10-2012

Estimation of High-Frequency Volatility: An Autoregressive Conditional Duration Approach

Yiu Kuen TSE

Singapore Management University, yktse@smu.edu.sg

Thomas Tao YANG

Boston College

DOI: https://doi.org/10.1080/07350015.2012.707582

Published in Journal of Business and Economic Statistics, Volume 30, Issue 4, 2012, Pages 533-545
https://doi.org/10.1080/07350015.2012.707582

Estimation of High-Frequency Volatility: An Autoregressive Conditional Duration Approach

Yiu-Kuen TSE
School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903 (*yktse@smu.edu.sg*)

Thomas Tao YANG
Department of Economics, Boston College, Chestnut Hill, MA 02467 (*yangta@bc.edu*)

We propose a method to estimate the intraday volatility of a stock by integrating the instantaneous conditional return variance per unit time obtained from the autoregressive conditional duration (ACD) model, called the ACD-ICV method. We compare the daily volatility estimated using the ACD-ICV method against several versions of the realized volatility (RV) method, including the bipower variation RV with subsampling, the realized kernel estimate, and the duration-based RV. Our Monte Carlo results show that the ACD-ICV method has lower root mean-squared error than the RV methods in almost all cases considered. This article has online supplementary material.

KEY WORDS: Market microstructure; Realized volatility; Semiparametric method; Transaction data.

1. INTRODUCTION

Since the seminal work by Andersen et al. (2001a, 2001b), the realized volatility (RV) method has been widely used for the estimation of daily volatility. The object of interest in the RV literature is the estimation of the integrated volatility (IV) of asset returns. Suppose the logarithmic asset price follows a diffusion process with instantaneous variance per unit time at time t being $\sigma^2(t)$. The IV of the asset return over the time interval $(0,\ t)$ is defined as

$$\mathrm{IV}_t = \int_0^t \sigma^2(u)\, du. \qquad (1)$$

In the RV literature, $\sigma^2(t)$ is typically assumed to be stochastic. The basic RV method makes use of asset-price data sampled at very high frequency, such as every 5 min or higher, and is computed as the sum of the squared differenced logarithmic asset prices. However, as the efficient prices may be contaminated by market microstructure noise and price jumps, other RV methods incorporating various improvements and modifications have been proposed. An advantage of the RV methods is that no specific functional form of the instantaneous variance $\sigma^2(t)$ is assumed and the method is sometimes described as nonparametric.

In this article, we propose to estimate high-frequency (daily) or ultra-high-frequency (intraday) return volatility *parametrically*. The object of interest in this approach is the price duration, which is defined as the time taken for the cumulative change in the logarithmic transaction price to reach or exceed a given threshold δ, called the price range. The occurrence of this incident is called a price event. As shown by Engle and Russell (1998), the instantaneous conditional return variance per unit time depends on δ and the conditional hazard rate function of the duration distribution. We model the price-duration process parametrically using an extended version of the autoregressive conditional duration (ACD) model by Engle and Russell (1998), namely, the augmented ACD (AACD) model by Fernandes and Grammig (2006). The variance over a given intraday time interval is estimated by calculating the integrated conditional variance (ICV) over the interval, and we call this the ACD-ICV method.

An important difference between the RV estimate and the ACD-ICV estimate of volatility is that the former estimates the integrated volatility over a time interval while the latter estimates the integrated instantaneous conditional variance. While instantaneous variance in the RV framework is stochastic, the instantaneous conditional variance in our approach is deterministic. This comparison is analogous to the stochastic volatility approach versus the conditional heteroscedasticity approach in the literature of volatility modeling.

The ACD-ICV method has several advantages over the RV approach. First, the RV approach is based on the asset prices, which may be affected by market microstructure noise, price discreteness, and price jumps. On the other hand, price data are used in the ACD-ICV method only for the determination of the price events, and their numerical values are not used in computation. This feature introduces some robustness in the ACD-ICV estimate, which is shared by the Andersen, Dobrev, and Schaumburg (2008) method. Second, unlike the RV methods, which sample data over regular time intervals, the ACD-ICV method samples data randomly, depending on the price movements. More data are used in periods of active trading, resulting in more efficient sampling. Third, the RV methods use only local data for the period of interest (daily or intraday). In contrast, the ACD-ICV method makes use of data outside the period of interest, based on the assumption that the transaction durations in the sample follow an autoregressive process. As empirical studies in the literature support this regularity in transaction duration, using data outside the local interval may improve the volatility estimation.

Finally, to invoke the consistency of the RV estimates, a large amount of infill data must be used. For short intraday intervals such as an hour or 15 min, it is doubtful if the infill sample size is large enough to justify the applicability of the asymptotics of the RV estimates. In contrast, the ACD-ICV method depends on the conditional expected duration, which can be consistently estimated by the maximum likelihood method with data extended beyond the period of interest. Thus, the ACD-ICV method may produce better estimates of volatility over short intraday intervals.

The balance of this article is as follows. In Section 2, we review the ACD model and its estimation. We then outline the use of the ACD model for the estimation of high-frequency volatility. In Section 3, we report some Monte Carlo (MC) results for comparing the performance of the RV methods and the ACD-ICV method. Our results show that the ACD-ICV method has smaller root mean-squared error (RMSE) than the RV estimates in almost all cases considered. Section 4 reports some results on out-of-sample one-day ahead volatility forecast and ultra-high-frequency (intraday) volatility estimation. Some empirical results using New York Stock Exchange (NYSE) data are presented in Section 5. Finally, Section 6 concludes. Supplementary materials can be found in the online Appendix, posted on the journal web site.

2. ACD MODEL AND HIGH-FREQUENCY VOLATILITY

The ACD model was proposed by Engle and Russell (1998) to analyze the duration of transactions of financial assets. A recent review of the literature on the ACD model and its applications to finance can be found in the article by Pacurar (2008). As the instantaneous conditional variance per unit time derived from the ACD model can be integrated over a given time interval (between two trades or over a day) to obtain a measure of the volatility over the interval, we propose to estimate the integral parametrically to obtain an estimate of high-frequency volatility.

2.1 ACD Model

Let $s(t)$ be the price of a stock at time t and $\tilde{s}(t)$ be its logarithmic price. Consider a sequence of times $t_0, t_1, \ldots, t_N$ with $t_0 < t_1 < \cdots < t_N$, for which t_i denotes the time of occurrence of the ith *price event* of the stock. A price event occurs if the cumulative change in $\tilde{s}(t)$ reaches or exceeds an amount δ, called the *price range*, whether upward or downward. Thus, $x_i = t_i - t_{i-1}$, for $i = 1, 2, \ldots, N$, are the intervals between consecutive price events, called the *price durations*, and are the data for analysis in the ACD model. Unlike the RV methods, which assume the transaction price follows a Brownian semimartingale (BSM) with possible contamination due to market microstructure noise and/or price jumps, our object of analysis is the price duration x_i.

Let Φ_i denote the information set upon the price event at time t_i, which may include lagged price duration, volume, and order flow. In this article, however, we only consider lagged price durations. We denote $\psi_{i+1} = \mathrm{E}(x_{i+1} \mid \Phi_i)$ as the conditional expectation of the price duration, and assume that the standardized durations $\epsilon_i = x_i/\psi_i$, $i = 1, \ldots, N$, are iid positive random variables with mean 1 and density function $f(\cdot)$. Thus, the hazard function of ϵ_i is $\lambda(\cdot) = f(\cdot)/S(\cdot)$, where $S(\cdot)$ is the survival function of ϵ_i. Assuming ψ_{i+1} to be known given Φ_i, the conditional hazard function (also called the conditional intensity) of $x = t - t_i$, for $t > t_i$, denoted by $\lambda_x(x \mid \Phi_i)$, is

$$\lambda_x(x \mid \Phi_i) = \lambda\left(\frac{t - t_i}{\psi_{i+1}}\right)\frac{1}{\psi_{i+1}}. \quad (2)$$

A popular model for the conditional duration ψ_i is the ACD(1, 1) model defined by

$$\psi_i = \omega + \alpha x_{i-1} + \beta\psi_{i-1}, \quad (3)$$

with the restrictions α, β, and $w \geq 0$, and $\alpha + \beta < 1$. Recently, Fernandes and Grammig (2006) proposed an extension called the AACD model, which is defined by

$$\psi_i^{\lambda} = \omega + \alpha\psi_{i-1}^{\lambda}\,[|\epsilon_{i-1} - b| + c(\epsilon_{i-1} - b)]^{v} + \beta\psi_{i-1}^{\lambda}. \quad (4)$$

The parameters λ and v determine the shape of the transformation. Asymmetric responses in duration shocks are permitted through the shift parameter b and the rotation parameter c. As in the case of the ACD(1, 1) model, the parameters α, β, and w are assumed to be nonnegative. The empirical study reported by Fernandes and Grammig (2006) showed that the AACD model performs better than the ACD(1, 1) model and provides a good fit for the data. Due to its flexibility, we adopt the AACD model as our operating ACD model (generically defined) for price duration.

