a risk model that perfectly predicts the realized volatilities would deliver a realized utility of $8\% - 4\% = 4\%$. Or said differently, the value of trading the risky asset with the perfect risk model is worth 4% percent of wealth. Importantly, the expected returns do not enter this expression. As such, this circumvents the invariable noise stemming from random return realizations, thereby allowing for a more pointed and meaningful comparison of the different risk models solely based on the actual realized volatilities and the different $E_t^\theta(RV_{t+1})$ risk model forecasts.

5.2 Quantifying the value of a risk model: Realized utility comparisons

Table 6 reports the realized utility for the same set of risk models studied above. Guided by the out-of-sample forecasting results in Section 4.2, we

⁴¹ We note that, while a proportional change in the Sharpe ratio and the relative risk aversion coefficient would imply the exact same risk target SR/γ , the resultant utility equals $SR^2/(2\gamma)$, so the risk target alone is not a sufficient statistic for the utility in general.

Table 6
Realized utility

		Static	21-daily	21-day RV	HAR	MIDAS	HExp	HExpGI	Future RV
Realized utility									
Commodities	Zero	3.48%	3.60%	3.74%	3.77%	3.77%	3.79%	3.79%	4.00%
	Full	3.48%	3.29%	3.60%	3.59%	3.66%	3.62%	3.62%	3.85%
	Gradual	3.47%	3.49%	3.65%	3.71%	3.72%	3.73%	3.73%	3.91%
Equities	Zero	2.93%	3.46%	3.53%	3.55%	3.57%	3.62%	3.64%	4.00%
	Full	2.92%	3.26%	3.34%	3.46%	3.51%	3.53%	3.55%	3.90%
	Gradual	2.90%	3.34%	3.41%	3.48%	3.49%	3.54%	3.56%	3.91%
Fixed income	Zero	3.59%	3.56%	3.73%	3.78%	3.77%	3.80%	3.81%	4.00%
	Full	3.58%	2.78%	3.34%	3.30%	3.49%	3.38%	3.36%	3.60%
	Gradual	3.58%	3.27%	3.53%	3.66%	3.67%	3.69%	3.68%	3.80%
Foreign exchange	Zero	3.30%	3.66%	3.79%	3.74%	3.76%	3.79%	3.82%	4.00%
	Full	3.30%	3.04%	3.49%	3.30%	3.51%	3.42%	3.44%	3.70%
	Gradual	3.28%	3.43%	3.63%	3.64%	3.65%	3.67%	3.69%	3.85%
All assets	Zero	3.27%	3.55%	3.67%	3.69%	3.70%	3.73%	3.74%	4.00%
	Full	3.26%	3.17%	3.48%	3.46%	3.56%	3.52%	3.53%	3.81%
	Gradual	3.25%	3.40%	3.54%	3.61%	3.62%	3.64%	3.66%	3.88%
DM <i>t</i> -tests									
Commodities	Zero	-6.30	-13.37	-4.18	-6.53	-6.16	NA	1.98	15.30
	Full	-2.92	-19.31	-1.35	-7.98	14.26	NA	1.10	16.09
	Gradual	-5.19	-13.86	-6.27	-4.98	-4.13	NA	1.50	12.21
Equities	Zero	-3.30	-3.51	-1.98	-4.55	-5.22	NA	2.99	8.51
	Full	-2.82	-6.48	-2.46	-4.59	-2.05	NA	2.27	7.92
	Gradual	-3.08	-3.97	-3.01	-3.61	-3.77	NA	2.27	6.32
Fixed income	Zero	-6.12	-3.97	-2.50	-2.95	-2.99	NA	1.22	9.18
	Full	7.42	-8.00	-0.87	-7.48	11.72	NA	-4.99	8.90
	Gradual	-2.43	-5.13	-3.84	-2.19	-1.21	NA	-1.64	4.79
Foreign exchange	Zero	-2.38	-5.80	-0.96	-2.13	-3.34	NA	1.79	5.33
	Full	-0.69	-12.47	1.03	-4.05	9.96	NA	1.00	5.45
	Gradual	-1.97	-9.62	-2.99	-2.80	-3.08	NA	1.25	3.13
All assets	Zero	-4.50	-9.90	-4.01	-5.11	-6.44	NA	2.79	12.90
	Full	-2.41	-17.47	-2.97	-6.06	10.09	NA	1.15	12.47
	Gradual	-3.86	-11.40	-6.09	-4.38	-4.13	NA	1.88	8.71

The top panel reports the average 20-day realized utilities from holding volatility-targeted positions based on the rolling out-of-sample predictions from the same set of risk models analyzed in Table 5. All of the results are based on the mega models that restrict the coefficients to be the same across all assets. The last (infeasible) column assumes knowledge of the true future 20-day realized volatility. We report the results for three scenarios: no transaction costs (“zero”); transaction costs equal to the full average spreads for each of the different assets with the positions fully rebalanced at the close of each business day (“full”); and transaction costs equal to the average spread with the positions rebalanced only gradually as described in the main text (“gradual”). The bottom panel reports Diebold-Mariano (DM) *t*-statistics for testing the significance of the average utility differentials relative to the HExp model under the identical transaction cost and trading speed assumptions. A positive (negative) DM-test statistic indicates that the alternative procedure outperforms (underperforms) the mega HExp model in terms of the out-of-sample realized utility gains.

only consider the mega versions that restrict the estimated coefficients in the dynamic models to be the same for all assets. In addition to the results for the practically feasible risk models, we also report the utility from exactly knowing the future volatility (the column labeled “Future RV”). We begin by studying the utility benefits in the absence of transaction costs (the rows denoted “Zero”). We discuss the effects of transaction costs in Section 5.3 below.

The relative utility-based ordering of the different risk models echoes that of the out-of-sample R^2 's reported in Table 5: the static model delivers the lowest utility, the first-generation risk model based on the 21-day rolling daily sample variance is better, the simple 21-day RV is better yet, and the more sophisticated dynamic RV models perform the best, with the highest realized utility achieved by the HExp models. Of course, the highest utility arises from knowing the future RV , which results in the maximum possible utility of 4%.

