Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

The Information Content of High-Frequency Data for Estimating Equity Return Models and Forecasting Risk

Dobrislav P. Dobrev and Pawel J. Szerszen

2010-45

order of the true quantile forecasts from low (representing bad times) to high (representing good times), as indicated in the first column. In our model this is equivalent to sorting by terminal volatility state from high (representing bad times) to low (representing good times). Each three rows reported for ranks 1 (low) to 5 (high) of the true quantile forecasts contain results for three different sample lengths T equal to 2 years, 5 years and 20 years (as given in the second column), taking parameter and volatility estimation uncertainty into account. For each quantile and forecast horizon we report results for the two alternative Bayesian estimation procedures in adjacent column pairs: either with (right column denoted "HF 5-min") or without (left column denoted "Daily only") augmenting the underlying state-space formulation with our daily volatility measurement equation based on high frequency intraday data.

As a graphical summary of the results reported in Table 8, Figure 1 plots the one-percent VaR (top graph) and five-percent VaR (bottom graph) relative forecast errors at a five-day horizon as a function of the rank order of the underlying true forecasts from low (representing bad times) to high (representing good times). The resulting VaR forecast errors without utilizing our high-frequency volatility measures are plotted as a solid line (denoted "Daily"), while those incorporating the information content of intraday data for the latent daily volatility are plotted as a dashed line (denoted "HF 5-min"). The reported relative errors of conditional return quantile forecasts can be interpreted also as the percentage overestimation or understimation of the implied capital charge for market risk based on one-percent (quantile 0.01) and five-percent (quantile 0.05) VaR.

Both Table 8 and Figure 1 reveal that risk forecasts stemming from traditional model inference on daily data tend to be overly conservative in good times (e.g. overestimating risk by as much as 30%) but they are not conservative enough in bad times (e.g. underestimating risk by as much as 10%). By contrast, risk forecasts based on our approach to exploiting high-frequency data are considerably closer to the truth in both bad and good times.

Leaving the reported magnitudes aside, this result is very intuitive as the use of volatility measures based on high frequency data allows for considerably faster and more precise incorporation of major changes in the current volatility level compared to daily data alone. For example, in bad times when volatility goes up it should take a longer sequence of daily returns alone than in conjunction with high-frequency volatility measures to deliver volatility state estimates that are not downward biased. Similarly, in good times when volatility goes down it should take longer for daily data alone than in conjunction with high-frequency volatility measures to produce volatility state estimates that are not upward biased. Thus, the observed differences between the risk forecast errors in bad versus good times (Table 8 and Figure 1) are completely in line with the asymmetric increase in volatility state uncertainty, coupled also with higher parameter uncertainty (see Section 4.2 above), characterizing traditional daily estimation in comparison to the proposed approach utilizing

also high-frequency data. In sum, thanks to incorporating the strong information content of high-frequency volatility measures, we are able to better curb risk taking exactly when needed the most, i.e. early on in times of crisis, while avoiding unnecessary overstatement of risk in normal times.

5 Empirical Illustration

Conditional return quantile forecasts play important role in risk management as they represent value-at-risk (VaR) forecasts. A key testable implication from our analysis in the previous section is that during bad times, e.g. early on in times of crisis, VaR forecast time-series based on our approach to exploiting high-frequency data will tend to "cross from above" the VaR forecast time-series stemming from traditional model inference on daily data. This is because, as explained above, the daily-based VaR forecasts are downward biased in bad times (when risk is elevated) and upward biased in good times (when risk is minimal), while our HF-based VaR forecasts are considerably closer to the truth in both bad and good times.

In order to test the empirical validity of this important risk management implication, we study the dynamics of five-day ahead VaR forecasts for S&P 500 and Google returns throughout the financial crisis of 2007-2008. Our goal is to illustrate the potentially large economic value from the proposed approach to incorporating the information content of high-frequency volatility measures. It is beyond the scope of this paper, though, to run a horse race between many viable alternative VaR forecasting techniques. We limit ourselves strictly to evaluating the empirical validity of our main testable implication with regard to HF-based versus daily-based VaR forecasts in the context of popular equity return models such as the fairly general two-factor log-SV model with jumps analyzed in the previous sections.

5.1 Data and estimation

In our empirical illustration we consider S&P 500 daily futures returns for the period October 2, 1985 - February 26, 2009 and Google daily equity returns for the period August 30, 2004 - July 31, 2009.[23] We exclude from each series holidays and shortened trading days. Our high-frequency measurement equation is constructed from five-minute intraday returns following the procedures given in section 2.2, while model estimation and forecasting is conducted as detailed in sections 3 and 4. We study the dynamics of five-day ahead VaR forecasts for the last 120 business weeks in each sample, both of which cover the financial crisis of 2007-2008. To produce each forecast we re-estimate our two-factor log-SV model

[23]The data for S&P 500 is provided by Tick Data, while the data for Google is from NYSE TAQ.

with all available data going back to the beginning of each sample. Thus, the sample for S&P 500 roughly corresponds to 20 years of data in our Monte Carlo study (Section 4). The sample for Google, on the other hand, represents 2-5 years of data and cannot be extended further back as it starts ten days after Google's IPO.

5.2 Forecasting risk throughout the 2007-2008 financial crisis

On Figures 2 and 3 we plot one-percent (top graph) and five-percent (bottom graph) VaR forecasts without overlapping at five-day horizon for S&P 500 futures returns (Figure 2) and Google equity returns (Figure 3) based on a two-factor log-SV model with jumps in returns. The model is estimated at a daily discretization interval by Bayesian MCMC methods either without or with augmenting the underlying state-space formulation with our daily volatility measurement equation based on high frequency intraday data. The resulting VaR forecasts without utilizing high-frequency volatility measures are plotted as a solid line (denoted "VaR with daily data"), those incorporating the information content of intraday data for the latent daily volatility are plotted as a dashed line (denoted "VaR with HF 5-min data"), while the corresponding actual observed returns are plotted as vertical bars (denoted "Return realizations").

As clearly seen from the graphs, the VaR forecasts with HF 5-min data seemingly correctly predict more risk and "cross from above" the VaR forecasts with daily data exactly around major turmoil events during the financial crisis of 2007-2008. These include the Bear Sterns turmoil in July 2007, the Countrywide turmoil in January 2008, the Fannie Mae and Freddie Mac turmoil in July 2008, and most notably, the Lehman Brothers collapse followed by the TARP Legislation turmoil in October 2008. The gap between the two alternative VaR forecasts around these events implies sizeable underestimation of risk by the traditional approach based on daily data. This is more pronounced for Google in line with the fact that individual stocks tend to be more risky than stock indices. At the same time, before the summer of 2007 and on many occasions afterwards the VaR forecasts with HF 5-min data predict a bit less risk than the VaR forecasts with daily data. Nonetheless, the number of incurred violations (given by the number of times the return realizations, plotted as vertical bars, go below the VaR forecasts) remains completely in line with the expected number of violations at the 1% and 5% VaR levels across 120 (non-overlapping) forecasts.

Overall, the observed dynamics of VaR forecasts for S&P 500 and Google returns throughout the financial crisis of 2007-2008 is in striking agreement with the key testable implication from our analysis in the previous sections. We obtain strong empirical support that not only in theory but also in important real-world examples our approach to incorporating the information content of high frequency volatility measures can help better curb risk taking exactly when needed the most, i.e. early on in times of crisis, while avoiding unnecessary overstatement of risk in normal times.

6 Conclusion

In this paper, we have developed a method for estimating popular equity return models relying not only on daily returns but also on nowadays ubiquitous high-frequency intraday return data. The essence of our approach is to borrow asymptotic results from the growing realized volatility literature and cast them as precise volatility measurement equations directly within the standard state-space representation of popular equity return models estimated at daily frequency. In this way, we avoid specifying explicitly the intraday return dynamics, while considerably improving estimation efficiency of such models at daily or monthly frequency. In particular, we utilize daily returns along with high-frequency jump-robust realized volatility measures within a standard Bayesian MCMC estimation framework. This allows us to take explicitly into account the resulting substantial reduction in parameter uncertainty. Thus, we are able to show sizeable economic gains when forecasting risk, compared to inference based on the more limited information provided by daily returns alone.

In this way, we depart from previous studies geared primarily towards specification testing that have focused on the use of such high-frequency volatility measures in classical rather than Bayesian estimation procedures. Instead, we demonstrate that across a variety of equity return models estimated at daily frequency the parameters controlling skewness and kurtosis can be obtained almost as precisely as if volatility is observable by incorporating the strong information content of realized volatility measures extracted from high-frequency data. In particular, we show that not only the parameters controlling volatility of volatility but also those controlling leverage effects can be estimated several times more precisely by exploiting high-frequency volatility measures. Furthermore, we show that our highly efficient estimates lead in turn to substantial gains for forecasting various risk measures at horizons ranging from a few days to a few months ahead when taking also into account parameter uncertainty. In fact, our approach not only reduces the root mean square prediction error but also shrinks and almost eliminates the forecast bias, which inevitably arises from the pronounced nonlinearities in the involved transformation of parameter and volatility estimates. As a practical rule of thumb we find that two years of high frequency data often suffice to obtain the same level of precision as twenty years of daily data, thereby making our approach particularly useful in finance applications where only short data samples are available or economically meaningful to use. Last, but perhaps most important in risk management applications, we find that risk forecasts based on our approach to exploiting high-frequency data are considerably closer to the truth in both bad and good times relative to those stemming from traditional model inference on daily data, which we find can overestimate risk by as much as 30% in good times or underestimate it by as much as 10% in bad times. We support our findings both with extensive simula-

tions and an empirical illustration on VaR forecasts for S&P500 and Google returns during the financial crisis of 2007-2008. Thanks to incorporating the strong information content of high-frequency volatility measures, we are able to better curb risk taking exactly when needed the most, i.e. early on in times of crisis (rather than with a delay), while avoiding unnecessary overstatement of risk in normal times. Qualitatively, our findings are robust both across different models and jump-robust volatility measures on high frequency data that we analyze.

