1-2010

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

Yiu Kuen Tse
Singapore Management University, yktse@smu.edu.sg

Tao Yang

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

Yiu-Kuen Tse
School of Economics, Singapore Management University

Tao Yang
School of Economics, Singapore Management University

May 2010

Abstract: 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) models. We compare the daily volatilities estimated using the ACD models against several versions of the realized volatility (RV) method, including the bipower variation realized volatility with subsampling, the realized kernel estimate and the duration-based realized volatility. The ACD volatility estimates correlate highly with and perform very well against the RV estimates. Our Monte Carlo results show that our method has lower root mean-squared error than the RV methods in most cases. A clear advantage of our method is that it can be used to estimate intraday volatilities over intervals such as an hour or 15 minutes.

JEL Codes: C410, G120

Keywords: Autoregressive Conditional Duration, Market Microstructure, Realized Volatility, Semi-parametric Method, Transaction Data

Acknowledgment: The authors gratefully acknowledge research support from a MOE Tier 2 research grant T206B4301-RS.

Corresponding Author: Yiu-Kuen Tse, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, email: `yktse@smu.edu.sg`.

1 Introduction

Since the seminal work of Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold and Labys (2001), realized volatility (RV) 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 at time t follows a diffusion process with instantaneous variance per unit time $\sigma^2(t)$, the IV of the asset return over the time interval $(0, t)$ is defined as

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

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 five minutes or higher. As an estimator of IV, RV 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, various RV methods incorporating some improvements and modifications have been proposed. These include the subsampling technique due to Zhang, Mykland and Aït-Sahalia (2005), the bipower variation method by Barndorff-Nielsen and Shephard (2004), the realized kernel method by Barndorff-Nielsen, Hansen, Lunde and Shephard (2008), and the duration-based RV method by Andersen, Dobrev and Schaumburg (2008). 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.[1]

In this paper we propose to estimate high-frequency (daily or 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 transaction price to reach or exceed a given threshold δ, called the price range. The price duration is then analyzed as a *point process.* As shown by Engle and Russell (1998), the instantaneous conditional variance per unit time of the price-duration process depends on δ and the conditional hazard rate function of the duration distribution. We model the price-duration process

[1] It should be noted that many of the theoretical asymptotic results of the RV methods depend on the Brownian semimartingale assumption. In the case of the duration-based RV method, some constants that are derived from the Brownian-motion assumption are used for the computation of the estimates.

parametrically using the autoregressive conditional duration (ACD) model of Engle and Russell (1998) and the augmented ACD (AACD) model suggested by Fernandes and Grammig (2006). The parametric formulation of the ACD and AACD models determines the dynamics of the conditional expected duration and hence the conditional hazard rate function. To estimate the volatility over a given time interval, we integrate the estimated instantaneous conditional variance per unit time over the interval.

To adopt the above approach some econometric issues have to be resolved. First, the ACD equation has to be appropriately specified, and a flexible functional form of the dynamics of the conditional expected duration is desired. Fernandes and Grammig (2006) showed that the AACD model has superior performance over its competitors, and this model will be adopted in this paper. Second, the distribution of the standardized price duration has to be appropriately specified. In a study examining different specifications of the conditional duration distribution, Bauwens, Giot, Grammig and Veredas (2004) found that none of the distributions they considered pass all the tests. To overcome this difficulty we propose to use a semiparametric (SP) method to estimate the AACD parameters, in which the conditional duration distribution is estimated using a nonparametric technique.

An important difference between the RV estimate and our parametric 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. The conditional information in our method, however, consists of information available in transaction data, which is updated as and when a trade occurs. This comparison is analogous to the stochastic volatility approach versus the conditional heteroscedastic approach in the literature of volatility modeling.

There are some advantages in our method over the RV approach. First, in the RV approach the sampling of prices is over fixed intervals and returns are computed as the differenced logarithmic prices. In contrast, in our approach the sampling of the price events is random. The (absolute) returns over the sampled intervals are the price range divided by the price at the beginning of the interval, which will be more robust to market microstructure noise and jumps. This advantage is shared by the Andersen, Dobrev and Schaumburg (2008) method in a similar context. Second, when the ACD-equation parameters are obtained, the instantaneous conditional variance can be calculated and integrated over any time interval. The viability of this computation is not dependent on the length of the interval, as, unlike the RV methods, our method does not depend on convergence based on a large number of squared-return observations. Thus, our method can be used to estimate the volatility on ultra-high frequency, such as hourly.[2] Currently, many studies use absolute return as a proxy for such ultra-high-frequency volatility. Some authors, however, found that absolute returns contain a substantial amount of measurement errors and hence do not provide a basis for reliable inference (see, e.g., Chan and Fong (2006)).

The contributions of this paper are twofold. First, we propose a new parametric method to estimate high-frequency volatility. We adopt the semiparametric (SP) method for the estimation of the ACD models, and show how the results can be used to estimate high-frequency volatility. Our Monte Carlo (MC) results demonstrate superior performance of the SP estimates over the quasi maximum likelihood estimates (QMLE). Second, we compare empirically the performance of our estimates of daily volatility versus several RV methods. Thus, apart from examining the performance of the new method, we also report a comparison of some recently developed RV methods.

The balance of this paper is as follows. In Section 2 we outline the ACD models and discuss their estimation. We conduct MC experiments to study the performance of the MLE, QMLE and SP methods. We then discuss the use of the ACD models for the estimation of high-frequency volatility. In Section 3 we review the RV methods considered in our empirical study. Section 4 reports some MC results for comparing the performance of the RV methods versus our method. The results show that our method has smaller root mean-squared errors (RMSE) than the RV estimates in almost all cases. Section 5 contains an empirical study using NYSE data, which shows that the daily volatility estimates based on the AACD model correlate highly with the RV estimates. Finally, Section 6 concludes the paper.

[2]Our focus is on the estimation of volatility over a time interval. For the estimation of spot volatility, see Bandi and Renò (2008), Kristensen (2010) and Malliavin and Mancino (2009).

2 ACD Models and High-Frequency Volatility

The ACD model was first proposed by Engle and Russell (1998) to analyze the duration of transaction data. A recent review of the literature on the ACD models and their applications to finance can be found in Pacurar (2008). Analogous to the generalized autoregressive conditional heteroscedasticity (GARCH) models, which capture the clustering of volatility, the ACD model analyzes the clustering of transaction duration. Following Engle and Russell (1998), the instantaneous conditional variance per unit time derived from the ACD model may be integrated over a time interval (e.g., 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 a high-frequency estimate of volatility.

2.1 ACD Models

Consider a sequence of times $t_0, t_1, \cdots, t_N$ with $t_0 < t_1 < \cdots < t_N$, in which t_i denotes the time of the ith transaction. Thus, $x_i = t_i - t_{i-1}$, for $i = 1, 2, \cdots, N$, are the intervals between consecutive transactions, called transaction durations. In this paper we consider the price duration, which is defined as the time interval needed to observe a cumulative change in the transacted price of at least δ. Thus, from time t_{i-1} to t_i, the transaction price changes by at least an amount δ, whether upwards or downwards. The occurrence of this incident is called a *price event*. The data for analysis in the ACD model are the durations x_i. Unlike the RV methods, which assume the transaction price follows a Brownian semimartingale (BSM) with possible contaminations due to market microstructure noise and/or price jumps, our object of analysis is the price duration x_i resulting from a price event.[3] As shown by Engle and Russell (1998), the instantaneous conditional variance of the return defined by the point process is a function of the conditional expected duration and the conditional intensity of the duration distribution. Exploiting this result, we can derive a parametric method to estimate the integrated conditional variance of the asset return over an interval.

[3] In the RV literature the efficient log-price of an asset is typically assumed to follow a diffusion process with a stochastic instantaneous variance, and RV estimates the integrated variance of the efficient price. The observed transaction price, however, is contaminated by market microstructure noise and may exhibit price jumps.

Let Φ_i denote the information set upon the transaction at time t_i.[4] We denote $\psi_i = \mathrm{E}(x_i \,|\, \Phi_{i-1})$, which is the conditional expectation of the price duration. We assume that

$$\epsilon_i = \frac{x_i}{\psi_i}, \qquad i = 1, \cdots, N, \tag{2}$$

are a sequence of i.i.d. positive random variables with mean 1 and density function $f(\cdot)$. Thus, the hazard function of ϵ_i is

$$\lambda(\cdot) = \frac{f(\cdot)}{S(\cdot)}, \tag{3}$$

where $S(\cdot)$ is the survival function of ϵ_i. Assuming ψ_i to be known given Φ_{i-1}, the conditional hazard function (also called the conditional intensity) of x_i, denoted by $\lambda_x(x_i \,|\, \Phi_{i-1})$, is

$$\begin{aligned} \lambda_x(x_i \,|\, \Phi_{i-1}) &= \lambda\left(\frac{x_i}{\psi_i}\right)\frac{1}{\psi_i} \\ &= \lambda\left(\frac{t_i - t_{i-1}}{\psi_i}\right)\frac{1}{\psi_i}, \end{aligned} \tag{4}$$

which is related to the *base* hazard function $\lambda(\cdot)$.