Given the density function $f(\cdot)$ of the standardized duration ϵ_i, the maximum likelihood estimates (MLE) of the parameters of the ACD equation can be computed. A simple case is when ϵ_i are assumed to be standard exponential, giving rise to the quasi MLE (QMLE) method. As discussed by Drost and Werker (2004), provided the conditional expected-duration equation is correctly specified, the QMLE is consistent for the parameters of the equation regardless of the true distribution of ϵ_i. However, misspecification in the conditional expected duration may induce inconsistency in the QMLE if the wrong density function $f(\cdot)$ is used. To resolve this problem, the semiparametric (SP) method may be adopted. This method was proposed by Engle and Gonzalez-Rivera (1991) to estimate the autoregressive conditional heteroscedasticity (ARCH) model. Drost and Werker (2004) discussed the conditions under which the SP method for estimating conditional-duration models is adaptive and efficient. In the Appendix (posted online), we report some MC results for comparing the performance of the MLE, QMLE, and SP methods for the ACD model. Our results support the consistency of these estimates when the ACD equation is correctly specified and demonstrate the relative efficiency of the SP method over the QMLE method.

2.2 Estimation of High-Frequency Volatility Using the ACD Model

Given the information Φ_i at time t_i, the conditional intensity function $\lambda_x(x \mid \Phi_i)$ determines the probability that the next price event will occur at time $t > t_i$. Specifically, given Φ_i, $\lambda_x(x \mid \Phi_i)\Delta x$ is the probability that the next price event after time t_i occurs in the interval $(t_i + x,\ t_i + x + \Delta x)$, for $x > 0$. The conditional instantaneous return variance per unit time at time t is defined as

$$\sigma^2(t \mid \Phi_i) = \lim_{\Delta t \to 0} \left\{ \frac{1}{\Delta t} \text{var}\,[\tilde{s}(t+\Delta t) - \tilde{s}(t) | \Phi_i] \right\}, \quad t > t_i, \tag{5}$$

where $\tilde{s}(t+\Delta t) - \tilde{s}(t)$ takes possible values $-\delta$, 0, and δ. In particular, $|\tilde{s}(t+\Delta t) - \tilde{s}(t)|$ is δ with probability $\lambda_x(x \mid \Phi_i)\Delta x$ and 0 with probability $1 - \lambda_x(x \mid \Phi_i)\Delta x$. Thus, Equation (5) can be evaluated as

$$\sigma^2(t \mid \Phi_i) = \delta^2 \lambda_x(x \mid \Phi_i), \tag{6}$$

where $x = t - t_i$, $t > t_i$. Using Equation (2), we have

$$\sigma^2(t \mid \Phi_i) = \frac{\delta^2}{\psi_{i+1}} \lambda \left(\frac{t - t_i}{\psi_{i+1}} \right), \qquad t > t_i. \tag{7}$$

Hence, the ICV over the interval $(t_i,\ t_{i+1})$, denoted by ICV_i, is

$$\begin{aligned} \text{ICV}_i &= \int_{t_i}^{t_{i+1}} \sigma^2(t \mid \Phi_i)\, dt \\ &= \frac{\delta^2}{\psi_{i+1}} \int_{t_i}^{t_{i+1}} \lambda \left(\frac{t - t_i}{\psi_{i+1}} \right) dt. \end{aligned} \tag{8}$$

If ϵ_i are iid standard exponential, $\lambda(\cdot) \equiv 1$ and we have

$$\text{ICV}_i = \delta^2 \left[\frac{t_{i+1} - t_i}{\psi_{i+1}} \right]. \tag{9}$$

Furthermore, if $t_0 < t_1 < \cdots < t_N$ denote the price events in a day, the ICV of the day is

$$\text{ICV} = \delta^2 \sum_{i=0}^{N-1} \frac{t_{i+1} - t_i}{\psi_{i+1}}. \tag{10}$$

Under the exponential assumption for ϵ_i we can estimate the parameters of the ACD model by the QMLE, from which we obtain the estimated conditional expected duration $\hat{\psi}_{i+1}$. The ACD-ICV estimate, denoted by V_A, is then computed as

$$V_A = \delta^2 \sum_{i=0}^{N-1} \frac{t_{i+1} - t_i}{\hat{\psi}_{i+1}}. \tag{11}$$

In the case when no specific distribution is assumed for ϵ_i, we may compute the SP estimate of ψ_{i+1}, denoted by ψ^*_{i+1}, and

Table 1. Monte Carlo results for stochastic volatility models with NSR = 0.25

	Volatility model											
	Heston			Heston with jumps			LV			LV with jumps		
Estimation method	ME	SE	RMSE	ME	SE	RMSE	ME	SE	RMSE	ME	SE	RMSE
Panel A: Transaction price with white noise												
V_A^*	−0.063	0.934	0.936	−0.206	1.147	1.166	−0.096	0.790	0.796	−0.094	0.787	0.792
V_A	−0.862	0.967	1.296	−1.238	1.173	1.705	−0.913	0.773	1.196	−0.910	0.770	1.192
V_B	−0.160	1.251	1.261	−0.230	1.667	1.683	−0.163	1.316	1.326	−0.177	1.311	1.322
V_D	−0.811	1.458	1.668	−1.460	1.868	2.371	−0.891	1.490	1.736	−0.828	1.493	1.708
V_K	−0.056	1.354	1.355	−0.083	1.806	1.808	−0.046	1.426	1.426	−0.064	1.420	1.422
V_R	−0.265	2.285	2.301	−0.376	3.050	3.073	−0.278	2.383	2.399	−0.277	2.405	2.421
Panel B: Transaction price with autocorrelated noise												
V_A^*	−0.057	0.933	0.935	−0.200	1.144	1.162	−0.084	0.792	0.796	−0.085	0.785	0.790
V_A	−0.861	0.966	1.294	−1.236	1.171	1.702	−0.903	0.774	1.189	−0.903	0.769	1.186
V_B	−0.165	1.247	1.258	−0.222	1.680	1.694	−0.163	1.316	1.326	−0.177	1.311	1.322
V_D	−0.812	1.465	1.675	−1.462	1.873	2.377	−0.882	1.492	1.733	−0.826	1.494	1.707
V_K	−0.050	1.347	1.347	−0.093	1.808	1.810	−0.047	1.426	1.426	−0.064	1.420	1.422
V_R	−0.264	2.288	2.303	−0.381	3.066	3.090	−0.268	2.402	2.417	−0.289	2.387	2.404
Panel C: Transaction price with noise correlated with efficient price												
V_A^*	−0.068	0.939	0.941	−0.211	1.137	1.156	−0.103	0.796	0.803	−0.096	0.781	0.787
V_A	−0.869	0.971	1.304	−1.241	1.164	1.701	−0.918	0.778	1.203	−0.913	0.764	1.190
V_B	−0.165	1.250	1.261	−0.219	1.670	1.684	−0.163	1.316	1.326	−0.178	1.311	1.323
V_D	−0.814	1.458	1.670	−1.460	1.873	2.375	−0.912	1.497	1.753	−0.845	1.490	1.713
V_K	−0.052	1.354	1.355	−0.075	1.799	1.800	−0.047	1.426	1.426	−0.065	1.420	1.422
V_R	−0.276	2.282	2.298	−0.359	3.038	3.059	−0.299	2.394	2.413	−0.268	2.404	2.419
Panel D: Transaction price with autocorrelated noise correlated with efficient price												
V_A^*	−0.065	0.951	0.954	−0.205	1.156	1.174	−0.094	0.783	0.789	−0.095	0.788	0.793
V_A	−0.871	0.983	1.314	−1.236	1.182	1.711	−0.911	0.766	1.191	−0.912	0.771	1.194
V_B	−0.167	1.253	1.264	−0.232	1.669	1.685	−0.169	1.313	1.324	−0.177	1.318	1.330
V_D	−0.816	1.468	1.679	−1.454	1.875	2.372	−0.897	1.502	1.749	−0.836	1.496	1.714
V_K	−0.062	1.347	1.348	−0.078	1.805	1.807	−0.046	1.423	1.424	−0.057	1.417	1.418
V_R	−0.278	2.291	2.308	−0.366	3.059	3.081	−0.284	2.386	2.403	−0.277	2.405	2.421

NOTES: ME = mean error; SE = standard error (standard deviation of MC samples); RMSE = root mean-squared error. The results are based on 1000 MC replications of 60-day daily volatility estimates. All figures are in percentage. V_A^* is computed using Equation (11) with δ being the average price range conditional on a price event being observed, which is defined as the cumulative change in the logarithmic price exceeding the threshold. V_A is computed from Equation (11) with δ being the threshold price range. The Heston model with jumps is the Heston diffusion model with jumps in the volatility, while the LV model with jumps is the LV model with jumps in the price.

estimate ICV by

$$V_A = \delta^2 \sum_{i=0}^{N-1} \frac{1}{\psi_{i+1}^*} \int_{t_i}^{t_{i+1}} \hat{\lambda}\left(\frac{t-t_i}{\psi_{i+1}^*}\right) dt, \tag{12}$$

where $\hat{\lambda}(\cdot)$ is the base hazard function calculated using the empirical density function $\hat{f}(\cdot)$ of the estimated standardized duration obtained from the QMLE. The computation of $\hat{\lambda}(\cdot)$ requires the numerical integration of $\hat{f}(\cdot)$ (to obtain the estimated survival function). Another round of numerical integration is then required to compute the integrals in Equation (12).

While the SP estimates of ψ_i are theoretically superior to the QMLE, they are computationally very demanding. On the other hand, our results for the ACD-ICV estimates using the QMLE and SP methods are found to be quite similar in both the MC experiments and empirical applications. Thus, we shall only report the results based on the QMLE method in this article. Some results based on the SP method, however, can be found in the online Appendix.

Stock prices may have jumps and market frictions may induce price discreteness. Thus, price events may occur with the actual price range exceeding the threshold. To estimate the intraday volatility, we may replace δ in Equations (11) and (12) by the average price range of the sample observations conditional on the threshold being exceeded. We shall denote this estimate of the ICV by V_A^*.