In view of the documented substantial precision gains in forecasting risk of equity returns, the estimation approach we propose can directly add value in different areas of risk management and asset pricing. Beyond equity returns, the method can be applied also to other financial data such as foreign exchange rates, bonds and interest rates. It can be easily geared also towards model specification testing. More generally, we establish a promising and tractable way to incorporate additional sources of information, such as alternative high frequency volatility measures, into models in state space form.

References

Aït-Sahalia, Y. and J. Jacod (2007). Volatility estimators for discretely sampled lévy processes. *Ann. Stat. 35*(1), 355–392.

Alizadeh, S., M. Brandt, and F. Diebold (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance 57*(3).

Andersen, T. G., L. Benzoni, and J. Lund (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance 57*(3), 1239–1284.

Andersen, T. G., T. Bollerslev, P. Christoffersen, and F. X. Diebold (2007). Practical volatility and correlation modeling for financial market risk management. In *The Risks of Financial Institutions*, NBER Chapters, pp. 513–548. National Bureau of Economic Research, Inc.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2009). *Parametric and nonparametric volatility measurement*. North Holland. In Handbook of Financial Econometrics, Yacine Aït-Sahalia, Lars P. Hansen, and Jose A. Scheinkman (Eds.).

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*(1), 43–76.

Andersen, T. G., T. Bollerslev, and D. Dobrev (2007). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *Journal of Econometrics 138*(1), 125–80.

Andersen, T. G., D. P. Dobrev, and E. Schaumburg (2009). Jump-robust volatility estimation using nearest neighbor truncation. *NBER Working Paper* (15533).

Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *The Journal of Finance 52*(5), 2003–2049.

Bandi, F. M. and R. Reno (2009). Nonparametric stochastic volatility. Global COE Hi-Stat Discussion Paper Series gd08-035, Institute of Economic Research, Hitotsubashi University.

Bandi, F. M. and J. R. Russell (2007). *Volatility*. Elsevier Science, New York. in Handbook of Financial Engineering, V. Linetski and J. Birge (Eds.).

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*, 253–280.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 2*(1), 1–37.

Barndorff-Nielsen, O. E. and N. Shephard (2005). How accurate is the asymptotic approximation to the distribution of realised volatility? In D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models. A Festschrift in Honour of T.J. Rothenberg*, pp. 306–331. Cambridge: Cambridge University Press.

Barndorff-Nielsen, O. E. and N. Shephard (2007). *Variation, Jumps, Market Frictions and High Frequency Data in Fnancial Econometrics*. Cambridge University Press. In Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress, R. Blundell, T. Persson, and W. Newey (Eds.).

Barndorff-Nielsen, O. E., N. Shephard, and M. Winkel (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Processes and Their Applications 116*, 796–806.

Bates, D. S. (2000). Post-'87 crash fears in the s&p 500 futures option market. *Journal of Econometrics 94*(1-2), 181–238.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*, 637–659.

Bollerslev, T. and H. Zhou (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics 109*, 33–65.

Broadie, M., M. Chernov, and M. Johannes (2007). Model specification and risk premia: Evidence from futures options. *Journal of Finance 62*(3), 1453–1490.

Chernov, M., A. Gallant, E. Ghysels, and G. Tauchen (2003). Alternative models for stock price dynamics. *Journal of Econometrics 116*, 225–257.

Chernov, M. and E. Ghysels (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of option valuation. *Journal of Financial Economics 56*, 407–458.

Chib, S. and E. Greenberg (1996). Markov chain monte carlo simulation methods in econometrics. *Econometric Theory 12*(3), 409–431.

Christensen, K., R. Oomen, and M. Podolskij (2008). Realized quantile-based estimation of integrated variance. *Working paper*.

Corradi, V. and W. Distaso (2006). Semiparametric comparison of stochastic volatility models via realized measures. *Review of Economic Studies 73*, 635–667.

Corsi, F., D. Pirino, and R. Renò (2008). Volatility forecasting: The jumps do matter. *SSRN Working Paper*.

Das, S. R. and R. K. Sundaram (1999). Of smiles and smirks: A term structure perspective. *The Journal of Financial and Quantitative Analysis 34*(2), 211–239.

Eraker, B. (2004). Do stock prices and volatility jump? reconciling evidence from spot and option prices. *The Journal of Finance 59*(3), 1367–1403.

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *The Journal of Finance 58*(3), 1269–1300.

Fleming, J., C. Kirby, and B. Ostdiek (2003, March). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics 67*(3), 473–509.

Hansen, P. R. and A. Lunde (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics 3*(4), 525–554.

Huang, X. and G. Tauchen (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics 3*(4), 456–99.

Jacquier, E., N. Polson, and P. Rossi (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics 122*, 185–212.

Johannes, M. and N. Polson (2002). Mcmc methods for financial econometrics. In *Handbook of Financial Econometrics*. North-Holland. Forthcoming.

Jones, C. (1998). Bayesian estimation of continuous-time finance models. *Working Paper*.

Jones, C. (2003). The dynamics of stochastic volatility: evidence from underlying and options market. *Journal of Econometrics 116*, 181–224.

Mancini, C. (2006). Estimating the integrated volatility in stochastic volatility models with levy type jumps. *Working paper, University of Firenze*.

McAleer, M. and M. C. Medeiros (2008). Realized volatility: A review. *Econometric Reviews 27*(1-3), 10–45.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics 51*(3), 247–257.

Pan, J. (2002). The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics 63*(1), 3–50.

Podolskij, M. and M. Vetter (2009). Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes Appl. 119*(9), 2803–2831.

Szerszen, P. (2009). Bayesian analysis of stochastic volatility models with lévy jumps: application to risk analysis. *Board of Governors of the Federal Reserve System Working Paper, 40*.

Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *Journal of Econometrics 148*(2), 131–148.

A Analytical Results In Toy Model: A Motivating Example

In this paper we introduce asymptotically exact volatility measurement equations in state space form and propose a Bayesian MCMC estimation approach that we use to demonstrate the efficiency gains when estimating key parameters in various popular SV models. Although our expressions for the complete conditional posteriors given in section 3.3.2 provide an intuitive explanation where the documented efficiency gains are coming from, it is useful to provide some extra analytical support for the obtained results also by using classical estimation methods in a suitable "toy model".

To this end, here we restrict attention on estimating the kurtosis parameter σ_h in the following simplified log-SV model in state space form augmented with a volatility measurement equation based on high-frequency data:

$$\begin{aligned} r_{t+1} &= \exp(\frac{h_t}{2})\ \varepsilon_{t+1}^{(r)} && (21)\\ h_{t+1} &= \beta_h h_t + \sigma_h\ \varepsilon_{t+1}^{(h)} && (22)\\ \log(\widehat{IV}_{t+1;M}) &= h_t + \sqrt{\frac{\nu}{M}}\ \varepsilon_{t+1}^{(IV)} && (23) \end{aligned}$$

where all error terms are i.i.d. Gaussian. Note that equations (21)-(22) represent a canonical log-SV model for daily returns (and log-variances) extensively studied in the literature, see e.g. Taylor (1986), Nelson (1988), Harvey, Ruiz, and Shephard (1994), Ruiz (1994), Andersen and Sorensen (1996), Francq and Zakoïan (2006), among others.[24] Without loss of generality, the log-variance process is zero mean with persistence controlled by $\beta_h = 1 - \kappa_h$. Equation (23) represents our volatility measurement equation based on high-frequency intraday data in its simplest form (for implicitly assumed Brownian intraday dynamics), where $M >> 1$ is the intraday sample frequency and ν is an efficiency factor depending on the chosen volatility measure $\widehat{IV}_{t+1;M}$ as detailed in section 2.2.

As usual, it is convenient to substitute the return measurement equation in this canonical log-SV model with the one obtained after taking the logarithm of squared returns (without incurring any information loss when the distribution of $\varepsilon_t^{(r)}$ is symmetric):

$$\begin{aligned} \log(r_{t+1}^2) &= h_t + \log(\varepsilon_{t+1}^{(r)})^2 && (24)\\ h_{t+1} &= \beta_h h_t + \sigma_h\ \varepsilon_{t+1}^{(h)} && (25)\\ \log(\widehat{IV}_{t+1;M}) &= h_t + \sqrt{\frac{\nu}{M}}\ \varepsilon_{t+1}^{(IV)} && (26) \end{aligned}$$

It is convenient to further simplify notation by redefining the measurements and their errors as $x_t = \log(r_t^2) - \mathbb{E}[\log(\varepsilon_t^{(r)})^2]$, $\varepsilon_t^{(x)} = \varepsilon_t^{(r)} - \mathbb{E}[\log(\varepsilon_t^{(r)})^2]$, $y_t = \log(\widehat{IV}_{t+1;M})$, $\varepsilon_t^{(y)} = \sqrt{\nu}\ \varepsilon_t^{(IV)}$. This yields the following representation of the model in state space form:

$$\begin{aligned} x_{t+1} &= h_t + \varepsilon_{t+1}^{(x)} && (27)\\ h_{t+1} &= \beta_h h_t + \sigma_h\ \varepsilon_{t+1}^{(h)} && (28)\\ y_{t+1} &= h_t + \frac{1}{\sqrt{M}}\ \varepsilon_{t+1}^{(y)} && (29) \end{aligned}$$

[24] Comprehensive surveys of the literature on SV models and estimation include Andersen, Bollerslev, and Diebold (2009), Ghysels, Harvey and Renault (1996), Shephard (1996), Taylor (1994), among others.