To model the conditional duration ψ_i, Engle and Russell (1998) proposed the ACD(p, q) model defined by

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}, \tag{5}$$

where the analogous relationship with the GARCH models is obvious. For model parsimony, small values of p and q are preferred. Setting $p = q = 1$, we obtain the ACD$(1, 1)$ model

$$\psi_i = \omega + \alpha x_{i-1} + \beta \psi_{i-1}, \tag{6}$$

with the restrictions α, β and $w \geq 0$, and $\alpha + \beta < 1$.

Recently, Fernandes and Grammig (2006) proposed some extensions of the ACD$(1, 1)$ model. We shall consider their AACD model, which is defined by

$$\psi_i^{\lambda} = \omega + \alpha \psi_{i-1}^{\lambda} \left[|\epsilon_{i-1} - b| + c(\epsilon_{i-1} - b)\right]^{v} + \beta \psi_{i-1}^{\lambda}. \tag{7}$$

[4] Φ_i may include transaction-related variables such as transaction duration, volume and order flow (buy- or sell-initiated orders). In this paper, however, we shall only consider past transaction durations.

The AACD model nests the ACD$(1,1)$ model as a special case and provides a more flexible model for the conditional expected duration ψ_i. 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. Various possible shapes of the shocks impact curve are illustrated by Fernandes and Grammig (2006).

The empirical study reported in 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. Indeed, the restrictions imposed by the simpler ACD$(1,1)$ model are rejected empirically. Due to its flexibility, we shall adopt the AACD model as our operating model. In subsequent discussions we refer to the ACD$(1,1)$ and AACD models generically as ACD models.[5]

2.2 Estimation of ACD Models

Given an assumed density function $f(\cdot)$, the maximum likelihood estimates (MLE) of the parameters of the ACD equation can be computed straightforwardly. A particularly simple model is the case when ϵ_i are assumed to be standard exponential. Under this assumption the hazard function is constant and does not vary with the duration. Furthermore, the MLE computed using the exponential assumption is consistent (provided the conditional expected duration equation is correctly specified), regardless of the true distribution of the error (see Drost and Werker (2004)). The MLE computed based on the exponential working assumption is called the QMLE method.[6] However, under the exponential distribution assumption, the conditional intensity is constant. When this assumption is used, misspecification errors in the conditional intensity may induce errors in the integrated conditional variance. To resolve this issue, we propose to use the SP method, in which the conditional duration distribution is estimated

[5] Some discussions of tests for and selection of ACD models can be found in Fernandes and Grammig (2005). An alternative ACD model can be found in Grammig and Maurer (2000).

[6] Following the work of Lee and Hansen (1994), Engle and Russell (1998) showed that the MLE of the ACD(1, 1) model is consistent and asymptotically normal. For the AACD model, Fernandes and Grammig (2006) established sufficient conditions for the existence of higher-order moments, stationarity and β-mixing. While a rigorous proof of the asymptotic normality of the MLE of the AACD model has not been established, Fernandes and Grammig (2006) used the asymptotic normal approximation and computed the standard errors based on the outer-product of the gradient.

nonparametrically.

Engle and Gonzalez-Rivera (1991) introduced the use of the SP method for the ARCH model. They showed that the SP estimator is consistent and more efficient than the QMLE, which assumes normal errors. Drost and Werker (2004) discussed the conditions under which the QMLE of the ACD model is consistent. They also derived conditions under which the SP method is adaptive and efficient. Although the SP method is only efficient under very restrictive conditions, there may yet be improvements in efficiency over the QMLE. Furthermore, the empirical distribution of the conditional duration distribution is required for the computation of the conditional hazard function.

To compute the SP estimates of the ACD model, we perform the computation as follows. First, we estimate the ACD model using the QMLE method (i.e., assuming ϵ_i are i.i.d. standard exponential). We denote the parameter vector of the ACD model by θ and its QMLE by $\hat{\theta}$. Next, we calculate the estimated conditional expected duration $\hat{\psi}_i$ using $\hat{\theta}$ and the standardized duration $\hat{\epsilon}_i$ as

$$\hat{\epsilon}_i = \frac{x_i}{\hat{\psi}_i}. \tag{8}$$

The unknown density function $f(\cdot)$ of ϵ_i is then estimated using $\hat{\epsilon}_i$ by the discrete maximum penalized likelihood estimation (DMPLE) technique (see Tapia and Thompson (1978) for the details), which was also used by Engle and Gonzalez-Rivera (1991).[7] The empirical density function of $\hat{\epsilon}_i$ is denoted by $\hat{f}(\cdot)$ and the parameters of the ACD equation are then estimated again assuming ϵ_i to be i.i.d. with density function $\hat{f}(\cdot)$, resulting in the SP estimates of the ACD model, denoted by θ^*.

2.3 Some Monte Carlo Results for the Estimation of ACD Models

To examine the performance of the estimators of θ in the ACD models, we conduct a MC experiment. We consider both the ACD$(1,1)$ model and the AACD model. For the conditional duration distribution, we consider the Weibull distribution and the Burr distribution. The Weibull distribution has density

[7] The sample of estimated standardized durations $\hat{\epsilon}_i$ are scaled to have unit sample mean before the application of the DMPLE procedure. Apart from the DMPLE method, an alternative is to use the gamma kernel method (see Chen (2000)). We find, however, that the gamma kernel method is generally computationally more costly and is inferior to the DMPLE method for its relative instability. Thus, the DMPLE method is adopted in the computation of the SP estimates in this paper.

function

$$f(\epsilon) = \frac{\kappa}{\mu}\left(\frac{\epsilon}{\mu}\right)^{\kappa-1} e^{-\left(\frac{\epsilon}{\mu}\right)^{\kappa}}, \qquad \epsilon > 0, \tag{9}$$

with parameters μ and $\kappa > 0$. We let $\kappa = 2$ and $\mu = 1/\Gamma(1.5)$ so that the mean of ϵ is unity.[8] In the MLE computation we constrain the expected value of ϵ to be unity by setting $\hat{\mu}$ equal to $1/\Gamma(1 + 1/\hat{\kappa})$.

The density function of the Burr distribution is

$$f(\epsilon) = \mu\kappa\epsilon^{\kappa-1}(1 + \mu\gamma\epsilon^{\kappa})^{-\frac{1}{\gamma}-1}, \qquad \epsilon > 0, \tag{10}$$

with positive parameters μ, κ and γ. We let $\kappa = 2$, $\gamma = 0.2906$ and $\mu = 1$, so that the expected value of ϵ is unity. In the MLE computation, we set the free parameters to be κ and γ, and constrain the value of μ so that the mean of ϵ is unity.

As the results for the ACD and AACD models are similar, we report the results of the AACD model only. We consider samples of N = 1,000, 10,000 and 50,000 price observations generated from the AACD models. The MC replications are 500, 100 and 50, depending on the values of N. The following estimators are considered: the MLE (assuming the true density function), the QMLE (assuming $f(\cdot)$ is standard exponential) and the SP method (as described in Section 2.2).

In Tables 1 and 2 we report the means of the MC samples as well as the standard errors (the standard deviations of the MC samples) and root mean-squared error (RMSE) of the parameter estimates. Only results for the parameters of the AACD equation are summarized. For the Weibull errors we observe that the RMSE of all estimators decreases as N increases, suggesting the estimators are consistent. The RMSE of the SP estimator is uniformly lower than that of the QMLE, showing that the SP estimator is more efficient than the QMLE. For the Burr errors, the results show that the RMSE of all estimators generally decreases as N increases (when N increases from 10,000 to 50,000 the only exception is for the MLE value of c; when N increases from 1,000 to 10,000 the only exception is for the SP estimate of α). This again suggests the consistency of all estimators. The superiority of the SP estimator over the QMLE is quite clear, although not as strong as for the case of the Weibull errors.

[8] $\Gamma(\cdot)$ denotes the gamma function.

2.4 Estimation of High-Frequency Volatility using ACD Models

Given the information Φ_i at time t_i, the conditional intensity function defined in equation (4) characterizes the probability that the next price event will occur at time $t > t_i$. Specifically, $\lambda_x(x \,|\, \Phi_i)\Delta x$ is the probability that the next price event after time t_i occurs in the interval $(t_i + x,\, t_i + x + \Delta x)$ given the information at time t_i. The instantaneous return variance per unit time at time t is defined as

$$\sigma^2(t) = \lim_{\Delta t \to 0} \left\{ \frac{1}{\Delta t} \mathrm{Var}\left[\frac{s(t + \Delta t) - s(t)}{s(t)} \right] \right\}, \tag{11}$$

where $s(t)$ is the price at time t. From Engle and Russell (1998), we see that the instantaneous conditional variance per unit time given information Φ_i at time t_i, denoted by $\sigma^2(t \,|\, \Phi_i)$, is

$$\sigma^2(t \,|\, \Phi_i) = \left(\frac{\delta}{s_i} \right)^2 \lambda_x(x \,|\, \Phi_i), \tag{12}$$

where $x = t - t_i$, $t > t_i$ and $s_i = s(t_i)$.[9] Using equation (4), we have

$$\sigma^2(t \,|\, \Phi_i) = \left(\frac{\delta}{s_i} \right)^2 \lambda \left(\frac{t - t_i}{\psi_{i+1}} \right) \frac{1}{\psi_{i+1}}, \qquad t > t_i, \tag{13}$$

where $\psi_{i+1} = \mathrm{E}(x_{i+1} \,|\, \Phi_i)$ is the conditional expected duration of the next price event given Φ_i. Thus, the integrated conditional variance (ICV) over the interval $(t_i,\, t_{i+1})$, denoted by ICV_i, is

$$\begin{aligned} \mathrm{ICV}_i &= \int_{t_i}^{t_{i+1}} \sigma^2(t \,|\, \Phi_i)\, dt \\ &= \left(\frac{\delta}{s_i} \right)^2 \frac{1}{\psi_{i+1}} \int_{t_i}^{t_{i+1}} \lambda \left(\frac{t - t_i}{\psi_{i+1}} \right) dt. \end{aligned} \tag{14}$$

If ϵ_i are i.i.d. standard exponential variates, then $\lambda(\cdot) \equiv 1$ and we have

$$\mathrm{ICV}_i = \left(\frac{\delta}{s_i} \right)^2 \left[\frac{t_{i+1} - t_i}{\psi_{i+1}} \right]. \tag{15}$$

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

$$\mathrm{ICV} = \delta^2 \sum_{i=0}^{N-1} \frac{t_{i+1} - t_i}{\psi_{i+1} s_i^2}. \tag{16}$$

[9] Note that $s(t)$ is equal to the price recorded in the last trade prior to time t. Hence, conditional on Φ_i, $s(t)$ is a constant. On the other hand, the price difference $s(t + \Delta t) - s(t)$ is either the price range δ or zero. Thus, equation (12) can be obtained by evaluating the variance of the price difference and deleting terms smaller than Δt.