3. MONTE CARLO COMPARISON OF ACD-ICV AND RV ESTIMATES

We perform some MC experiments to compare the ACD-ICV estimates against various RV estimates. As the two methods are based on different notions of volatility, we consider both deterministic and stochastic volatility models. In the online Appendix, we report the MC results of a deterministic volatility set-up. In this section, we focus on the estimation results for stochastic volatility models. Our MC study for the stochastic volatility models follows closely the experiments designed by Aït-Sahalia and Mancini (2008).

Table 2. Monte Carlo results for stochastic volatility models with NSR = 0.6

	Volatility model											
	Heston			Heston with jumps			LV			LV with jumps		
Estimation method	ME	SE	RMSE	ME	SE	RMSE	ME	SE	RMSE	ME	SE	RMSE
Panel A: Transaction price with white noise												
V_A^*	0.056	1.154	1.155	−0.230	1.523	1.541	−0.031	0.934	0.934	−0.019	0.943	0.943
V_A	−0.525	1.159	1.273	−1.021	1.525	1.835	−0.637	0.917	1.117	−0.625	0.927	1.118
V_B	−0.118	1.206	1.211	−0.210	1.716	1.728	−0.144	1.304	1.312	−0.138	1.316	1.323
V_D	−0.576	1.425	1.537	−1.354	1.929	2.357	−0.700	1.498	1.654	−0.698	1.500	1.654
V_K	0.020	1.303	1.303	−0.040	1.852	1.853	0.015	1.420	1.420	0.008	1.418	1.418
V_R	−0.244	2.209	2.222	−0.382	3.126	3.149	−0.264	2.384	2.399	−0.268	2.406	2.421
Panel B: Transaction price with autocorrelated noise												
V_A^*	0.089	1.142	1.145	−0.209	1.526	1.541	0.000	0.923	0.923	−0.004	0.931	0.931
V_A	−0.498	1.148	1.251	−1.004	1.528	1.828	−0.613	0.906	1.094	−0.616	0.914	1.102
V_B	−0.121	1.195	1.201	−0.216	1.711	1.724	−0.142	1.308	1.316	−0.154	1.305	1.314
V_D	−0.564	1.433	1.540	−1.341	1.935	2.354	−0.688	1.510	1.660	−0.706	1.504	1.662
V_K	0.017	1.301	1.301	−0.047	1.852	1.853	0.020	1.416	1.416	−0.006	1.416	1.416
V_R	−0.251	2.203	2.218	−0.387	3.125	3.149	−0.263	2.396	2.411	−0.285	2.393	2.410
Panel C: Transaction price with noise correlated with efficient price												
V_A^*	0.044	1.146	1.147	−0.243	1.522	1.541	−0.042	0.955	0.956	−0.032	0.941	0.942
V_A	−0.534	1.152	1.269	−1.029	1.525	1.839	−0.646	0.938	1.139	−0.637	0.924	1.123
V_B	−0.120	1.210	1.216	−0.202	1.710	1.722	−0.154	1.307	1.316	−0.141	1.314	1.322
V_D	−0.570	1.433	1.542	−1.352	1.925	2.353	−0.729	1.503	1.670	−0.716	1.510	1.671
V_K	0.027	1.311	1.311	−0.027	1.845	1.846	−0.009	1.416	1.416	0.013	1.428	1.428
V_R	−0.250	2.207	2.221	−0.364	3.107	3.129	−0.295	2.398	2.416	−0.261	2.411	2.425
Panel D: Transaction price with autocorrelated noise correlated with efficient price												
V_A^*	0.059	1.150	1.152	−0.229	1.523	1.540	−0.011	0.933	0.933	−0.009	0.933	0.933
V_A	−0.526	1.156	1.270	−1.021	1.525	1.835	−0.620	0.917	1.107	−0.617	0.916	1.105
V_B	−0.129	1.207	1.214	−0.216	1.713	1.727	−0.133	1.306	1.312	−0.137	1.306	1.313
V_D	−0.565	1.430	1.538	−1.350	1.939	2.362	−0.696	1.496	1.651	−0.691	1.506	1.657
V_K	0.018	1.302	1.302	−0.034	1.852	1.852	0.007	1.419	1.419	0.017	1.418	1.418
V_R	−0.251	2.205	2.219	−0.375	3.137	3.160	−0.263	2.394	2.408	−0.255	2.396	2.409

NOTES: ME = mean error; SE = standard error (standard deviation of MC samples); RMSE = root mean-squared error. The results are based on 1000 MC replications of 60-day daily volatility estimates. All figures are in percentage. V_A^* is computed using Equation (11) with δ being the average price range conditional on a price event being observed, which is defined as the cumulative change in the logarithmic price exceeding the threshold. V_A is computed from Equation (11) with δ being the threshold price range. The Heston model with jumps is the Heston diffusion model with jumps in the volatility, while the LV model with jumps is the LV model with jumps in the price.

Table 3. Monte Carlo results for stochastic volatility models with NSR = 1.00

Estimation method	Heston ME	Heston SE	Heston RMSE	Heston with jumps ME	Heston with jumps SE	Heston with jumps RMSE	LV ME	LV SE	LV RMSE	LV with jumps ME	LV with jumps SE	LV with jumps RMSE
Panel A: Transaction price with white noise												
V_A^*	0.443	1.216	1.294	0.088	1.474	1.476	0.409	0.888	0.978	0.414	0.888	0.979
V_A	−0.146	1.214	1.223	−0.639	1.469	1.602	−0.209	0.872	0.896	−0.205	0.870	0.894
V_B	−0.050	1.240	1.241	−0.138	1.628	1.634	−0.074	1.300	1.302	−0.067	1.312	1.313
V_D	−0.248	1.480	1.501	−0.827	1.864	2.039	−0.319	1.513	1.547	−0.319	1.519	1.552
V_K	0.171	1.351	1.362	0.092	1.761	1.764	0.157	1.423	1.432	0.150	1.421	1.429
V_R	−0.229	2.286	2.298	−0.330	2.989	3.007	−0.237	2.387	2.398	−0.240	2.409	2.420
Panel B: Transaction price with autocorrelated noise												
V_A^*	0.485	1.210	1.304	0.145	1.462	1.469	0.456	0.886	0.996	0.442	0.882	0.987
V_A	−0.114	1.208	1.214	−0.590	1.459	1.574	−0.172	0.868	0.885	−0.186	0.865	0.885
V_B	−0.045	1.236	1.237	−0.142	1.637	1.643	−0.071	1.304	1.306	−0.083	1.300	1.303
V_D	−0.222	1.478	1.495	−0.792	1.852	2.015	−0.299	1.528	1.557	−0.317	1.522	1.555
V_K	0.175	1.345	1.357	0.082	1.771	1.773	0.162	1.420	1.429	0.136	1.419	1.426
V_R	−0.217	2.274	2.284	−0.346	3.006	3.025	−0.235	2.398	2.410	−0.257	2.395	2.409
Panel C: Transaction price with noise correlated with efficient price												
V_A^*	0.402	1.214	1.279	0.061	1.463	1.464	0.368	0.891	0.963	0.367	0.881	0.954
V_A	−0.185	1.212	1.226	−0.664	1.461	1.605	−0.247	0.873	0.907	−0.246	0.864	0.898
V_B	−0.053	1.234	1.235	−0.147	1.636	1.643	−0.086	1.302	1.305	−0.073	1.310	1.312
V_D	−0.264	1.473	1.497	−0.852	1.863	2.049	−0.366	1.520	1.563	−0.353	1.527	1.567
V_K	0.167	1.355	1.365	0.077	1.773	1.775	0.127	1.419	1.425	0.149	1.432	1.439
V_R	−0.229	2.263	2.274	−0.343	2.987	3.007	−0.268	2.400	2.415	−0.235	2.413	2.424
Panel D: Transaction price with autocorrelated noise correlated with efficient price												
V_A^*	0.458	1.210	1.294	0.102	1.467	1.471	0.425	0.884	0.981	0.421	0.885	0.980
V_A	−0.137	1.209	1.217	−0.629	1.464	1.593	−0.196	0.867	0.889	−0.200	0.868	0.891
V_B	−0.052	1.240	1.241	−0.142	1.635	1.641	−0.064	1.301	1.303	−0.067	1.302	1.304
V_D	−0.236	1.482	1.501	−0.819	1.863	2.035	−0.318	1.517	1.550	−0.316	1.520	1.552
V_K	0.167	1.348	1.359	0.084	1.764	1.766	0.145	1.422	1.429	0.155	1.421	1.430
V_R	−0.226	2.277	2.289	−0.321	2.984	3.001	−0.235	2.396	2.407	−0.228	2.398	2.409

NOTES: ME = mean error; SE = standard error (standard deviation of MC samples); RMSE = root mean-squared error. The results are based on 1000 MC replications of 60-day daily volatility estimates. All figures are in percentage. V_A^* is computed using Equation (11) with δ being the average price range conditional on a price event being observed, which is defined as the cumulative change in the logarithmic price exceeding the threshold. V_A is computed from Equation (11) with δ being the threshold price range. The Heston model with jumps is the Heston diffusion model with jumps in the volatility, while the LV model with jumps is the LV model with jumps in the price.