In what follows, we study the efficiency of estimating σ_h taking β_h as given in the following two specifications: (i) Standard "Daily" given by the first two equations (27)-(28); (ii) Augmented "Daily + HF" given by the full system (27)-(29). Intuitively, it is clear that the relative difference in the precision of the two measurement equations (27) and (29) determines the attainable efficiency gains from using both measurement equations in the proposed "Daily + HF" specification as opposed to using only the first measurement equation in the standard "Daily" specification. Clearly, increasing the sample frequency M improves the precision of the additional volatility measurement equation (29) and in the limit as $M \to \infty$ it yields perfect measurements of the volatility states. This means that the maximum attainable efficiency in the case of perfectly observed volatility states can be closely achieved by increasing the sample frequency M sufficiently.

A GMM estimation approach provides a straightforward formalization of these intuitive observations. Let $m_q^{\varepsilon(x)} = \mathbb{E}[(\varepsilon_t^{(x)})^q]$, $q = 2, 4$ and $m_q^{\varepsilon(y)} = \mathbb{E}[(\varepsilon_t^{(y)})^q]$, $q = 2, 4$ denote the known second and fourth unconditional moments of the two measurement error terms. Consider the following two moment conditions:

$$g_1(\sigma_h, \beta_h) = x_t^2 - \frac{\sigma_h^2}{1-\beta_h^2} - m_2^{\varepsilon(x)} \tag{30}$$

$$g_2(\sigma_h, \beta_h) = y_t^2 - \frac{\sigma_h^2}{1-\beta_h^2} - \frac{1}{M}\, m_2^{\varepsilon(y)} \tag{31}$$

It is easy to confirm that these are valid moments:

$$\mathbb{E}[g_1] = 0 \tag{32}$$

$$\mathbb{E}[g_2] = 0 \tag{33}$$

The corresponding variance of each moment and the covariance between them is given by:

$$\mathbb{V}[g_1] = \frac{2\sigma_h^4}{(1-\beta_h^2)^2} + \frac{4\sigma_h^2}{(1-\beta_h^2)}\, m_2^{\varepsilon(x)} + m_4^{\varepsilon(x)} - (m_2^{\varepsilon(x)})^2 \tag{34}$$

$$\mathbb{V}[g_2] = \frac{2\sigma_h^4}{(1-\beta_h^2)^2} + \frac{4\sigma_h^2}{(1-\beta_h^2)}\, \frac{m_2^{\varepsilon(y)}}{M} + \frac{m_4^{\varepsilon(y)} - (m_2^{\varepsilon(y)})^2}{M^2} \tag{35}$$

$$\mathbb{C}[g_1, g_2] = \frac{2\sigma_h^4}{(1-\beta_h^2)^2} \tag{36}$$

Note that the unconditional second moment of the log-variance process is given by $m_2^h = \mathbb{E}[h_t^2] = \frac{\sigma_h^2}{(1-\beta_h^2)}$. Hence, the above variance and covariance expressions take the following form:

$$\mathbb{V}[g_1] = 2\,(m_2^h)^2 + 4\, m_2^h\, m_2^{\varepsilon(x)} + m_4^{\varepsilon(x)} - (m_2^{\varepsilon(x)})^2 \tag{37}$$

$$\mathbb{V}[g_2] = 2\,(m_2^h)^2 + 4\, m_2^h\, \frac{m_2^{\varepsilon(y)}}{M} + \frac{m_4^{\varepsilon(y)} - (m_2^{\varepsilon(y)})^2}{M^2} \tag{38}$$

$$\mathbb{C}[g_1, g_2] = 2\,(m_2^h)^2 \tag{39}$$

The resulting optimal GMM weighting matrix is given by:

$$\begin{pmatrix} \mathbb{V}[g_1] & \mathbb{C}[g_1, g_2] \\ \mathbb{C}[g_1, g_2] & \mathbb{V}[g_2] \end{pmatrix}^{-1} = \frac{1}{\mathbb{V}[g_1]\,\mathbb{V}[g_2] - \mathbb{C}[g_1, g_2]^2} \begin{pmatrix} \mathbb{V}[g_2] & -\mathbb{C}[g_1, g_2] \\ -\mathbb{C}[g_1, g_2] & \mathbb{V}[g_1] \end{pmatrix} \tag{40}$$

It follows that the ratio between the variance of an estimator of σ_h based on the first moment condition ("Daily" specification) and the variance of the optimal GMM estimator of σ_h combining both moment conditions ("Daily + HF" specification) is given by:

$$\frac{\mathbb{V}[g_1]}{\left(\frac{\mathbb{V}[g_1]+\mathbb{V}[g_2]-2\,\mathbb{C}[g_1,g_2]}{\mathbb{V}[g_1]\,\mathbb{V}[g_2]-\mathbb{C}[g_1,g_2]^2}\right)^{-1}} = 1 + \frac{\left(\frac{\mathbb{V}[g_1]}{\mathbb{C}[g_1,g_2]}-1\right)^2}{\frac{\mathbb{V}[g_1]}{\mathbb{C}[g_1,g_2]}\,\frac{\mathbb{V}[g_2]}{\mathbb{C}[g_1,g_2]}-1} \tag{41}$$

$$= 1 + \frac{2\,\frac{m_2^{\varepsilon(x)}}{m_2^h} + \frac{1}{2}\,[\,\frac{m_4^{\varepsilon(x)}}{(m_2^h)^2} - (\frac{m_2^{\varepsilon(x)}}{m_2^h})^2]}{1 + \left(\frac{2}{M}\,\frac{m_2^{\varepsilon(y)}}{m_2^h} + \frac{1}{2M^2}[\frac{m_4^{\varepsilon(y)}}{(m_2^h)^2} - (\frac{m_2^{\varepsilon(y)}}{m_2^h})^2]\right)\left(1 + 2\,\frac{m_2^{\varepsilon(x)}}{m_2^h} + \frac{1}{2}\,[\,\frac{m_4^{\varepsilon(x)}}{(m_2^h)^2} - (\frac{m_2^{\varepsilon(x)}}{m_2^h})^2]\right)} \tag{42}$$

Expressed in this form, the variance reduction factor is a function of the variance of each measurement error relative to the variance of the state variable and, hence, the intraday sample frequency M affecting the precision of the second measurement equation based on high-frequency intraday data.

Two important conclusions follow. First, as $M \to \infty$ this variance reduction factor approaches

$$1 + 2\,\frac{m_2^{\varepsilon(x)}}{m_2^h} + \frac{1}{2}\,[\frac{m_4^{\varepsilon(x)}}{(m_2^h)^2} - (\frac{m_2^{\varepsilon(x)}}{m_2^h})^2] = \frac{\mathbb{V}[g_1]}{\mathbb{C}[g_1,g_2]} \equiv \frac{\mathbb{V}[g_1]}{\lim_{M\to\infty}\mathbb{V}[g_2]}\ , \tag{43}$$

which means that the Hausman principle applies in the limit, in the sense that when volatility is perfectly observed in the second measurement equation then it alone achieves minimum variance, i.e. maximum efficiency of the estimator. Second, for values of M typically used in empirical work such as $M = 78$ (five-minute returns) and $M = 195$ (two-minute returns), the above variance reduction factor (42) is very close to its limiting value (43) for $M \to \infty$ since the denominator in (42) would be close to unity.

This proves analytically in the considered simplified setting that augmenting the state space form of the model with a volatility measurement equation based on high frequency data yields an estimator with several times smaller variance compared to the one without a volatility measurement equation. For typical values of M in the order of 100 the variance reduction factor is fairly close to its limiting value (43). In particular, based on the derived formulas it is easy to see that for parameter values in the neighborhood of those used in prior studies of the same model (see for example Ruiz (1994) or Andersen and Sorensen (1997)) implies variance reduction factor somewhere in the range 5 to 30 times, which roughly translates into 2 to 5 times smaller standard deviation. This is quite in line with the RMSE reduction documented in our Monte Carlo study for popular non-analytically tractable models for which we propose Baysian MCMC estimation methods with the added benefit of more fully exploiting information via the model state space form.

B Figures and Tables

Table 1: **Parameter estimates for a one-factor log-SV model with leverage effects.** For select model parameters we report the mean, bias, and RMSE of the estimates obtained across 1,000 Monte Carlo replications. The state-space form of the model is as follows:

$$\begin{aligned} Y_{t+1} - Y_t &= \mu + \exp(\frac{h_t}{2})\ \varepsilon_{t+1}^{(1)} \\ h_{t+1} &= h_t + \kappa_h(\theta_h - h_t) + \sigma_h\ (\rho_h \cdot \varepsilon_{t+1}^{(1)}\ +\ \sqrt{(1-\rho_h^2)} \cdot \varepsilon_{t+1}^{(2)}) \\ \log(\widehat{IV}_{t,t+1;M}) &\approx h_t + \sqrt{\frac{1}{M}\widehat{\Omega}_{t,t+1;M}}\ \varepsilon_{t+1}^{(IV)} \end{aligned}$$

Columns represent results for alternative estimation procedures depending on weather our volatility measurement equation based on high-frequency log integrated variance measures $\log(\widehat{IV}_{t,t+1;M})$ is used (HF 5-min with M=78; HF 2-min with M=195) or not (daily only), as well as for the infeasible case of perfect observability (known volatility). The rows in each block contain results for different yearly sample lengths (2, 5, or 20 years).