In the general case when ϵ_i takes an arbitrary distribution, the daily ICV derived from equation (14) is

$$\text{ICV} = \delta^2 \sum_{i=0}^{N-1} \frac{1}{\psi_{i+1} s_i^2} \int_{t_i}^{t_{i+1}} \lambda \left(\frac{t - t_i}{\psi_{i+1}} \right) dt. \tag{17}$$

To estimate ICV, we first use the SP estimate θ^* of the ACD-equation parameters to calculate ψ_i^*, the estimate of ψ_i. Our proposed ACD estimate of ICV (ACD-ICV), denoted by V_A, is then given by

$$V_A = \delta^2 \sum_{i=0}^{N-1} \frac{1}{\psi_{i+1}^* s_i^2} \int_{t_i}^{t_{i+1}} \hat{\lambda} \left(\frac{t - t_i}{\psi_{i+1}^*} \right) dt, \tag{18}$$

where $\hat{\lambda}(\cdot)$ is the base hazard function computed using the empirical density function $\hat{f}(\cdot)$, i.e.,

$$\hat{\lambda}(\cdot) = \frac{\hat{f}(\cdot)}{\hat{S}(\cdot)}, \tag{19}$$

and $\hat{S}(\cdot)$ is the survival function of $\hat{f}(\cdot)$.[10] If the exponential assumption is made and the QMLE $\hat{\theta}$ is adopted, the ACD-ICV estimate is simplified to

$$V_A = \delta^2 \sum_{i=0}^{N-1} \frac{t_{i+1} - t_i}{\hat{\psi}_{i+1} s_i^2}. \tag{20}$$

We conclude this section with some remarks. First, stock prices may have jumps and market frictions may produce price discreteness. Thus, price events may occur with the actual price range exceeding the threshold. We shall replace δ in equations (18) and (20) by the average price range of the sample observations conditional on the threshold being exceeded. For comparison, the quantities in these equations for which the nominal threshold is used will also be computed and denoted by V_A^*. Second, the volatility of the efficient prices is often the object of interest. As we use transaction prices to compute the price durations, measurement errors will be incurred in the estimation of the volatility. Third, we compute the squared return over each interval by $(\delta/s_i)^2$, which will have lower fluctuations than the squared returns computed as the squared differenced logarithmic prices. Indeed, if we sample the price events based on the cumulative change in the logarithmic prices exceeding the threshold, the returns over each interval will be constant. In this paper, however, we follow Engle and Russell (1998)

[10]We use the MATLAB function "quadv" to compute the integrals in equation (18). The algorithm used is a recursive adaptive Simpson quadrature. This procedure is applied sequentially, as the survival function is required for the computation of the hazard function. Thus, "quadv" is first applied to compute the survival function and then calculate the integrals of the hazard function.

and Engle (2000) to model durations of prices. Fourth, in our sampling scheme the price events are sampled more frequently when the market is more active and prices are more volatile. In contrast, the RV methods typically sample prices at equal time intervals.

3 Review of RV Methods

Since the first appearance of the basic RV method in the literature many enhanced methods have been proposed, with the purpose of overcoming the contamination due to the market microstructure noise (see Aït-Sahalia, Mykland and Zhang (2009)) and jumps (see Aït-Sahalia (2004) and Barndorff-Nielsen and Shephard (2004)). Many of these methods are dependent on specific algorithmic choices such as the sampling interval, subsampling frequency, smoothing function and price range. Although knowledge has been gained in the properties of these methods, the technique that would generally be preferred in empirical applications has yet to emerge. We shall do an empirical comparison of some selected RV estimators and investigate the performance of the parametric ACD-ICV estimates against these methods.[11]

We first define some notations before summarizing various RV methods. Let $s_0, s_1, \cdots, s_N$ denote the prices of a stock at times $t_0, t_1, \cdots, t_N$, for $N+1$ prices on the day. Thus, t_i denotes the time when price s_i is observed, and $x_i = t_i - t_{i-1}$ is the duration of the trade. The definition of t_i will be specified in each method below. In line with many studies in the literature, we use the mid-quote at the time of trade t_i as s_i.[12]

The basic RV estimate, denoted by V_R, is defined as

$$V_R = \sum_{i=1}^{N} (\log s_i - \log s_{i-1})^2. \tag{21}$$

Aït-Sahalia, Mykland and Zhang (2005) discussed the optimal sampling frequency of the price data in the presence of market microstructure noise. Further results based on finite-sample properties can

[11] Empirical comparison of different RV estimates is lacking in the literature. A notable exception is the study by Andersen, Dobrev and Schaumburg (2008), who compared the V_B, V_D and V_R estimates (see the definitions to follow) using the blue-chip companies of the Dow Jones Index. They also reported some MC comparisons of the efficiency of the these RV estimators. Our results will add to this empirical literature.

[12] For the realized kernel method we use the transaction prices instead, which follows the practice of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008).

be found in Bandi and Russell (2006a, 2006b and 2008). In our empirical study, we fix the sampling interval to be 5 minutes and adopt the practice of using the mid-quote at the time of trade immediately prior to the end of each 5-min interval.

The bipower variation RV estimate, suggested by Barndorff-Nielsen and Shephard (2004), is denoted by V_B. It is computed as

$$V_B = \mu_1^2 \sum_{i=1}^{N-1} |\log s_i - \log s_{i-1}||\log s_{i+1} - \log s_i|, \tag{22}$$

where $\mu_1 = \sqrt{2}/\Gamma(1/2)$. This method is found to be robust to rare jumps in prices. Following the suggestion of Andersen, Dobrev and Schaumburg (2008), we sample the price data over 2-min intervals. In addition, we apply the subsampling method proposed by Zhang, Mykland and Aït-Sahalia (2005) to V_B using subsampling intervals of 5 seconds.

The realized kernel estimate, denoted by V_K, is proposed by Barndorff-Nielsen, Hansen, Lunde and Shephard (2008). It is computed as

$$V_K = \gamma_0 + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h + \gamma_{-h}), \tag{23}$$

where

$$\gamma_h = \sum_{i=h+1}^{N} (\log s_i - \log s_{i-1})(\log s_{i-h} - \log s_{i-h-1}), \qquad h = 0, \cdots, H, \tag{24}$$

and the non-stochastic function $k(x)$ for $x \in [0, 1]$ is a weight function. In this paper we adopt the Tukey-Hanning weighting function.

Finally, a duration-based estimate of RV, denoted by V_D, has been proposed recently by Andersen, Dobrev and Schaumburg (2008). We adopt the event of price exiting a range δ for the definition of the passage-time duration. V_D is then computed as

$$V_D = \sum_{i=0}^{N-1} \hat{\sigma}_\delta^2(t_i)(t_{i+1} - t_i), \tag{25}$$

where $\hat{\sigma}_\delta^2(t_i)$ is the local variance estimate given in Andersen, Dobrev and Schaumburg (2008).

The similarity between equations (25) and (20) should be noted. While the grid points $t_0, t_1, \cdots, t_N$ in V_D are fixed, these values are the observed price-event times in V_A. In V_D, $\hat{\sigma}_\delta^2(t_i)$ estimates the local variance, whereas in V_A the time-scaled squared return (i.e., $(\delta/s_i)^2$ scaled by the estimated expected conditional duration $\hat{\psi}_{i+1}$) estimates the instantaneous conditional variance per unit time.