3.1 Heston Model

We assume the following price generation process by Heston (1993):

$$d\log s(t) = \left(\mu - \frac{\sigma^2(t)}{2}\right)dt + \sigma(t)\,dW_1(t), \qquad (13)$$

$$d\sigma^2(t) = \kappa\left(\alpha - \sigma^2(t)\right)dt + \gamma\sigma(t)\,dW_2(t), \qquad (14)$$

with $\mu = 0.05$, $\kappa = 5$, $\alpha = 0.04$, and $\gamma = 0.5$. The correlation coefficient between the two Brownian motions $W_1(t)$ and $W_2(t)$ is −0.5. We generate second-by-second data with initial value of $\sigma(0)$ equaling 0.3. We also incorporate the inclusion of a jump component into the volatility process with a Poisson jump intensity of $2/(6.5 \times 3600)$ and an exponential jump size with a mean of 0.0007.

3.2 Log-Volatility (LV) Model

Let V_t denote the integrated variance for day t and $l(t) = \log(V_t^{\frac{1}{2}})$, which follows the process

$$l(t) = \phi_0 + \sum_{i=1}^{5}\phi_i\, l(t-i) + u(t), \qquad (15)$$

where $u(t)$ is a white noise. The logarithmic return over each 1-sec interval in day t is generated randomly as $V_t^{\frac{1}{2}}z$, where z is normal with mean 0 and standard deviation $1/(3600 \times 6.5 \times 252)^{\frac{1}{2}}$. The parameter set is $\{\phi_0, \phi_1, \phi_2, \phi_3, \phi_4, \phi_5\} = \{-0.0161, -0.35, 0.25, 0.20, 0.10, 0.09\}$, and the standard deviation of $u(t)$ is 0.02. Similar to the Heston model with volatility jumps, we also consider a LV model with a jump component.

3.3 Noise Structure

Given the logarithmic efficient price $\tilde{s}(t)$ generated from Equation (13), we add a noise component $\varepsilon(t)$ to obtain the logarithmic transaction price. The following noise structures are considered. First, we assume a white noise, so that $\varepsilon(t)$ are iid normal variates. Second, we consider the case where $\varepsilon(t)$ follows an AR(1) process with a correlation coefficient of −0.2. Third, we generate serially uncorrelated $\varepsilon(t)$, which are correlated with the return, with $\mathrm{corr}\{\varepsilon(t), \tilde{s}(t) - \tilde{s}(t-\Delta t)\} = -0.2$. Fourth, we consider noises that are autocorrelated as well as correlated with the return, where the correlation coefficients are as given in the second and third cases above. Finally, we consider the case where there is a jump in the price process for the LV model, with

0.4 jump per 5 min and jump sizes of −0.05, −0.03, 0.03, and 0.05 with equal probabilities. Based on the model specification, the annualized volatility is around 25%–30%. The noise-to-signal ratio (NSR) given by $\mathrm{NSR} = [\mathrm{var}\{\varepsilon(t)\}/\mathrm{var}\{\sigma(t)\}]^{\frac{1}{2}}$ is set to 0.25, 0.6, and 1.0 for the main experiments (see the set-up by Andersen, Dobrev, and Schaumburg 2008).

3.4 RV Estimates

We consider the basic RV estimate, denoted by V_R, sampled at 5-min intervals. We also compute the bipower variation RV estimate (Barndorff-Nielsen and Shephard 2004), denoted by V_B. For this method, we sample the price data over 2-min intervals and apply the subsampling method proposed by Zhang, Mykland, and Aït-Sahalia (2005) using subsampling intervals of 5 sec. Next, we consider the realized kernel estimate V_K proposed by Barndorff-Nielsen et al. (2008) using the Tukey-Hanning weighting function. Finally, we calculate the duration-based estimate, denoted by V_D, which was proposed by Andersen, Dobrev, and Schaumburg (2008). We adopt the event of price exiting a range δ for the definition of the passage-time duration, for which V_D is computed as

$$V_D = \sum_{i=0}^{N-1} \hat{\sigma}_\delta^2(t_i)(t_{i+1} - t_i), \tag{16}$$

where $\hat{\sigma}_\delta^2(t_i)$ is the local variance estimate given by Andersen, Dobrev, and Schaumburg (2008). The similarity between Equations (11) and (16) should be noted. While the grid points $t_0, t_1, \ldots, t_N$ in V_D are fixed, these values are the observed price-event times in V_A. In V_D the logarithmic stock price is assumed to follow a local Brownian motion with $\hat{\sigma}_\delta^2(t_i)$ being an estimate of the local variance, whereas in V_A we use $\delta^2/\hat{\psi}_{i+1}$ to estimate the instantaneous conditional variance per unit time within the interval (t_i, t_{i+1}). To compute V_D, we set $\delta = 0.24\%$, which is three times a presumed log spread of 0.08%.

3.5 Monte Carlo Results

In our MC study, we consider the performance of both V_A and V_A^*. We target the average price duration to be 2 min, 4 min, and 5 min for models with NSR of 0.25, 0.6, and 1.0, respectively, so that the average price duration increases with NSR. In each MC replication, δ is determined so as to obtain the approximate desired average duration. The results for the Heston model, Heston model with volatility jumps, LV model, and LV model with price jumps are summarized in Tables 1–3. For the Heston, LV, and LV-with-jumps models, the average value of δ is about 0.15%, 0.17%, and 0.2% for NSR of 0.25, 0.6, and 1.0, respectively. For the Heston model with volatility jumps, the average value of δ is larger at about 0.2%, 0.24%, and 0.26%.

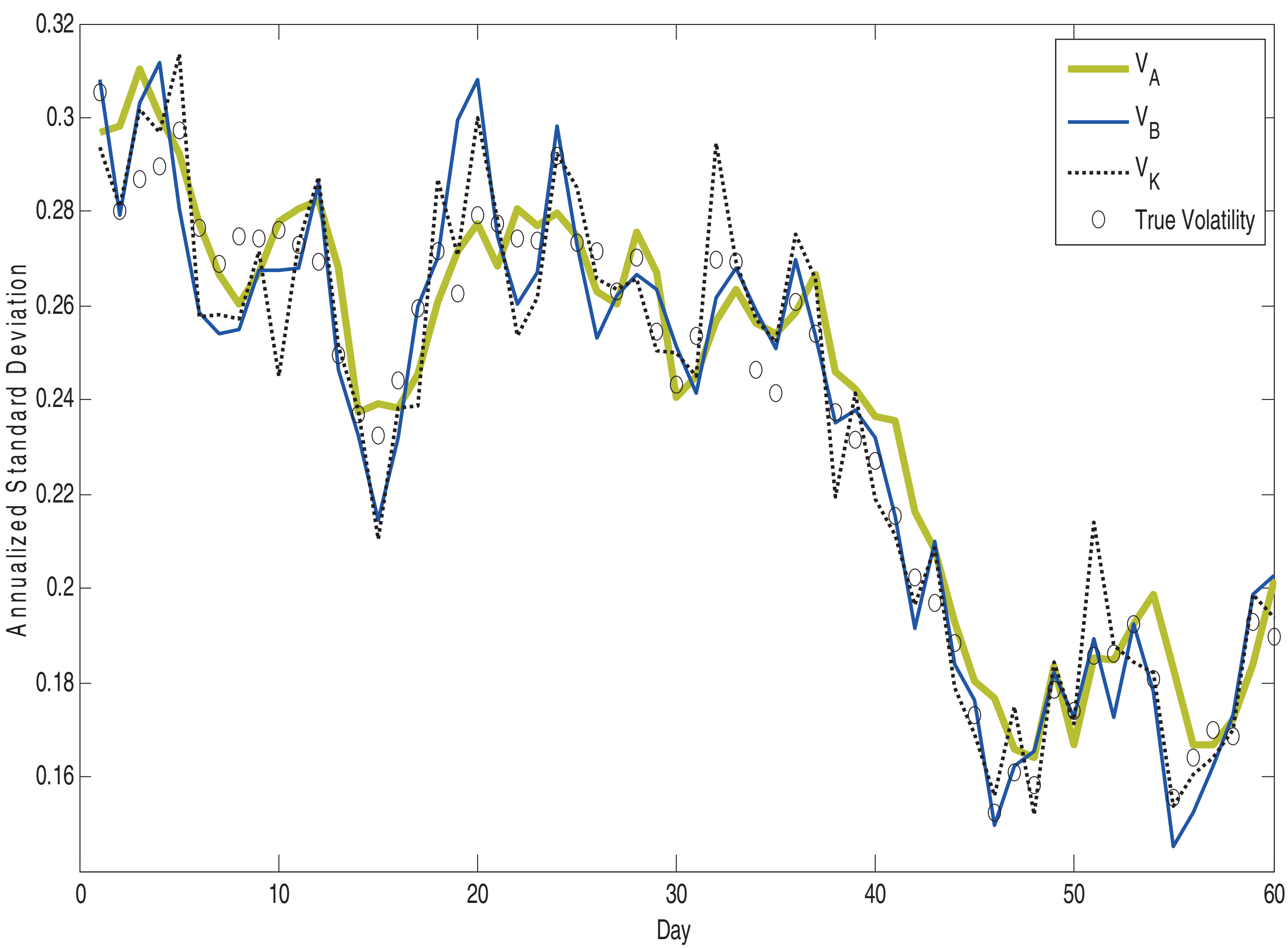

Figure 1. A sample path of daily stochastic volatility (Heston model) and its estimates. The online version of this figure is in color.

The mean error (ME), standard error (SE), and RMSE of the daily volatility estimates (annualized standard deviation in percent) calculated using MC experiments of 1000 replications are summarized.

