

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 1005

September 2010

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

Dobrislav Dobrev

Pawel Szerszen

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at www.federalreserve.gov/pubs/ifdp/. This paper can be downloaded without charge from Social Science Research Network electronic library at www.ssrn.com.

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

Dobrislav Dobrev, Pawel Szerszen *

First version: December 30, 2009

This version: August 25, 2010

Abstract

We demonstrate that the parameters controlling skewness and kurtosis in popular equity return models estimated at daily frequency can be obtained almost as precisely as if volatility is observable by simply incorporating the strong information content of realized volatility measures extracted from high-frequency data. For this purpose, we introduce asymptotically exact volatility measurement equations in state space form and propose a Bayesian estimation approach. 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. 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. Moreover, we find that compared to model inference without high-frequency data, our approach largely eliminates underestimation of risk during bad times or overestimation of risk during good times. We assess the attainable improvements in VaR forecast accuracy on simulated data and provide an empirical illustration on stock returns during the financial crisis of 2007-2008.

JEL classification: C11; C13; C14; C15; C22; C53; C80; G17;

Keywords: Equity return models; Parameter uncertainty; Bayesian estimation; MCMC; High-frequency data; Jump-robust volatility measures; Value at Risk; Forecasting.

*Dobrislav Dobrev: Federal Reserve Board of Governors, Dobrislav.P.Dobrev@frb.gov
Pawel Szerszen: Federal Reserve Board of Governors, Pawel.J.Szerszen@frb.gov

We are grateful to Torben Andersen, Federico Bandi, Luca Benzoni, Neil Ericsson, Michael Gordy, Erik Hjalmarsson, Michael Johannes, Matthew Pritsker, and Viktor Todorov for helpful discussions and comments. We also thank conference participants at the 2010 International Risk Management Conference, the 30th International Symposium on Forecasting, the 16th International Conference on Computing in Economics and Finance, as well as seminar participants at the Federal Reserve Board, Johns Hopkins University, the Office of the Comptroller of the Currency, and University of Maryland for their feedback. Excellent research assistance has been provided by Patrick Mason as well as Erica Reisman and Raymond Zhong. Any errors or omissions are our sole responsibility.

The views in this paper are solely those of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

1 Introduction

Modeling equity returns is central to risk management, derivatives pricing, portfolio choice, and asset pricing in general. Continuous time jump-diffusion models succeeding those pioneered by Merton (1969) and Black and Scholes (1973) are now commonplace. Typically, the inherent time-varying and stochastic nature of continuous market activity is represented by a combination of persistent and non-persistent latent stochastic volatility factors. The pronounced asymmetric return-volatility relation in equities, known also as leverage or volatility feedback effects, is captured by correlated return and volatility innovations. Sudden price revisions due to news and other market surprises give rise to jumps in returns, while the often abrupt changes in the level of market activity and risk has justified the introduction of jumps in volatility. The latent nature of volatility in such rich models, however, poses serious challenges for reliable inference based solely on daily or monthly return series, even the longest existing ones. It is thus critical to develop estimation methods exploiting relevant additional information that could help reduce the severe parameter and volatility estimation uncertainty.

Two different approaches have emerged to improve estimation efficiency in this regard. The first approach relies on the cross section of option prices over time.¹ However, as pointed out by Eraker, Johannes, and Polson (2003), it is unclear whether the inclusion of option price data leads to decrease or increase of parameter uncertainty given that the risk premia embedded in option prices introduce additional parameters, which are typically difficult to estimate. The second and seemingly more viable approach avoids such complications by exclusively relying on daily realized volatility measures extracted from nowadays ubiquitous high-frequency intraday return data.^{2,3} Our paper contributes to this second line of research by utilizing high-frequency realized volatility measures within a standard Bayesian Markov Chain Monte Carlo (MCMC) estimation framework of popular equity return models. In particular, we take explicitly into account the resulting substantial reduction in parameter uncertainty and are able to show sizeable economic gains when forecasting risk.

The most closely related studies to our work such as Alizadeh, Brandt, and Diebold (2002), Barndorff-Nielsen and Shephard (2002), Bollerslev and Zhou (2002), Corradi and Distaso (2006), Todorov (2009) among others have used classical rather than Bayesian estimation methods and have focused on using high-frequency volatility measures for assessing the goodness of fit of alternative model specifications without explicitly analyzing the economic value of reducing parameter uncertainty. These studies have largely ruled out the

¹See for example, Chernov and Ghysels (2000), Pan (2002), and Eraker (2004) among others.

²Recent surveys of the realized volatility literature include Andersen, Bollerslev, and Diebold (2009), Bandi and Russell (2007), Barndorff-Nielsen and Shephard (2007), McAleer and Medeiros (2008).

³Thorough empirical evidence pointing more broadly towards the value of realized volatility for modeling equity returns can be found in Andersen, Bollerslev, Diebold, and Ebens (2001).

simplest known single factor stochastic volatility models with Poisson jumps in returns in favor of more complex specifications including one or more extra features such as a second stochastic volatility factor, pronounced non-linearities, jumps both in returns and volatility (possibly even of infinite activity).⁴ But rather than reconciling or refining such findings, our main goal is to go a step beyond specification testing and clearly demonstrate the economic gains from harnessing the information content of high-frequency volatility measures regardless of the underlying model.

To this end, we exploit recent advances in jump robust volatility estimation from high-frequency data such as Andersen, Dobrev, and Schaumburg (2009), Barndorff-Nielsen, Shephard, and Winkel (2006), Podolskij and Vetter (2009) and references therein to formally introduce an asymptotically precise volatility measurement equation directly within the standard state-space representation of popular equity return models estimated at daily or lower frequency. Then we adopt a standard Bayesian MCMC estimation framework allowing us to exploit the strong information content of such volatility measurement equation across a wide range of models featuring stochastic volatility, leverage effects, and jumps in returns and volatility. In terms of efficiency, our approach considerably improves on Bayesian estimation methods based on an identical state-space representation at a daily or lower frequency but without a volatility measurement equation such as Eraker, Johannes, and Polson (2003) and Jacquier, Polson, and Rossi (2004), among others.⁵ In terms of generality, we overcome major limitations of the quasi-maximum likelihood estimation methods for state-space formulations with volatility measurement equation pursued by Barndorff-Nielsen and Shephard (2002) who consider non jump-robust realized volatility measures and Alizadeh, Brandt, and Diebold (2002) who consider non jump-robust and less efficient range-based volatility measures. In particular, our approach incorporates leverage effects and jumps, necessary for modeling equity returns, as well as possibly two (one persistent and one non-persistent) stochastic volatility factors. We also offer an attractive alternative to existing moment-based estimation approaches such as Bollerslev and Zhou (2002), Corradi and Distaso (2006) and Todorov (2009) in terms of more fully exploiting the information content of high frequency volatility measures in various model settings via their state-space formulations. In particular, unlike these studies, the Bayesian estimation approach we propose allows us to easily account for parameter uncertainty and demonstrate the economic gains from using high-frequency volatility measures for model estimation and risk forecasting across a range of popular equity return models.⁶

⁴Results in the same spirit have been obtained also in studies based solely on daily returns or in combination with options data such as Broadie, Chernov, and Johannes (2007). Other non-parametric studies based on high-frequency data include Andersen, Bollerslev, and Dobrev (2007), Bandi and Reno (2009).

⁵Cf. Andersen, Benzoni, and Lund (2002) and Chernov, Gallant, Ghysels, and Tauchen (2003).

⁶As such, our results add to the growing body of evidence showing the economic value of high-frequency realized volatility measures in finance applications, e.g. Fleming, Kirby, and Ostdiek (2003) among others.

Our main contributions can be summarized as follows. First, we demonstrate theoretically and empirically that the parameters controlling skewness and kurtosis in popular equity return models estimated at daily and monthly frequency 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 extend the empirical findings in Alizadeh, Brandt, and Diebold (2002) by showing 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. Second, 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. Third, and most important in risk management applications, our simulation results 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. 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. Finally, our findings are robust both across different models and jump-robust volatility measures on high frequency data that we analyze. This allows us to remain largely agnostic about the best suited ones, while making a strong case for the potentially large economic value of our approach to using high-frequency volatility measures in model estimation and risk forecasting or other closely related finance applications such as derivatives pricing.

The rest of the paper is organized as follows. Section 2 introduces our volatility measurement equations in detail. Section 3 incorporates such equations within the state space formulation of popular equity return models and develops appropriate Bayesian estimation methods. Section 4 documents the resulting gains in estimation efficiency and risk forecasting accuracy. Section 5 provides an empirical comparison of Value-at-Risk forecasts on S&P 500 and Google returns during the financial crisis of 2007-2008. Section 6 concludes.

2 Volatility measurement equations

Jumps in returns have been recognized as an important feature for continuous-time modeling of equity returns within standard no-arbitrage semimartingale setting. Moreover, recent progress in non-parametric volatility measurement based on high-frequency intraday data has made it possible to separate ex-post the daily continuous part of the volatility process from the daily return variation induced by discontinuities or jumps. Originally pioneered by Barndorff-Nielsen and Shephard (2004), jump-robust volatility estimators with different asymptotic and finite sample properties have been proposed by Andersen, Dobrev, and Schaumburg (2009), Barndorff-Nielsen, Shephard, and Winkel (2006), Podolskij and Vetter (2009) among others. A common feature among these and other high-frequency volatility estimators is that as the intraday sampling frequency increases, the arising measurement error shrinks to zero and converges to a known mixed normal asymptotic distribution.⁷

For our purposes, suitable asymptotic results of this kind directly imply asymptotically precise measurement equations that formally capture the extent to which the continuous and jump parts of daily total variance become ex-post nearly observable when high frequency intraday data is available. Such separation of the continuous and jump components of volatility can be directly utilized in state space form. In this section, we formally introduce a general form of the jump-robust volatility measurement equations that play a key role in our approach to estimating models in state space form and allow us to tackle considerably more general settings than those considered by Alizadeh, Brandt, and Diebold (2002) and Barndorff-Nielsen and Shephard (2002) in the absence of jump-robust volatility measures nearly a decade ago.

2.1 Jump-robust estimators of diffusive volatility

On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ we consider an adapted process $Y = \{Y_t\}_{t \geq 0}$, providing the following jump-diffusion representation of the evolution of the logarithmic price of an asset in continuous time:

$$dY_t = \mu_t dt + \sigma_t dB_t + dJ_t \tag{1}$$

Here μ is a locally bounded and predictable process, σ is cadlag and bounded away from zero almost surely, while J is a jump process so that dJ_t , whenever different from zero,

⁷The rate of convergence is typically square-root. It is slower for high-frequency volatility measures that are robust also to market microstructure noise, empirically found to matter at sample frequencies higher than a few minutes.

represents the size of a jump at time t . Without loss of generality, we restrict attention to finite activity jumps.⁸

For a day of unit length with $M + 1$ discrete observations of the logarithmic price process $\{Y_t\}_{0 \leq t \leq 1}$ on $0 \leq t_0 < t_1 < \dots < t_M \leq 1$ we denote the intraday time intervals and corresponding returns as $\Delta t_i = t_i - t_{i-1}$ and $\Delta Y_i = Y_{t_i} - Y_{t_{i-1}}$, $i = 1, \dots, M$. In what follows, we consider standard continuous record in-fill asymptotics where the time intervals characterizing the intraday sampling scheme uniformly shrink towards zero as the sampling frequency M increases.