The lower panel of the table reports DM-tests for testing the statistical significance of these differences in utilities obtained by comparing the realized utilities for the HExp model to the realized utilities for each of the other models and procedures. As the table shows, when averaged across all assets, the HExp model does indeed result in significantly higher utility than the static and simple 21-day risk models, as well as the more sophisticated dynamic HAR and MIDAS models.⁴² Further, consistent with the ordering of the risk models based on the out-of-sample R^2 's in Table 5, the HExpG1 model results in significantly higher utility than the simple HExp model, underscoring the additional improvements afforded by incorporating a common global risk factor.

It is instructive to also consider the economic significance (in addition to the statistical significance) of these findings. Note first that, while the ordering is the same, the magnitudes of the differences in utilities appear smaller than those of the R^2 's. To appreciate this difference, note that the relation between utility and R^2 is highly nonlinear (and, in fact, not one-to-one) as seen in Figure A3 in the appendix. Intuitively, since the utility is maximized at $U(x^*)=4\%$, the marginal utility is zero around this optimum, $U'(x^*)=0$. Hence, locally around the optimum, small deviations in the input will have only small effects on the utility; that is, $U(x) \cong U(x^*) + U'(x^*)(x - x^*) = U(x^*)$. As such, first-order differences in R^2 may appear to have only a second-order impact on the utility. Nevertheless, the utility differences are still economically significant when properly viewed.⁴³

Specifically, consider the economic magnitude in the context of the results for "all assets." The difference in the utility of the HExp model over and above that of having a static risk model is 0.46%, or 46 bps (and even more, 48 bps, for the HExpG1 model). This means that under our maintained assumptions,

⁴² Even though the differences in the average realized utilities appear numerically small, the HExp model almost always performs the best, as seen in, for example, Figure 7, and discussed further below, and thus explains the significance of the t -statistics.

⁴³ In contrast to our results, Campbell and Thompson (2008) find large utility gains despite low R^2 's in return-prediction regressions. The difference is that forecasting the future return potentially increases the maximum utility (e.g., from 4% with no return predictability to a higher number) while the benefit of forecasting the volatility, as we study, is bounded from above (by 4% in our example). Further, when volatility is relatively stable over time for a given asset, then even a static risk model (with close to zero R^2) can realize a utility close to the upper bound as seen in Figure A3 in the appendix. See also Cenesizoglu and Timmermann (2012), who find a weak relation between "a return prediction model's out-of-sample statistical performance and its ability to add economic value."

an investor would be willing to pay 46 bps to have access to the HExp risk model rather than using the expanding sample mean of RV . This magnitude is of the same order as typical institutional fees for active asset management, but these fees are often thought of as compensation for “alpha,” that is, the asset manager’s extra sources of returns, rather than a good risk model. For example, studying predictors of the future mean return (rather than predictors of return volatility), Campbell and Thompson (2008) document “large benefits for investors” of order of “50 basis points per year.” However, many managers and investors would be surprised to learn that a “good” risk model can also have such a large value.

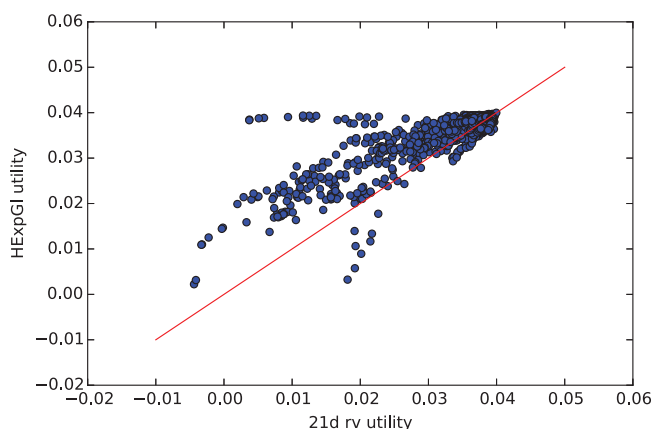
Clearly, a large portion of this “extra utility” can be achieved through a simpler dynamic risk model. Indeed, the utility increases by 29 bps when moving from a static risk model to the first-generation risk model (21-Daily). Utility further increases by 12 bps when moving from the daily risk model to the simple 21-day RV -based model, showing the value of high-frequency data. Finally, moving from the use of 21-day RV to the HExp model leads to a utility increase of 6 bps, while the HExpGI model adds an additional 2 bps.

To further appreciate the economic significance of these numbers, recall that firms like Blackrock and Vanguard have created big businesses around various products with fees of single-digit basis points so, in investments, improvements of the order of multiple basis points are noticeable, tens of basis points are large, hundreds of basis points seem unrealistic, while less than 1 bps is relatively insignificant.

As an even more concrete example, suppose that a pension plan hires an asset manager to a mandate of managing \$1B of assets. Based on the above numbers, the pension plan should be willing to pay up to \$4.6M per year to the manager if he uses the HExp model rather than a static risk model. If the daily (first-generation) risk model is an industry standard that is expected to be delivered for free, then the pension plan should be willing to pay up to \$1.8M per year for the HExp model. Similarly, if the 21-day RV is considered the benchmark, then the pension fund should be willing to pay up to an extra \$0.6M per year for the HExp model.

These specific numbers, of course, depend on the assumed constant Sharpe ratio and risk target (or, equivalently, risk aversion). As seen in equation (15), the utility scales proportionally with SR and the risk target. For example, an investor with access to an asset with a Sharpe ratio of 0.8, double the Sharpe ratio of 0.4 underlying the calculations in Table 6, would be willing to pay twice as much for having access to any of the more sophisticated dynamic risk models.