	Sample length	κ_h = 0.0163				σ_h = 0.1648			
	T (years)	Daily Only	HF 5-min	HF 2-min	Known Volatility	Daily Only	HF 5-min	HF 2-min	Known Volatility
MEAN	2	0.0249	0.0176	0.0174	0.0173	0.1739	0.1655	0.1647	0.1649
	5	0.0188	0.0171	0.0171	0.0171	0.1694	0.1649	0.1650	0.1649
	20	0.0168	0.0165	0.0165	0.0165	0.1653	0.1647	0.1647	0.1647
BIAS	2	0.0086	0.0014	0.0011	0.0010	0.0091	0.0007	0.0000	0.0001
	5	0.0025	0.0009	0.0009	0.0008	0.0046	0.0002	0.0002	0.0001
	20	0.0006	0.0002	0.0002	0.0002	0.0005	-0.0001	0.0000	0.0000
RMSE	2	0.0309	0.0092	0.0090	0.0089	0.0260	0.0087	0.0069	0.0047
	5	0.0071	0.0050	0.0049	0.0048	0.0183	0.0058	0.0045	0.0030
	20	0.0028	0.0021	0.0021	0.0021	0.0100	0.0028	0.0023	0.0015
	Sample length	θ_h = - 9.4243				ρ_h = - 0.6716			
	T (years)	Daily Only	HF 5-min	HF 2-min	Known Volatility	Daily Only	HF 5-min	HF 2-min	Known Volatility
MEAN	2	-9.4024	-9.4837	-9.5006	-9.5067	-0.5840	-0.6618	-0.6651	-0.6679
	5	-9.4481	-9.4468	-9.4478	-9.4467	-0.6360	-0.6668	-0.6683	-0.6704
	20	-9.4308	-9.4299	-9.4297	-9.4295	-0.6599	-0.6709	-0.6703	-0.6706
BIAS	2	0.0219	-0.0594	-0.0764	-0.0825	0.0875	0.0098	0.0065	0.0037
	5	-0.0239	-0.0225	-0.0235	-0.0224	0.0355	0.0047	0.0032	0.0012
	20	-0.0065	-0.0057	-0.0054	-0.0053	0.0116	0.0006	0.0012	0.0010
RMSE	2	0.9659	0.8943	0.8897	0.8720	0.1383	0.0395	0.0321	0.0210
	5	0.3026	0.2906	0.2917	0.2911	0.0776	0.0247	0.0214	0.0140
	20	0.1416	0.1390	0.1395	0.1398	0.0392	0.0132	0.0110	0.0073

Table 2: **Parameter estimates for a two-factor log-SV model with leverage effects and jumps.** For select model parameters we report the mean, bias, and RMSE of the estimates obtained across 1,000 Monte Carlo replications. The state-space form of the model is as follows:

$$
\begin{aligned}
Y_{t+1} - Y_t &= \mu + \exp(\frac{h_t + f_t}{2})\ \varepsilon_{t+\Delta}^{(1)} + q_{t+1} \cdot J_{t+1} \\
h_{t+1} &= h_t + \kappa_h(\theta_h - h_t) + \sigma_h\ (\rho_h \cdot \varepsilon_{t+1}^{(1)}\ +\ \sqrt{(1-\rho_h^2)} \cdot \varepsilon_{t+1}^{(2)}) \\
f_{t+1} &= f_t + \kappa_f(\theta_f - f_t) + \sigma_f\ (\rho_f \cdot \varepsilon_{t+1}^{(1)}\ +\ \sqrt{(1-\rho_f^2)} \cdot \varepsilon_{t+1}^{(3)}) \\
\log(\widehat{IV}_{t,t+1;M}) &\approx h_t + f_t + \sqrt{\frac{1}{M}\widehat{\Omega}_{t,t+1;M}}\ \varepsilon_{t+1}^{(IV)}
\end{aligned}
$$

Columns represent results for alternative estimation procedures depending on weather our volatility measurement equation based on high-frequency log integrated variance measures $\log(\widehat{IV}_{t,t+1;M})$ is used (HF 5-min with M=78; HF 2-min with M=195) or not (daily only), as well as for the infeasible case of perfect observability (known volatility). The rows in each block contain results for different yearly sample lengths (2, 5, or 20 years).

	Sample length	κ_h = 0.0130				κ_f = 0.6724			
	T (years)	Daily Only	HF 5-min	HF 2-min	Known Volatility	Daily Only	HF 5-min	HF 2-min	Known Volatility
MEAN	2	0.0254	0.0203	0.0202	0.0154	0.9868	0.7011	0.6930	0.6750
	5	0.0195	0.0157	0.0156	0.0142	0.9146	0.6870	0.6774	0.6719
	20	0.0150	0.0136	0.0136	0.0134	0.7303	0.6810	0.6742	0.6721
BIAS	2	0.0124	0.0073	0.0072	0.0024	0.3144	0.0287	0.0206	0.0026
	5	0.0065	0.0027	0.0026	0.0012	0.2422	0.0146	0.0050	-0.0005
	20	0.0020	0.0006	0.0006	0.0004	0.0579	0.0086	0.0018	-0.0003
RMSE	2	0.0206	0.0156	0.0156	0.0101	0.3794	0.0866	0.0775	0.0424
	5	0.0121	0.0079	0.0079	0.0059	0.3400	0.0512	0.0444	0.0275
	20	0.0042	0.0027	0.0027	0.0025	0.2444	0.0260	0.0231	0.0131
	Sample length	σ_h = 0.1320				σ_f = 0.3876			
	T (years)	Daily Only	HF 5-min	HF 2-min	Known Volatility	Daily Only	HF 5-min	HF 2-min	Known Volatility
MEAN	2	0.1553	0.1426	0.1421	0.1321	0.1685	0.3818	0.3795	0.3862
	5	0.1507	0.1367	0.1362	0.1322	0.1752	0.3883	0.3849	0.3864
	20	0.1400	0.1333	0.1329	0.1320	0.2643	0.3919	0.3879	0.3872
BIAS	2	0.0233	0.0106	0.0101	0.0001	-0.2191	-0.0058	-0.0081	-0.0014
	5	0.0187	0.0047	0.0042	0.0002	-0.2124	0.0007	-0.0027	-0.0012
	20	0.0080	0.0013	0.0009	0.0000	-0.1233	0.0043	0.0003	-0.0004
RMSE	2	0.0338	0.0208	0.0205	0.0042	0.2203	0.0196	0.0184	0.0121
	5	0.0285	0.0143	0.0139	0.0026	0.2159	0.0118	0.0117	0.0081
	20	0.0154	0.0072	0.0070	0.0013	0.1423	0.0075	0.0056	0.0039
	Sample length	ρ_h = - 0.2831				ρ_f = - 0.1109			
	T (years)	Daily Only	HF 5-min	HF 2-min	Known Volatility	Daily Only	HF 5-min	HF 2-min	Known Volatility
MEAN	2	-0.2225	-0.2615	-0.2610	-0.2802	-0.1443	-0.1130	-0.1172	-0.1095
	5	-0.2509	-0.2769	-0.2755	-0.2815	-0.1739	-0.1096	-0.1116	-0.1105
	20	-0.2669	-0.2849	-0.2824	-0.2824	-0.1712	-0.1104	-0.1112	-0.1098
BIAS	2	0.0606	0.0216	0.0221	0.0029	-0.0334	-0.0021	-0.0063	0.0014
	5	0.0322	0.0062	0.0076	0.0016	-0.0630	0.0013	-0.0007	0.0004
	20	0.0162	-0.0018	0.0007	0.0007	-0.0603	0.0005	-0.0003	0.0011
RMSE	2	0.1965	0.1316	0.1276	0.0411	0.2543	0.0628	0.0613	0.0439
	5	0.1376	0.0864	0.0857	0.0256	0.2305	0.0399	0.0375	0.0292
	20	0.0774	0.0461	0.0453	0.0121	0.1185	0.0190	0.0178	0.0143

Table 3: **Volatility state estimates for a two-factor log-SV model with leverage effects and jumps.** We report the mean, bias, and RMSE of the terminal volatility state estimates obtained across 1,000 Monte Carlo replications. The state-space form of the model is as follows:

$$\begin{aligned} Y_{t+1} - Y_t &= \mu + \exp(\frac{h_t + f_t}{2})\ \varepsilon^{(1)}_{t+\Delta} + q_{t+1} \cdot J_{t+1} \\ h_{t+1} &= h_t + \kappa_h(\theta_h - h_t) + \sigma_h\ (\rho_h \cdot \varepsilon^{(1)}_{t+1}\ +\ \sqrt{(1-\rho_h^2)} \cdot \varepsilon^{(2)}_{t+1}) \\ f_{t+1} &= f_t + \kappa_f(\theta_f - f_t) + \sigma_f\ (\rho_f \cdot \varepsilon^{(1)}_{t+1}\ +\ \sqrt{(1-\rho_f^2)} \cdot \varepsilon^{(3)}_{t+1}) \\ \log(\widehat{IV}_{t,t+1;M}) &\approx h_t + f_t + \sqrt{\frac{1}{M}\widehat{\Omega}_{t,t+1;M}}\ \varepsilon^{(IV)}_{t+1} \end{aligned}$$

Columns represent results for alternative estimation procedures depending on weather our volatility measurement equation based on high-frequency log integrated variance measures $\log(\widehat{IV}_{t,t+1;M})$ is used (HF 5-min with M=78; HF 2-min with M=195) or not (daily only), as well as for the infeasible case of perfect observability (known volatility). The rows in each block contain results for different yearly sample lengths (2, 5, or 20 years).