In this setting, the daily quadratic variation (QV) of the observed process consists of the sum of its continuous and jump parts, $QV = \int_0^1 \sigma_u^2 du + \sum_{0 \leq u \leq 1} (dJ_u)^2$, and is estimated consistently by the well established realized volatility (RV) measure:⁹

$$RV_M = \sum_{i=1}^M (\Delta Y_i)^2. \quad (2)$$

Our main object of interest, though, is the diffusive part of the quadratic variation defined as the integrated variance (IV), $IV = \int_0^1 \sigma_u^2 du$. It can be conveniently estimated by various multipower variation measures developed by Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen, Shephard, and Winkel (2006) or more recent analogous measures based on nearest neighbor truncation developed by Andersen, Dobrev, and Schaumburg (2009).¹⁰ In the case of finite activity jumps, the most efficient multipower variation measure that allows for an asymptotic mixed normal limit theory is the realized tripower variation (TV) based on the product of triplets of adjacent absolute returns:¹¹

$$TV_M = \mu_{2/3}^{-3} \left(\frac{M}{M-2} \right) \sum_{i=2}^{M-1} |\Delta Y_{i-1}|^{2/3} |\Delta Y_i|^{2/3} |\Delta Y_{i+1}|^{2/3} \quad (3)$$

The TV estimator is only marginally less efficient than the corresponding MedRV estimator based on (two-sided) nearest neighbor truncation, taking the median instead of the product of triplets of adjacent absolute returns:¹²

⁸Our subsequent analysis remains valid to the extent that the utilized asymptotic results are unaffected by jumps of possibly infinite activity (but still finite variation).

⁹For recent surveys of the realized volatility literature see, e.g., Andersen, Bollerslev, and Diebold (2009), Bandi and Russell (2007), Barndorff-Nielsen and Shephard (2007), McAleer and Medeiros (2008).

¹⁰Other approaches that involve potentially delicate threshold or bandwidth choices include the truncated RV of Mancini (2006) and Ait-Sahalia and Jacod (2007), the truncated bipower variation of Corsi, Pirino, and Renò (2008), as well as the quantile RV estimator of Christensen, Oomen, and Podolskij (2008).

¹¹The scaling factor μ_p is defined as $\mu_p = E|U|^p = 2^{p/2} \frac{\Gamma((p+1)/2)}{\Gamma(1/2)}$, $U \sim N(0, 1)$.

¹²The asymptotic variance factor for TV is 3.06 as opposed to 2.96 for MedRV. Also by design, MedRV is somewhat more robust than TV not only to jumps but also to the occurrence of “zero” returns in finite

$$\text{MedRV}_M = \frac{\pi}{6 - 4\sqrt{3} + \pi} \left(\frac{M}{M-2} \right) \sum_{i=2}^{M-1} \text{med} (|\Delta Y_{i-1}|, |\Delta Y_i|, |\Delta Y_{i+1}|)^2 \quad (4)$$

Hence, in our empirical analysis we rely on both TV and MedRV, allowing us to conclude that our main results are not sensitive to the particular jump-robust volatility measures that we use to derive volatility measurement equations. By presenting these equations below in generic form, we are able to abstract from the chosen jump-robust estimators that we are going to utilize in the state space formulation of various models for the sake of reducing parameter and volatility estimation uncertainty.

2.2 Generic asymptotic results and volatility measurement equations

Let \widehat{IV}_M be some jump-robust volatility estimator applicable in the considered setting such as TV and MedRV defined above. Then a central limit theorem (CLT) of the following generic form holds:

$$\sqrt{M}(\widehat{IV}_M - IV) \xrightarrow{\mathcal{D}} N \left(0, \nu \int_0^1 \sigma_u^4 du \right), \quad (5)$$

where ν is a known asymptotic variance factor depending on the particular estimator (e.g. 3.06 for TV and 2.96 for MedRV), while $IQ = \int_0^1 \sigma_u^4 du$ is the integrated quarticity controlling the precision of all such estimators. Moreover, since the convergence in (5) is stable, it is possible to apply the delta method to derive feasible asymptotic results based on any consistent jump-robust estimator \widehat{IQ}_M of IQ .¹³ In particular,

$$\sqrt{M} \frac{\widehat{IV}_M - IV}{\nu \widehat{IQ}_M} \xrightarrow{\mathcal{D}} N(0, 1), \quad (6)$$

and

$$\sqrt{M} \frac{\log(\widehat{IV}_M) - \log(IV)}{\nu \frac{\widehat{IQ}_M}{\widehat{IV}_M}} \xrightarrow{\mathcal{D}} N(0, 1). \quad (7)$$

samples.

¹³Without loss of generality, in our empirical analysis we focus on the popular realized quad-power quarticity estimator $QQ_M = \frac{\pi^2 M}{4} \left(\frac{M}{M-3} \right) \sum_{i=1}^{M-3} |\Delta Y_i| |\Delta Y_{i+1}| |\Delta Y_{i+2}| |\Delta Y_{i+3}|$ of Barndorff-Nielsen and Shephard (2004) as well as the slightly more efficient (and robust to both jumps and zero returns) median realized quarticity estimator $\text{MedRQ}_M = \frac{3\pi M}{9\pi + 72 - 52\sqrt{3}} \left(\frac{M}{M-2} \right) \sum_{i=2}^{M-1} \text{med} (|\Delta Y_{i-1}|, |\Delta Y_i|, |\Delta Y_{i+1}|)^4$, of Andersen, Dobrev, and Schaumburg (2009).

The log transformation in (7) results in better finite sample approximation than (6), as already noted by Barndorff-Nielsen and Shephard (2005) and Huang and Tauchen (2005). This is especially useful for our purposes as we will focus our subsequent analysis exactly on logarithmic SV models.

In what follows, we denote the feasible estimate of the asymptotic variance of $\log(\widehat{IV}_M)$ implied by (7) as $\widehat{\Omega}_M = \nu \widehat{IQ}_M / \widehat{IV}_M^2$ to obtain the following logarithmic volatility measurement equation that we are going to utilize in the state space representation of various logarithmic SV models (with leverage effects and jumps) to improve estimation efficiency:

$$\log(\widehat{IV}_M) \approx \log(IV) + \sqrt{\frac{1}{M} \widehat{\Omega}_M} \varepsilon_t, \quad (8)$$

where $\varepsilon_t \sim N(0, 1)$ is independent of the underlying process and the measurement error vanishes as the intraday sampling frequency M increases. More generally, to make explicit distinction between different days, we rewrite this key equation as:

$$\log(\widehat{IV}_{t,t+1;M}) \approx \log(IV_{t,t+1}) + \sqrt{\frac{1}{M} \widehat{\Omega}_{t,t+1;M}} \varepsilon_t, \quad (9)$$

where $\varepsilon_t \sim N(0, 1)$ as above, while $\log(IV_{t,t+1})$, $\log(\widehat{IV}_{t,t+1;M})$, and $\widehat{\Omega}_{t,t+1;M}$ stand, respectively, for the true daily diffusive variance, its available jump-robust estimate at any sample frequency M , and the corresponding asymptotic variance on a given day of unit length represented by the interval $(t, t + 1]$.

We restrict attention to moderate sample frequencies such as two or five minutes (e.g. $M = 195$ or $M = 78$ over a typical trading day of six and a half hours) in order to avoid complications arising from various market microstructure effects that cannot be safely ignored.¹⁴ Alternatively, for jump-robust volatility estimation at higher frequencies one can resort to noise-reduction techniques such as pre-averaging, introduced in the context of multipower variations by Podolskij and Vetter (2009).¹⁵

Quite similarly to the way we obtained equation (9) above, it is possible to single out also the jump part of volatility by using available asymptotic results for the difference between non jump-robust and jump-robust high frequency volatility measures, such as those exploited for moment-based estimation by Todorov (2009). What is important to keep in mind is that any such volatility measurement equations based on high-frequency data similar to (9) do not require knowledge of the exact intraday dynamics of the logarithmic price

¹⁴Volatility measures obtained at higher frequencies can incur biases due to market imperfections such as bid-ask bounce effects, stale quotes, price discreteness, and intraday patterns.

¹⁵The extra robustness comes at the cost of lower convergence rate that can be easily accommodated by our generic volatility measurement equation (9) by changing the power of M accordingly.

process. This observation is crucial for our analysis as it allows us to largely abstract from modeling complications due to non-trivial intraday market microstructure effects. Thus, in the next section we focus entirely on the estimation of popular parametric models for equity returns at daily or lower frequencies by directly bringing our generic daily volatility measurement equations based on high-frequency intraday data to the state space form of each model.

3 Equity return models and estimation

With this extra machinery at hand, our goal is to demonstrate the ease and importance of utilizing high-frequency data for more efficient estimation of a broad range of commonly used equity return models. On one side of the spectrum we consider a basic continuous-time diffusion model similar to the setting of Jacquier, Polson, and Rossi (2004) with log-volatility specification, leverage effect and no jumps. On the other side of the spectrum we also study a two-factor logarithmic SV model with leverage effects and compound Poisson jumps in returns. It offers a less restrictive setting than the two-factor models studied by Alizadeh, Brandt, and Diebold (2002) and Bollerslev and Zhou (2002) thanks to incorporating both leverage effects and jumps. Moreover, like the single-factor model, it can still be successfully fitted using information on daily data only, which we use as a natural benchmark for gauging the attainable efficiency gains from our approach to incorporating high-frequency data. Formally, by relying on Bayesian estimation methods, we are able to fully exploit the information content of high frequency volatility measures within the standard state-space form of the models. Hence, we can obtain a clean measure of the incremental value of high-frequency data compared to estimation based on daily data only.

As shown in Das and Sundaram (1999) among others, models with stochastic volatility, leverage effects and jumps allow for skewness and excess kurtosis of returns and make it possible to closely match stylized facts of empirical asset return distributions that have been extensively studied under both physical and risk-neutral measures. For example, Andersen, Benzoni, and Lund (2002) find that adding jumps in returns to single-factor stochastic volatility models can help better fit stock return skewness and kurtosis and better reproduce volatility smiles in option prices. Eraker, Johannes, and Polson (2003) further extend single-factor jump-diffusion models by adding jumps not only in returns but also in volatility, which Broadie, Chernov, and Johannes (2007) show to be important for fitting volatility skewness and kurtosis. Studies of stochastic volatility models with similar findings under risk-neutral measure include Bakshi, Cao, and Chen (1997), Bates (2000), among others. A unified approach using both returns and options data, pursued by Chernov and Ghysels (2000), Eraker (2004) and Jones (2003), has also stressed the importance of properly fitting the conditional skewness and kurtosis of return distributions at various

return horizons.

From such broader modeling point of view, it is not our goal to use high-frequency data for the sake of improved specification testing as done, for example, by Alizadeh, Brandt, and Diebold (2002), Bollerslev and Zhou (2002), Corradi and Distaso (2006) and Todorov (2009). Instead, we go a step beyond specification testing and attempt to clearly demonstrate first the efficiency and then the economic gains from harnessing the information content of high-frequency volatility measures regardless of the underlying model.