Among the RV methods, V_B gives the best performance, having the lowest RMSE in all cases considered, followed by V_K. Generally, V_K has smaller absolute ME than V_B but larger SE, with the final result that it is inferior to V_B in RMSE. V_A^* (with δ defined as the conditional mean price range exceeding the threshold) gives lower RMSE than V_B in all cases except for the Heston model without volatility jumps with NSR $= 1$. On the other hand, V_A (with δ defined as the threshold) outperforms V_B for all models except for the Heston models with and without volatility jumps with lower NSR of 0.25 and 0.6. In addition, both V_A and V_A^* outperform V_K in all cases. Among the two ACD-ICV estimates, V_A^* gives better results than V_A when NSR $= 0.25$ and 0.6 for all models, as well as for the Heston model with volatility jumps when NSR $= 1$. However, the over-estimation of V_A^* becomes more serious when NSR is large.

By construction, V_A is always smaller than V_A^*. Since empirically price moves in discrete amounts, it may be theoretically more appropriate to apply the conditional mean price range to compute the ACD-ICV estimate, thus using V_A^*. However, when the microstructure noise and/or price jumps are large, the excess amount in δ in computing V_A^* is mainly due to the noise. The MC results demonstrate this effect clearly, as we can see that when NSR is large the upward bias in V_A^* is large. We should note that as transaction price is used to obtain the price duration, the ACD-ICV method estimates the volatility of the transaction price and not the efficient price. Thus, the positive bias in V_A^* is magnified when NSR is large, because the volatility of the transaction price exceeds that of the efficient price by a larger amount. On the other hand, our MC results also show that when NSR is small V_A^* performs better than V_A. This is because when the microstructure noise is small, the theoretical justification for using V_A^* dominates. Overall, we recommend the use of V_A, which is less contaminated by microstructure noise and/or price jumps. Further MC results for robustness check can be found in the online Appendix.

Figure 1 presents an example of a daily volatility plot over a 60-day period for the Heston model with white noise. It can be seen that all estimates trace the true volatility quite closely, although the RV estimates appear to have larger fluctuations. Figure 2 illustrates a sample of one-day instantaneous volatility path and its intraday volatility estimates. The true instantaneous variance is generated from the Heston model with an intraday periodicity function superimposed to describe the stylistic fact of intraday variation in volatility, with the details provided in the online Appendix. Estimates of 1-hr and 15-min integrated

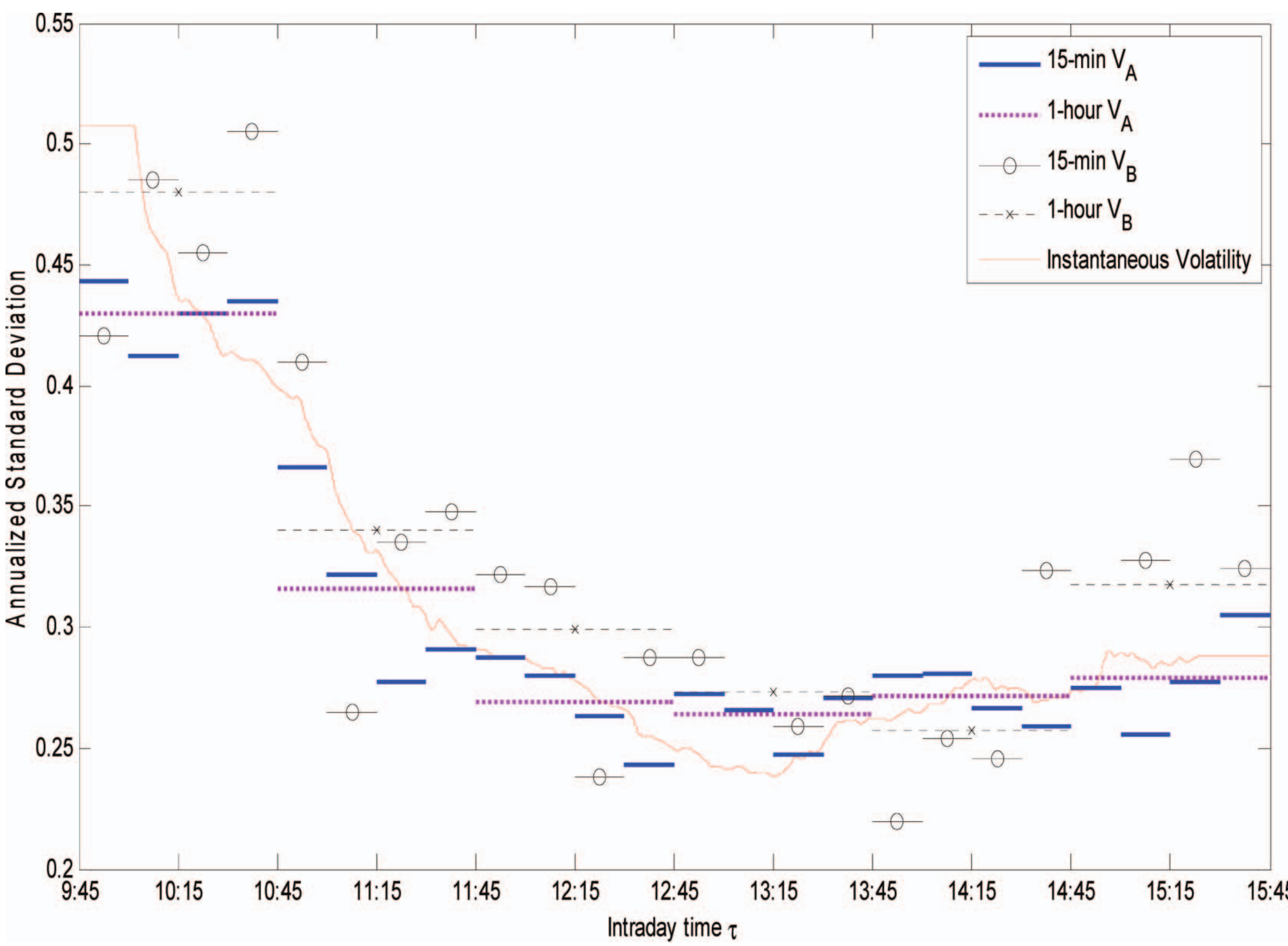

Figure 2. A sample path of intraday volatility and its estimates. The online version of this figure is in color.

volatility using V_A and V_B are presented. It can be seen that both estimates track the true volatility path and exhibit an intraday volatility smile. V_B, however, has clearly larger fluctuations than V_A. This demonstrates the advantage of the ACD-ICV method over the RV method in estimating integrated volatility over ultra-short (intraday) intervals. We will pursue the investigation of this issue in the next section in a MC study.

4. FORECAST PERFORMANCE AND INTRADAY VOLATILITY ESTIMATION

We now consider the performance of the out-of-sample one-day ahead forecast of daily volatility using the ACD-ICV and RV estimates, following the MC design by Aït-Sahalia and Mancini (2008). We generate $61 \times 23{,}400$ second-by-second (61 days) stochastic volatility and stock-price data using the stochastic volatility models (Heston diffusion model and LV model). The first 60 days of data are used to estimate daily volatility using the ACD-ICV and RV methods. We fit AR(1) models to the time series of daily volatility estimates and use these models to forecast the volatility of the 61st day. We run the MC experiment with $M = 1000$ replications and follow Aït-Sahalia and Mancini (2008) to assess the performance of the volatility forecasts by running the following regressions

$$y_j = b_0 + b_1 x_{1j} + b_2 x_{2j}, \qquad j = 1, \ldots, M, \tag{17}$$

where y_j is the integrated volatility of the 61st day, x_{1j} is the one-day ahead forecast of the volatility using V_A, and x_{2j} is the one-day ahead forecast of the volatility using the RV method. As V_B and V_K are found to have good performance in estimation,

Table 4. Performance of out-of-sample one-day ahead forecasts of daily volatility