4 Monte Carlo Comparison of ACD-ICV and RV Estimates

We conduct some MC experiments to compare the performance of the ACD-ICV estimates and various RV estimates. As the two approaches are based on different notions of volatility, we consider both deterministic and stochastic volatility models. We report the results of a deterministic volatility set-up in Section 4.1, followed by some results on stochastic volatility models in Section 4.2. Our focus is on the estimation of daily volatilities.[13]

4.1 Deterministic Volatility Models

We consider a deterministic volatility set-up in which the volatility function is fixed with intraday variation. Following the assumptions underlying the theoretical derivation of the RV estimates that the logarithmic stock prices follow a Brownian semimartingale (BSM), we set up an artificial data generation process along this line. We assume that the observed price $s(u)$ at time u (we now use u for time, for reasons to be made clearer later) consists of a BSM component $s_B(u)$ and a jump component $s_J(u)$. Let $\tilde{s}_B(u) = \log s_B(u)$, which is generated from the following BSM

$$\tilde{s}_B(u) = \int_0^u \mu(u)\, du + \int_0^u \sigma(u)\, dW(u), \tag{26}$$

where $\mu(u)$ is the instantaneous drift, $\sigma^2(u)$ is the instantaneous variance and $W(u)$ is a standard Brownian process. We assume $\tilde{s}_B(u)$ has no drift so that $\mu(u) \equiv 0$. To specify the time u, we denote t as the day of trade and τ as the intraday time. Thus, we write $\sigma(u) = \sigma(t, \tau)$ and let $\sigma(t, \tau) = \sigma_1(t)\sigma_2(\tau)$, so that the instantaneous volatility at time u depends on the component $\sigma_1(t)$ (representing the *average* volatility of day t) and the component $\sigma_2(\tau)$ (capturing the intraday variations). In our MC study, we consider a period of 150 days. We set $\sigma_1(t) = 20\%$ (annualized standard deviation of returns) for $t = 1$, which increases linearly with t over 50 days to reach 30%. It then remains level for 50 days, and

[13] Aït-Sahalia and Mancini (2008) considered the forecasting performance of the V_R estimates with subsampling. In our MC study we focus on the estimation performance of various methods.

after that decreases linearly to 20% over 50 days. The intraday variation function $\sigma_2(\tau)$ is estimated empirically from the data.[14]

Having determined $\sigma_1(t)$ and $\sigma_2(\tau)$, and thus $\sigma(u)$, we generate observations of $\tilde{s}_B(u)$ by the equation

$$\tilde{s}_B(u + \Delta u) = \tilde{s}_B(u) + \sigma(u)\varepsilon, \tag{27}$$

where $\varepsilon \sim N(0, 1)$ and are independent for different values of u. We take Δu to be one second and use equation (27) to compute a second-by-second series of observations of $s_B(u) = \exp(\tilde{s}_B(u))$, with the starting price of \$65. We further add to the series $s_B(u)$ a jump component $s_J(u)$, which is assumed to follow a Poisson process with a mean of 0.4 per five minutes. When a jump occurs, it takes value of –\$0.05, –\$0.03, \$0.03 and \$0.05 with probabilities of 0.25 each. We also consider a jump component with a higher jump frequency of 2.72 per five minutes and jump sizes of –\$0.02, –\$0.01, \$0.01 and \$0.02 with equal probabilities. Finally, we consider a price process consisting of a BSM and a white noise. In sum, we consider four experiments. Experiment 1 has a pure BSM price process, Experiment 2 consists of a BSM process plus infrequent jumps of large sizes, Experiment 3 uses a BSM with more frequent jumps and smaller jump sizes, and Experiment 4 has a BSM process with white noise.

We calculate ICV using the AACD model with SP estimates for a price range of δ = \$0.08, and compare it against the RV methods. Figure 1 presents the daily volatility estimates of one MC sample of Experiments 1 and 2.[15] It can be seen that the V_A estimates track the true daily volatilities quite closely. The RV estimates also follow the volatility trends, albeit apparently with larger variations. We further simulate 50 samples of 150-day data and estimate the daily volatilities.[16] The mean error (ME) and RMSE of the volatity estimates are summarized in Table 3. It can be seen that V_A performs the best, with a RMSE of less than 0.009 for all experiments. In contrast, V_R have the largest RMSE, which is more than double that of V_A in all experiments. Among the RV estimates, V_K produces the best results and is marginally better than V_B, although its RMSE is still about 50% higher than that of

[14] See Figure 2 for a plot of the intraday variation function.

[15] All volatilities are expressed as annualized standard deviation of returns.

[16] As the SP estimation of the AACD model is very computer intensive, the MC sample size is maintained to be small. Increasing the sample size for some of the models showed that the results are qualitatively similar. Larger MC sample size is more feasible for QMLE estimates, and will be used in Section 5.2.

V_A. For comparison, we also consider V_A^*, which uses the SP method with the nominal threshold, and V_A', which uses the QMLE method with the conditional price range. It can be seen that V_A' performs better than the RV estimates, while V_A^* ranks behind V_A'. The results suggest that the conditional price range should be adopted, and the use of the QMLE estimate performs quite well albeit inferior to the SP estimate.

We further estimate intraday ICV using V_A. Figure 2 presents the results of the intraday ICV estimates on days 1, 75 and 150 of a simulated sample based on Experiment 2 (BSM with infrequent large jumps). We consider ICV over two time intervals: 15 minutes and one hour, beginning from 9:45 and ending at 15:45. The graphs show that V_A performs very well, in particular, for $t = 1$ and 75. In contrast, the absolute returns over the intervals, which are often used as proxies for intraday volatilities, are unable to trace the true instantaneous volatility accurately. Indeed, many of the absolute return plots are outside the range of the graphs and are not shown.

4.2 Stochastic Volatility Models

We now consider the set-up when the volatility function is stochastic. We follow closely the experiments designed by Aït-Sahalia and Mancini (2008). For completeness, we describe briefly the models adopted. Further details can be found in Aït-Sahalia and Mancini (2008).

4.2.1 Heston Model

We assume the following price generation process where the stochastic volatility process follows Heston's (1993) model

$$d \log s(t) = \left(\mu - \frac{\sigma^2(t)}{2}\right) dt + \sigma(t)\, dW_1(t), \tag{28}$$

$$d\sigma^2(t) = \kappa\left(\alpha - \sigma^2(t)\right) dt + \gamma\sigma(t)\, dW_2(t). \tag{29}$$

The parameters are set as follows: $\mu = 0.05$, $\kappa = 5$, $\alpha = 0.04$ and $\gamma = 0.5$. The correlation coefficient between the two Brownian motions $W_1(t)$ and $W_2(t)$, ρ, is -0.5. We generate 100-day data second by second with initial value of $\sigma(t)$ equalling 0.3. We also consider the inclusion of a jump component into the volatility process. Thus, equation (29) is modified as follows:

$$d\sigma^2(t) = \kappa \left(\alpha - \sigma^2(t)\right) dt + \gamma\sigma(t)\, dW_2(t) + J(t)\, dq(t), \tag{30}$$

where $q(t)$ is a Poisson process with intensity $\lambda = 2/(6.5 \times 3600)$ and $J(t)$ is assumed to be exponentially distributed with intensity $\xi = 0.0007$.

4.2.2 Log-volatility (LV) Model

Let $\mathrm{IV}_t^{(d)}$ denote the integrated daily volatility so that $\mathrm{IV}_t^{(d)} = \int_{t-1}^{t} \sigma^2(u)\, du$ and $l(t) = \frac{1}{2}\log(\mathrm{IV}_t^{(d)})$, where d denotes daily measure. We let $l(t)$ follow an AR(5) process so that

$$l(t) = \phi_0 + \sum_{i=1}^{5} \phi_i\, l(t-i) + u(t), \tag{31}$$

where $u(t)$ is a white noise. The log-return $r(t)$ is computed as $r(t) = [\mathrm{IV}_t^{(d)}]^{\frac{1}{2}}\, z(t)$, where $z(t)$ is normal with mean zero and standard deviation $1/(3600 \times 6.5 \times 252)^{\frac{1}{2}}$. The parameters are set as: $\phi_0 = -0.0161$, $\phi_1 = -0.35$, $\phi_2 = 0.25$, $\phi_3 = 0.20$, $\phi_4 = 0.10$, $\phi_5 = 0.09$, and the standard deviation of $u(t)$ is 0.02.

4.2.3 The Noise Structure

Given the logarithmic efficient price $\log s(t)$, 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 i.i.d. normal variates. Second, we consider the case where $\varepsilon(t)$ have serial dependence so that it follows an AR(1) process with a correlation coefficient of -0.2. Third, we generate serially uncorrelated $\varepsilon(t)$, which are correlated with the latent return process so that $\mathrm{Corr}\{\varepsilon(t), \log s(t) - \log s(t-\Delta)\} = -0.2$. Fourth, we consider noises that are autocorrelated as well as correlated with the latent price process. The correlation coefficients are as given in Case 2 and 3 aforementioned. Finally, we also consider the case where there is a jump in the price process for the LV model, with 0.4 jump per 5 minutes 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% to 30%. We define the noise-to-signal (NSR) ratio as $\mathrm{NSR} = [\mathrm{Var}\{\varepsilon(t)\}/\mathrm{Var}\{\sigma(t)\}]^{\frac{1}{2}}$, which is set equal to 0.25, 0.4 and 0.6. The variance of $\varepsilon(t)$ is then determined given the value of NSR and the volatility process.

4.3 The Results

As the SP estimation is computationally very intensive, we conduct our MC study of the performance of V_A and V_A^* using the QMLE estimates of the AACD model. This is encouraged by the results in Section 4.1, which suggest the good performance of the QMLE for the estimation of ICV. Figure 3 presents two samples of the volatility paths for the Heston model and LV model, and various volatility estimates. It can be seen that V_A performs very well against the RV estimates. It tracks the volatility function very well with relatively small estimation errors.