As our workhorse for analysis, in this section we develop appropriate Bayesian estimation methods that allow us to easily incorporate high frequency volatility measurement equations (such as those presented in the previous section) directly in state space form of any model. In this regard, our estimation approach is closest to Barndorff-Nielsen and Shephard (2002), although they do not allow for jumps and use quasi-maximum likelihood rather than Bayesian estimation methods. By following a Bayesian Markov Chain Monte Carlo (MCMC) approach to estimation, we are able to easily take parameter uncertainty into account and demonstrate that high-frequency information helps greatly increase precision in the parameter estimates governing skewness and kurtosis of returns, which in turn leads to considerably more precise and less biased Value-at-Risk forecasts for multi-day returns. Thus, our study contributes directly to the growing body of evidence that high frequency returns are an important source of information in asset pricing and risk management.

Without loss of generality, here we restrict our exposition to one and two-factor models on opposite sides of the spectrum in terms of complexity. We impose a logarithmic specification for the stochastic volatility components in our models directly in line with Andersen, Bollerslev, Christoffersen, and Diebold (2007) who point out that lognormal/normal mixture models show great appeal in financial risk management in view of the empirically observed near lognormality of realized volatility coupled with the near normality of daily returns standardized by realized volatility.

3.1 One-factor log-SV model with leverage effects

We consider a standard one-factor log-SV model that provides a high level of simplicity and transparency, while it is still rich enough to allow for both skewness and excess kurtosis of asset returns. Our contribution consists in extending the equations of the model in state space form with our extra volatility measurement equation derived in generic form in Section 2.2, which is the only difference compared to the standard specification in Jacquier, Polson, and Rossi (2004). It is worth noting that Jones (2003) has studied a similar system of equations with extra measurement equation coming from option implied volatilities. In our model we use high-frequency volatility measures as extra information, which in contrast to implied volatilities allows to theoretically derive the variance of measurement noise and does not require estimation of risk-premia related parameters. In order to facilitate the

exposition, we first present the part of our single-factor model identical with Jacquier, Polson, and Rossi (2004). After standard first order Euler discretization as in Kloeden and Platen (1992), or cast directly as a discrete-time model, the system of equations takes the following form:

$$Y_{t+\Delta} - Y_t = \mu\Delta + \exp\left(\frac{h_t}{2}\right) \sqrt{\Delta} \varepsilon_{t+\Delta}^{(1)} \quad (10)$$

$$h_{t+\Delta} = h_t + \kappa_h(\theta_h - h_t)\Delta + \sigma_h \sqrt{\Delta} (\rho_h \cdot \varepsilon_{t+\Delta}^{(1)} + \sqrt{(1 - \rho_h^2)} \cdot \varepsilon_{t+\Delta}^{(2)}) \quad (11)$$

where $t = 0, \Delta, 2\Delta, \dots, T\Delta$ is a sequence of discrete times, $\{\varepsilon_t^{(j)}\}_{t \geq 0}$, $j = 1, 2$ are sequences of jointly independent i.i.d. $N(0, 1)$ random variables, $\{Y_t\}_{t \geq 0}$ denotes the logarithmic asset price or index level at time t , $\mu \in \mathbb{R}$ is the drift part of the return process, $\kappa_h \in (0, 2)$ defines the speed of mean reversion¹⁶ of the log-volatility process h_t towards its mean $\theta_h \in \mathbb{R}$, $\sigma_h > 0$ defines the volatility of volatility parameter, $\rho_h \in (-1, 1)$ defines the typically negative correlation between returns and volatility increments known as leverage effect, and finally $\Delta > 0$ is a discretization parameter. In this paper we consider dynamics at a daily frequency and fix accordingly $\Delta = 1$.¹⁷

We next consider a version of the model new to the literature, where the discretized system of equations (10)-(11) is augmented by our additional daily volatility measurement equation based on high-frequency data, given by (9) above for this model as:

$$\log(\widehat{IV}_{t,t+\Delta;M}) \approx \alpha_0 + h_t + \sqrt{\frac{1}{M} \widehat{\Omega}_{t,t+\Delta;M}} \varepsilon_{t+\Delta}^{(IV)}, \quad (12)$$

where $\{\varepsilon_t^{(IV)}\}_{t \geq 0}$ is a sequence of i.i.d. $N(0, 1)$ random variables independent of $\{\varepsilon_t^{(j)}\}$ for $j = 1, 2$, while $\{\widehat{IV}_{t,t+\Delta;M}\}_{t \geq 0}$ is some integrated variance measure such as MedRV or TV with measurement error determined by the sampling frequency M and efficiency $\{\widehat{\Omega}_{t,t+\Delta;M}\}_{t \geq 0}$ as described in Section 2.2. Note that both $\{\widehat{IV}_{t,t+\Delta;M}\}_{t \geq 0}$ and $\{\widehat{\Omega}_{t,t+\Delta;M}\}_{t \geq 0}$ are treated as daily observations and are directly calculated as functions of the available high frequency intraday returns at any suitable sample frequency M . As part of the volatility measurement equation (12) we also introduce an optional auxiliary parameter α_0 , which serves the purpose of correcting for the discrepancy between the log integrated variance measures $\log(\widehat{IV}_{t,t+\Delta;M})$ calculated using open-to-close intraday data and the corresponding log-variances of close-to-close daily returns represented by h_t .¹⁸ Note, that such correction is not required, though, if we use open-to-close data for the daily returns, in which case we simply impose $\alpha_0 = 0$. To complete the probabilistic set-up of the one-factor model, we assume that all random variables are constructed on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with a

¹⁶As usual, we restrict κ_h to satisfy standard stationarity conditions for h_t .

¹⁷As noted by Eraker, Johannes, and Polson (2003), the discretization bias for daily data is not significant.

¹⁸This is a standard correction of realized volatility measures as given in more detail, for example, by Hansen and Lunde (2005).

given filtration $\{\mathcal{F}_t\}_{t \geq 0}$ and all processes are adapted to the filtration.

We keep the daily dynamics given by (10) and (11) in the center of our analysis, while (12) serves the sole purpose of incorporating the information content of high-frequency data without incurring modeling complications due to market microstructure effects and other features of intraday data not relevant for modeling of daily returns, as discussed in Section 2.2. Thus, the use of non-parametric high-frequency volatility measures $\{\widehat{IV}_{t,t+\Delta;M}\}_{t \geq 0}$ designed to be robust to known irregularities of intraday data gives us an additional degree of freedom to implicitly allow for high-frequency returns to follow possibly different dynamics from that of daily returns.

In order to find the contribution of high frequency information, we consider the above two versions of the model in state space form: (i) the one with daily returns only; (ii) the one including both daily returns and a daily volatility measurement equation from high frequency intraday data. The former is given by the system of equations (10)-(11), while the latter consists of all equations from the “daily only” model augmented by our additional volatility measurement equation (12).

3.2 Two-factor log-SV model with leverage effects and jumps

Alizadeh, Brandt, and Diebold (2002) and Bollerslev and Zhou (2002) provide strong support in favor of two-factor models of foreign exchange rates by utilizing high frequency data as part of non-Bayesian estimation procedures for specifications without leverage effects. Their first factor mimics the long-memory component in volatility, while the second factor has considerably smaller degree of persistence. Bollerslev and Zhou (2002) further find that even in the presence of a second short-memory stochastic volatility factor, it is still important to include also a jump component in the model. Therefore, we consider a two factor log-SV model with compound Poisson jumps in returns. Moreover, we extend the specification by incorporating leverage effects, which allows us to model also the negative correlation between return and volatility innovations typical for equity returns.

Thus, our two-factor logarithmic stochastic volatility model with Poisson jumps in returns represents a very general setting in the current literature. It still allows, though, successful estimation with the use of only daily data, for the sake of comparison to our approach with an extra volatility measurement equation. Similarly to our one-factor specification above, the discretized version of our two-factor model is given by the following set of equations in state space form, where the probabilistic setup and notation are analogous

to those of our one-factor model:

$$Y_{t+\Delta} - Y_t = \mu\Delta + \exp\left(\frac{h_t + f_t}{2}\right) \sqrt{\Delta} \varepsilon_{t+\Delta}^{(1)} + q_{t+\Delta} \cdot J_{t+\Delta} \quad (13)$$

$$h_{t+\Delta} = h_t + \kappa_h(\theta_h - h_t)\Delta + \sigma_h \sqrt{\Delta} (\rho_h \cdot \varepsilon_{t+\Delta}^{(1)} + \sqrt{(1 - \rho_h^2)} \cdot \varepsilon_{t+\Delta}^{(2)}) \quad (14)$$

$$f_{t+\Delta} = f_t + \kappa_f(\theta_f - f_t)\Delta + \sigma_f \sqrt{\Delta} (\rho_f \cdot \varepsilon_{t+\Delta}^{(1)} + \sqrt{(1 - \rho_f^2)} \cdot \varepsilon_{t+\Delta}^{(3)}) \quad (15)$$

$$\log(\widehat{IV}_{t,t+\Delta;M}) \approx \alpha_0 + h_t + f_t + \sqrt{\frac{1}{M} \widehat{\Omega}_{t,t+\Delta;M}} \varepsilon_{t+\Delta}^{(IV)} \quad (16)$$

We assume without loss of generality that $\kappa_h < \kappa_f$ and denote the persistent and non-persistent volatility factors as h_t and f_t respectively. Other than that, the parameters κ_f and σ_f governing the short-memory factor f_t have similar domain and interpretation as their counterparts κ_h and σ_h for the long-memory factor h_t . We further assume for identification purposes $\theta_f = 0$, since only the total (unconditional) mean log-volatility is identified in the model. Also by construction, $\{\varepsilon_t^{(j)}\}_{t \geq 0}$, $j = 1, 2, 3$ and $\{\varepsilon_t^{(IV)}\}_{t \geq 0}$ are sequences of jointly independent i.i.d. $N(0, 1)$ random variables. Thus, we allow for leverage effects in both factors, which is more explicitly seen by defining the innovations specific to h_t and f_t as:

$$\varepsilon_{t+\Delta}^{(h)} = (\rho_h \cdot \varepsilon_{t+\Delta}^{(1)} + \sqrt{(1 - \rho_h^2)} \cdot \varepsilon_{t+\Delta}^{(2)}) \quad (17)$$

$$\varepsilon_{t+\Delta}^{(f)} = (\rho_f \cdot \varepsilon_{t+\Delta}^{(1)} + \sqrt{(1 - \rho_f^2)} \cdot \varepsilon_{t+\Delta}^{(3)}) \quad (18)$$

In particular, the instantaneous covariance matrix between return and volatility innovations is given by:

$$\Sigma_{t+1|t} \equiv E\left(\begin{pmatrix} \varepsilon_{t+1}^{(1)} \\ \varepsilon_{t+1}^{(h)} \\ \varepsilon_{t+1}^{(f)} \end{pmatrix} \begin{pmatrix} \varepsilon_{t+1}^{(1)} \\ \varepsilon_{t+1}^{(h)} \\ \varepsilon_{t+1}^{(f)} \end{pmatrix}'\right) = \begin{pmatrix} 1 & \rho_h & \rho_f \\ \rho_h & 1 & 0 \\ \rho_f & 0 & 1 \end{pmatrix},$$

where we impose the positive definite restriction $1 - \rho_h^2 - \rho_f^2 > 0$.