	Sample length	h_T : E[h_T] = -9.8998				f_T : E[f_T] = 0			
	T (years)	Daily Only	HF 5-min	HF 2-min	Known Volatility	Daily Only	HF 5-min	HF 2-min	Known Volatility
MEAN	2	-9.8259	-9.8616	-9.8560	-9.9200	0.0017	0.0014	-0.0013	-0.0065
	5	-9.8631	-9.9379	-9.9238	-9.8467	-0.0028	-0.0170	-0.0172	-0.0084
	20	-9.9611	-9.9978	-9.9893	-9.9037	-0.0042	-0.0028	-0.0029	0.0013
BIAS	2	0.0030	-0.0326	-0.0270	0.0000	0.0207	0.0204	0.0177	0.0000
	5	0.0520	-0.0229	-0.0087	0.0000	0.0299	0.0157	0.0155	0.0000
	20	0.0255	-0.0112	-0.0027	0.0000	0.0184	0.0199	0.0197	0.0000
RMSE	2	0.4748	0.3049	0.2958	0.0000	0.4207	0.4075	0.4070	0.0000
	5	0.4687	0.3116	0.2997	0.0000	0.4343	0.4197	0.4189	0.0000
	20	0.4491	0.2736	0.2662	0.0000	0.3980	0.3813	0.3800	0.0000

Table 4: **Forecasts of conditional return moments at horizons of one, five, ten and twenty days ahead for a one-factor log-SV model with leverage effects.** We report the mean, bias and RMSE of forecasts of the conditional return mean, standard deviation, skewness and kurtosis at horizons of one (panel A), five (panel B), ten (panel C) and twenty (panel D) days ahead, obtained across 1,000 Monte Carlo replications. Each three rows reported for the mean, bias, and RMSE of these moment forecasts contain results for three different sample lengths T equal to 2, 5 and 20 years (as indicated in the second column), taking parameter and volatility estimation uncertainty into account. For each moment and forecast horizon we report results for two alternative Bayesian estimation procedures in adjacent column pairs: either without (left column denoted "Daily only") or with (right column denoted "HF 5-min") augmenting the underlying state-space formulation with a daily volatility measurement equation based on high frequency intraday data, as proposed in this paper.

Panel A: 1-day horizon

		Mean Forecast		StdDev Forecast		Skewness Forecast		Kurtosis Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	0.0001	0.0001	0.0089	0.0085	0.0004	0.0001	3.6083	3.0767
	5	0.0001	0.0001	0.0090	0.0087	0.0000	0.0000	3.5487	3.0757
	20	0.0001	0.0001	0.0088	0.0087	0.0000	0.0000	3.5070	3.0752
BIAS	2	0.0000	0.0000	0.0004	0.0000	0.0003	0.0001	0.6141	0.0826
	5	0.0000	0.0000	0.0003	0.0000	0.0000	0.0001	0.5546	0.0815
	20	0.0000	0.0000	0.0001	-0.0001	0.0000	0.0000	0.5129	0.0811
RMSE	2	0.0003	0.0003	0.0019	0.0008	0.0023	0.0008	0.6347	0.0831
	5	0.0002	0.0002	0.0018	0.0008	0.0011	0.0007	0.5671	0.0818
	20	0.0001	0.0001	0.0018	0.0008	0.0009	0.0007	0.5214	0.0812

Panel B: 5-day horizon

		Mean Forecast		StdDev Forecast		Skewness Forecast		Kurtosis Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	0.0009	0.0008	0.0238	0.0228	-0.0747	-0.0728	3.3648	3.1203
	5	0.0009	0.0009	0.0239	0.0232	-0.0756	-0.0729	3.3232	3.1185
	20	0.0010	0.0010	0.0235	0.0231	-0.0761	-0.0733	3.2938	3.1175
BIAS	2	-0.0001	-0.0002	0.0011	0.0001	-0.0020	-0.0001	0.2857	0.0411
	5	-0.0001	-0.0001	0.0008	0.0001	-0.0030	-0.0003	0.2442	0.0396
	20	0.0000	0.0000	0.0002	-0.0002	-0.0036	-0.0008	0.2149	0.0386
RMSE	2	0.0024	0.0022	0.0050	0.0021	0.0352	0.0108	0.3032	0.0441
	5	0.0015	0.0014	0.0047	0.0020	0.0234	0.0072	0.2534	0.0408
	20	0.0007	0.0007	0.0047	0.0020	0.0125	0.0037	0.2204	0.0390

Panel C: 10-day horizon

		Mean Forecast		StdDev Forecast		Skewness Forecast		Kurtosis Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	0.0017	0.0015	0.0327	0.0313	-0.0953	-0.0929	3.3719	3.1644
	5	0.0017	0.0016	0.0328	0.0318	-0.0957	-0.0926	3.3232	3.1614
	20	0.0018	0.0018	0.0322	0.0317	-0.0959	-0.0929	3.2909	3.1598
BIAS	2	-0.0002	-0.0004	0.0016	0.0002	-0.0032	-0.0008	0.2449	0.0374
	5	-0.0001	-0.0003	0.0012	0.0002	-0.0038	-0.0007	0.1966	0.0349
	20	0.0000	0.0000	0.0003	-0.0002	-0.0041	-0.0011	0.1643	0.0332
RMSE	2	0.0045	0.0042	0.0067	0.0029	0.0453	0.0141	0.2717	0.0457
	5	0.0027	0.0026	0.0063	0.0027	0.0297	0.0093	0.2092	0.0384
	20	0.0013	0.0013	0.0062	0.0027	0.0154	0.0047	0.1708	0.0343

Panel D: 20-day horizon

		Mean Forecast		StdDev Forecast		Skewness Forecast		Kurtosis Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	0.0031	0.0027	0.0451	0.0433	-0.1375	-0.1347	3.5016	3.2777
	5	0.0031	0.0029	0.0451	0.0438	-0.1370	-0.1335	3.4183	3.2694
	20	0.0034	0.0034	0.0441	0.0434	-0.1364	-0.1334	3.3711	3.2646
BIAS	2	-0.0003	-0.0007	0.0024	0.0006	-0.0051	-0.0022	0.2698	0.0460
	5	-0.0003	-0.0005	0.0017	0.0004	-0.0049	-0.0014	0.1879	0.0390
	20	0.0000	0.0000	0.0004	-0.0002	-0.0045	-0.0015	0.1410	0.0345
RMSE	2	0.0082	0.0077	0.0091	0.0042	0.0669	0.0219	0.3366	0.0760
	5	0.0050	0.0048	0.0084	0.0037	0.0429	0.0142	0.2169	0.0539
	20	0.0025	0.0023	0.0079	0.0035	0.0214	0.0071	0.1525	0.0387

Table 5: **Forecasts of conditional return moments at horizons of one, five, ten and twenty days ahead for a two-factor log-SV model with leverage effects and jumps.** We report the mean, bias and RMSE of forecasts of the conditional return mean, standard deviation, skewness and kurtosis at horizons of one (panel A), five (panel B), ten (panel C) and twenty (panel D) days ahead, obtained across 1,000 Monte Carlo replications. Each three rows reported for the mean, bias, and RMSE of these moment forecasts contain results for three different sample lengths T equal to 2, 5 and 20 years (as indicated in the second column), taking parameter and volatility estimation uncertainty into account. For each moment and forecast horizon we report results for two alternative Bayesian estimation procedures in adjacent column pairs: either without (left column denoted "Daily only") or with (right column denoted "HF 5-min") augmenting the underlying state-space formulation with a daily volatility measurement equation based on high frequency intraday data, as proposed in this paper.

Panel A: 1-day horizon

		Mean Forecast		StdDev Forecast		Skewness Forecast		Kurtosis Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	0.0005	0.0005	0.0099	0.0098	-0.5831	-0.6173	92.9059	96.2450
	5	0.0006	0.0005	0.0097	0.0095	-0.5923	-0.6520	90.2971	102.7384
	20	0.0006	0.0006	0.0095	0.0094	-0.6382	-0.6956	98.1776	110.3568
BIAS	2	0.0000	0.0000	0.0000	-0.0001	0.0501	0.0158	-6.8173	-3.4781
	5	0.0000	0.0000	0.0002	0.0000	0.0824	0.0226	-17.8110	-5.3698
	20	0.0000	0.0000	0.0001	0.0000	0.0871	0.0297	-19.4314	-7.2521
RMSE	2	0.0003	0.0003	0.0020	0.0012	0.3128	0.2007	114.2795	36.3367
	5	0.0002	0.0002	0.0017	0.0012	0.3042	0.1981	60.8291	39.1525
	20	0.0001	0.0001	0.0015	0.0010	0.3085	0.1924	64.5270	37.8032

Panel B: 5-day horizon

		Mean Forecast		StdDev Forecast		Skewness Forecast		Kurtosis Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	0.0038	0.0038	0.0264	0.0261	-0.2473	-0.2722	17.1562	15.8092
	5	0.0039	0.0038	0.0258	0.0253	-0.2617	-0.2840	15.0475	16.6968
	20	0.0040	0.0039	0.0251	0.0249	-0.2800	-0.2973	16.1293	17.7012
BIAS	2	-0.0001	-0.0001	0.0002	0.0000	0.0286	0.0036	0.9160	-0.4310
	5	-0.0001	-0.0001	0.0005	-0.0001	0.0287	0.0066	-2.3048	-0.6776
	20	0.0000	-0.0001	0.0002	0.0000	0.0257	0.0084	-2.5075	-0.9356
RMSE	2	0.0022	0.0020	0.0052	0.0031	0.1766	0.0621	51.3547	4.9173
	5	0.0013	0.0012	0.0044	0.0030	0.1009	0.0642	8.1963	5.2467
	20	0.0006	0.0006	0.0039	0.0025	0.1001	0.0589	8.6529	5.0507