Tables 4 through 6 report the performance of various estimation methods for different volatility models and noise structures, each based on MC experiments with 1,000 replications. It can be seen that the RMSE of V_A increases as NSR increases. In contrast, the RV estimates hardly vary with NSR and the structure of the noise. For NSR of 0.25 and 0.4, V_A provides the lowest RMSE among all estimates. It also gives the lowest RMSE for the Heston model with jump when NSR is 0.6. Among all RV estimates, V_B has the best performance, followed by V_K, V_D and then V_R. When NSR is 0.6, V_B gives the lowest RMSE for the Heston model. For the RV estimates, volatility jumps cause the RMSE to increase, while price jumps hardly have any effect on the RMSE. As expected, V_A^* has lower ME than V_A (by construction, $V_A < V_A^*$), and its performance seems to rank behind V_K but ahead of V_R.[17] This suggests that the use of the sample price range conditional upon the threshold being exceeded should be adopted for the ACD-ICV estimates.

5 Empirical Comparison of ACD-ICV and RV Methods

We now consider the use of the ACD-ICV method for the estimation of daily volatility on empirical data, and compare the results against the RV methods. While the ACD-ICV and RV methods are based on different theoretical set-ups, their objectives of estimating daily volatility are similar.

[17] An exception occurs for the LV models with NSR = 0.60, in which case V_A^* provides the lowest RMSE with V_A being the second best.

5.1 The Data

The data we used are extracted and compiled from the NYSE Trade and Quote (TAQ) Database provided through the Wharton Research Data Services. We downloaded the following data from the Consolidated Trade (CT) file: date, trading time, price and number of shares traded. To select the stocks, we rank the 500 component stocks of the S&P500 Index by market capitalization as of September 2008. We then divide the stocks into three groups and select the largest 10 stocks from each group. We call these the large stocks, medium stocks and small stocks. The selected companies and their stock codes are summarized in Table 7.

We consider three periods of trades characterized by the price movements as periods of upward market, downward market and sideways market. The begin and end dates of the periods, as well as the index values of these periods, are summarized in Table 8. Period 1 is a sideways market; Period 2 is an upward market and Period 3 is a downward market, which has a high volatility of 21.76%.

On each day transaction data from 9:30 to 16:00 were downloaded. We compute x_i as the mean-diurnally-adjusted duration, which will be used for analysis. Some features of the data are summarized as follows. First, the average duration increases when the size of the stock decreases. Also, the average duration decreases from Period 1 to Period 3, indicating that trades in the market are more frequent in later periods. Second, the changes in price per trade are the lowest for large stocks, and generally lower for medium stocks than small stocks (except for Period 1). Third, the daily volatility is the highest in Period 3. For many of the small stocks the annualized standard deviation exceeded 40%.

To investigate the pattern of price duration in relation to the price range δ, we first define δ^* as the relative price range so that $\delta = \bar{s}\delta^*$, where $\bar{s}$ is the mean stock price in the period. After experimenting with different values of δ^* we set δ^* to 0.001 in the first two periods and 0.002 in the third period. The compiled price durations have a mean of approximately 5 minutes. We compute the autocorrelation coefficients of the price durations and the Q statistics. The autocorrelation coefficients are generally significant even up to order 30. In addition, the Q statistics show significant serial correlation in duration, which suggests duration clustering and supports modeling using the ACD models.

5.2 Empirical Results of NYSE Stocks

We estimate the daily variance by various methods, which are then multiplied by 252 to obtain the annual variance. The square root of this variance, which is the annualized return standard deviation based on the open-to-close price changes, is then plotted against time. Figures 4 through 12 plot the daily volatility estimates of the four methods for different sizes of stocks over three sample periods. To conserve space, only the first five stocks in each stock-size group are included and V_R is excluded. The results of other stocks are qualitatively similar and their graphs are not presented.

Several characteristics emerging from the graphs can be summarized. It can be seen that the ACD-ICV estimates track the RV estimates very closely, and they do not appear to have any systematic biases versus the RV estimates. This is especially true for the stocks in Period 1 (sideways market). During this period there appears to be little interday variation in volatility. All estimates track each other quite closely, although there are more variations for the medium and small stocks, especially for the V_D estimates. In Period 2 (upward market), the overall volatility level is higher. We can also see more daily fluctuations in the volatility estimates. In particular, the V_D estimates fluctuate quite significantly over this period, especially for medium and small stocks. Generally, there appear to be a ranking of V_D being higher than V_A, which is in turn higher than V_B and V_K. The latter two estimates appear to be closest to each other. For the medium and small stocks it happens quite a significant number of times that V_D is more than double the values of the other estimates. Finally, Period 3 (downward market) exhibits an interesting feature that the volatilities of all stocks trend upwards during this 25-day period in which the overall market dropped about 9%. Most stocks start with an annualized volatility of about 20% and trend upwards to 40% or over. Again, the V_D estimates appear to be the highest on average, while V_B and V_K remain close to each other. Rather interestingly, V_A seem to moderate between V_D and the pair of estimates V_B and V_K.

We further compute two statistics to measure the closeness of the volatility estimates. First, we calculate the correlation coefficients between pairs of volatility estimates. The results are plotted in Figure 13. The first panel summarizes the results of V_A against other estimates. Panels 2, 3 and 4 summarize, respectively, the correlations of V_B, V_K and V_D against other estimates. The horizontal axis orders the stocks by period and then by size within each period. Thus, the 30 stocks in Period 1, ordered according to large, medium and then small stocks, come first. This is followed by the 30 stocks in Period 2 and then in Period 3. The graphs show that V_A generally correlates quite highly with other estimates. V_D appears to be the estimate that has the lowest correlation with other estimates (see Panel 4).[18] V_B and V_K correlate higher with each other than with V_A. We also observe that the correlations are the highest in Period 3 (see the last 30 data points). This is due to the trending (upwards) of volatility in this period.

Second, we compute the pairwise root mean-squared differences (RMSD) between the estimates. The results are presented in Figure 14 in a manner similar to Figure 13. Again, the RMSD between V_K and V_B appears to be the lowest amongst all paired comparisons. Also, the RMSD of V_D against other estimates are the highest. Although the correlations between the estimates are the highest in Period 3, the RMSE in Period 3 are also the largest, which is due to the higher general level of volatility in Period 3 for all stocks.

In Figure 15 we present the intraday volatility estimates of three stocks, namely GE (large stock) in Period 1 for $t = 18$, TJX (medium stock) in Period 2 for $t = 43$ and CSC (small stock) in Period 3 for $t = 20$. These stocks and times are chosen for illustration because in these cases the volatility estimates of various methods agree very closely with each other. Intraday ACD-ICV estimates over time intervals of 15 minutes and one hour are considered, and we also present the absolute returns over these intervals. It can be seen that the volatility estimates over one-hour intervals are subject to less fluctuation than the 15-min counterparts, as may be expected. For V_A there is an intraday pattern of a U-shape volatility, especially for the estimates over hourly intervals and for GE and TJX. In contrast, the absolute return values have large intraday fluctuations (indeed many intraday estimates are not shown as they are outside the range of the graph), which again verifies that this proxy is very noisy. In summary, the results suggest that V_A provides a workable and superior estimate of intraday volatility.

[18] There are cases of V_D having negative correlations with other estimates. In order to maintain a good scale for the graphs, such cases are deleted.

We now investigate whether V_A is sensitive to the specification of the ACD model and the method of estimation of the model parameters. To this effect, we compute the daily ACD-ICV estimates of the NYSE stocks using the ACD$(1,1)$ and AACD models estimated by the QMLE method, denoted by ACD-QMLE and AACD-QMLE, respectively. These are compared against the estimates computed using the AACD-SP method. We calculate the RMSD of the two QMLE daily ICV estimates versus the AACD-SP estimates for the thirty stocks in the three periods. For Period 1, the average RMSD is less than 1 percentage point (in annualized return standard deviation). Overall, the RMSD appears to be small and suggests that the IV estimates are not sensitive to the choice of the ACD$(1,1)$ model versus the AACD model.

We further examine the effects of the relative price range δ^* on the ACD-ICV estimates V_A. For this purpose, we compute the V_A estimates of the large and small stocks in Period 1 using the ACD-QMLE method. We vary δ^* from 0.00050 to 0.00175 in increments of 0.00025, and calculate the RMSD between the daily V_A estimates for δ^* and $\delta^* + 0.00025$. The RMSD appears to be quite stable and is below 0.01 for almost all cases when δ^* does not exceed 0.001.[19] Thus, the results suggest that the estimates V_A are not sensitive to the choice of the price range, provided δ^* is not too large so as to induce infrequent sampling of price events. Further research, however, has to be conducted for the determination of an optimal price range.[20]

6 Conclusion

We propose a method to estimate intraday volatility by integrating the instantaneous conditional variance per unit time obtained from the ACD models, and consider the estimation of the ACD model using a SP method. Our Monte Carlo results verify that the SP method is more efficient than the QMLE and compares favorably against the MLE. We compare the daily ACD-ICV estimate against several versions of realized volatility method. Our Monte Carlo results show that the ACD-ICV estimate provides the smallest RMSE, while the realized kernel RV gives the best results among the RV estimates.

[19]For $\delta^* = 0.001$, the mean duration varies from 90 second to 513 second for the cases considered.

[20]The issue of the optimal price range is analogous to the problem of determining the optimal sampling frequency for the V_R method as studied by Bandi and Russell (2006a, 2006b and 2008).