Our compound Poisson jump specification with normally distributed jump sizes draws on Andersen, Benzoni, and Lund (2002), Eraker, Johannes, and Polson (2003), and Johannes and Polson (2002). In particular, we assume a maximum of one jump per day. The jump increments in the interval $(t, t+\Delta]$ follow the law of $q_{t+\Delta} \cdot J_{t+\Delta}$, where the jump times $\{q_t\}_{t \geq 0}$ are i.i.d. $Bernoulli(\lambda)$ and the jump sizes $\{J_t\}_{t \geq 0}$ are i.i.d. $N(\mu_J, \sigma_J^2)$. The parameters $\lambda > 0$, $\mu_J \in \mathbb{R}$ and $\sigma_J > 0$ denote respectively the jump intensity, mean and standard deviation of jump sizes. Since at a daily frequency the jump intensity parameter λ is close to zero, our assumption of maximum one jump per day is not binding.

Most importantly, we extend the state-space form of the model with our volatility measurement equation (16), which is a direct counterpart to equation (12) in the one-factor model and specializes equation (9) given in general form in section 2.2. Here the high fre-

quency measure of log integrated variance $\log(\widehat{IV}_{t,t+\Delta;M})$ is an estimate of $h_t + f_t$ as the total diffusive variance in the two-factor model. The extra parameter α_0 serves the same purpose as in the one-factor model. It provides standard correction for the discrepancy between log integrated variance measures $\log(\widehat{IV}_{t,t+\Delta;M})$ calculated using open-to-close intraday data and the log variance of close-to-close daily returns modeled by $h_t + f_t$. For modeling the log variance of open-to-close daily returns we simply restrict $\alpha_0 = 0$.

In order to find the contribution of high frequency information, similarly to our one-factor model, we consider two versions of the two factor model: (i) the one with only daily returns; (ii) the one including both daily returns and a daily volatility measurement equation from high frequency intraday data. The former is given by the system of equations (13)-(15), while the latter consists of all equations from the “daily” model augmented by our additional volatility measurement equation (16).

3.3 Estimation

3.3.1 Markov chain Monte Carlo methods

We first briefly describe the general principles of Markov chain Monte Carlo (MCMC) methods, with more detailed exposition in Chib and Greenberg (1996), Johannes and Polson (2002) and Jones (1998). Let Y denote the vector of observations, X be the vector of latent state variables and Θ be the vector of model parameters. In Bayesian inference we utilize the prior information on the parameters to derive the joint posterior distribution for both parameters and state variables. By the Bayes rule, we have:

$$p(\Theta, X|Y) \propto p(Y|X, \Theta) \cdot p(X|\Theta) \cdot p(\Theta) ,$$

where $p(Y|X, \Theta)$ is the likelihood function of the model, $p(X|\Theta)$ is the probability distribution of state variables conditional on the parameters and $p(\Theta)$ is the prior probability distribution on the parameters of the model. Ideally we would like to know the analytical properties of the joint posterior distribution of X and Θ , however, this is hardly feasible. The highly multidimensional joint posterior distribution is very often too complicated to work with and analytically intractable and hence even direct simulation from the joint posterior distribution is hard to perform.

In the sequel we base our exposition on Jones (2003). The idea behind MCMC methods is to break the highly dimensional vectors of latent variables X and parameters Θ into smaller pieces. The Gibbs sampler developed in Geman and Geman (1984) considers partitioning of X and Θ into respectively I^X and I^Θ subvectors $X^{(1)}, X^{(2)}, \dots, X^{(I^X)}$ and $\Theta^{(1)}, \Theta^{(2)}, \dots, \Theta^{(I^\Theta)}$. Then the Markov chain is constructed by first defining starting values

of the chain X_0 and Θ_0 and then iteratively forming the chain

$$(X_n, \Theta_n) = (X_n^{(1)}, X_n^{(2)}, \dots, X_n^{(I^X)}, \Theta_n^{(1)}, \Theta_n^{(2)}, \dots, \Theta_n^{(I^\Theta)})$$

The draws of (X_n, Θ_n) are performed for each $i = 1, \dots, I^X$ and each $j = 1, \dots, I^\Theta$ by drawing from the following transition densities:

$$p(X_n^{(i)} | X_n^{(-i)}, \Theta_{n-1}, Y), i = 1, 2, \dots, I^X \quad (19)$$

$$p(\Theta_n^{(j)} | \Theta_n^{(-j)}, X_n, Y), j = 1, 2, \dots, I^\Theta \quad (20)$$

where $X_n^{(-i)} \equiv (X_n^{(k)}; k < i) \cup (X_{n-1}^{(k)}; k > i)$ and $\Theta_n^{(-j)} \equiv (\Theta_n^{(k)}; k < j) \cup (\Theta_{n-1}^{(k)}; k > j)$ It can be shown that under mild conditions the chain (X_n, Θ_n) converges to its invariant distribution $p(\Theta, X|Y)$ that is by construction a joint posterior distribution of the model under consideration. The proof of the Gibbs sampler convergence to invariant distribution, sufficient conditions and some applications can be found in Chib and Greenberg (1996).

The Gibbs sampler algorithm provides a tractable method to draw from multidimensional and complicated distributions only if one can draw from all complete conditional distributions in equations (19) and (20). However, even one-dimensional complete conditional distributions can be in practice difficult if not impossible to draw from. In this case we replace a particular Gibbs sampler step by the Metropolis-Hastings (MH) step in Metropolis, Rosenbluth and Rosenbluth (1953). Chib and Greenberg (1996) provide further details about the MH algorithm. The main building block of our estimation method is based on the Gibbs sampler algorithm with some blocks replaced by MH steps.

After discarding a “burn-in” period of the first N draws, the discrete approximation $\{(X_n, \Theta_n)\}_{n>N}$ of the joint posterior density $p(\Theta, X|Y)$ allows one to compute various statistics. For example, the sample mean of the posterior distributions can be taken to obtain parameter estimates for our models. Likewise, one can estimate statistics of particular interest in applications such as moment and quantile forecasts for multi-horizon returns as well as associated risk measures such as Value-at-Risk (VaR) or any other function of the conditional multi-horizon return density such as the price of a derivative contract. Moreover, parameter uncertainty is taken automatically into account by integrating over the entire joint posterior distribution of parameters and state variables. This important property of MCMC estimation methods is especially valuable for our purposes, as it allows us to show how increasing the precision of parameter and volatility state estimation (by including our volatility measurement equations (12) and (16)) gets translated into more accurate conditional return density forecasts and moments/quantiles in particular.

3.3.2 Bayesian MCMC inference for models with high frequency volatility measurement equations

We limit our exposition to describing our MCMC estimation procedure for the two-factor stochastic volatility models from Section (3.2).¹⁹ We put special emphasis on how to estimate models including our high-frequency measurement equations by offering a straightforward extension of estimation methods based only on daily returns.

Following the notation from the previous section, we need to specify the vector of observations Y , the vector of latent state variables X and the vector of parameters Θ along with their appropriate subdivision in line with the construction of the Gibbs sampler algorithm. In particular, we define the following vectors, where “Daily” stays for estimation based only on daily returns (equations (13)-(15)) and “HF” stays for estimation incorporating also volatility measures based on high-frequency intraday data (equations (13)-(16)):

$$\begin{aligned}
 Y^{(Daily)} &= \{\{Y_t\}_{t=1,\dots,T}\} \\
 Y^{(HF)} &= \{\{Y_t\}_{t=1,\dots,T}, \{\widehat{IV}_{t,t+1;M}\}_{t=1,\dots,T-1}, \{\widehat{\Omega}_{t,t+1;M}\}_{t=1,\dots,T-1}\} \\
 X &= \{\{h_t\}_{t=1,\dots,T}, \{f_t\}_{t=1,\dots,T}, \{q_t\}_{t=2,\dots,T}, \{J_t\}_{t=2,\dots,T}\} \\
 \Theta &= \{\mu, \kappa_h, \theta_h, (\sigma_h, \rho_h), \kappa_f, (\sigma_f, \rho_f), \lambda, \mu_J, \sigma_J, \alpha_0\} .
 \end{aligned}$$

The partitions of Θ and X are given by $\Theta^{(1)} = \mu$, $\Theta^{(2)} = \kappa_h$, $\Theta^{(3)} = \theta_h$, $\Theta^{(4)} = (\sigma_h, \rho_h)$, $\Theta^{(5)} = \kappa_f$, $\Theta^{(6)} = (\sigma_f, \rho_f)$, $\Theta^{(7)} = \lambda$, $\Theta^{(8)} = \mu_J$, $\Theta^{(9)} = \sigma_J$ and $X^{(i)} = h_i$, $X^{(i+T)} = f_i$, $X^{(j+2T)} = q_{j+1}$, $X^{(j+(3T-1))} = J_{j+1}$ where $i = 1, 2, \dots, T$, $j = 1, \dots, T - 1$. Thus, we treat each element of the state vector X as a single block. For the vector of parameters Θ all elements are treated as a single block with the exception of (σ_h, ρ_h) and (σ_f, ρ_f) . These parameters are drawn jointly as in Jacquier, Polson, and Rossi (2004). Finally, the extra parameter $\Theta^{(10)} = \alpha_0$ in equation (16) appears only in the “HF” model including high frequency information and is estimated along with the rest of the parameters or it can be exogenously specified following standard approaches in the realized volatility literature to obtain variances for the whole day such as Hansen and Lunde (2005). It is set to zero when modeling open-to-close daily returns.

Having defined above all blocks for the latent state variables X and parameters Θ , we apply the MCMC algorithm based on the Gibbs sampler presented in Section 3.3.1. Since draws of all parameters and jump related latent variables are standard in the literature, we directly refer to Szerszen (2009) for the imposed prior distributions on the model parameters Θ and all other details.

Here we focus on addressing the fundamental difference between estimation of the stan-

¹⁹One-factor models can be viewed as a special case by restricting $f_t = 0$ for all t , omitting the parameters $\kappa_f, \theta_f, \sigma_f, \rho_f$ for the f factor and imposing the constraint $\rho_f = 0$ in the instantaneous correlation matrix.

dard “Daily” and our “HF” version of the model, which differ just by the additional volatility measurement equation (16) based on high-frequency data. The information provided by this extra equation affects only the complete conditional posteriors of the volatility states h_t and f_t . In particular, the MCMC update for h_t is given by

$$p(h_t|\{f_t\}, h_{t+1}, h_{t-1}, q, J, \Theta, Y) \propto p(Y_{t+1}|Y_t, \{f_t\}, \{h_t\}, \Theta, q, J) \cdot p(Y_t|Y_{t-1}, \{f_t\}, \{h_t\}, \Theta, q, J) \\ \cdot p(h_{t+1}|h_t, \Theta) \cdot p(h_t|h_{t-1}, \Theta) \cdot p(h_t|\widehat{IV}_{t,t+1;M}, \widehat{\Omega}_{t,t+1;M}, f_t, \Theta)$$

for $t=1, 2, \dots, T$, where the second and fourth kernels on the right hand side are omitted for $t=1$, while the first, third and last kernels are omitted for $t=T$. The MCMC update for the second factor f_t is performed analogously.