Panel C: 10-day horizon

		Mean Forecast		StdDev Forecast		Skewness Forecast		Kurtosis Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	0.0071	0.0071	0.0362	0.0358	-0.1987	-0.2149	10.8792	9.6985
	5	0.0072	0.0071	0.0353	0.0346	-0.2070	-0.2220	9.3137	10.1363
	20	0.0073	0.0072	0.0344	0.0341	-0.2198	-0.2306	9.8631	10.6367
BIAS	2	-0.0002	-0.0003	0.0005	0.0001	0.0169	0.0007	1.0048	-0.1759
	5	-0.0001	-0.0003	0.0007	0.0000	0.0184	0.0036	-1.1204	-0.3090
	20	0.0000	-0.0001	0.0003	0.0000	0.0158	0.0050	-1.2186	-0.4450
RMSE	2	0.0040	0.0037	0.0069	0.0042	0.0986	0.0443	35.4541	2.5309
	5	0.0024	0.0022	0.0059	0.0040	0.0710	0.0436	4.1702	2.6744
	20	0.0012	0.0011	0.0051	0.0032	0.0679	0.0395	4.3685	2.5495

Panel D: 20-day horizon

		Mean Forecast		StdDev Forecast		Skewness Forecast		Kurtosis Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	0.0132	0.0130	0.0498	0.0492	-0.1718	-0.1811	7.0727	6.5231
	5	0.0133	0.0130	0.0484	0.0474	-0.1765	-0.1848	6.3465	6.7111
	20	0.0136	0.0133	0.0470	0.0466	-0.1842	-0.1892	6.5902	6.9316
BIAS	2	-0.0004	-0.0005	0.0010	0.0004	0.0073	-0.0021	0.5331	-0.0166
	5	-0.0002	-0.0005	0.0010	0.0001	0.0084	0.0003	-0.4614	-0.1021
	20	0.0000	-0.0002	0.0004	0.0000	0.0070	0.0019	-0.5289	-0.1875
RMSE	2	0.0074	0.0067	0.0092	0.0058	0.0729	0.0353	14.6681	1.2715
	5	0.0044	0.0041	0.0077	0.0052	0.0515	0.0298	2.0322	1.3205
	20	0.0022	0.0020	0.0065	0.0041	0.0432	0.0252	2.0915	1.2267

Table 6: **Forecasts of conditional return quantiles at horizons of one, five, ten and twenty days ahead for a one-factor log-SV model with leverage effects.** We report the mean, bias and RMSE of forecasts of the 0.01, 0.05, 0.95 and 0.99 conditional return quantiles at horizons of one (panel A), five (panel B), ten (panel C) and twenty (panel D) days ahead, obtained across 1,000 Monte Carlo replications. Each three rows reported for the mean, bias, and RMSE of these quantile forecasts contain results for three different sample lengths T equal to 2, 5 and 20 years (as indicated in the second column), taking parameter and volatility estimation uncertainty into account. For each quantile and forecast horizon we report results for two alternative Bayesian estimation procedures in adjacent column pairs: either without (left column denoted "Daily only") or with (right column denoted "HF 5-min") augmenting the underlying state-space formulation with a daily volatility measurement equation based on high frequency intraday data, as proposed in this paper.

Panel A: 1-day horizon

		Quantile 0.01 Forecast		Quantile 0.05 Forecast		Quantile 0.95 Forecast		Quantile 0.99 Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	-0.0215	-0.0199	-0.0144	-0.0139	0.0146	0.0141	0.0218	0.0201
	5	-0.0217	-0.0202	-0.0146	-0.0142	0.0148	0.0144	0.0220	0.0205
	20	-0.0213	-0.0202	-0.0143	-0.0141	0.0146	0.0144	0.0215	0.0205
BIAS	2	-0.0018	-0.0002	-0.0005	0.0000	0.0005	-0.0001	0.0018	0.0001
	5	-0.0017	-0.0002	-0.0004	0.0000	0.0004	0.0000	0.0016	0.0002
	20	-0.0011	0.0000	-0.0001	0.0001	0.0001	-0.0001	0.0011	0.0000
RMSE	2	0.0048	0.0019	0.0032	0.0014	0.0031	0.0014	0.0048	0.0019
	5	0.0046	0.0018	0.0030	0.0013	0.0030	0.0013	0.0046	0.0018
	20	0.0044	0.0018	0.0030	0.0013	0.0030	0.0013	0.0044	0.0018

Panel B: 5-day horizon

		Quantile 0.01 Forecast		Quantile 0.05 Forecast		Quantile 0.95 Forecast		Quantile 0.99 Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	-0.0571	-0.0539	-0.0384	-0.0371	0.0394	0.0378	0.0566	0.0532
	5	-0.0574	-0.0548	-0.0387	-0.0376	0.0397	0.0385	0.0568	0.0541
	20	-0.0562	-0.0545	-0.0379	-0.0374	0.0390	0.0385	0.0558	0.0541
BIAS	2	-0.0038	-0.0006	-0.0017	-0.0003	0.0015	-0.0001	0.0036	0.0002
	5	-0.0032	-0.0005	-0.0013	-0.0002	0.0012	0.0000	0.0030	0.0002
	20	-0.0015	0.0002	-0.0003	0.0003	0.0003	-0.0003	0.0015	-0.0002
RMSE	2	0.0128	0.0056	0.0088	0.0042	0.0082	0.0040	0.0119	0.0052
	5	0.0118	0.0050	0.0080	0.0036	0.0078	0.0035	0.0113	0.0047
	20	0.0113	0.0049	0.0078	0.0035	0.0076	0.0034	0.0108	0.0047

Panel C: 10-day horizon

		Quantile 0.01 Forecast		Quantile 0.05 Forecast		Quantile 0.95 Forecast		Quantile 0.99 Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	-0.0785	-0.0745	-0.0525	-0.0507	0.0544	0.0522	0.0777	0.0733
	5	-0.0787	-0.0754	-0.0528	-0.0514	0.0547	0.0530	0.0779	0.0745
	20	-0.0768	-0.0748	-0.0516	-0.0509	0.0538	0.0530	0.0763	0.0744
BIAS	2	-0.0053	-0.0012	-0.0026	-0.0007	0.0022	0.0000	0.0048	0.0004
	5	-0.0041	-0.0008	-0.0019	-0.0005	0.0016	0.0000	0.0037	0.0003
	20	-0.0017	0.0002	-0.0004	0.0003	0.0004	-0.0004	0.0017	-0.0003
RMSE	2	0.0179	0.0085	0.0124	0.0066	0.0113	0.0059	0.0162	0.0075
	5	0.0160	0.0072	0.0109	0.0053	0.0106	0.0050	0.0151	0.0066
	20	0.0151	0.0066	0.0104	0.0047	0.0100	0.0045	0.0142	0.0062

Panel D: 20-day horizon

		Quantile 0.01 Forecast		Quantile 0.05 Forecast		Quantile 0.95 Forecast		Quantile 0.99 Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	-0.1092	-0.1041	-0.0720	-0.0697	0.0754	0.0723	0.1075	0.1016
	5	-0.1089	-0.1048	-0.0720	-0.0702	0.0754	0.0732	0.1070	0.1028
	20	-0.1058	-0.1035	-0.0701	-0.0692	0.0741	0.0731	0.1047	0.1024
BIAS	2	-0.0078	-0.0026	-0.0040	-0.0017	0.0034	0.0003	0.0069	0.0011
	5	-0.0057	-0.0016	-0.0028	-0.0010	0.0023	0.0001	0.0049	0.0006
	20	-0.0020	0.0003	-0.0005	0.0004	0.0005	-0.0004	0.0019	-0.0003
RMSE	2	0.0258	0.0140	0.0182	0.0111	0.0159	0.0094	0.0223	0.0114
	5	0.0219	0.0109	0.0150	0.0082	0.0143	0.0073	0.0202	0.0094
	20	0.0197	0.0089	0.0136	0.0064	0.0128	0.0059	0.0181	0.0081

Table 7: **Forecasts of conditional return quantiles at horizons of one, five, ten and twenty days ahead for a two-factor log-SV model with leverage effects and jumps.** We report the mean, bias and RMSE of forecasts of the 0.01, 0.05, 0.95 and 0.99 conditional return quantiles at horizons of one (panel A), five (panel B), ten (panel C) and twenty (panel D) days ahead, obtained across 1,000 Monte Carlo replications. Each three rows reported for the mean, bias, and RMSE of these quantile forecasts contain results for three different sample lengths T equal to 2, 5 and 20 years (as indicated in the second column), taking parameter and volatility estimation uncertainty into account. For each quantile and forecast horizon we report results for two alternative Bayesian estimation procedures in adjacent column pairs: either without (left column denoted "Daily only") or with (right column denoted "HF 5-min") augmenting the underlying state-space formulation with a daily volatility measurement equation based on high frequency intraday data, as proposed in this paper.