Regressor(s)	b_0	b_1	b_2	R^2	b_0	b_1	b_2	R^2
				Panel A: NSR = 0.6				
		Heston diffusion model				Heston diffusion model with volatility jumps		
V_A	−0.005 (0.003)	1.010 (0.010)		0.903	0.007 (0.004)	1.001 (0.009)		0.925
V_B	−0.008 (0.003)		1.011 (0.011)	0.894	0.010 (0.004)		0.994 (0.010)	0.906
V_K	−0.017 (0.004)		1.045 (0.014)	0.841	0.010 (0.005)		1.004 (0.012)	0.871
$V_A + V_B$	−0.006 (0.002)	0.433 (0.040)	0.586 (0.040)	0.913	0.008 (0.003)	0.765 (0.035)	0.245 (0.035)	0.950
$V_A + V_K$	−0.007 (0.002)	0.513 (0.037)	0.509 (0.038)	0.913	0.009 (0.003)	0.845 (0.029)	0.164 (0.030)	0.949
		LV Model				LV Model with price jumps		
V_A	0.008 (0.001)	0.960 (0.007)		0.956	0.005 (0.001)	0.974 (0.006)		0.960
V_B	0.013 (0.002)		0.932 (0.007)	0.946	0.009 (0.002)		0.954 (0.007)	0.950
V_K	0.024 (0.002)		0.890 (0.010)	0.893	0.021 (0.002)		0.900 (0.010)	0.891
$V_A + V_B$	0.003 (0.001)	0.795 (0.018)	0.199 (0.018)	0.985	0.000 (0.001)	0.774 (0.018)	0.231 (0.018)	0.986
$V_A + V_K$	0.003 (0.001)	0.826 (0.016)	0.165 (0.016)	0.985	0.000 (0.001)	0.791 (0.015)	0.213 (0.015)	0.986
				Panel B: NSR = 1.0				
		Heston diffusion model				Heston diffusion model with volatility jumps		
V_A	−0.017 (0.003)	1.053 (0.011)		0.899	0.005 (0.003)	1.002 (0.009)		0.931
V_B	−0.021 (0.003)		1.058 (0.012)	0.886	0.005 (0.004)		1.001 (0.010)	0.910
V_K	−0.033 (0.004)		1.107 (0.016)	0.835	0.010 (0.005)		1.002 (0.012)	0.867
$V_A + V_B$	−0.018 (0.003)	0.390 (0.041)	0.670 (0.041)	0.908	0.005 (0.003)	0.674 (0.033)	0.335 (0.033)	0.951
$V_A + V_K$	−0.020 (0.003)	0.502 (0.038)	0.562 (0.039)	0.906	0.005 (0.003)	0.785 (0.028)	0.225 (0.029)	0.951
		LV Model				LV Model with price jumps		
V_A	0.008 (0.001)	0.957 (0.006)		0.963	0.008 (0.001)	0.955 (0.006)		0.962
V_B	0.011 (0.002)		0.939 (0.007)	0.949	0.007 (0.002)		0.953 (0.007)	0.950
V_K	0.018 (0.002)		0.912 (0.009)	0.905	0.023 (0.002)		0.890 (0.010)	0.895
$V_A + V_B$	0.003 (0.001)	0.765 (0.019)	0.213 (0.018)	0.986	0.003 (0.001)	0.715 (0.020)	0.263 (0.019)	0.984
$V_A + V_K$	0.003 (0.001)	0.817 (0.016)	0.159 (0.016)	0.986	0.002 (0.001)	0.759 (0.017)	0.222 (0.016)	0.984
				Panel C: NSR = 1.5				
		Heston diffusion model				Heston diffusion model with volatility jumps		
V_A	−0.014 (0.003)	1.031 (0.011)		0.902	0.006 (0.003)	0.997 (0.008)		0.936
V_B	−0.023 (0.003)		1.047 (0.011)	0.895	0.003 (0.004)		0.997 (0.009)	0.928
V_K	−0.026 (0.004)		1.073 (0.016)	0.825	0.013 (0.005)		0.995 (0.012)	0.873
$V_A + V_B$	−0.015 (0.003)	0.261 (0.041)	0.772 (0.042)	0.906	0.001 (0.003)	0.598 (0.035)	0.410 (0.035)	0.950
$V_A + V_K$	−0.022 (0.003)	0.341 (0.039)	0.704 (0.041)	0.902	−0.000 (0.003)	0.658 (0.032)	0.352 (0.032)	0.950
		LV Model				LV Model with price jumps		
V_A	0.007 (0.001)	0.949 (0.005)		0.970	0.006 (0.001)	0.953 (0.005)		0.969
V_B	0.008 (0.001)		0.934 (0.006)	0.966	0.006 (0.001)		0.946 (0.005)	0.968
V_K	0.020 (0.002)		0.893 (0.009)	0.910	0.021 (0.002)		0.889 (0.009)	0.902
$V_A + V_B$	0.004 (0.001)	0.694 (0.021)	0.260 (0.021)	0.986	0.003 (0.001)	0.657 (0.020)	0.301 (0.020)	0.985
$V_A + V_K$	0.004 (0.001)	0.688 (0.017)	0.263 (0.017)	0.986	0.003 (0.001)	0.645 (0.018)	0.312 (0.018)	0.985

NOTES: The results are based on 1000 Monte Carlo replications of 61-day daily volatility estimates. The first 60 days of data are used to estimate the AR(1) model for the daily volatility, and the 61st day is used for forecasting. The parameters b_0, b_1, b_2 are defined in Equation (17). Values in parentheses are standard errors. All results are for models with stock prices following BSM with white noise.

Table 5. Results for ultra-high-frequency intraday volatility estimation

		Volatility model											
		Heston			Heston with volatility jumps			LV			LV with price jumps		
NSR	Method	ME	SE	RMSE	ME	SE	RMSE	ME	SE	RMSE	ME	SE	RMSE
Panel A: Estimates over 15-min intervals, 9:45–15:45													
0.6	V_A	−0.689	1.686	1.822	−1.064	2.139	2.389	−0.723	1.351	1.533	−0.726	1.353	1.535
	V_B	−0.051	5.462	5.462	−0.212	7.726	7.729	−0.051	5.893	5.893	−0.053	5.881	5.881
	V_K	−1.411	9.670	9.772	−3.631	11.860	12.403	−0.740	7.247	7.284	−0.744	7.229	7.267
1.0	V_A	−0.230	2.008	2.021	−0.575	2.432	2.499	−0.241	1.648	1.666	−0.248	1.626	1.645
	V_B	0.048	5.578	5.578	−0.093	7.347	7.348	0.050	5.868	5.869	0.048	5.856	5.856
	V_K	−1.950	9.743	9.937	−1.798	11.599	11.737	−0.599	7.260	7.285	−0.601	7.246	7.271
1.5	V_A	0.439	2.385	2.425	−0.038	2.728	2.728	0.460	1.876	1.931	0.415	1.852	1.898
	V_B	0.281	5.425	5.432	0.040	7.489	7.489	0.257	5.815	5.821	0.225	5.821	5.826
	V_K	−1.065	9.601	9.660	−2.401	11.596	11.842	−0.325	7.289	7.297	−0.359	7.248	7.257
Panel B: Estimates over 30-min intervals, 9:45–15:45													
0.6	V_A	−0.689	1.508	1.658	−1.064	1.895	2.173	−0.723	1.130	1.342	−0.726	1.127	1.340
	V_B	0.133	4.589	4.591	0.032	6.450	6.450	0.152	4.927	4.929	0.141	4.908	4.910
	V_K	−1.022	8.547	8.608	−3.097	10.124	10.587	−0.294	5.113	5.122	−0.299	5.099	5.108
1.0	V_A	−0.230	1.697	1.712	−0.575	2.046	2.125	−0.241	1.319	1.341	−0.248	1.290	1.314
	V_B	0.205	4.683	4.687	0.118	6.163	6.164	0.226	4.911	4.916	0.215	4.892	4.897
	V_K	−1.561	8.621	8.761	−1.264	9.809	9.890	−0.153	5.123	5.126	−0.156	5.112	5.114
1.5	V_A	0.439	1.925	1.974	−0.038	2.242	2.242	0.460	1.464	1.535	0.415	1.438	1.497
	V_B	0.381	4.568	4.584	0.210	6.272	6.276	0.386	4.875	4.891	0.344	4.863	4.875
	V_K	−0.675	8.443	8.470	−1.867	9.792	9.969	0.121	5.144	5.146	0.082	5.114	5.115
Panel C: Estimates over 60-min intervals, 9:45–15:45													
0.6	V_A	−0.689	1.404	1.564	−1.064	1.757	2.054	−0.723	1.003	1.237	−0.726	0.998	1.234
	V_B	0.133	3.311	3.314	0.032	4.630	4.630	0.152	3.502	3.505	0.141	3.502	3.505
	V_K	−0.830	7.925	7.968	−2.830	9.122	9.551	−0.077	3.635	3.635	−0.081	3.604	3.605
1.0	V_A	−0.230	1.525	1.542	−0.575	1.832	1.920	−0.241	1.133	1.158	−0.248	1.098	1.126
	V_B	0.205	3.394	3.400	0.118	4.414	4.415	0.226	3.490	3.498	0.215	3.491	3.498
	V_K	−1.369	8.002	8.118	−0.997	8.766	8.822	0.064	3.641	3.641	0.062	3.613	3.614
1.5	V_A	0.439	1.666	1.723	−0.038	1.976	1.977	0.460	1.229	1.312	0.415	1.198	1.268
	V_B	0.381	3.308	3.330	0.210	4.503	4.508	0.386	3.481	3.502	0.344	3.475	3.493
	V_K	−0.483	7.800	7.815	−1.600	8.739	8.884	0.338	3.655	3.670	0.300	3.612	3.624

NOTES: ME = mean error; SE = standard error (standard deviation of MC samples); RMSE = root mean-squared error. The results are based on 1000 Monte Carlo replications of 60-day intraday (15-min, 30-min, or 60-min) volatility estimates. Intraday volatility is computed as annualized standard deviation in percentage. Stock prices follow BSM with white noise.

they are considered for the forecasting study. We only present the results for the case of stock prices following a BSM with white noise, as results for other pricing errors are similar. The results are summarized in Table 4. It can be seen that when a single forecast is considered V_A provides the best performance, giving the highest R^2 in the evaluation regression for all models. The performance of V_B comes in the second, followed by V_K. Using two forecasts does not improve the performance in terms of the incremental R^2.

We further consider the performance of the volatility estimates for ultra-high-frequency (intraday) integrated volatility. We divide the interval 9:45 through 15:45 into subintervals of 15 min, 30 min, and 60 min. Integrated stochastic volatility over each subinterval is computed and compared against the estimates V_A, V_B, and V_K. The results based on 1000 MC replications over 60 days are summarized in Table 5. It can be seen that V_A provides estimates with the lowest RMSE, followed by V_B and then V_K, and this ranking is consistent over all models. While V_B provides the lowest ME in most cases it has a higher SE than V_A, thus resulting in a higher RMSE. Indeed, the RMSE of V_A is less than half of those of V_B and V_K in all cases. As expected, V_B and V_K improve in giving lower RMSE when the intraday interval increases. For intraday intervals of 15 min, the RMSE of V_B and V_K are generally larger than 5 percentage points, with some cases exceeding 10 percentage points. In contrast, the RMSE of V_A are less than 3 percentage points for all cases. Overall, the superior performance of the ACD-ICV method in estimating intraday volatility is very clear.