Thus, an inspection of the above update expression reveals that the only kernel affected by the high frequency information with $Y = Y^{(HF)}$ is the last one $p(h_t|\widehat{IV}_{t,t+1;M}, \widehat{\Omega}_{t,t+1;M}, f_t)$ for the h factor and, similarly, $p(f_t|\widehat{IV}_{t,t+1;M}, \widehat{\Omega}_{t,t+1;M}, h_t)$ for the f factor. The rest of the kernels are exactly those coming from inference based on daily returns only, i.e. with $Y = Y^{(Daily)}$, which appear also with $Y = Y^{(HF)}$. This is of key importance for understanding how the extra information provided by high-frequency data improves estimation efficiency in our “HF” versus “Daily” approaches. The extra kernels $p(h_t|\widehat{IV}_{t,t+1;M}, \widehat{\Omega}_{t,t+1;M}, f_t)$ and $p(f_t|\widehat{IV}_{t,t+1;M}, \widehat{\Omega}_{t,t+1;M}, h_t)$ in the MCMC updates of h and f , respectively, are very spiked around the mode for dates with low values of $\frac{1}{M}\widehat{\Omega}_{t,t+1;M}$ in the volatility measurement equation (16) and, hence, they are very informative about the latent volatility states. The attainable precision improvements increase with the sample frequency M and depend also on $\widehat{\Omega}_{t,t+1;M}$, being a function of the underlying volatility paths and the chosen high-frequency integrated variance and quarticity measures as detailed in Section 2.2. By contrast, the use of only daily data is equivalent to artificially setting $\frac{1}{M}\widehat{\Omega}_{t,t+1;M}$ to infinity in order to suppress the strong information content of high frequency data provided by our volatility measurement equation. In what follows, we analyze the gains in estimation efficiency and risk forecasting accuracy from our “HF” versus traditional “Daily” estimation as a natural benchmark for comparison.

4 Estimation efficiency and risk forecasting accuracy

The ability to estimate parameters and volatility states more efficiently directly translates into more accurate risk forecasts. Moreover, the highly non-linear nature of the underlying transformation from noisy parameter and volatility estimates to risk forecasts implies reduction not only in the variance but also in the bias of the prediction errors. Our analysis in this section is designed to study the interplay between longer sample size and higher intraday frequency as an additional source of information introduced by our volatility measurement equation for the purpose of reducing estimation uncertainty. We document that

even for the longest sample lengths encountered in practice there is a substantial efficiency gain from incorporating the extra information provided by high-frequency volatility measures. Moreover, for key model parameters controlling skewness and kurtosis we find that two or five years of high frequency data would suffice to obtain the same level of precision as twenty years of daily data. This suggests that our approach can be particularly useful in finance applications where only short data samples are available or economically meaningful to use.

It is possible to derive analytical results along these lines in certain more restrictive settings. An instructive example for a canonical log-SV model is given in the appendix. Monte Carlo analysis is the only viable option, though, for models that are not analytically tractable. Hence, we take a Monte Carlo approach to study estimation efficiency and the impact of parameter uncertainty on risk forecasting accuracy. We conduct considerably more thorough and extensive simulations than usual in order to properly document the substantial efficiency gains and improved precision of risk forecasts at horizons of up to a few months ahead regardless of the chosen model when high frequency information is included in the model. Perhaps the most important of our findings is that there is considerable asymmetry between bad and good times when it comes to the attainable improvements in risk forecasting accuracy: in good times we are able to largely eliminate overstatement of risk, while in bad times our approach helps avoid understatement of risk. From a practical point of view, this implies imposing an appropriate larger risk cushion exactly when needed the most, e.g. early on in times of crisis (rather than with a delay), while at the same time avoiding excessive risk cushion requirements in normal times. In this sense, our main purpose in what follows is to document both the efficiency gains for model estimation and forecasting and the implied potentially large economic value of our approach to incorporating the information content of high-frequency volatility measures for model estimation and risk forecasting.

4.1 Monte Carlo setup

In order to set-up the stage for Monte Carlo analysis we first describe how to draw sample paths consistent with the data generating process implied by our model specifications. Daily dynamics of both returns and volatility are based on equations (10)-(11) and (13)-(15) respectively for the one-factor and two-factor log-SV models that we consider. The intraday dynamics is based on a Brownian bridge connecting consecutive daily sample points and producing valid integrated variance measures $\{\widehat{IV}_{t,t+1;M}\}_{t \geq 0}$ and corresponding scaled integrated quarticity measures $\{\widehat{\Omega}_{t,t+1;M}\}_{t \geq 0}$ that govern our additional volatility measurement equations in (12) and (16) as described in Section 2.2. In this way, we allow for potentially richer intraday dynamics than the one at the daily frequency, possibly including also realistic intraday market-microstructure effects that many novel high-frequency volatility

measures are designed to be robust to when sampled at two to five minute frequency.²⁰

We draw 1,000 sample paths for each of the considered one- and two-factor log-SV models. For each sample path we estimate the underlying model parameters using different information sets: (i) daily data only; (ii) daily data with additional high frequency volatility measurements based on 5-minute or 2-minute intraday returns; (iii) the “infeasible” case of perfectly observed volatility.²¹ In order to study the interplay between additional information coming from more high frequency data and longer sample size in terms of number of days, we consider three sample windows of 2, 5 and 20 years. This gives a total of twelve one-factor and twelve two-factor specifications for the information sets used for model estimation. We estimate all specifications using the Bayesian MCMC methods described in Section 3 with 250,000 draws, where the first 50,000 draws are discarded as the burn-in sample. For the purposes of forecasting conditional return moments and quantiles, based on the obtained 200,000 draws of the posterior distribution of parameters and volatility states, we approximate multi-period conditional density forecasts by a cloud of 25,000,000 points. We then compare moments and quantiles of the obtained conditional density forecasts for the two different estimation procedures that we consider, depending on whether a daily volatility measurement equation based on high-frequency data is used or not.

4.2 Efficiency gains in parameter and volatility estimation

In Tables 1 and 2 we report parameter estimates, bias and root mean squared error (RMSE) of volatility related parameters governing equations (11) and (14)-(15) for our one-factor and two-factor specifications respectively. The true parameter values in each table represent our estimates on S&P 500 daily futures returns for the period October 2, 1985 - February 26, 2009.

For the one-factor model (Table 1) we attain up to few times better precision when using high frequency data compared to only daily data for estimating the parameters governing skewness and kurtosis. This translates into RMSE reduction of as much as 70%. In particular, we find that the information content of high-frequency volatility measures improves the most the estimation efficiency of the volatility of volatility parameter σ_h and the leverage effect parameter ρ_h in the model.²² Moreover, the gains are consistent across different sample lengths, even for the longest ones typically encountered in practice such as

²⁰In particular, in our analysis we focus on the MedRV estimator of Andersen, Dobrev, and Schaumburg (2009) and the tri-power variation measure of Barndorff-Nielsen, Shephard, and Winkel (2006). We report results only for the former as the results obtained for the latter are in the same spirit.

²¹In the one-factor model the case of perfectly observed volatility can be viewed as the limiting case of our volatility measurement equation when the intraday sample frequency grows to infinity. In the two-factor model, though, the volatility measurement equation provides information only about the sum of the two volatility factors without separating them as in the infeasible case of full observability.

²²In this sense, our results extend those obtained by Alizadeh, Brandt, and Diebold (2002) in a considerably more restrictive range-based analysis of a model without leverage effects.

20 years, when daily estimation is more likely to produce satisfactory results. As a practical rule of thumb we find that two years of high frequency data often suffice to obtain the same level of precision for these parameters as twenty years of daily data. At the same time, a comparison between the attainable improvement by switching from daily to 5-minute estimation and any further increase in the intraday sample frequency from 5 to 2 minutes and beyond (up to the infeasible case of perfectly known volatility) reveals a rapid decrease in the additional efficiency gains that can be obtained. We also observe a substantial RMSE reduction for the parameter governing persistence of volatility κ_h for the shortest sample sizes, while still dominating the estimation efficiency with only daily data across all sample sizes.

For the richer two-factor log-SV model (Table 2) these substantial efficiency gains from incorporating high-frequency volatility measures naturally get even larger. Moreover, a somewhat larger part of the gains is due to bias reduction. It is important to note that here skewness and kurtosis are driven not only by a persistent volatility factor but also by a second non-persistent factor. For the non-persistent factor we find that the gains from incorporating high frequency information are more pronounced than those from increasing the yearly sample length. We do not find such evidence for the persistent factor, where both sources of information play an important role in parameter estimation. This implies bigger efficiency gains from incorporating high frequency information for the parameters ρ_f and σ_f governing skewness and kurtosis arising from the non-persistent factor f . The reduction of parameter uncertainty for the persistent factor h is somewhat smaller but still very visible.

The quality of risk forecasts depends not only on the degree of parameter uncertainty but also on the degree of volatility estimation uncertainty. In particular, it is important to assess the impact of incorporating additional high frequency information on the accuracy of estimation of terminal volatility states as they play important role in forecasting risk. In Table 3 we report mean estimates, bias and RMSE for the terminal volatility states h_T and f_T of the two-factor log-SV model. Thus, we conclude that our volatility measurement equation helps in estimating better not only model parameters but also latent volatility states. Considerable efficiency gains are obtained mainly for the persistent volatility factor, while for the non-persistent factor we still observe slight improvements. Similarly to parameter estimates, our findings for volatility states are consistent across all considered sample sizes. Moreover, the biggest efficiency gains take place when moving from estimation based only on daily data to estimation incorporating our volatility measurement equation based on 5-minute returns. Further increase of the intraday sample frequency from 5-minutes to 2-minutes leads to additional efficiency gains of much smaller magnitude. Overall, for the estimation of volatility states adding high frequency information has somewhat bigger importance than increasing the yearly sample length. This plays a major role especially for

short-term risk forecasting.

4.3 Precision improvements in risk forecasting accuracy

The documented substantial decrease in parameter and volatility estimation uncertainty implies non-trivial improvements in the accuracy of forecasts of conditional return moments and quantiles. We compare forecasts resulting from inference based on daily data to those utilizing 5-minute high frequency volatility measures. We restrict attention to the 5-minute frequency in accordance with our finding that it offers essentially the bulk of the attainable improvements based on our volatility measurement equation. We perform our analysis incorporating parameter and volatility estimation uncertainty for all three considered sample lengths of 2, 5 and 20 years.

In Tables 4 and 5 we report forecasts of conditional return moments respectively for one-factor and two-factor models. In Tables 6 and 7 we also report forecasts of conditional return quantiles. The considered forecast horizons are 1, 5, 10 and 20 days ahead and are presented in separate panels in each table. These forecast horizons are of primary interest in many finance applications.

Our main finding is that our more efficient parameter and state estimates incorporating the strong information content of high-frequency volatility measures translate into equally better conditional return density forecasts not only in terms of RMSE but also in terms of bias. The bias reduction is due to the pronounced non-linearities in the underlying transformation of parameters and state variables. The main message from our analysis summarized in Tables 4-7 is that for any model, any estimation sample length, and across all forecast horizons of interest, the forecasts incorporating the extra information from our volatility measurement equation clearly dominate those based only on daily data. Moreover, these results strengthen our rule of thumb that model specifications estimated with two years of high frequency data perform at least as good as the same model specifications estimated with twenty years of daily data, which in turn are considerably outperformed if estimated on twenty years of high-frequency data.