Our empirical results using 30 NYSE stocks show that the ACD-ICV estimate correlates highly with and performs very well against the realized volatility estimates. It frequently moderates between the realized kernel and duration based RV estimates. A clear advantage of our method is that it can be used to estimate intraday volatilities over intervals such as 15 minutes or an hour. Our robustness check shows that the ACD-ICV estimate is not sensitive to the selection of the ACD model and the method of estimation. Also, it is not sensitive to the choice of the price range, provided the price events sampled are not too infrequent. The optimal choice of the price range, however, remains an important topic for future research.

References

[1] Aït-Sahalia, Y., 2004, Disentangling diffusion from jumps, Journal of Financial Economics, 74, 487-528.

[2] Aït-Sahalia, Y., and L. Mancini, 2008, Out of sample forecasts of quadratic variation, Journal of Econometrics, 147, 17-33.

[3] Aït-Sahalia, Y., P. A. Mykland, and L. Zhang, 2005, How often to sample a continuous-time process in the presence of market microstructure noise, Review of Financial Studies, 18, 351-416.

[4] Aït-Sahalia, Y., P. A. Mykland, and L. Zhang, 2009, Ultra high frequency volatility estimation with dependent microstructure noise, Journal of Econometrcis, forthcoming.

[5] Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, The distribution of realized stock return volatility, Journal of Financial Economics, 61, 43-76.

[6] Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2001, The distribution of exchange rate volatility, Journal of the American Statistical Association 96, 42-55. Correction published in 2003, volume 98, page 501.

[7] Andersen, T. G., D. Dobrev, and E. Schaumburg, 2008, Duration-based volatility estimation, Working paper.

[8] Bandi, F. M., and R. Renò, 2008, Nonparametric stochastic volatility, working paper.

[9] Bandi, F. M., and J. R. Russell, 2006a, Separating market microstructure noise from volatility, Journal of Financial Economics, 79, 655-692.

[10] Bandi, F. M., and J. R. Russell, 2006b, Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations, Journal of Econometrics, forthcoming.

[11] Bandi, F. M., and J. R. Russell, 2008, Microstructure noise, realized variance, and optimal sampling, Review of Economic Studies, 75, 339-369.

[12] Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 2008, Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise, Econometrica, 76, 1481-1536.

[13] Barndorff-Nielsen, O. E., and N. Shephard, 2004, Power and bipower variation with stochastic volatility and jumps, Journal of Financial Econometrics, 2, 1-37.

[14] Bauwens, L., P. Giot, J. Grammig, and D. Veredas, 2004, A comparison of financial duration models via density forecasts, International Journal of Forecasting, 20, 589-609.

[15] Chan, C. C., and W. M. Fong, 2006, Realized volatility and transactions, Journal of Banking and Finance, 30, 2063-2085.

[16] Chen, S. X., 2000, Probability density function estimation using gamma kernels, Annals of Institute of Statistical Mathematics, 52, 471-480.

[17] Drost, F. C., and B. J. M. Werker, 2004, Semiparametric duration models, Journal of Business and Economic Statistics, 22, 40-50.

[18] Engle, R., 2000, The econometrics of ultra-high frequency data, Econometrica, 68, 1-22.

[19] Engle, R., and G. Gonzalez-Rivera, 1991, Semiparametric ARCH Models, Journal of Business and Economic Statistics, 9, 345-359.

[20] Engle R., and J. R. Russell, 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, Econometrica, 66, 1127-1162.

[21] Fernandes, M., and J. Grammig, 2005, Non-parametric specification tests for conditional duration models, Journal of Econometrics, 127, 35-68.

[22] Fernandes, M., and J. Grammig, 2006, A family of autoregressive conditional duration models, Journal of Econometrics, 130, 1-23.

[23] Grammig, J., and K. O. Maurer, 2000, Non-monotonic hazard functions and the autoregressive conditional duration model, Econometrics Journal, 3, 16-38.

[24] Kristensen, D., 2010, Nonparametric filtering of the realized spot volatility: A kernel-based approach, Econometric Theory, 26, 60-93.

[25] Lee, S., and B. Hansen, 1994, Asymptotic theory for the GARCH(1, 1)quasi-maximum likelihood estimator, Econometric Theory, 10, 29-52.

[26] Malliavin, P., and M. E. Mancino, 2009, A Fourier transfrom method for nonparametric estimation of multivariate volatility, Annals of Statistics 37, 1983-2010.

[27] Pacurar, M., 2008, Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature, Journal of Economic Surveys, 22, 711-751.

[28] Tapia, R. A., and J. R. Thompson, 1978, *Nonparametric Probability Density Estimation*, Johns Hopkins University Press.

[29] Zhang, L., P. A. Mykland, and Y. Aït-Sahalia, 2005, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, Journal of the American Statistical Association, 100, 1394-1411.

[30] Zhang, M. Y., J. R. Russell, and R. S. Tsay, 2001, A nonlinear autoregressive conditional duration model with applications to financial transaction data, Journal of Econometrics, 104, 179-207.

Table 1: Monte Carlo results for the AACD model with Weibull errors

Parameter	MLE	QMLE	SP
$N = 1{,}000$, MC replications = 500			
ω (0.04)	0.0423 (0.0081) (0.0084)	0.0492 (0.0170) (0.0193)	0.0444 (0.0120) (0.0128)
α (0.13)	0.1308 (0.0260) (0.0260)	0.1550 (0.0794) (0.0832)	0.1563 (0.0577) (0.0633)
β (0.80)	0.7972 (0.0213) (0.0215)	0.7781 (0.0792) (0.0821)	0.7741 (0.0587) (0.0641)
λ (0.26)	0.2525 (0.0255) (0.0266)	0.1070 (0.0663) (0.1667)	0.2134 (0.0822) (0.0944)
v (0.50)	0.4860 (0.1032) (0.1041)	0.1873 (0.0965) (0.3272)	0.3741 (0.1136) (0.1695)
b (0.10)	0.1006 (0.0324) (0.0323)	0.0317 (0.0263) (0.0732)	0.0679 (0.0323) (0.0455)
c (0.20)	0.2109 (0.1823) (0.1825)	0.5020 (0.1960) (0.3599)	0.2780 (0.1448) (0.1643)
$N = 10{,}000$, MC replications = 100			
ω (0.04)	0.0407 (0.0033) (0.0033)	0.0421 (0.0052) (0.0056)	0.0407 (0.0021) (0.0022)
α (0.13)	0.1288 (0.0078) (0.0078)	0.1305 (0.0119) (0.0118)	0.1351 (0.0092) (0.0105)
β (0.80)	0.8000 (0.0083) (0.0083)	0.7922 (0.0146) (0.0165)	0.7987 (0.0096) (0.0096)
λ (0.26)	0.2568 (0.0162) (0.0165)	0.2354 (0.0514) (0.0567)	0.2602 (0.0177) (0.0177)
v (0.50)	0.4956 (0.0376) (0.0377)	0.4346 (0.0744) (0.0987)	0.4959 (0.0276) (0.0278)
b (0.10)	0.0998 (0.0112) (0.0112)	0.0732 (0.0340) (0.0432)	0.0945 (0.0087) (0.0103)
c (0.20)	0.2105 (0.0699) (0.0704)	0.3629 (0.2464) (0.2944)	0.2010 (0.0098) (0.0098)
$N = 50{,}000$, MC replications = 50			
ω (0.04)	0.0403 (0.0015) (0.0015)	0.0409 (0.0034) (0.0034)	0.0411 (0.0006) (0.0012)
α (0.13)	0.1295 (0.0038) (0.0038)	0.1297 (0.0126) (0.0125)	0.1360 (0.0015) (0.0062)
β (0.80)	0.8006 (0.0050) (0.0050)	0.7983 (0.0110) (0.0111)	0.7964 (0.0013) (0.0038)
λ (0.26)	0.2584 (0.0091) (0.0092)	0.2537 (0.0270) (0.0275)	0.2636 (0.0030) (0.0046)
v (0.50)	0.4985 (0.0212) (0.0210)	0.4883 (0.0251) (0.0274)	0.5088 (0.0082) (0.0120)
b (0.10)	0.1008 (0.0063) (0.0063)	0.0916 (0.0059) (0.0102)	0.0918 (0.0015) (0.0083)
c (0.20)	0.1948 (0.0518) (0.0516)	0.2218 (0.0599) (0.0632)	0.2006 (0.0021) (0.0021)

Notes: The AACD equation is $\psi_i^\lambda = \omega + \alpha\psi_{i-1}^\lambda[|\epsilon_{i-1} - b| + c(\epsilon_{i-1} - b)]^v + \beta\psi_{i-1}^\lambda$. The figures in the first column are the true parameter values. The figures in other columns are the means of the parameter estimates, with standard deviations in the first parentheses and RMSE in the second parentheses.