The stochastic volatilities generated in the MC experiments above do not have intraday periodicity. As a robustness check we consider the case when there is an intraday periodicity function superimposed onto the stochastic volatility process. The results are reported in Table 6. Although the RMSE of the estimators generally increases (except for V_K), the RMSE of V_A remains the lowest. The performance of V_K, however, is now better than that of V_B. In the MC experiments without imposing intraday periodicity, the true volatility process is relatively flat. In this case, V_B with subsampling increases the estimation efficiency significantly. On the other hand, due to the small number of

Table 6. Results for ultra-high-frequency intraday volatility estimation with intraday periodicity

		Volatility model											
		Heston			Heston with jumps			LV			LV with jumps		
NSR	Method	ME	SE	RMSE	ME	SE	RMSE	ME	SE	RMSE	ME	SE	RMSE
Panel A: Estimates over 15-min intervals, 9:45–15:45													
0.6	V_A	0.346	3.813	3.829	0.279	5.493	5.500	0.328	4.020	4.033	0.319	4.036	4.049
	V_B	0.204	9.697	9.700	0.190	14.779	14.780	0.200	10.882	10.884	0.204	10.875	10.877
	V_K	−0.954	7.927	7.984	−1.432	11.257	11.347	−1.086	8.713	8.780	−1.075	8.724	8.790
1.0	V_A	0.514	4.505	4.578	0.512	5.784	5.828	0.437	4.546	4.623	0.463	4.580	4.660
	V_B	0.357	10.052	10.058	0.236	13.928	13.930	0.312	10.858	10.862	0.314	10.850	10.855
	V_K	−0.867	8.240	8.286	−1.298	10.716	10.794	−0.986	8.726	8.782	−0.974	8.741	8.795
1.5	V_A	0.830	4.523	4.674	0.810	6.033	6.073	0.912	4.667	4.706	0.881	4.604	4.737
	V_B	0.592	10.001	10.019	0.415	13.890	13.896	0.526	10.810	10.823	0.528	10.802	10.815
	V_K	−0.647	8.272	8.297	−1.132	10.740	10.799	−0.788	8.757	8.792	−0.776	8.773	8.808
Panel B: Estimates over 30-min intervals, 9:45–15:45													
0.6	V_A	0.346	3.437	3.455	0.279	4.983	4.991	0.328	3.619	3.634	0.319	3.634	3.648
	V_B	0.452	9.680	9.690	0.534	14.797	14.807	0.476	10.890	10.901	0.471	10.877	10.887
	V_K	−0.361	5.564	5.576	−0.579	7.858	7.879	−0.419	6.089	6.103	−0.405	6.088	6.102
1.0	V_A	0.514	4.055	4.135	0.512	5.194	5.243	0.437	4.105	4.190	0.463	4.142	4.230
	V_B	0.588	10.058	10.075	0.538	13.954	13.964	0.557	10.873	10.888	0.552	10.860	10.874
	V_K	−0.250	5.784	5.790	−0.491	7.497	7.513	−0.320	6.098	6.106	−0.303	6.098	6.106
1.5	V_A	0.830	4.020	4.101	0.810	5.413	5.500	0.912	4.189	4.354	0.881	4.118	4.276
	V_B	0.760	10.023	10.052	0.669	13.928	13.944	0.714	10.841	10.865	0.709	10.827	10.850
	V_K	−0.030	5.807	5.807	−0.324	7.513	7.520	−0.122	6.119	6.120	−0.104	6.119	6.120
Panel C: Estimates over 60-min intervals, 9:45–15:45													
0.6	V_A	0.346	2.962	2.982	0.279	4.340	4.349	0.328	3.105	3.123	0.319	3.116	3.133
	V_B	0.452	8.526	8.538	0.534	13.256	13.267	0.476	9.661	9.673	0.471	9.654	9.665
	V_K	−0.038	3.948	3.949	−0.126	5.596	5.598	−0.066	4.340	4.341	−0.050	4.337	4.337
1.0	V_A	0.514	3.508	3.602	0.512	4.474	4.530	0.437	3.558	3.656	0.463	3.593	3.696
	V_B	0.588	8.890	8.909	0.538	12.489	12.501	0.557	9.647	9.663	0.552	9.641	9.656
	V_K	0.085	4.099	4.100	−0.057	5.336	5.336	0.033	4.343	4.343	0.052	4.342	4.343
1.5	V_A	0.830	3.419	3.746	0.810	4.667	4.747	0.912	3.611	3.815	0.881	3.516	3.715
	V_B	0.760	8.862	8.894	0.669	12.469	12.486	0.714	9.621	9.648	0.709	9.615	9.641
	V_K	0.304	4.114	4.125	0.109	5.346	5.347	0.231	4.357	4.363	0.250	4.355	4.362

NOTES: ME = mean error; SE = standard error (standard deviation of MC samples); RMSE = root mean-squared error. The results are based on 1000 Monte Carlo replications of 60-day intraday (15-min, 30-min, and 60-min) volatility estimates. Volatility is computed as annualized standard deviation in percentage. Stock prices follow BSM with white noise. An intraday periodicity function is superimposed on the stochastic volatility process. The Heston model with jumps is the Heston diffusion model with jumps in the volatility, while the LV model with jumps is the LV model with jumps in the price.

lagged terms taken over short intervals, V_K is adversely affected by some negative values, which are forced to be zero. For the experiments with intraday periodicity, the intraday volatility paths exhibit far more variability, which reduces the efficiency of subsampling in V_B. In contrast, the relative magnitude of the lagged terms in V_K is smaller, thus alleviating the problem of negative values in the estimates.

5. EMPIRICAL ESTIMATES USING NYSE DATA

We consider the use of the ACD-ICV method for the estimation of daily volatility with empirical data from the NYSE. Our data consist of 30 stocks, with 10 stocks each classified as large, medium, and small (all are component stocks of the S&P500), sampled over three different periods in 2006 and 2007, ranging from 25 days to 58 days in each period. Period 1 is a sideways market, Period 2 is an upward market, and Period 3 is a downward market. Some statistics of the sample periods are given in Table 7. To account for intraday periodicity, we estimate diurnal factors by applying a smoothing spline to the average duration over different periods of the day, and compute the mean-diurnally adjusted duration for use in calibrating the ACD models.

We estimate the daily volatility using V_A (QMLE with δ being the threshold price range) and various RV methods. To target an average price duration of 4–5 min, we set δ to 0.1% in Periods 1 and 2, and 0.2% in Period 3. The annualized return standard deviations, which are the square root of 252 multiplied by the daily variance estimates, are compared across different methods. To save space, we select the results of two stocks for presentation, with the full set of results for all stocks over the three different periods summarized in the online Appendix. Figures 3 and 4 exhibit the results for JP Morgan and Moody's.

It can be seen that V_B and V_K track each other very closely, while V_D appears to have more fluctuations, especially in Period 2. V_A frequently moderates between the RV estimates and exhibits smaller fluctuations. It is remarkable that the estimates are quite similar in some turbulent periods. For example, in Period 2,