4.4 Forecast error reductions in good versus bad times

From risk management perspective it is important to know how the improvements in risk forecasting accuracy vary across good and bad times. To this end, in Table 8 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. The reported relative errors are 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. 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 underestimation 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

²³The 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 Financial 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:

$$r_{t+1} = \exp\left(\frac{h_t}{2}\right) \varepsilon_{t+1}^{(r)} \quad (21)$$

$$h_{t+1} = \beta_h h_t + \sigma_h \varepsilon_{t+1}^{(h)} \quad (22)$$

$$\log(\widehat{IV}_{t+1;M}) = h_t + \sqrt{\frac{\nu}{M}} \varepsilon_{t+1}^{(IV)} \quad (23)$$

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 $M \gg 1$ 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):

$$\log(r_{t+1}^2) = h_t + \log(\varepsilon_{t+1}^{(r)})^2 \quad (24)$$

$$h_{t+1} = \beta_h h_t + \sigma_h \varepsilon_{t+1}^{(h)} \quad (25)$$

$$\log(\widehat{IV}_{t+1;M}) = h_t + \sqrt{\frac{\nu}{M}} \varepsilon_{t+1}^{(IV)} \quad (26)$$

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:

$$x_{t+1} = h_t + \varepsilon_{t+1}^{(x)} \quad (27)$$

$$h_{t+1} = \beta_h h_t + \sigma_h \varepsilon_{t+1}^{(h)} \quad (28)$$

$$y_{t+1} = h_t + \frac{1}{\sqrt{M}} \varepsilon_{t+1}^{(y)} \quad (29)$$

²⁴Comprehensive 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 $M \rightarrow \infty$ 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:

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

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

It is easy to confirm that these are valid moments:

$$\mathbb{E}[g_1] = 0 \quad (32)$$

$$\mathbb{E}[g_2] = 0 \quad (33)$$

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

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

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

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

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:

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

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

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

The resulting optimal GMM weighting matrix is given by:

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

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:

$$\begin{aligned} \frac{\mathbb{V}[g_1]}{\left(\frac{\mathbb{V}[g_1]+\mathbb{V}[g_2]-2\mathbb{C}[g_1,g_2]}{\mathbb{V}[g_1]\mathbb{V}[g_2]-\mathbb{C}[g_1,g_2]^2}\right)^{-1}} &= 1 + \frac{\left(\frac{\mathbb{V}[g_1]}{\mathbb{C}[g_1,g_2]} - 1\right)^2}{\frac{\mathbb{V}[g_1]}{\mathbb{C}[g_1,g_2]} \frac{\mathbb{V}[g_2]}{\mathbb{C}[g_1,g_2]} - 1} \\ &= 1 + \frac{2\frac{m_2^{\varepsilon(x)}}{m_2^h} + \frac{1}{2}\left[\frac{m_4^{\varepsilon(x)}}{(m_2^h)^2} - \left(\frac{m_2^{\varepsilon(x)}}{m_2^h}\right)^2\right]}{1 + \left(\frac{2}{M}\frac{m_2^{\varepsilon(y)}}{m_2^h} + \frac{1}{2M^2}\left[\frac{m_4^{\varepsilon(y)}}{(m_2^h)^2} - \left(\frac{m_2^{\varepsilon(y)}}{m_2^h}\right)^2\right]\right) \left(1 + 2\frac{m_2^{\varepsilon(x)}}{m_2^h} + \frac{1}{2}\left[\frac{m_4^{\varepsilon(x)}}{(m_2^h)^2} - \left(\frac{m_2^{\varepsilon(x)}}{m_2^h}\right)^2\right]\right)} \end{aligned} \quad (41)$$

$$(42)$$

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 $M \rightarrow \infty$ this variance reduction factor approaches

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

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 $M \rightarrow \infty$ 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 Bayesian 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:

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

Columns represent results for alternative estimation procedures depending on whether 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 | | $\kappa_h = 0.0163$ | | | | $\sigma_h = 0.1648$ | | | |
|---------------|----|---------------------|-------------|-------------|---------------------|---------------------|-------------|-------------|---------------------|
| 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 | | $\theta_h = -9.4243$ | | | | $\rho_h = -0.6716$ | | | |
|---------------|----|----------------------|-------------|-------------|---------------------|--------------------|-------------|-------------|---------------------|
| 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:

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

Columns represent results for alternative estimation procedures depending on whether 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 | | $\kappa_h = 0.0130$ | | | | $\kappa_f = 0.6724$ | | | |
|---------------|-----------|---------------------|----------|----------|------------------|---------------------|----------|----------|------------------|
| | | Daily Only | HF 5-min | HF 2-min | Known Volatility | Daily Only | HF 5-min | HF 2-min | Known Volatility |
| MEAN | T (years) | | | | | | | | |
| | 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 |
| BIAS | 20 | 0.0150 | 0.0136 | 0.0136 | 0.0134 | 0.7303 | 0.6810 | 0.6742 | 0.6721 |
| | 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 |
| RMSE | 20 | 0.0020 | 0.0006 | 0.0006 | 0.0004 | 0.0579 | 0.0086 | 0.0018 | -0.0003 |
| | 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 | | $\sigma_h = 0.1320$ | | | | $\sigma_f = 0.3876$ | | | |
|---------------|-----------|---------------------|----------|----------|------------------|---------------------|----------|----------|------------------|
| | | Daily Only | HF 5-min | HF 2-min | Known Volatility | Daily Only | HF 5-min | HF 2-min | Known Volatility |
| MEAN | T (years) | | | | | | | | |
| | 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 |
| BIAS | 20 | 0.1400 | 0.1333 | 0.1329 | 0.1320 | 0.2643 | 0.3919 | 0.3879 | 0.3872 |
| | 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 |
| RMSE | 20 | 0.0080 | 0.0013 | 0.0009 | 0.0000 | -0.1233 | 0.0043 | 0.0003 | -0.0004 |
| | 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 | | $\rho_h = -0.2831$ | | | | $\rho_f = -0.1109$ | | | |
|---------------|-----------|--------------------|----------|----------|------------------|--------------------|----------|----------|------------------|
| | | Daily Only | HF 5-min | HF 2-min | Known Volatility | Daily Only | HF 5-min | HF 2-min | Known Volatility |
| MEAN | T (years) | | | | | | | | |
| | 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 |
| BIAS | 20 | -0.2669 | -0.2849 | -0.2824 | -0.2824 | -0.1712 | -0.1104 | -0.1112 | -0.1098 |
| | 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 |
| RMSE | 20 | 0.0162 | -0.0018 | 0.0007 | 0.0007 | -0.0603 | 0.0005 | -0.0003 | 0.0011 |
| | 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:

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

Columns represent results for alternative estimation procedures depending on whether 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 | | | | | | | | | | | | |
|-------------------------------|------------------|--------|--------|--------------------|---------|--------|----------------------|--------|--------|----------------------|--------|--------|
| T (years) | Mean Forecast | | | StdDev Forecast | | | Skewness Forecast | | | Kurtosis Forecast | | |
| | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF |
| | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min |
| MEAN | 2 | 0.0001 | 0.0001 | 0.0089 | 0.0085 | 0.0004 | 0.0001 | 0.0004 | 0.0001 | 3.6083 | 3.0767 | 3.1203 |
| | 5 | 0.0001 | 0.0001 | 0.0090 | 0.0087 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 3.5487 | 3.0757 | 3.1185 |
| | 20 | 0.0001 | 0.0001 | 0.0088 | 0.0087 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 3.5070 | 3.0752 | 3.1175 |
| BIAS | 2 | 0.0000 | 0.0000 | 0.0004 | 0.0000 | 0.0003 | 0.0001 | 0.0003 | 0.0001 | 0.6141 | 0.0826 | 0.0411 |
| | 5 | 0.0000 | 0.0000 | 0.0003 | 0.0000 | 0.0000 | 0.0001 | 0.0000 | 0.0001 | 0.5546 | 0.0815 | 0.0396 |
| | 20 | 0.0000 | 0.0000 | 0.0001 | -0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.5129 | 0.0811 | 0.0386 |
| RMSE | 2 | 0.0003 | 0.0003 | 0.0019 | 0.0008 | 0.0023 | 0.0008 | 0.0023 | 0.0008 | 0.6347 | 0.0831 | 0.0441 |
| | 5 | 0.0002 | 0.0002 | 0.0018 | 0.0008 | 0.0011 | 0.0007 | 0.0011 | 0.0007 | 0.5671 | 0.0818 | 0.0408 |
| | 20 | 0.0001 | 0.0001 | 0.0018 | 0.0008 | 0.0009 | 0.0007 | 0.0009 | 0.0007 | 0.5214 | 0.0812 | 0.0390 |

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

| Panel C: 10-day horizon | | | | | | | | | | | | |
|--------------------------------|------------------|---------|---------|--------------------|---------|---------|----------------------|--------|---------|----------------------|--------|--------|
| T (years) | Mean Forecast | | | StdDev Forecast | | | Skewness Forecast | | | Kurtosis Forecast | | |
| | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF |
| | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min |
| MEAN | 2 | 0.0017 | 0.0015 | 0.0327 | 0.0313 | -0.0953 | -0.0929 | 0.0327 | 0.0313 | 3.3719 | 3.1644 | 3.1644 |
| | 5 | 0.0017 | 0.0016 | 0.0328 | 0.0318 | -0.0957 | -0.0926 | 0.0328 | 0.0318 | 3.3252 | 3.1614 | 3.1614 |
| | 20 | 0.0018 | 0.0018 | 0.0322 | 0.0317 | -0.0959 | -0.0929 | 0.0322 | 0.0317 | 3.2909 | 3.1598 | 3.1598 |
| BIAS | 2 | -0.0002 | -0.0004 | 0.0016 | 0.0002 | -0.0032 | -0.0008 | 0.0016 | 0.0002 | 0.2449 | 0.0374 | 0.0374 |
| | 5 | -0.0001 | -0.0003 | 0.0012 | 0.0002 | -0.0038 | -0.0007 | 0.0012 | 0.0002 | 0.1966 | 0.0349 | 0.0349 |
| | 20 | 0.0000 | 0.0000 | 0.0003 | -0.0002 | -0.0041 | -0.0011 | 0.0003 | -0.0002 | 0.1643 | 0.0332 | 0.0332 |
| RMSE | 2 | 0.0045 | 0.0042 | 0.0067 | 0.0029 | 0.0453 | 0.0141 | 0.0067 | 0.0029 | 0.2717 | 0.0457 | 0.0457 |
| | 5 | 0.0027 | 0.0026 | 0.0063 | 0.0027 | 0.0297 | 0.0093 | 0.0063 | 0.0027 | 0.2092 | 0.0384 | 0.0384 |
| | 20 | 0.0013 | 0.0013 | 0.0062 | 0.0027 | 0.0154 | 0.0047 | 0.0062 | 0.0027 | 0.1708 | 0.0343 | 0.0343 |