Another indication of the economic significance and the value of better risk models per se stem from the uses of the models during turbulent times and periods with especially large volatility shocks. To illustrate, Figure 7 presents a scatter plot of the period-by-period utility arising from the simple 21-day RV and the most sophisticated dynamic risk model that we consider, the HExpGI

**Figure 7****Realized utility: HExpGI versus 21-day RV**

This figure shows the daily realized 20-day utility obtained using the mega-based HExpGI model on the y-axis versus the realized utility of using a simple 21-day RV on the x-axis. Realized utilities are calculated without transactions costs, trading all the way to the desired positions each day.

model. As the figure clearly shows, most of the observations are above the 45-degree line, meaning that the HExpGI model almost always delivers the highest utility. At the same time, while a lot of the observations are clustered around (4%, 4%), close to the maximum possible utility, the more sophisticated HExpGI model tends to outperform by more in situations when both models deliver utility substantially below 4%, that is, during periods of large volatility shocks that are difficult to predict.

The specific numbers discussion above all pertain to the realized utilities averaged across all assets. However, as the more detailed numbers in Table 6 show, the utility gains vary substantially across asset classes. To understand the reason behind these differences, recall that the benefits from risk targeting are intimately related to the volatility-of-volatility: if the volatility was constant through time, the optimal positions would also be constant. As such, fixed income, which had the most stable volatility over the sample period, also shows the smallest relative utility gains from dynamic risk modeling. By comparison, equities, which exhibited the highest asset class volatility-of-volatility over the sample, show much larger relative gains from using the more sophisticated dynamic risk models.

5.3 Realized utility comparisons with transaction costs

The previous comparisons ignore the cost of implementing the risk-targeted positions. In actuality, of course, trading is costly. Since the realized utility in (19) is effectively expressed in units of returns, it is easy to incorporate the effect of transaction costs by simply subtracting the simulated costs of implementing the positions. For simplicity, we assume that the costs of trading are linear in

the absolute magnitude of the change in the positions, $|x_{t-1}^\theta - x_t^\theta|$.⁴⁴ As our benchmark transaction cost estimate, we use the median bid-ask spread for each of the assets over the last nine months of the sample.⁴⁵ This is twice the so-called “half-spread,” which is the difference between the mid-quote and the bid or ask price. We use the full spread (rather than the half spread) because transaction costs could be higher for a large trader because of market impact. On the other hand, the transaction costs could be lower due to the possibility of strategic trading and transactions occurring inside the spread combined with the netting of other positions.⁴⁶

The second row in each section of Table 6 (labeled “Full”) reports the resultant realized utilities net of transaction costs. All of the utilities are obviously lower than the ones reported in the first row (the only exception being the static risk model based on the expanding sample mean of RV , which is effectively constant over the majority of the sample). The incorporation of transaction costs also change the ordering of the different risk models, as they imply different amounts of trading. In particular, while the use of a static risk model results in little transaction costs, the first-generation daily model has the largest cost, and, the RV models are in between. Among the more sophisticated risk models, the MIDAS model results in the lowest turnover due to its “slower” response to new information, as seen in the previous Figure 6.

A commonly employed approach to help mitigate the impact of transaction costs is to slow down trading, allowing the investor to deviate from the zero-cost optimal positions. The formal development of optimal trading strategies for the different risk models that explicitly incorporate transaction costs is beyond the scope of this paper. Instead, we rely on the strategy discussed by Garleanu and Pedersen (2013, 2016) of trading only partially toward the desired position.⁴⁷ The third row in each section of Table 6 (labeled “gradual”), in particular, reports the realized utilities when the “speed” of trading is reduced, and the positions are only traded 15% toward the zero-cost optimal targets each day. This “slowing down” of the models does indeed restore some of the “lost” utility vis-à-vis the results in the first row that abstracts from transaction costs and the results in the second row based on trading all the way to the targeted positions. The HExp and HExpGI models now also again significantly outperform all of the other practically feasible risk models.

⁴⁴ In practice, of course, the cost will generally increase with the magnitude of the trade and will likely also depend on the volatility. We leave a more thorough analysis of these issues for future work.

⁴⁵ The only exception is USDSEK (Sweden FX), where we use spread data from all of 2013 because of data availability.

⁴⁶ The Online Appendix reports the additional results for other transaction cost assumptions.

⁴⁷ Garleanu and Pedersen (2013, 2016) show that this strategy is optimal under certain conditions, including quadratic transaction costs due to, for example, market impact. Thus, even though this strategy is not fully optimal here, it nevertheless gives a sense of the benefits that may be obtained by reducing the “speed” of trading.

6. Conclusion

This paper studies risk across commodities, currencies, global equities, and fixed-income securities using a new extensive data set of intraday data. Based on this rich data set, we find that risk dynamics are surprisingly similar across assets, asset classes, and countries. Normalized by the average levels, risk measures everywhere have similar unconditional distributions and autocorrelation structures. This commonality in risk structures can be exploited when estimating risk models by aggregating information across assets using panel estimation methods. Beyond common structures in asset risk levels and dynamics, we also find that a common normalized “global” risk factor contains information on the future volatilities of individual assets not already contained in the asset-specific realized volatility histories, and that models that include this common factor outperform models that do not.

In addition to the robustness generated by exploiting commonality in risk dynamics, we add further robustness by developing new “smooth” and “centered” realized volatility models that enforce a natural continuous and monotonic dependence on lagged realized volatilities. We show that these new robust risk models perform well in out-of-sample risk forecasting.

Lastly, we develop a simple framework for quantifying the utility benefits of risk models for risk-targeting investors. We show that, under empirically realistic assumptions, our robust dynamic risk models are worth close to 0.5% of total wealth per year relative to a static risk model.

Appendix

A.1. Data and Data Cleaning

Our data for the different assets come from a few different data sources. The data start at different points in time, but all end on September 30, 2014. Table A1 provides a summary.

A.1.1 Contract rolling

Our returns for commodities, equities, and fixed income are all constructed from futures contract prices. Unlike spot prices (like our foreign exchange data), individual futures contracts are listed and subsequently expire multiple times per year.