Panel A: 1-day horizon

		Quantile 0.01 Forecast		Quantile 0.05 Forecast		Quantile 0.95 Forecast		Quantile 0.99 Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	-0.0211	-0.0207	-0.0133	-0.0131	0.0144	0.0142	0.0221	0.0217
	5	-0.0205	-0.0198	-0.0129	-0.0125	0.0141	0.0136	0.0216	0.0208
	20	-0.0199	-0.0193	-0.0124	-0.0122	0.0136	0.0134	0.0210	0.0204
BIAS	2	-0.0008	-0.0003	-0.0001	0.0001	0.0001	-0.0001	0.0007	0.0003
	5	-0.0011	-0.0004	-0.0004	0.0000	0.0003	-0.0001	0.0011	0.0003
	20	-0.0011	-0.0005	-0.0002	0.0000	0.0002	0.0000	0.0011	0.0005
RMSE	2	0.0058	0.0034	0.0037	0.0022	0.0037	0.0022	0.0058	0.0034
	5	0.0050	0.0033	0.0032	0.0022	0.0032	0.0022	0.0050	0.0034
	20	0.0045	0.0028	0.0029	0.0018	0.0029	0.0018	0.0045	0.0028

Panel B: 5-day horizon

		Quantile 0.01 Forecast		Quantile 0.05 Forecast		Quantile 0.95 Forecast		Quantile 0.99 Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	-0.0618	-0.0605	-0.0352	-0.0348	0.0420	0.0414	0.0650	0.0633
	5	-0.0603	-0.0582	-0.0342	-0.0332	0.0410	0.0398	0.0635	0.0611
	20	-0.0585	-0.0570	-0.0328	-0.0323	0.0397	0.0391	0.0617	0.0600
BIAS	2	-0.0023	-0.0010	-0.0005	-0.0001	0.0005	-0.0002	0.0024	0.0006
	5	-0.0029	-0.0009	-0.0011	-0.0001	0.0010	-0.0002	0.0030	0.0006
	20	-0.0024	-0.0009	-0.0007	-0.0002	0.0007	0.0000	0.0025	0.0008
RMSE	2	0.0139	0.0082	0.0098	0.0062	0.0098	0.0061	0.0138	0.0082
	5	0.0120	0.0078	0.0086	0.0058	0.0084	0.0059	0.0118	0.0079
	20	0.0104	0.0064	0.0076	0.0048	0.0074	0.0047	0.0104	0.0065

Panel C: 10-day horizon

		Quantile 0.01 Forecast		Quantile 0.05 Forecast		Quantile 0.95 Forecast		Quantile 0.99 Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	-0.0897	-0.0885	-0.0481	-0.0475	0.0608	0.0598	0.0963	0.0945
	5	-0.0875	-0.0857	-0.0466	-0.0453	0.0593	0.0576	0.0942	0.0918
	20	-0.0854	-0.0845	-0.0446	-0.0439	0.0576	0.0566	0.0921	0.0907
BIAS	2	-0.0023	-0.0011	-0.0011	-0.0005	0.0009	-0.0001	0.0024	0.0006
	5	-0.0024	-0.0006	-0.0017	-0.0003	0.0015	-0.0002	0.0026	0.0001
	20	-0.0013	-0.0004	-0.0010	-0.0003	0.0010	0.0000	0.0016	0.0003
RMSE	2	0.0165	0.0100	0.0132	0.0087	0.0132	0.0085	0.0166	0.0101
	5	0.0133	0.0088	0.0115	0.0077	0.0112	0.0079	0.0133	0.0093
	20	0.0108	0.0069	0.0100	0.0063	0.0098	0.0062	0.0111	0.0072

Panel D: 20-day horizon

		Quantile 0.01 Forecast		Quantile 0.05 Forecast		Quantile 0.95 Forecast		Quantile 0.99 Forecast	
	T (years)	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min	Daily Only	HF 5-min
MEAN	2	-0.1224	-0.1208	-0.0656	-0.0647	0.0889	0.0875	0.1388	0.1366
	5	-0.1189	-0.1169	-0.0632	-0.0616	0.0867	0.0844	0.1355	0.1327
	20	-0.1158	-0.1151	-0.0605	-0.0597	0.0844	0.0831	0.1327	0.1314
BIAS	2	-0.0039	-0.0023	-0.0022	-0.0013	0.0017	0.0003	0.0034	0.0012
	5	-0.0030	-0.0010	-0.0024	-0.0007	0.0020	-0.0003	0.0027	-0.0001
	20	-0.0012	-0.0004	-0.0013	-0.0005	0.0013	0.0000	0.0012	-0.0001
RMSE	2	0.0219	0.0143	0.0179	0.0125	0.0178	0.0122	0.0214	0.0138
	5	0.0169	0.0114	0.0149	0.0101	0.0145	0.0105	0.0162	0.0117
	20	0.0130	0.0084	0.0126	0.0079	0.0123	0.0079	0.0128	0.0085

Table 8: **Relative errors of conditional return quantile forecasts sorted by rank order of the underlying true forecasts at horizons of one, five, ten and twenty days ahead for a two-factor log-SV model with leverage effects and jumps.** We report relative errors of forecasts of the 0.01 and 0.05 conditional return quantiles at horizons of one (panel A), five (panel B), ten (panel C) and twenty (panel D) days ahead, calculated across 1,000 Monte Carlo replications as the mean of the percentage difference between a forecast based on parameter and state estimates and the forecast based on the corresponding true values. The results are sorted by the rank order of the true quantile forecasts from low (representing bad times) to high (representing good times), as indicated in the first column. Each three rows reported for ranks 1 (low) to 5 (high) of the true quantile forecasts contain results for three different sample lengths T equal to 2 years, 5 years and 20 years (as given in the second column), taking parameter and volatility estimation uncertainty into account. For each quantile and forecast horizon we report results for two alternative Bayesian estimation procedures in adjacent column pairs: either without (left column denoted "Daily only") or with (right column denoted "HF 5-min") augmenting the underlying state-space formulation with a daily volatility measurement equation based on high frequency intraday data, as proposed in this paper. The reported relative errors of conditional return quantile forecasts can be interpreted also as the percentage overestimation or understimation of the implied capital charge for market risk based on one-percent (quantile 0.01) and five-percent (quantile 0.05) VaR.

		Panel A: 1-day horizon				**Panel B: 5-day horizon**				**Panel C: 10-day horizon**				**Panel D: 20-day horizon**			
True Forecast	**Sample Size**	**Quantile 0.01 Forecast Error**		**Quantile 0.05 Forecast Error**		**Quantile 0.01 Forecast Error**		**Quantile 0.05 Forecast Error**		**Quantile 0.01 Forecast Error**		**Quantile 0.05 Forecast Error**		**Quantile 0.01 Forecast Error**		**Quantile 0.05 Forecast Error**	
Rank (quintiles)	**T (years)**	**Daily Only**	**HF 5-min**	**Daily Only**	**HF 5-min**	**Daily Only**	**HF 5-min**	**Daily Only**	**HF 5-min**	**Daily Only**	**HF 5-min**	**Daily Only**	**HF 5-min**	**Daily Only**	**HF 5-min**	**Daily Only**	**HF 5-min**
1st (Low)	**2**	-6.49%	-2.45%	-9.18%	-4.75%	-4.58%	-1.64%	-7.91%	-3.30%	-2.96%	-0.56%	-6.45%	-2.02%	-0.46%	1.66%	-3.80%	0.39%
	5	-6.67%	-4.03%	-9.31%	-6.27%	-5.69%	-3.58%	-9.00%	-5.33%	-4.74%	-2.81%	-8.37%	-4.51%	-3.70%	-1.61%	-7.22%	-3.08%
	20	-4.14%	-1.49%	-7.56%	-3.79%	-4.26%	-1.93%	-7.57%	-3.61%	-3.95%	-1.79%	-7.42%	-3.50%	-3.99%	-1.76%	-7.19%	-3.28%
2nd	**2**	-1.83%	-1.09%	-4.84%	-3.26%	-0.57%	-0.61%	-4.01%	-2.35%	0.84%	0.13%	-3.01%	-1.66%	2.16%	0.97%	-1.30%	-0.45%
	5	0.81%	1.01%	-1.82%	-1.20%	1.38%	0.87%	-1.24%	-0.35%	1.82%	1.09%	-0.68%	0.12%	2.06%	1.35%	0.20%	0.87%
	20	-0.25%	-0.58%	-4.17%	-2.95%	-0.48%	-0.88%	-3.98%	-2.66%	-0.22%	-0.57%	-3.76%	-2.53%	-0.46%	-0.57%	-3.52%	-2.26%
3rd	**2**	7.88%	4.67%	5.02%	2.53%	6.88%	3.87%	5.33%	3.23%	6.09%	3.46%	5.56%	3.56%	5.82%	3.51%	5.79%	4.00%
	5	11.19%	4.27%	8.29%	1.93%	9.31%	3.23%	8.51%	2.40%	7.42%	2.71%	8.40%	2.49%	6.38%	2.37%	7.95%	2.56%
	20	7.61%	4.67%	2.64%	2.24%	5.97%	3.29%	3.71%	2.71%	4.68%	2.52%	3.56%	2.77%	3.46%	2.03%	3.06%	2.41%
4th	**2**	15.20%	6.75%	12.15%	4.37%	11.87%	5.25%	12.19%	4.98%	8.17%	3.47%	11.88%	5.05%	6.58%	2.80%	10.86%	4.78%
	5	20.04%	9.40%	17.02%	7.05%	15.53%	6.58%	17.36%	7.36%	9.53%	3.85%	16.39%	6.53%	7.37%	2.75%	14.84%	6.01%
	20	17.81%	8.47%	14.69%	6.22%	13.82%	6.36%	14.73%	6.80%	6.75%	2.87%	14.47%	6.74%	5.21%	2.28%	13.27%	6.41%
5th (High)	**2**	30.80%	11.24%	27.24%	8.81%	19.70%	6.51%	28.48%	8.99%	6.55%	1.76%	27.54%	8.12%	5.33%	1.12%	23.39%	5.83%
	5	30.95%	11.75%	27.66%	9.26%	20.28%	7.60%	29.71%	9.75%	5.51%	1.45%	29.46%	9.63%	4.59%	1.12%	25.11%	7.22%
	20	37.23%	12.91%	32.52%	10.56%	22.04%	7.29%	35.70%	11.92%	5.45%	1.38%	35.05%	11.81%	4.52%	1.20%	29.98%	10.18%

Figure 1: **Relative error plots for five day VaR forecasts against the rank order of the underlying true forecasts for a two-factor log-SV model with leverage effects and jumps.** We plot one-percent VaR (top graph) and five-percent VaR (bottom graph) relative forecast errors at a five-day horizon as a function of the rank order of the underlying true forecasts from low (representing bad times) to high (representing good times). The errors are calculated as the mean of the percentage difference between a forecast based on parameter and state estimates and the forecast based on the corresponding true values across 1,000 Monte Carlo replications. The model is estimated at a daily discretization interval by Bayesian MCMC methods either without or with augmenting the underlying state-space formulation with a daily volatility measurement equation based on high frequency intraday data, as proposed in this paper. The resulting VaR forecast errors without utilizing high-frequency volatility measures are plotted as a solid line (denoted "Daily"), while those incorporating the information content of intraday data for the latent daily volatility are plotted as a dashed line (denoted "HF 5-min").