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

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

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

| Panel C: 10-day horizon | | | | | | | | | | | | |
|--------------------------------|---------------------------|---------|---------|---------------------------|---------|--------|---------------------------|--------|---------|---------------------------|---------|---------|
| T (years) | Quantile 0.01 Forecast | | | Quantile 0.05 Forecast | | | Quantile 0.95 Forecast | | | Quantile 0.99 Forecast | | |
| | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF |
| | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min |
| MEAN | 2 | -0.0785 | -0.0745 | -0.0525 | -0.0507 | 0.0544 | 0.0522 | 0.0777 | 0.0733 | 0.0777 | 0.0733 | 0.0733 |
| | 5 | -0.0787 | -0.0754 | -0.0528 | -0.0514 | 0.0547 | 0.0530 | 0.0779 | 0.0745 | 0.0779 | 0.0745 | 0.0745 |
| | 20 | -0.0768 | -0.0748 | -0.0516 | -0.0509 | 0.0538 | 0.0530 | 0.0763 | 0.0744 | 0.0763 | 0.0744 | 0.0744 |
| BIAS | 2 | -0.0053 | -0.0012 | -0.0026 | -0.0007 | 0.0022 | 0.0000 | 0.0048 | 0.0004 | 0.0048 | 0.0004 | 0.0004 |
| | 5 | -0.0041 | -0.0008 | -0.0019 | -0.0005 | 0.0016 | 0.0000 | 0.0037 | 0.0003 | 0.0037 | 0.0003 | 0.0003 |
| | 20 | -0.0017 | 0.0002 | -0.0004 | 0.0003 | 0.0004 | -0.0004 | 0.0017 | -0.0003 | 0.0017 | -0.0003 | -0.0003 |
| RMSE | 2 | 0.0179 | 0.0085 | 0.0124 | 0.0066 | 0.0113 | 0.0059 | 0.0162 | 0.0075 | 0.0162 | 0.0075 | 0.0075 |
| | 5 | 0.0160 | 0.0072 | 0.0109 | 0.0053 | 0.0106 | 0.0050 | 0.0151 | 0.0066 | 0.0151 | 0.0066 | 0.0066 |
| | 20 | 0.0151 | 0.0066 | 0.0104 | 0.0047 | 0.0100 | 0.0045 | 0.0142 | 0.0062 | 0.0142 | 0.0062 | 0.0062 |

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

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 | | | | | | | | | | | | | | | | | |
|-------------------------------|---------------------------|---------|---------|---------------------------|---------|--------|---------------------------|--------|--------|---------------------------|---------|---------|---------|--------|---------|--------|--------|
| T (years) | Quantile 0.01 Forecast | | | Quantile 0.05 Forecast | | | Quantile 0.95 Forecast | | | Quantile 0.99 Forecast | | | | | | | |
| | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF | | | | | |
| | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min | | | | | |
| MEAN | 2 | -0.0211 | -0.0207 | -0.0133 | -0.0131 | 0.0144 | 0.0142 | 0.0221 | 0.0217 | -0.0618 | -0.0605 | -0.0352 | -0.0348 | 0.0420 | 0.0414 | 0.0650 | 0.0633 |
| | 5 | -0.0205 | -0.0198 | -0.0129 | -0.0125 | 0.0141 | 0.0136 | 0.0216 | 0.0208 | -0.0603 | -0.0582 | -0.0342 | -0.0332 | 0.0410 | 0.0398 | 0.0635 | 0.0611 |
| | 20 | -0.0199 | -0.0193 | -0.0124 | -0.0122 | 0.0136 | 0.0134 | 0.0210 | 0.0204 | -0.0585 | -0.0570 | -0.0328 | -0.0323 | 0.0397 | 0.0391 | 0.0617 | 0.0600 |
| BIAS | 2 | -0.0008 | -0.0003 | -0.0001 | 0.0001 | 0.0001 | -0.0001 | 0.0007 | 0.0003 | -0.0023 | -0.0010 | -0.0005 | -0.0001 | 0.0005 | -0.0002 | 0.0024 | 0.0006 |
| | 5 | -0.0011 | -0.0004 | -0.0004 | 0.0000 | 0.0003 | -0.0001 | 0.0011 | 0.0003 | -0.0029 | -0.0009 | -0.0011 | -0.0001 | 0.0010 | -0.0002 | 0.0030 | 0.0006 |
| | 20 | -0.0011 | -0.0005 | -0.0002 | 0.0000 | 0.0002 | 0.0000 | 0.0011 | 0.0005 | -0.0024 | -0.0009 | -0.0007 | -0.0002 | 0.0007 | 0.0000 | 0.0025 | 0.0008 |
| RMSE | 2 | 0.0058 | 0.0034 | 0.0037 | 0.0022 | 0.0037 | 0.0022 | 0.0058 | 0.0034 | 0.0139 | 0.0082 | 0.0098 | 0.0062 | 0.0098 | 0.0061 | 0.0138 | 0.0082 |
| | 5 | 0.0050 | 0.0033 | 0.0032 | 0.0022 | 0.0032 | 0.0022 | 0.0050 | 0.0034 | 0.0120 | 0.0078 | 0.0086 | 0.0058 | 0.0084 | 0.0059 | 0.0118 | 0.0079 |
| | 20 | 0.0045 | 0.0028 | 0.0029 | 0.0018 | 0.0029 | 0.0018 | 0.0045 | 0.0028 | 0.0104 | 0.0064 | 0.0076 | 0.0048 | 0.0074 | 0.0047 | 0.0104 | 0.0065 |

| Panel B: 5-day horizon | | | | | | | | | | | | | | | | | |
|-------------------------------|---------------------------|---------|---------|---------------------------|---------|---------|---------------------------|---------|---------|---------------------------|---------|---------|---------|--------|---------|--------|--------|
| T (years) | Quantile 0.01 Forecast | | | Quantile 0.05 Forecast | | | Quantile 0.95 Forecast | | | Quantile 0.99 Forecast | | | | | | | |
| | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF | | | | | |
| | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min | | | | | |
| MEAN | 2 | -0.0618 | -0.0605 | -0.0352 | -0.0348 | -0.0618 | -0.0605 | -0.0352 | -0.0348 | -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 | -0.0603 | -0.0582 | -0.0342 | -0.0332 | -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 | -0.0585 | -0.0570 | -0.0328 | -0.0323 | -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 | -0.0023 | -0.0010 | -0.0005 | -0.0001 | -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 | -0.0029 | -0.0009 | -0.0011 | -0.0001 | -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 | -0.0024 | -0.0009 | -0.0007 | -0.0002 | -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 | 0.0139 | 0.0082 | 0.0098 | 0.0062 | 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 | 0.0120 | 0.0078 | 0.0086 | 0.0058 | 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 | 0.0104 | 0.0064 | 0.0076 | 0.0048 | 0.0104 | 0.0064 | 0.0076 | 0.0048 | 0.0074 | 0.0047 | 0.0104 | 0.0065 |

| Panel C: 10-day horizon | | | | | | | | | | | | | | | | | |
|--------------------------------|---------------------------|---------|---------|---------------------------|---------|--------|---------------------------|--------|--------|---------------------------|---------|---------|---------|--------|---------|--------|---------|
| T (years) | Quantile 0.01 Forecast | | | Quantile 0.05 Forecast | | | Quantile 0.95 Forecast | | | Quantile 0.99 Forecast | | | | | | | |
| | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF | | | | | |
| | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min | | | | | |
| MEAN | 2 | -0.0897 | -0.0885 | -0.0481 | -0.0475 | 0.0608 | 0.0598 | 0.0963 | 0.0945 | -0.1224 | -0.1208 | -0.0656 | -0.0647 | 0.0889 | 0.0875 | 0.1388 | 0.1366 |
| | 5 | -0.0875 | -0.0857 | -0.0466 | -0.0453 | 0.0593 | 0.0576 | 0.0942 | 0.0918 | -0.1189 | -0.1169 | -0.0632 | -0.0616 | 0.0867 | 0.0844 | 0.1355 | 0.1327 |
| | 20 | -0.0854 | -0.0845 | -0.0446 | -0.0439 | 0.0576 | 0.0566 | 0.0921 | 0.0907 | -0.1158 | -0.1151 | -0.0605 | -0.0597 | 0.0844 | 0.0831 | 0.1327 | 0.1314 |
| BIAS | 2 | -0.0023 | -0.0011 | -0.0011 | -0.0005 | 0.0009 | -0.0001 | 0.0024 | 0.0006 | -0.0039 | -0.0023 | -0.0022 | -0.0013 | 0.0017 | 0.0003 | 0.0034 | 0.0012 |
| | 5 | -0.0024 | -0.0006 | -0.0017 | -0.0003 | 0.0015 | -0.0002 | 0.0026 | 0.0001 | -0.0030 | -0.0010 | -0.0024 | -0.0007 | 0.0020 | -0.0003 | 0.0027 | -0.0001 |
| | 20 | -0.0013 | -0.0004 | -0.0010 | -0.0003 | 0.0010 | 0.0000 | 0.0016 | 0.0003 | -0.0012 | -0.0004 | -0.0013 | -0.0005 | 0.0013 | 0.0000 | 0.0012 | -0.0001 |
| RMSE | 2 | 0.0165 | 0.0100 | 0.0132 | 0.0087 | 0.0132 | 0.0085 | 0.0166 | 0.0101 | 0.0219 | 0.0143 | 0.0179 | 0.0125 | 0.0178 | 0.0122 | 0.0214 | 0.0138 |
| | 5 | 0.0133 | 0.0088 | 0.0115 | 0.0077 | 0.0112 | 0.0079 | 0.0133 | 0.0093 | 0.0169 | 0.0114 | 0.0149 | 0.0101 | 0.0145 | 0.0105 | 0.0162 | 0.0117 |
| | 20 | 0.0108 | 0.0069 | 0.0100 | 0.0063 | 0.0098 | 0.0062 | 0.0111 | 0.0072 | 0.0130 | 0.0084 | 0.0126 | 0.0079 | 0.0123 | 0.0079 | 0.0128 | 0.0085 |

| Panel D: 20-day horizon | | | | | | | | | | | | | | | | | |
|--------------------------------|---------------------------|---------|---------|---------------------------|---------|---------|---------------------------|---------|---------|---------------------------|---------|---------|---------|--------|---------|--------|---------|
| T (years) | Quantile 0.01 Forecast | | | Quantile 0.05 Forecast | | | Quantile 0.95 Forecast | | | Quantile 0.99 Forecast | | | | | | | |
| | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF | Daily | HF | HF | | | | | |
| | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min | Only | 5-min | 5-min | | | | | |
| MEAN | 2 | -0.1224 | -0.1208 | -0.0656 | -0.0647 | -0.1224 | -0.1208 | -0.0656 | -0.0647 | -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 | -0.1189 | -0.1169 | -0.0632 | -0.0616 | -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 | -0.1158 | -0.1151 | -0.0605 | -0.0597 | -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 | -0.0039 | -0.0023 | -0.0022 | -0.0013 | -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 | -0.0030 | -0.0010 | -0.0024 | -0.0007 | -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 | -0.0012 | -0.0004 | -0.0013 | -0.0005 | -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 | 0.0219 | 0.0143 | 0.0179 | 0.0125 | 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 | 0.0169 | 0.0114 | 0.0149 | 0.0101 | 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 | 0.0130 | 0.0084 | 0.0126 | 0.0079 | 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 procedures in adjacent column pairs: either without (left column 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 underestimation of the implied capital charge for market risk based on one-percent (quantile 0.01) and five-percent (quantile 0.05) VaR.