With multiple futures contracts for the same underlying asset trading at the same time, we need a rule for determining which contracts to actually use on any given day. In practice, an asset manager typically gains continuous exposure through these contracts by “rolling” their position from a contract that is set to expire to the next contract in line. For instance, if S&P 500 E-mini futures come in contracts that expire in mid-March and mid-June, then an investor can hold the contract expiring in March through the beginning of March, and then close that position while simultaneously opening a comparable position in the E-mini contract that does not expire until June. Naturally, many investors would like to roll their positions at the same time.⁴⁸ Rolling can

⁴⁸ Recall that all futures contract positions have offsetting positions on the other side of the contract, so if investor A is long one contract and investor B is short one contract and they both desire to keep those positions, they can mutually close their position in the expiring contract (the “near” contract) and open a new position in the contract with the next expiry (the “far” contract) without requiring outright liquidity from the market.

Table A1
Data Sources

Asset Class	Asset	Number of Assets	Total Days in Analysis	Primary Data Source	Used From	Secondary Data Source	Used From	Assumed T-Costs (in bps)
COMMODITIES		20	108149	TRTH		TDC		
	Brent Oil	1	4754	TRTH	1/3/1996	TDC	1/3/1996	1.0
	Cattle	1	5483	TRTH	12/20/2004	TDC	11/30/1992	3.2
	Cocoa	1	5471	TRTH	4/1/2008	TDC	11/11/1992	3.4
	Coffee	1	5469	TRTH	4/1/2008	TDC	11/17/1992	8.0
	Corn	1	5502	TRTH	8/1/2006	TDC	11/19/1992	5.7
	Cotton	1	5453	TRTH	4/1/2008	TDC	11/12/1992	4.6
	Crude (WTI) Oil	1	5480	TRTH	9/5/2006	TDC	11/10/1992	1.0
	Feeder Cattle	1	5513	TRTH	8/1/2007	TDC	10/29/1992	4.5
	Gas Oil	1	4754	TRTH	1/3/1996	TDC	1/3/1996	2.8
	Gold	1	5471	TRTH	12/4/2006	TDC	12/2/1992	0.8
	Heating Oil	1	5480	TRTH	9/5/2006	TDC	11/16/1992	1.7
	Lean Hogs	1	5486	TRTH	2/15/2005	TDC	11/30/1992	4.5
	Natural Gas	1	5442	TRTH	8/23/2006	TDC	1/5/1993	4.0
	Silver	1	5412	TRTH	12/4/2006	TDC	1/5/1993	2.6
	Soybeans	1	5522	TRTH	8/1/2006	TDC	10/22/1992	2.1
	Soymeal	1	5502	TRTH	8/1/2006	TDC	11/19/1992	4.1
	Soyoil	1	5501	TRTH	8/1/2006	TDC	11/19/1992	3.0
	Sugar	1	5481	TRTH	4/1/2008	TDC	11/3/1992	5.9
	Unleaded (RBOB)	1	5475	TRTH	8/22/2006	TDC	11/16/1992	2.0
	Wheat	1	5498	TRTH	8/1/2006	TDC	11/19/1992	4.4
EQUITIES		21	80042	TRTH		NONE		
	Australia (SPI 200)	1	3472	TRTH	12/18/2000	NA	NA	1.9
	Germany (DAX 30)	1	4732	TRTH	1/3/1996	NA	NA	1.0
	Brazil (BOVESPA)	1	4577	TRTH	2/27/1996	NA	NA	2.8
	China (Hang Seng CEI)	1	2667	TRTH	12/9/2003	NA	NA	2.0
	Canada (S&P/TSX 60)	1	3773	TRTH	9/14/1999	NA	NA	1.3
	Spain (IBEX 35)	1	4698	TRTH	1/4/1996	NA	NA	2.0
	Eurostoxx	1	4130	TRTH	6/23/1998	NA	NA	3.2
	France (CAC 40)	1	4007	TRTH	1/7/1999	NA	NA	1.1
	Hong Kong (Hang Seng)	1	4591	TRTH	1/3/1996	NA	NA	1.2
	India (SGX NIFTY)	1	2213	TRTH	10/11/2005	NA	NA	1.7
	Italy (FTSE MIB)	1	2617	TRTH	6/15/2004	NA	NA	2.4
	Japan (TOPIX)	1	4570	TRTH	1/5/1996	NA	NA	4.1
	South Korea (KOSPI 200)	1	4466	TRTH	5/6/1996	NA	NA	1.9
	Netherlands (AEX)	1	4499	TRTH	1/9/1997	NA	NA	1.3
	South Africa (ALSI)	1	2308	TRTH	7/7/2005	NA	NA	1.7
	Switzerland (SMI)	1	4027	TRTH	9/15/1998	NA	NA	1.2
	Taiwan (SGX-MSCI Taiwan)	1	4295	TRTH	2/24/1997	NA	NA	3.1
	UK (FTSE 100)	1	4706	TRTH	1/3/1996	NA	NA	0.8
	US (S&P 500 E-Mini)	1	4274	TRTH	9/10/1997	NA	NA	1.3
	US (Russell 2000 E-Mini)	1	2234	TRTH	12/13/2005	NA	NA	0.9
	US (S&P 400 Mid Cap E-Mini)	1	3186	TRTH	1/29/2002	NA	NA	1.5
FIXED INCOME		8	32333	TRTH		TDC		
	Australia 10y	1	4734	TRTH	1/3/1996	TDC	NA	3.9
	Germany 10y	1	4499	TRTH	1/5/1999	TDC	1/3/1997	0.7
	Germany 5y	1	4493	TRTH	2/1/1999	TDC	1/3/1997	0.8
	Canada 10y	1	2771	TRTH	9/26/2000	TDC	NA	0.8
	Japan 10y	1	3605	TRTH	1/5/1996	TDC	NA	0.7
	UK 10y	1	4711	TRTH	1/3/1996	TDC	NA	0.9
	US 10y	1	3993	TRTH	1/1/2001	TDC	10/20/1998	1.3
	US 5y	1	3527	TRTH	7/1/2001	TDC	9/5/2000	0.7

(continued)