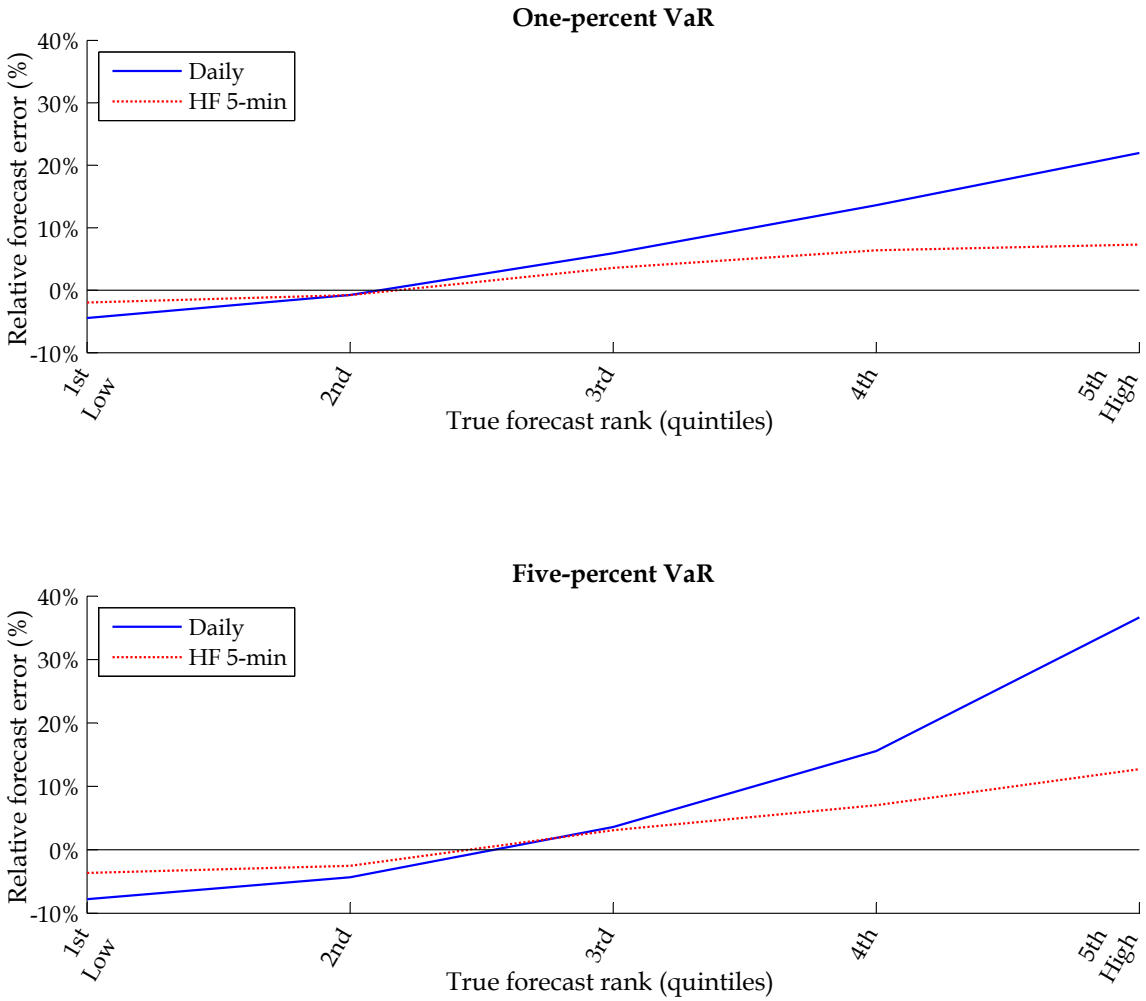

Figure 2: **One-percent and five-percent VaR forecasts for S&P 500 returns during the financial crisis of 2008.** We plot one-percent (top graph) and five-percent (bottom graph) VaR forecasts at five-day horizon without overlapping for S&P 500 futures returns based on a two-factor log-SV model with jumps in returns. The model is estimated at a daily discretization interval by Bayesian MCMC methods either without or with augmenting the underlying state-space formulation with a daily volatility measurement equation based on high frequency intraday data, as proposed in this paper. The resulting VaR forecasts without utilizing high-frequency volatility measures are plotted as a solid line (denoted "VaR with daily data"), those incorporating the information content of intraday data for the latent daily volatility are plotted as a dashed line (denoted "VaR with HF 5-min data"), while the corresponding actual observed returns are plotted as vertical bars (denoted "Return realizations"). The VaR analysis is for the period July 6, 2006 - February 19, 2009 and involves re-estimating the model on each date with all past data going back to October 2, 1985.

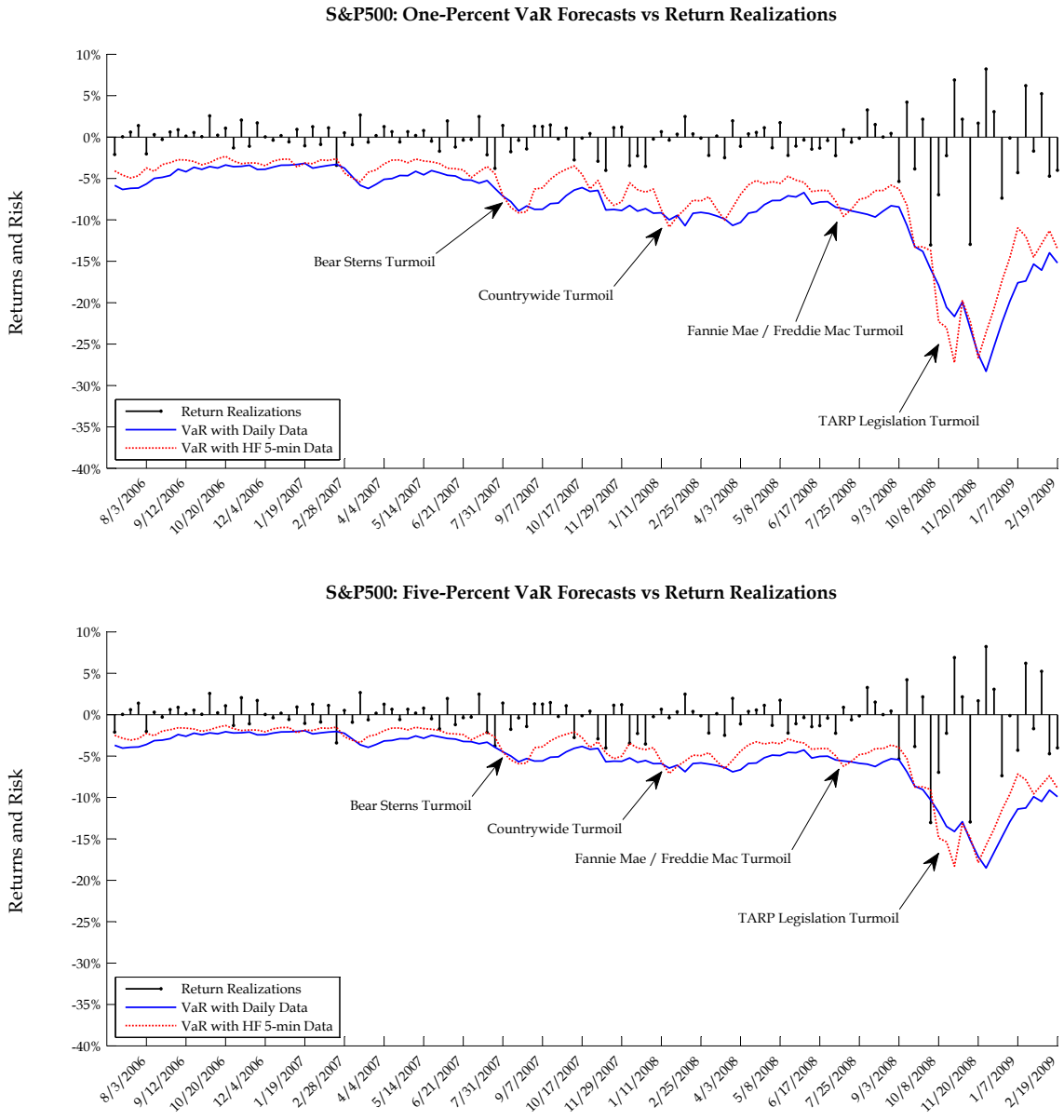

Figure 3: **One-percent and five-percent VaR forecasts for Google returns during the financial crisis of 2008.** We plot one-percent (top graph) and five-percent (bottom graph) VaR forecasts at five-day horizon without overlapping for Google returns based on a two-factor log-SV model with jumps in returns. The model is estimated at a daily discretization interval by Bayesian MCMC methods either without or with augmenting the underlying state-space formulation with a daily volatility measurement equation based on high frequency intraday data, as proposed in this paper. The resulting VaR forecasts without utilizing high-frequency volatility measures are plotted as a solid line (denoted "VaR with daily data"), those incorporating the information content of intraday data for the latent daily volatility are plotted as a dashed line (denoted "VaR with HF 5-min data"), while the corresponding actual observed returns are plotted as vertical bars (denoted "Return realizations"). The VaR analysis is for the period December 8, 2006 - July 24, 2009 and involves re-estimating the model on each date with all past data going back to August 30, 2004 (ten days after Google's IPO).