Table 2: Monte Carlo results for the AACD model with Burr errors

Parameter	MLE	QMLE	SP
N = 1,000, MC replications = 500			
ω (0.04)	0.0432 (0.0124) (0.0128)	0.0448 (0.0181) (0.0187)	0.0414 (0.0091) (0.0092)
α (0.13)	0.1329 (0.0374) (0.0375)	0.1192 (0.0547) (0.0557)	0.1283 (0.0222) (0.0222)
β (0.80)	0.7940 (0.0392) (0.0396)	0.8078 (0.0574) (0.0579)	0.8009 (0.0220) (0.0220)
λ (0.26)	0.2582 (0.0945) (0.0944)	0.2568 (0.1716) (0.1715)	0.2570 (0.0498) (0.0498)
v (0.50)	0.4862 (0.1382) (0.1388)	0.5850 (0.1773) (0.1965)	0.5189 (0.1029) (0.1045)
b (0.10)	0.1035 (0.0374) (0.0375)	0.0943 (0.0400) (0.0404)	0.0979 (0.0255) (0.0255)
c (0.20)	0.1969 (0.1697) (0.1696)	0.1878 (0.2611) (0.2611)	0.2155 (0.0931) (0.0943)
N = 10,000, MC replications = 100			
ω (0.04)	0.0410 (0.0033) (0.0034)	0.0418 (0.0045) (0.0048)	0.0389 (0.0035) (0.0036)
α (0.13)	0.1302 (0.0058) (0.0058)	0.1322 (0.0081) (0.0084)	0.1533 (0.0104) (0.0255)
β (0.80)	0.7981 (0.0083) (0.0085)	0.7966 (0.0100) (0.0105)	0.7923 (0.0089) (0.0118)
λ (0.26)	0.2575 (0.0181) (0.0182)	0.2403 (0.0247) (0.0314)	0.2226 (0.0173) (0.0411)
v (0.50)	0.4929 (0.0424) (0.0428)	0.6776 (0.0314) (0.1803)	0.5886 (0.0323) (0.0942)
b (0.10)	0.1013 (0.0142) (0.0141)	0.0808 (0.0274) (0.0334)	0.1005 (0.0226) (0.0225)
c (0.20)	0.2115 (0.0378) (0.0394)	0.2037 (0.0950) (0.0946)	0.1979 (0.0202) (0.0202)
N = 50,000, MC replications = 50			
ω (0.04)	0.0402 (0.0016) (0.0016)	0.0401 (0.0014) (0.0014)	0.0405 (0.0010) (0.0011)
α (0.13)	0.1302 (0.0026) (0.0025)	0.1301 (0.0023) (0.0023)	0.1324 (0.0019) (0.0031)
β (0.80)	0.7987 (0.0030) (0.0032)	0.7995 (0.0043) (0.0043)	0.8005 (0.0017) (0.0018)
λ (0.26)	0.2623 (0.0079) (0.0082)	0.2608 (0.0091) (0.0090)	0.2626 (0.0042) (0.0049)
v (0.50)	0.5001 (0.0179) (0.0177)	0.5018 (0.0216) (0.0214)	0.5020 (0.0158) (0.0158)
b (0.10)	0.1011 (0.0055) (0.0056)	0.1007 (0.0055) (0.0055)	0.0980 (0.0043) (0.0047)
c (0.20)	0.2104 (0.0569) (0.0573)	0.2026 (0.0496) (0.0492)	0.2022 (0.0028) (0.0035)

Notes: The AACD equation is $\psi_i^\lambda = \omega + \alpha\psi_{i-1}^\lambda[|\epsilon_{i-1} - b| + c(\epsilon_{i-1} - b)]^v + \beta\psi_{i-1}^\lambda$. The figures in the first column are the true parameter values. The figures in other columns are the means of the parameter estimates, with standard deviations in the first parentheses and RMSE in the second parentheses.

Table 3: Monte Carlo results for experiments with deterministic volatility

Estimation Method	Experiment 1 ME	Experiment 1 RMSE	Experiment 2 ME	Experiment 2 RMSE	Experiment 3 ME	Experiment 3 RMSE	Experiment 4 ME	Experiment 4 RMSE
V_A	−0.0024	0.0087	−0.0004	0.0089	−0.0013	0.0088	−0.0005	0.0089
V_A^*	−0.0124	0.0141	−0.0111	0.0143	−0.0114	0.0145	−0.0101	0.0143
V_A'	−0.0006	0.0110	0.0047	0.0121	0.0042	0.0120	−0.0052	0.0121
V_B	−0.0026	0.0137	0.0028	0.0139	0.0030	0.0140	0.0038	0.0141
V_D	−0.0120	0.0196	−0.0068	0.0170	−0.0066	0.0170	−0.0188	0.0157
V_K	−0.0050	0.0139	0.0006	0.0131	0.0008	0.0131	0.0014	0.0134
V_R	−0.0120	0.0249	−0.0068	0.0245	−0.0066	0.0248	−0.0118	0.0243

Notes: ME = mean error, RMSE = root mean-squared error. The results are based on 50 MC replications of 150-day daily volatility estimates. V_A and V_A' are computed using SP and QMLE methods, respectively, with δ being the average price range conditional on a price event being observed. V_A^* is computed using SP method, with δ being the threshold price range.

Table 4: Monte Carlo results for stochastic volatility models with NSR = 0.25

Estimation method	Heston ME	Heston RMSE	Heston with jump ME	Heston with jump RMSE	LV ME	LV RMSE	LV with jump ME	LV with jump RMSE
Panel A: Transaction price with white noise								
V_A	−0.0005	0.0111	0.0010	0.0141	−0.0007	0.0075	−0.0003	0.0074
V_A^*	−0.0162	0.0202	−0.0287	0.0335	−0.0144	0.0161	−0.0137	0.0155
V_B	−0.0017	0.0140	−0.0024	0.0194	−0.0016	0.0128	−0.0016	0.0128
V_D	−0.0095	0.0189	−0.0191	0.0296	−0.0079	0.0166	−0.0075	0.0165
V_K	−0.0004	0.0150	−0.0008	0.0209	−0.0003	0.0138	−0.0003	0.0138
V_R	−0.0029	0.0256	−0.0041	0.0356	−0.0028	0.0235	−0.0027	0.0234
Panel B: Transaction price with autocorrelated noise								
V_A	−0.0003	0.0110	0.0012	0.0141	−0.0005	0.0074	−0.0002	0.0074
V_A^*	−0.0161	0.0201	−0.0286	0.0334	−0.0143	0.0161	−0.0136	0.0154
V_B	−0.0017	0.0140	−0.0025	0.0195	−0.0016	0.0128	−0.0016	0.0128
V_D	−0.0094	0.0189	−0.0190	0.0295	−0.0078	0.0166	−0.0075	0.0165
V_K	−0.0004	0.0150	−0.0008	0.0209	−0.0003	0.0138	−0.0003	0.0138
V_R	−0.0029	0.0256	−0.0041	0.0356	−0.0028	0.0235	−0.0027	0.0235
Panel C: Transaction price with noise correlated with efficient price								
V_A	−0.0006	0.0109	0.0009	0.0141	−0.0008	0.0075	−0.0001	0.0074
V_A^*	−0.0163	0.0202	−0.0288	0.0336	−0.0145	0.0162	−0.0138	0.0156
V_B	−0.0017	0.0140	−0.0025	0.0195	−0.0016	0.0128	−0.0016	0.0128
V_D	−0.0095	0.0189	−0.0191	0.0296	−0.0080	0.0166	−0.0076	0.0165
V_K	−0.0004	0.0150	−0.0008	0.0209	−0.0004	0.0137	−0.0003	0.0137
V_R	−0.0029	0.0256	−0.0041	0.0356	−0.0027	0.0234	−0.0027	0.0234
Panel D: Transaction price with autocorrelated noise correlated with efficient price								
V_A	−0.0004	0.0110	0.0010	0.0141	−0.0007	0.0074	0.0003	0.0074
V_A^*	−0.0162	0.0202	−0.0288	0.0336	−0.0143	0.0160	−0.0136	0.0154
V_B	−0.0017	0.0140	−0.0025	0.0195	−0.0016	0.0128	−0.0016	0.0128
V_D	−0.0094	0.0189	−0.0190	0.0295	−0.0079	0.0166	−0.0075	0.0165
V_K	−0.0004	0.0150	−0.0008	0.0208	−0.0004	0.0138	−0.0003	0.0138
V_R	−0.0029	0.0255	−0.0041	0.0355	−0.0028	0.0235	−0.0028	0.0235

Notes: ME = mean error, RMSE = root mean-squared error. The results are based on 1,000 MC replications of 100-day daily volatility estimates. V_A and V_A^* are based on the QMLE of the AACD model. For V_A δ is the average price range conditional on a price event, while for V_A^* δ is the price-range threshold. The Heston-with-jump model is the Heston model with jumps in the volatility, while the LV-with-jump model has jumps in the price.