Table A1
Continued

Asset Class	Asset	Number of Assets	Total Days in Analysis	Primary Data Source	Used From	Secondary Data Source	Used From	Assumed T-Costs (in bps)
FOREIGN EXCHANGE		9	30161	Olsen Data		NONE		
	Australia (AUD-USD)	1	2802	OlsenData	1/1/2004	NA	NA	2.2
	Eurozone (EUR-USD)	1	4103	OlsenData	1/1/1999	NA	NA	0.7
	Canada (USD-CAD)	1	3061	OlsenData	1/1/2003	NA	NA	2.5
	Japan (USD-JPY)	1	3841	OlsenData	1/1/2000	NA	NA	1.0
	Norway (USD-NOK)	1	2801	OlsenData	1/1/2004	NA	NA	8.1
	New Zealand (NZD-USD)	1	2803	OlsenData	1/1/2004	NA	NA	4.7
	Sweden (USD-SEK)	1	3062	OlsenData	1/1/2003	NA	NA	7.7
	Switzerland (USD-CHF)	1	3844	OlsenData	1/1/2000	NA	NA	1.7
	UK (GBP-USD)	1	3844	OlsenData	1/1/2000	NA	NA	1.5

typically be done at far lower cost and market impact than outright trading for this reason, assuming that the trader rolls positions at the same time that many investors are seeking to roll their positions as well.

The dates when most investors roll positions (and therefore the times when the roll trade has the most liquidity) vary for different assets, but typically fall around 3–6 days prior to contract expiry or first notice date for noncommodities. Commodity roll times are more likely to be associated with the roll schedule of major commodity indices, such as the S&P GSCI. That most of our assets studied are futures contracts that need to be rolled necessitates a formal roll rule.⁴⁹ As an actual investor would, we determine our roll time based on a measure of liquidity in the near and far contracts. For each contract, we specify a roll period (typically 3–6 days prior to expiry for noncommodities, and the roll period for the GSCI for commodities), measuring the total number of minutes with at least one valid trade in them for both the near and far contract. Once the number of valid minute bars for the far contract exceeds the near contract over a trading day, or we reach the end of the roll period, we consider that day to be the roll date.⁵⁰

Once we have a roll date, we seek to line up the two contracts before switching our reference from “near” to “far.” Omitting the last 5 minutes of the trading day to avoid any possible end-of-day effects, we find the last (nearest to the close) minute bar where both contracts have at least one trade in order to switch between contracts. This ensures that both contracts have nonstale prices at the time of the change. We use all returns from the near contract up to and including the roll minute bar and then use returns from the far contract after that.

A.1.2 Liquidity plots and “sanity” Filters

To begin, we use the Financial Calendars (FinCal) to provide market open and close times for some of the market sessions dating back to the year 2000. For assets with data prior to 2000, or assets outside the FinCal database, we rely on so-called “liquidity plots,” in which for each minute of the day, we plot the proportion with at least one trade. As the resultant liquidity plots for the four representative assets and 3 years given in Figure A1 show, this effectively delineate the periods of the day when the markets are actively operating.

⁴⁹ For example, rather than having a single asset representing the S&P 500 index, we have data for a series of different S&P 500 index futures contracts and a roll rule that we specify to connect them into a single return series.

⁵⁰ Since we only seek to compute daily *RV*’s, it is possible to determine the occurrence of a roll date ex post and splice return sequences from both the near and the far contracts at the end of the trading day.

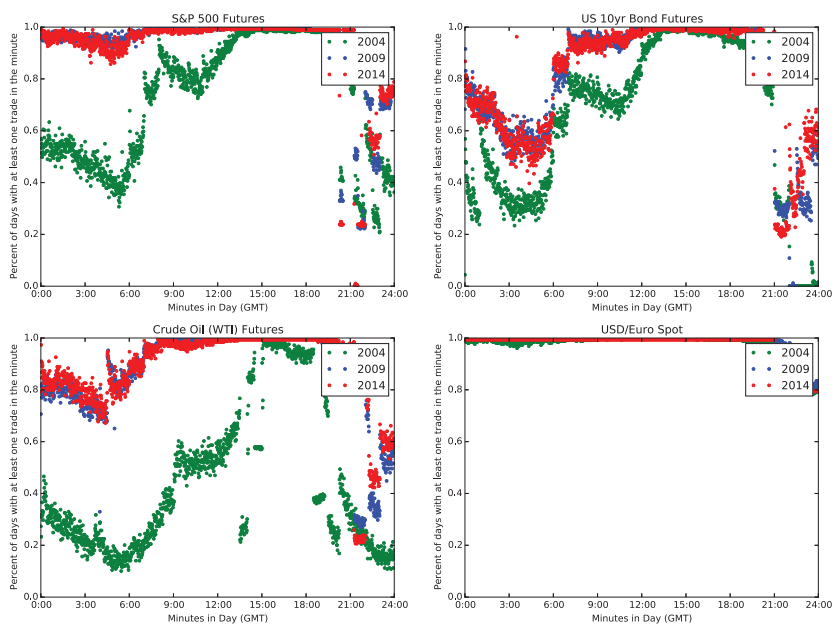


Figure A1
Liquidity plots

This figure shows the proportion of business days in the relevant year (2004, 2009, or 2014) that have at least one trade in any given minute in the 24-hour day (GMT). For instance, a value of 1.0 shows for that particular minute in the day, all business days in that year had at least one valid trade, and a value of 0.5 would mean that only half of the days had at least one valid trade in that minute.

We also apply some mild “sanity” filters to identify valid minute bars. To begin, all valid minute bars must have at least one recorded trade over the course of the minute. In addition, we omit data when:

- The bid or the ask price is less than zero.
- The ask price is less than or equal to the bid price.
- The bid-ask spread is greater than 1% of the bid price.
- The bid size or the ask size is less than zero.
- The recorded trade price is less than zero.

Finally, based on the rolling medians of the past three returns, we omit any data point associated with a return more than 1% away from the median centered on that minute bar. This effectively removes any data points that cause very large moves in one direction followed by an immediate large reversal. This filter generally removes fewer than five data points per year per asset.