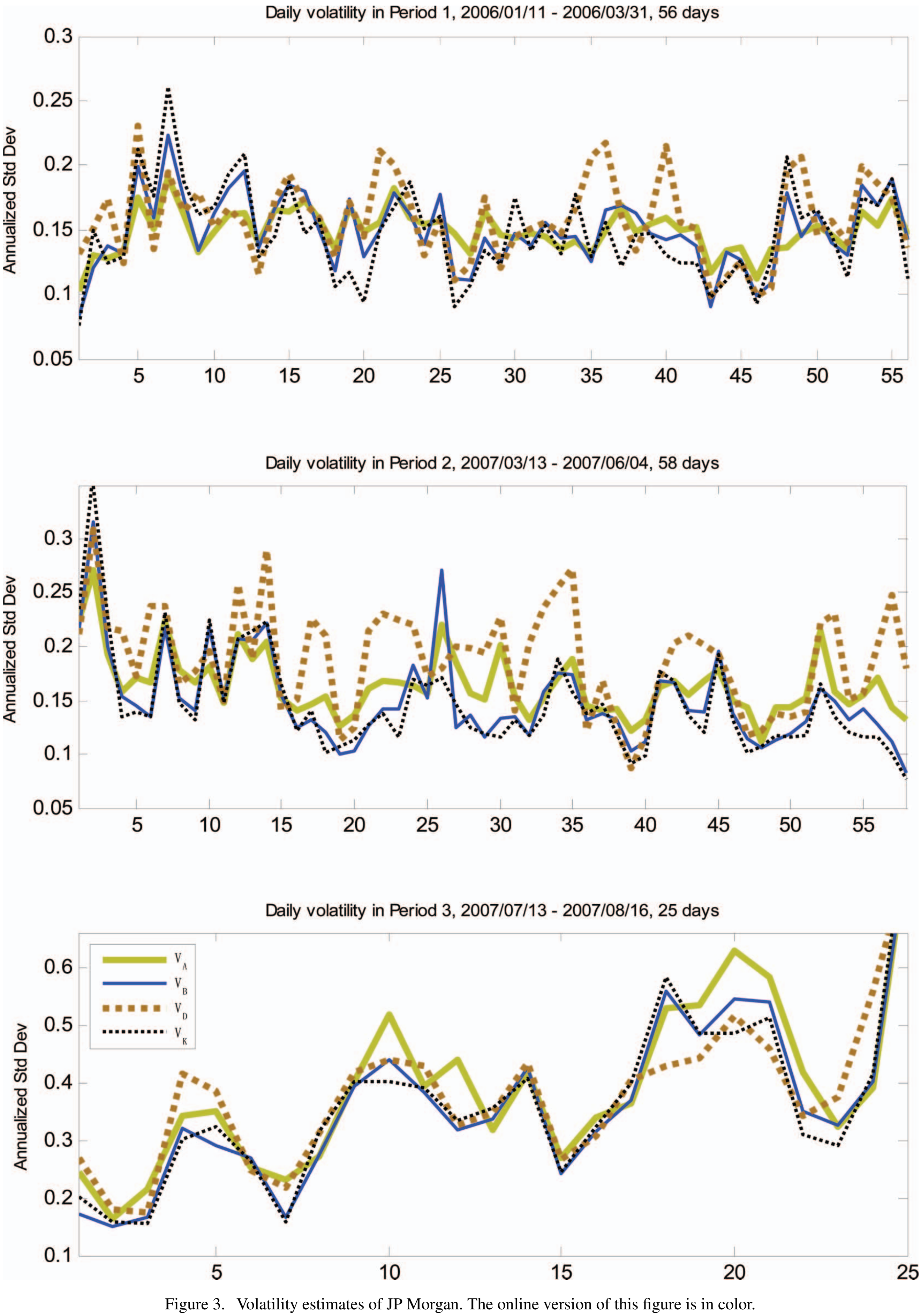

Figure 3. Volatility estimates of JP Morgan. The online version of this figure is in color.

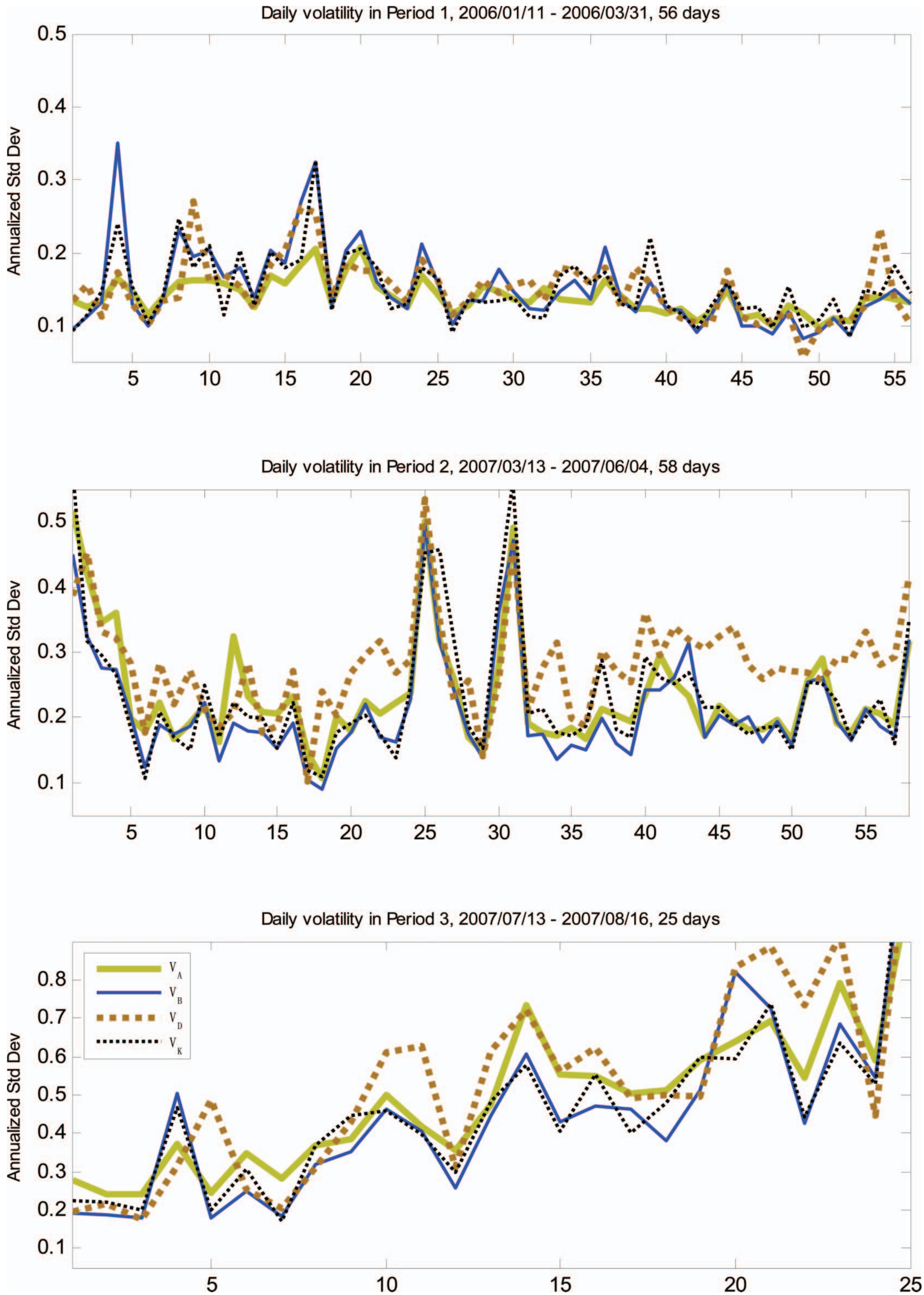

Figure 4. Volatility estimates of Moody's. The online version of this figure is in color.

Table 7. Sample period and summary of empirical data

Period	1	2	3
Dates	2006: 01/11–03/31	2007: 03/13–06/04	2007: 07/13–08/16
Number of days	56	58	25
Begin and end S&P500	1294.18–1294.87	1377.95–1539.18	1552.5–1411.27
Index return in period	0.05%	11.70%	−9.1%
Annualized standard deviation of daily index return	9.01%	9.85%	21.76%

all volatility estimates of Moody's give very similar results on several very volatile days. On the other hand, V_D appears to be quite erratic for JP Morgan in this period. In Period 3, the volatilities of both stocks clearly trend upward, with Moody's reaching over 80% toward the end of the period. Again, it is quite clear that all estimates follow the upward trend and track each other closely.

6. CONCLUSION

In this article, we propose a method to estimate high-frequency (daily) or ultra-high-frequency (intraday) volatility by integrating the instantaneous conditional return variance per unit time obtained from the ACD model, called the ACD-ICV method. Adopting the exponential-distribution assumption for the standardized duration, the ACD-ICV method can be computed easily using the QMLE of the ACD model. We compare the performance of this method against several RV methods using MC experiments. Our results show that the ACD-ICV estimates provide the smallest RMSE over a range of stochastic volatility models. Our MC results support the superior performance of the ACD-ICV method over the RV methods in estimating volatility over short intraday intervals. The accuracy of the RV estimates over short intraday intervals is clearly adversely affected by the lack of infill data.

Our evaluation of the out-of-sample one-day ahead forecast performance shows that the ACD-ICV method provides better forecasts than the RV methods. Empirical results using the data of 30 NYSE stocks show that the ACD-ICV estimates and the RV estimates generally track each other quite well, although there are larger fluctuations in the RV estimates across time. Overall, our results show that the ACD-ICV method is a useful tool for estimating high-frequency volatility.

ACKNOWLEDGMENT

The authors gratefully acknowledge research support from the Singapore Ministry of Education Tier 2 research grant T206B4301-RS. They are indebted to the referees, the associate editor, and the editor, Jonathan Wright, for their helpful and insightful comments and suggestions.

REFERENCES

Aït-Sahalia, Y., and Mancini, L. (2008), "Out of Sample Forecasts of Quadratic Variation," *Journal of Econometrics*, 147, 17–33. [536,540]

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001a), "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76. [533]

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b), "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55. (Correction: 2003, 98, 501.) [533]

Andersen, T. G., Dobrev, D., and Schaumburg, E. (2008), "Duration-Based Volatility Estimation," working paper, available online at *http://www.devsmith.umd.edu/finance/pdfs_docs/SeminarFall2008/Dobrev Dobrislav.pdf*. [533,538]

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008), "Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76, 1481–1536. [538]

Barndorff-Nielsen, O. E., and Shephard, N. (2004), "Power and Bipower Variation With Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37. [538]

Drost, F. C., and Werker, B. J. M. (2004), "Semiparametric Duration Models," *Journal of Business and Economic Statistics*, 22, 40–50. [534]

Engle, R., and Gonzalez-Rivera, G. (1991), "Semiparametric ARCH Models," *Journal of Business and Economic Statistics*, 9, 345–359. [534]

Engle R., and Russell, J. R. (1998), "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162. [533,534]

Fernandes, M., and Grammig, J. (2006), "A Family of Autoregressive Conditional Duration Models," *Journal of Econometrics*, 130, 1–23. [533,534]

Heston, S. L. (1993), "A Closed-Form Solution for Options With Stochastic Volatility With Application to Bond and Currency Options," *Review of Financial Studies*, 6, 327–343. [537]

Pacurar, M. (2008), "Autoregressive Conditional Duration Models in Finance: A Survey of the Theoretical and Empirical Literature," *Journal of Economic Surveys*, 22, 711–751. [534]

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005), "A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411. [538]