| True Forecast Rank (quantiles) | Sample Size T (years) | Panel A: 1-day horizon | | | | | | Panel B: 5-day horizon | | | | | | Panel C: 10-day horizon | | | | | | Panel D: 20-day horizon | | | | | |
|--------------------------------|-----------------------|------------------------|----------------|----------------|----------------|----------------|----------------|------------------------|----------------|----------------|----------------|----------------|----------------|-------------------------|----------------|----------------|----------------|----------------|----------------|-------------------------|----------------|--|--|--|--|
| | | Quantile 0.01 | | Quantile 0.05 | | Quantile 0.01 | | Quantile 0.05 | | Quantile 0.01 | | Quantile 0.05 | | Quantile 0.01 | | Quantile 0.05 | | Quantile 0.01 | | Quantile 0.05 | | | | | |
| | | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | Forecast Error | | | | |
| 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% | -2.96% | -0.56% | -6.45% | -2.02% | -0.46% | 1.66% | -0.46% | 1.66% | | | | |
| | 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% | -4.74% | -2.81% | -8.37% | -4.51% | -3.70% | -1.61% | -3.70% | -1.61% | | | | |
| 2nd | 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% | -3.95% | -1.79% | -7.42% | -3.50% | -3.99% | -1.76% | -3.99% | -1.76% | | | | |
| | 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% | 0.84% | 0.13% | -3.01% | -1.66% | 2.16% | 0.97% | 2.16% | 0.97% | | | | |
| 3rd | 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% | 1.82% | 1.09% | -0.68% | 0.12% | 2.06% | 1.35% | 2.06% | 1.35% | | | | |
| | 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% | -0.22% | -0.57% | -3.76% | -2.53% | -0.46% | -0.57% | -0.46% | -0.57% | | | | |
| 4th | 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% | 6.09% | 3.46% | 5.56% | 3.56% | 5.82% | 3.51% | 5.82% | 3.51% | | | | |
| | 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% | 7.42% | 2.71% | 8.40% | 2.49% | 6.38% | 2.37% | 6.38% | 2.37% | | | | |
| 5th (High) | 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% | 4.68% | 2.52% | 3.56% | 2.77% | 3.46% | 2.03% | 3.46% | 2.03% | | | | |
| | 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% | 8.17% | 3.47% | 11.88% | 5.05% | 6.58% | 2.80% | 6.58% | 2.80% | | | | |
| 5th (High) | 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% | 9.53% | 3.85% | 16.39% | 6.53% | 7.37% | 2.75% | 7.37% | 2.75% | | | | |
| | 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% | 6.75% | 2.87% | 14.47% | 6.74% | 5.21% | 2.28% | 5.21% | 2.28% | | | | |
| 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% | 6.55% | 1.76% | 27.54% | 8.12% | 5.33% | 1.12% | 5.33% | 1.12% | | | | |
| | 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% | 5.51% | 1.45% | 29.46% | 9.63% | 4.59% | 1.12% | 4.59% | 1.12% | | | | |
| 5th (High) | 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% | 5.45% | 1.38% | 35.05% | 11.81% | 4.52% | 1.20% | 4.52% | 1.20% | | | | |
| | 20 | | | | | | | | | | | | | | | | | | | | | | | | |

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”).

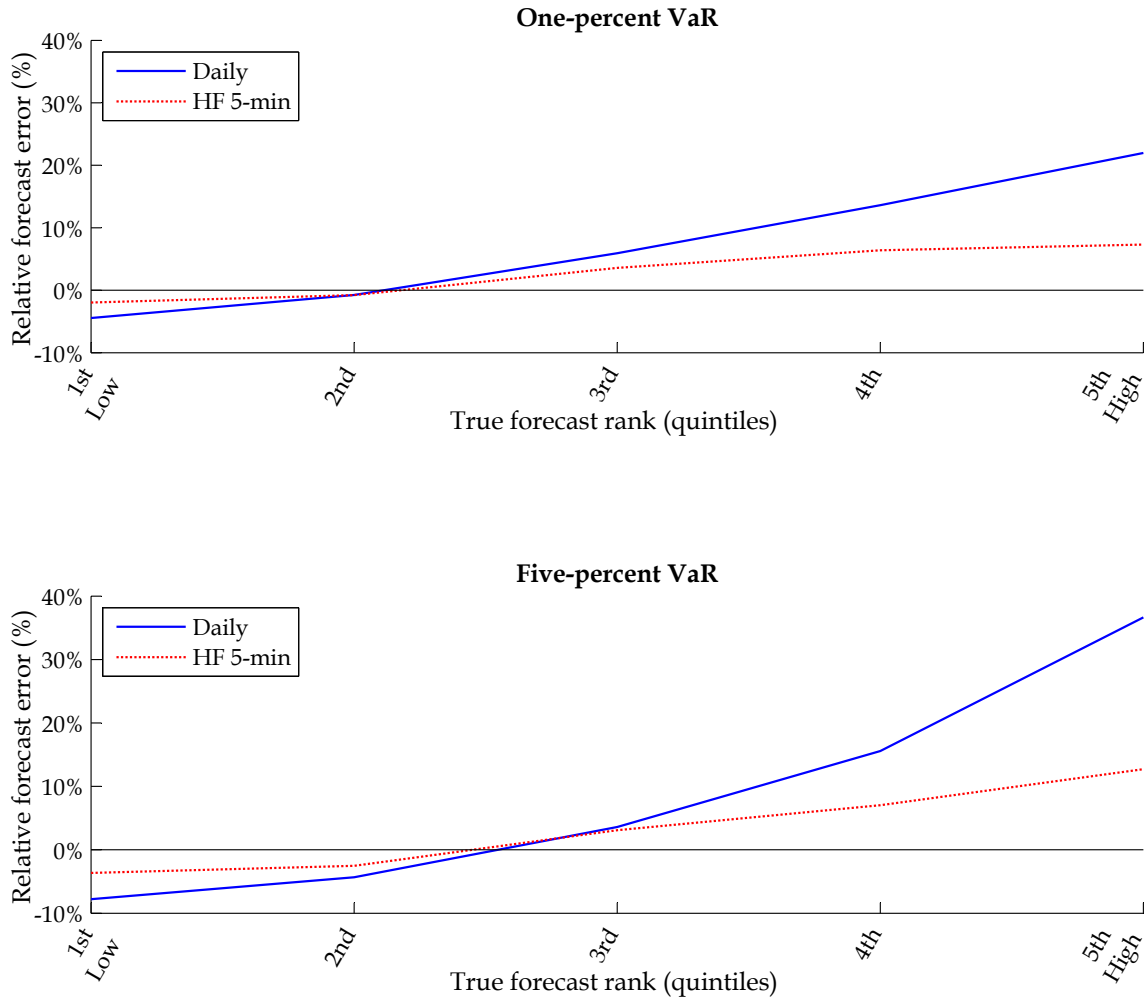


Figure 2: **One-percent and five-percent VaR forecasts for S&P 500 returns during the financial crisis of 2008.** We plot one-percent (top graph) and five-percent (bottom graph) VaR forecasts at five-day horizon without overlapping for S&P 500 futures returns based on a two-factor log-SV model with jumps 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.

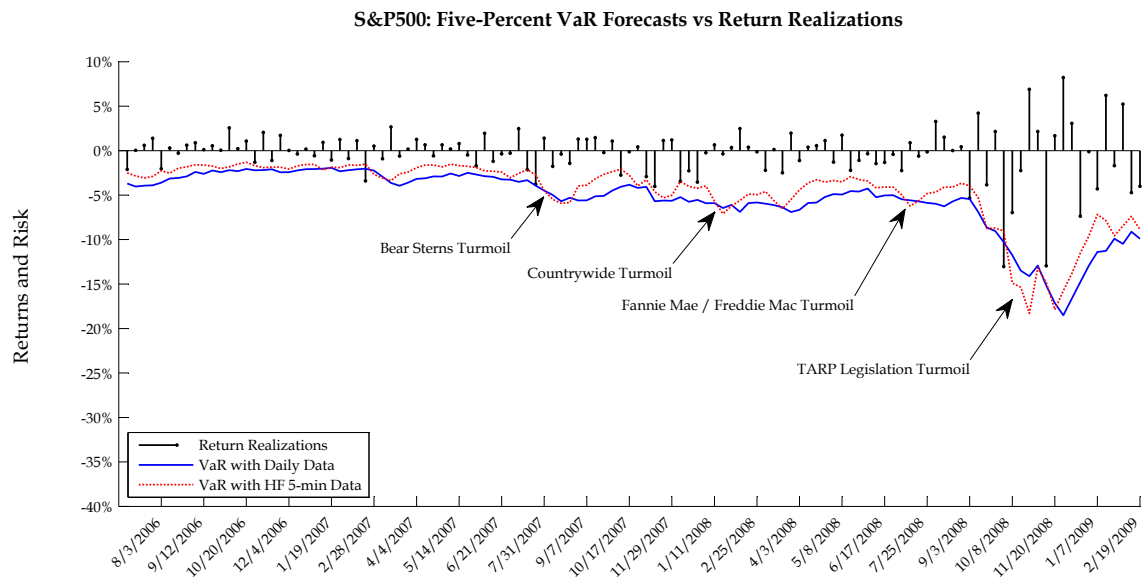
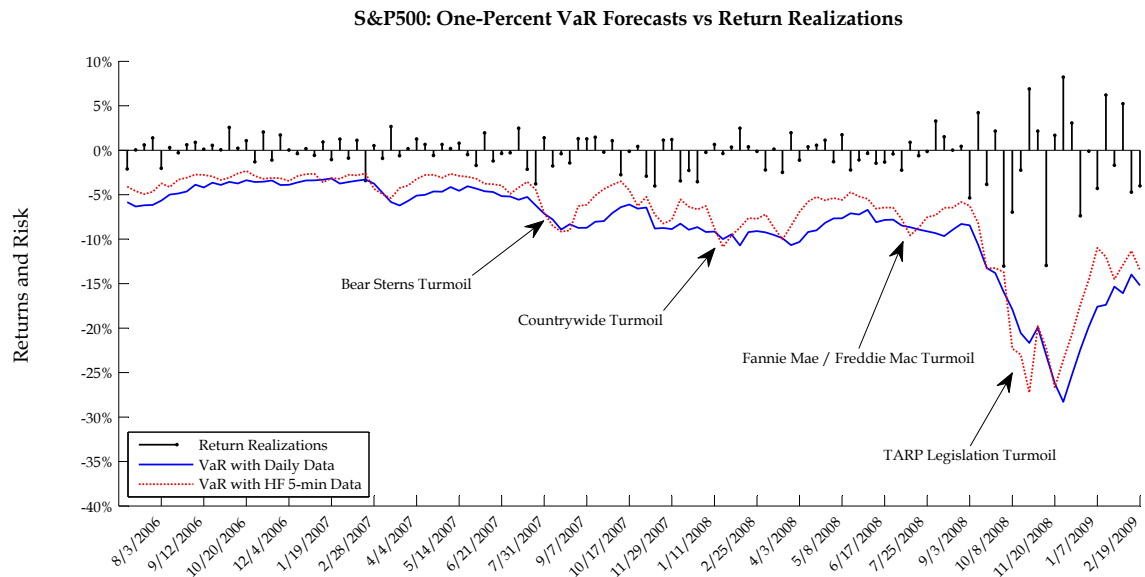


Figure 3: One-percent and five-percent VaR forecasts for Google returns during the financial crisis of 2008. We plot one-percent (top graph) and five-percent (bottom graph) VaR forecasts at five-day horizon without overlapping for Google returns based on a two-factor log-SV model with jumps in returns. 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).