A.1.3 Volatility signature plots

Our choice of intraday sampling frequency Δ for the high-frequency returns used in the construction of the realized volatilities is guided by the volatility signature plots first proposed by Andersen, Bollerslev, Diebold, and Labys (2000). Intuitively, in the absence of any contaminating market microstructure influences, the time-series sample mean of the daily RV_t 's provides an unbiased estimate for the true daily unconditional variance for the active part of the trading day. If, on the other hand, the RV_t 's are calculated from too high a sampling frequency $1/\Delta$ to render the basic assumption of an arbitrage-free price process a good description of the discretely sampled intraday price process, the time-series sample mean of RV_t will not provide an unbiased estimate of the true

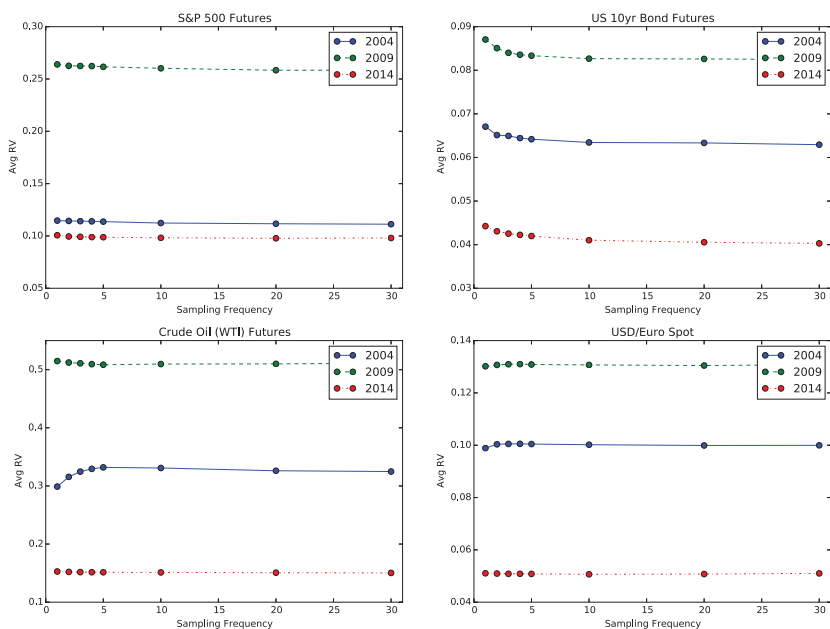


Figure A2
Signature plots

This figure shows the average daily realized volatility for a given year (2004, 2009, or 2014) as the sampling frequency varies.

daily variance. Hence, by plotting the sample mean of RV_t as a function of the sampling frequency of the underlying high-frequency returns, it is possible to determine an appropriate choice of Δ . This choice could obviously differ across assets and time. However, guided by the inflection points visible in most of the signature plots for each of the individual assets, as directly illustrated by the plots for the four representative assets and 3 years given in Figure A2, 5-minute naturally emerge as an appropriate common choice.

A.2. Robustness Checks and Additional Empirical Results

Our main empirical findings are based on “centered” risk models, forecasting the realized variation one month ahead, using the lagged RV s for the whole day. This appendix briefly discusses the sensitivity to each of these assumptions and modeling choices. We will focus our comparisons on the HExp model. The Online Appendix provides more detailed results for the other dynamic risk models along with additional robustness checks.

A.2.1 “Centered” versus “uncentered” risk model forecasts

Our “centering” of the dynamic risk models formally alter the models relative to their “uncentered” counterparts. Specifically, rewriting the generic “centered” risk model in (3) with only RV_{t+1} on the left-hand side, the “intercept” on the right-hand side becomes $(1 - b_1 - b_2 - \dots)RV_t^{LR} = (1 - b(1))RV_t^{LR}$. This time-varying “intercept” obviously differs from the time-invariant b_0 intercept in the more standard risk model (2). Only by using the unconditional sample variance for the full sample, say RV_T^{LR} , in place of the expanding RV_t^{LR} , would the full in-sample OLS and panel regression estimates of the dynamic b_i coefficients be the same for the two models.

Meanwhile, the information structure implicit in the “centered” risk models is more compatible with the practical uses of the models, in that only current information is used in the construction of forecasts. Importantly, this translates into the actual forecast performance of the differently formulated models. For instance, while the “uncentered” version of the HExp model performs better than the “centered” version in-sample with 0.8% higher average R^2 for the mega-based model, the out-of-sample R^2 for the same mega-based model is 0.6% higher for the “centered” model. This same general conclusion holds true for all of the other risk models.

A.2.2 Other forecast horizons

The results discussed in the main text all pertain to monthly predictions. We also implemented the same risk models for both shorter (daily and weekly) and longer (60 days) prediction horizons. Consistent with the results for the monthly predictions, the differences in the in-sample R^2 s are generally fairly small across the different risk models. Also, while the individually estimated models necessarily result in the highest in-sample R^2 s, the mega-based estimation that restricts the coefficients to be the same across all assets typically results in the best performing models out-of-sample. Interestingly, the largest (in a relative sense) improvements in the R^2 s compared to the original HAR model manifest over longer horizons, indirectly underscoring the importance of including longer-run volatility factors and properly “centering” the risk models when forecasting further into the future.

A.2.3 Risk models based on daily returns versus intraday realized volatilities

As noted above, the monthly out-of-sample predictive R^2 s for the RV -based risk models reported in Table 5 easily exceed those based on monthly past realized volatilities constructed from daily squared returns or simple EWMA of the lagged daily squared returns. To further investigate whether these improvements arise from better risk modeling or better risk measurements, or both, we implement the identical suite of risk models and predictive R^2 s using the daily squared returns in place of the daily RV s. Not surprisingly, benchmarking the performance of the models against the sum of the daily squared returns results in more “noisy” and systematically lower R^2 s. More important, the risk models based on realized volatilities systematically outperform the identical risk models based on risk factors constructed from daily squared returns. The out-of-sample R^2 for the mega HExp model, for example, improves from 44.0% with the daily squared returns to 47.3% with daily RV s, directly underscoring the informational advantages afforded by the use of the more accurate daily realized volatility measures.

A.2.4 Intraday and overnight risk models

As discussed in Section 1.3, ample empirical evidence dating back to at least French and Roll (1986) shows that volatility tends to be lower during nontrading hours than during the active part of the trading day.⁵¹ The daily RV s employed in our risk models simply add the intraday variation to the overnight squared returns to obtain an estimate of the volatility for the whole day. However, different dynamic dependencies may be at work intraday and overnight. To investigate whether implicitly restricting the dynamics to be the same result in inferior forecasts, we implement separate risk models for the intraday and overnight components. Our results confirm that it is generally more difficult to predict the overnight volatility compared to the volatility for the active part of the trading day. Meanwhile, the HExp model that performed the best for predicting the daily RV s, also performs the best for separately predicting the intraday and overnight volatility components.

⁵¹ French and Roll (1986) argue that private information revealed through the process of trading may help explain these differences. Recent work by Boudoukh, Feldman, Kogan, and Richardson (2016) uses textual analysis to identify specific types of public “news” and finds that the ratio of overnight variation to within-day variation also varies importantly with the intensity of the “news,” suggesting that public information account for a nontrivial portion of the return variation as well.

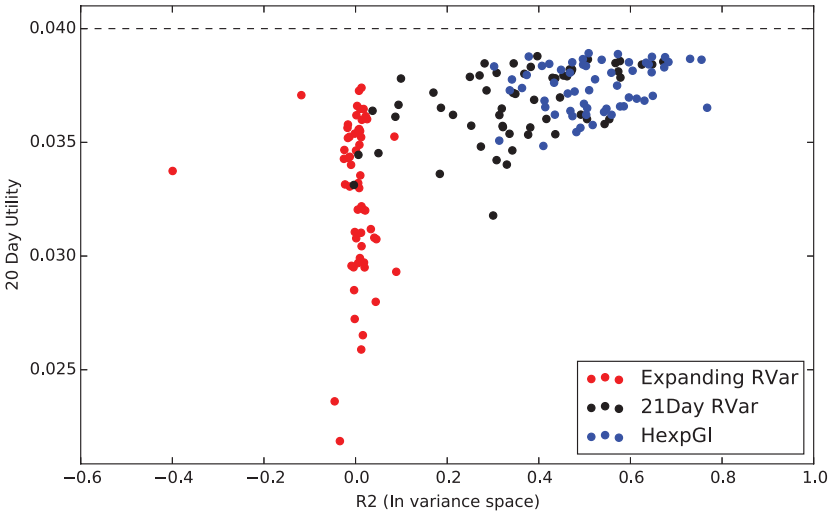


Figure A3
Utility vs. R^2 for static, 21-day RV , and HExpGI Models

This figure shows the average out-of-sample 20-day realized utility on the y-axis versus the out-of-sample R^2 on the x-axis. Each data point corresponds to an individual asset for one of the three color-coded risk models. The three different risk models considered are the static expanding window average RV (red), 21-day RV (black), and the HExpGI (blue) using mega-panel coefficients.

Importantly, combining the forecasts from the separately estimated models for the intraday and overnight volatilities to obtain a forecast for the whole day, or month, do not result in obviously superior out-of-sample forecasts compared to the risk models for the daily RV s. For example, combining individually estimated HExp models for the intraday and overnight volatilities results in an average out-of-sample R^2 of 46.8%, compared to the average R^2 of 47.3% for the single mega HExp model.

A.2.5 Volatility versus variance forecasting

Investment decisions and risk measurements often depend on functions of the future variance, as opposed to the future variance itself. As discussed in Section 3.3, the different risk models are easily modified to accommodate this by replacing RV_{t+h}^h on the left-hand side with the volatility object of interest. In so doing, the RV s and variance risk factors on the right-hand side in the models are also naturally replaced by the same transformations. Implementing, such transformed versions of the risk models for forecasting $\sqrt{RV_{t+h}^h}$ again confirms our general conclusions. While the individually estimated risk models necessarily result in the highest R^2 's in-sample, the mega models that restrict the dynamic coefficients to be the same systematically result in the best out-of-sample volatility forecasts.

A.3. Relation between Utility and R^2

The discussion in the main text formally highlights the nonlinear relationship that exists between the predictive R^2 s and the realized utilities. Figure A3 empirically illustrates this relationship for three of the different risk models.

References

- Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold. 2006. Volatility and correlation forecasting. In *Handbook of economic forecasting* Eds. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier.
- . 2013. Financial risk measurement for financial risk management. In *Handbook of the economics of finance*. Eds. G. Constantinides, M. Harris and R. Stulz. Elsevier: Amsterdam.
- Andersen, T. G., T. Bollerslev, and F. X. Diebold. 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics* 89:701–20.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys. 2000. Great realizations. *Risk* 13:105–8.
- . 2003. Modeling and forecasting realized volatility. *Econometrica* 71:579–625.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega. 2003. Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review* 93:38–62.
- Andersen, T. G., T. Bollerslev, and N. Meddahi. 2011. Realized volatility forecasting and market microstructure noise. *Journal of Econometrics* 160, 220–234.
- Asness, C. S., T. J. Moskowitz, and L. H. Pedersen. 2013. Value and momentum everywhere. *Journal of Finance* 68:929–85.
- Baker, M., and J. Wurgler. 2006. Investor sentiment and the cross-section of stock returns. *Journal of Finance* 61, 1645–1680.
- Baker, M., J. Wurgler, and Y. Yuan. 2012. Global, local, and contagious sentiment. *Journal of Financial Economics* 104:272–87.
- Bakshi, G., and D. Madan. 2006. A theory of volatility spreads. *Management Science* 52:1945–56.
- Balduzzi, P., and A. W. Lynch. 1999. Transaction costs and predictability: Some utility cost calculations. *Journal of Finance* 52:47–78.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard. 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76:1481–536.
- Barroso, P., and P. Santa-Clara. 2015. Momentum has its moments. *Journal of Financial Economics* 116:111–20.
- Bollerslev, T., G. Tauchen, and H. Zhou. 2009. Expected stock returns and variance risk premia. *Review of Financial Studies* 22:4463–92.
- Boudoukh, J., R. Feldman, S. Kogan, and M. Richardson. 2016. Information, trading, and volatility: Evidence from firm-specific news. Working Paper, Hebrew University, University of Texas at Austin, and New York University.
- Brown, G. W. 1999. Volatility, sentiment, and noise traders. *Financial Analyst Journal* 55:82–90.
- Campbell, J. Y., and S. B. Thompson. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21:s1509–31.
- Cenesizoglu, T., and A. Timmermann. 2012. Do return prediction models add economic value? *Journal of Banking & Finance* 36:2974–87.
- Chen, X., and E. Ghysels. 2011. News – good or bad – and its impact on volatility predictions over multiple horizons. *Review of Financial Studies* 24:46–81.
- Clements, M. P., A. B. Galvao, and J. H. Kim. 2008. Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance* 15:729–750.
- Corsi, F. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7:174–96.
- Daniel, K., and T. J. Moskowitz. 2016. Momentum crashes. *Journal of Financial Economics* 122:221–47.

- Diebold, F. 2015. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests. *Journal of Business and Economic Statistics* 33:1–9.
- Diebold, F., and R. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13:253–63.
- Ederington, L. H., and J. H. Lee. (1993). How markets process information: News releases and volatility. *Journal of Finance* 48:1161–91.
- Engle, R., T. Ito, and W. Lin. 1990. Meteor showers of heat waves? Heteroskedastic intradaily volatility in the foreign exchange market. *Econometrica* 58:525–42.
- Engle, R., and J. Mezrich. 1996. GARCH for groups. *Risk* 9:36–40.
- Engle, R., and J. Rangel. 2008. The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies* 21:1187–222.
- Fleming, J., C. Kirby, and B. Ostdiek. 2001. The economic value of volatility timing. *Journal of Finance* 56:329–52.
- . 2003. The economic value of volatility timing using realized volatility. *Journal of Financial Economics* 67:473–509.
- Foroni, C., M. Marcellino, and C. Schumacher. 2015. Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society* 178:57–82.
- French, K., and R. Roll. 1986. Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics* 17:5–26.
- French, K. R., G. W. Schwert, and R. F. Stambaugh. 1987. Expected stock returns and volatility. *Journal of Financial Economics* 19:3–30.
- Gallant, R. A., C. T. Hsu, and G. Tauchen. 1999. Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *Review of Economics and Statistics* 81:617–31.
- Garleanu, N., and L. H. Pedersen. 2013. Dynamic trading with predictably returns and transaction costs. *Journal of Finance* 68:2309–40.
- . 2016. Dynamic portfolios choice with frictions. *Journal of Economic Theory* 165:487–516.
- Ghysels, E., and H. Qian. 2016. Estimating MIDAS regressions via ols with polynomial parameter profiling. Working Paper, UNC.
- Ghysels, E., A. Rubia, and R. Valkanov. 2009. Multi-period forecasts of variance: Direct, iterated, and mixed-data approaches. Working Paper, Alicante University, UCSD and UNC.
- Ghysels, E., P. Santa-Clara, and R. Valkanov. 2006. Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131:59–96.
- Ghysels, E., and A. Sinko. 2011. Volatility forecasting and microstructure noise. *Journal of Econometrics* 160:257–71.
- Ghysels, E., A. Sinko, and R. Valkanov. 2007. MIDAS regressions: Further results and new directions. *Econometric Reviews* 26:53–90.
- Hansen, L. P., and R. J. Hodrick. 1980. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy* 88:829–53.
- Hansen, P. R., and A. Lunde. 2005a. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20:873–89.
- . 2005b. A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics* 3:525–54.
- Hansen, P. R., and A. Lunde. 2014. Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory* 30:60–93.

- Hastie, T., R. Tibshirani, and J. Friedman. 2009. *Elements of statistical learning: Data mining, inference and prediction*. New York: Springer.
- Hsieh, D. A. 1991. Chaos and nonlinear dynamics: Application to financial markets. *Journal of Finance* 46, 1839–1877.
- Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter. 2009. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications* 119:2249–76.
- Karolyi, A. G. 1993. A Bayesian approach to modeling stock return volatility for option valuation. *Journal of Financial and Quantitative Analysis* 28:579–93.
- . 1995. A multivariate GARCH model of international transmissions of stock returns and volatility: The case of the United States and Canada. *Journal of Business and Economic Statistics* 13:11–25.
- Karolyi, G., K. Lee, and M. van Dijk. 2012. Understanding commonality in liquidity around the world. *Journal of Financial Economics* 105:82–112.
- Karpoff, J. M. 1987. The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis* 22:109–26.
- Liu, L., A. J. Patton, and K. Sheppard. 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics* 187:293–311.
- Lynch, A. W. and S. Tan. 2010. Multiple risky assets, transaction costs, and return predictability: Allocation rules and implications for us investors. *Journal of Financial and Quantitative Analysis* 45:1015–53.
- Moreira, A., and T. Muir. 2017. Volatility-managed portfolios. *Journal of Finance* 72:1611–44.
- Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen. 2012. Time series momentum. *Journal of Financial Economics* 104:228–50.
- Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–8.
- Patton, A. J. 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160:246–56.
- Patton, A. J., and K. Sheppard. 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics* 97:683–97.
- Pedersen, L. H. 2015. *Efficiently inefficient: How smart money invests and market prices are determined*. Princeton, NJ: Princeton University Press.
- Sangvinatsos, A., and J. A. Wachter. 2005. Does the failure of the expectations hypothesis matter for long-term investors? *Journal of Finance* 60:179–230.
- Schwert, G. W. 1989. Why does stock market volatility change over time? *Journal of Finance* 44:1115–53.
- Sizova, N. 2011. Integrated variance forecasting: Model based vs. reduced form. *Journal of Econometrics* 162:294–311.
- Swanson, N. R., and H. White. 1997. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting* 13:439–61.
- Taylor, S. 1986. *Modelling financial time series*. New York: Wiley.
- . 2015. *Asset price dynamics, volatility, and prediction*. Princeton, NJ: Princeton University Press.
- Yu, J., and Y. Yuan. 2011. Investor sentiment and the mean-variance relation. *Journal of Financial Economics* 100:367–81.
- Zhang, L., P. A. Mykland, and Y. Ait-Sahalia. 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100:1394–411.