Table 5: Monte Carlo results for stochastic volatility models with NSR = 0.40

Estimation method	Heston ME	Heston RMSE	Heston with jump ME	Heston with jump RMSE	LV ME	LV RMSE	LV with jump ME	LV with jump RMSE
Panel A: Transaction price with white noise								
V_A	0.0028	0.0113	0.0044	0.0147	0.0027	0.0078	0.0036	0.0082
V_A^*	−0.0138	0.0184	−0.0263	0.0315	−0.0120	0.0139	−0.0113	0.0134
V_B	−0.0014	0.0138	−0.0022	0.0193	−0.0013	0.0128	−0.0013	0.0128
V_D	−0.0077	0.0180	−0.0171	0.0283	−0.0064	0.0160	−0.0060	0.0159
V_K	0.0002	0.0148	−0.0004	0.0207	0.0002	0.0138	0.0003	0.0138
V_R	−0.0027	0.0253	−0.0039	0.0352	−0.0026	0.0234	−0.0026	0.0234
Panel B: Transaction price with autocorrelated noise								
V_A	0.0032	0.0114	0.0048	0.0149	0.0030	0.0079	0.0039	0.0084
V_A^*	−0.0136	0.0182	−0.0261	0.0314	−0.0118	0.0138	−0.0111	0.0132
V_B	−0.0014	0.0138	−0.0022	0.0193	−0.0013	0.0128	−0.0013	0.0128
V_D	−0.0076	0.0179	−0.0170	0.0282	−0.0063	0.0160	−0.0059	0.0158
V_K	0.0002	0.0148	−0.0004	0.0207	0.0002	0.0137	0.0003	0.0138
V_R	−0.0027	0.0253	−0.0039	0.0352	−0.0026	0.0234	−0.0026	0.0234
Panel C: Transaction price with noise correlated with efficient price								
V_A	0.0026	0.0113	0.0042	0.0147	0.0024	0.0078	0.0034	0.0081
V_A^*	−0.0140	0.0185	−0.0265	0.0317	−0.0121	0.0141	−0.0114	0.0135
V_B	−0.0014	0.0138	−0.0022	0.0193	−0.0014	0.0128	−0.0013	0.0128
V_D	−0.0078	0.0180	−0.0172	0.0284	−0.0065	0.0160	−0.0061	0.0159
V_K	0.0001	0.0148	−0.0004	0.0207	0.0001	0.0137	0.0002	0.0138
V_R	−0.0027	0.0253	−0.0039	0.0352	−0.0026	0.0234	−0.0026	0.0234
Panel D: Transaction price with autocorrelated noise correlated with efficient price								
V_A	0.0029	0.0113	0.0046	0.0148	0.0027	0.0079	0.0037	0.0083
V_A^*	−0.0138	0.0184	−0.0263	0.0315	−0.0119	0.0139	−0.0112	0.0133
V_B	−0.0014	0.0138	−0.0022	0.0193	−0.0012	0.0128	−0.0012	0.0128
V_D	−0.0077	0.0180	−0.0171	0.0283	−0.0063	0.0160	−0.0059	0.0158
V_K	0.0002	0.0149	−0.0004	0.0207	0.0002	0.0137	0.0003	0.0137
V_R	−0.0027	0.0253	−0.0039	0.0352	−0.0025	0.0234	−0.0025	0.0234

Notes: ME = mean error, RMSE = root mean-squared error. The results are based on 1,000 MC replications of 100-day daily volatility estimates. V_A and V_A^* are based on the QMLE of the AACD model. For V_A δ is the average price range conditional on a price event, while for V_A^* δ is the price-range threshold. The Heston-with-jump model is the Heston model with jumps in the volatility, while the LV-with-jump model has jumps in the price.

Table 6: Monte Carlo results for stochastic volatility models with NSR = 0.60

Estimation method	Heston ME	Heston RMSE	Heston with jump ME	Heston with jump RMSE	LV ME	LV RMSE	LV with jump ME	LV with jump RMSE
Panel A: Transaction price with white noise								
V_A	0.0091	0.0143	0.0111	0.0178	0.0089	0.0117	0.0098	0.0124
V_A^*	−0.0091	0.0152	−0.0216	0.0276	−0.0073	0.0102	−0.0066	0.0098
V_B	−0.0008	0.0137	−0.0018	0.0193	−0.0008	0.0127	−0.0007	0.0127
V_D	−0.0048	0.0170	−0.0141	0.0267	−0.0035	0.0152	−0.0031	0.0151
V_K	0.0013	0.0149	0.0004	0.0207	0.0013	0.0138	0.0014	0.0139
V_R	−0.0025	0.0253	−0.0038	0.0352	−0.0024	0.0234	−0.0024	0.0234
Panel B: Transaction price with autocorrelated noise								
V_A	0.0098	0.0147	0.0119	0.0184	0.0096	0.0122	0.0105	0.0129
V_A^*	−0.0087	0.0148	−0.0211	0.0273	−0.0069	0.0100	−0.0062	0.0096
V_B	−0.0008	0.0137	−0.0018	0.0193	−0.0008	0.0127	−0.0007	0.0127
V_D	−0.0046	0.0169	−0.0138	0.0265	−0.0032	0.0151	−0.0029	0.0151
V_K	0.0013	0.0149	0.0004	0.0207	0.0013	0.0138	0.0014	0.0138
V_R	−0.0025	0.0253	−0.0038	0.0352	−0.0024	0.0234	−0.0024	0.0234
Panel C: Transaction price with noise correlated with efficient price								
V_A	0.0087	0.0139	0.0106	0.0174	0.0085	0.0113	0.0094	0.0121
V_A^*	−0.0094	0.0153	−0.0219	0.0279	−0.0076	0.0105	−0.0069	0.0100
V_B	−0.0008	0.0138	−0.0018	0.0193	−0.0008	0.0127	−0.0008	0.0127
V_D	−0.0050	0.0171	−0.0143	0.0268	−0.0037	0.0152	−0.0033	0.0152
V_K	0.0012	0.0149	0.0004	0.0207	0.0012	0.0138	0.0013	0.0138
V_R	−0.0025	0.0253	−0.0038	0.0352	−0.0024	0.0234	−0.0024	0.0234
Panel D: Transaction price with autocorrelated noise correlated with efficient price								
V_A	0.0093	0.0143	0.0113	0.0179	0.0091	0.0118	0.0100	0.0125
V_A^*	−0.0091	0.0151	−0.0216	0.0277	−0.0072	0.0102	−0.0065	0.0098
V_B	−0.0008	0.0138	−0.0018	0.0192	−0.0006	0.0127	−0.0006	0.0127
V_D	−0.0048	0.0171	−0.0140	0.0266	−0.0033	0.0152	−0.0030	0.0151
V_K	0.0013	0.0149	0.0004	0.0207	0.0014	0.0138	0.0014	0.0138
V_R	−0.0025	0.0252	−0.0037	0.0351	−0.0023	0.0234	−0.0023	0.0234

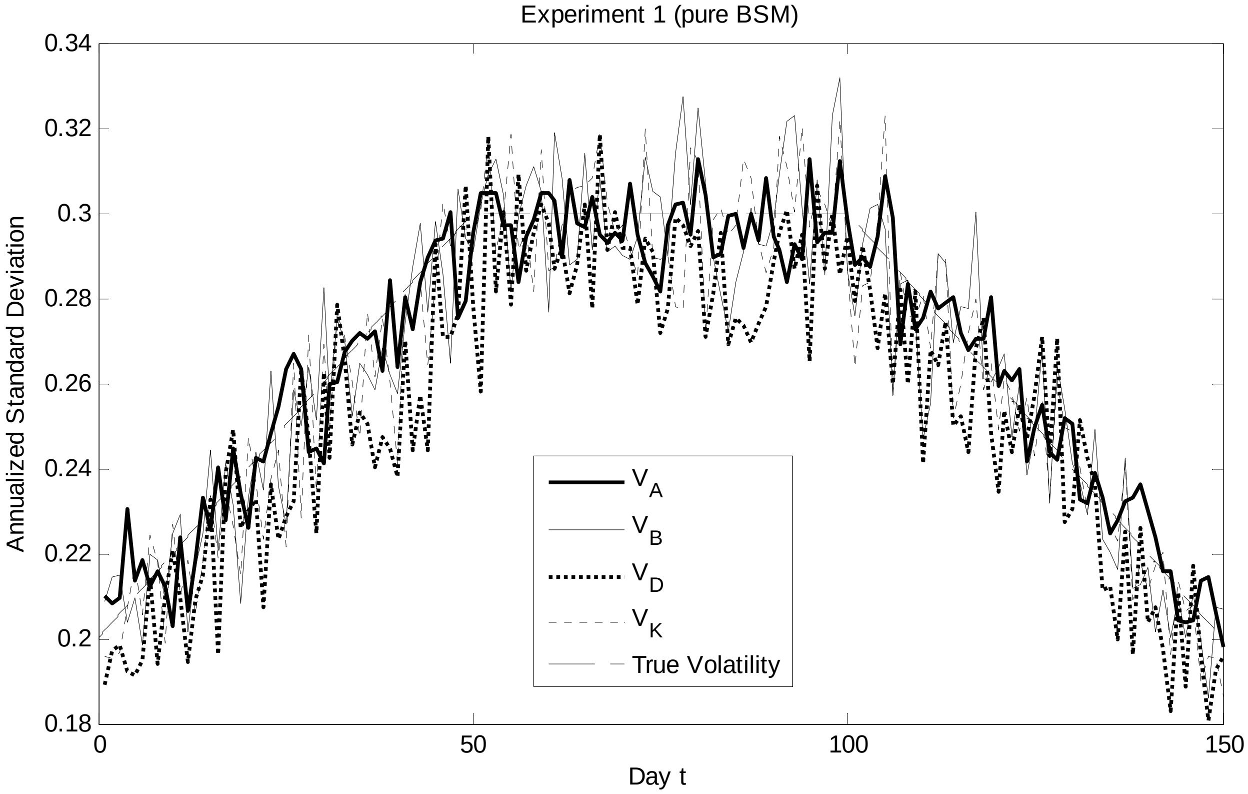
Notes: ME = mean error, RMSE = root mean-squared error. The results are based on 1,000 MC replications of 100-day daily volatility estimates. V_A and V_A^* are based on the QMLE of the AACD model. For V_A δ is the average price range conditional on a price event, while for V_A^* δ is the price-range threshold. The Heston-with-jump model is the Heston model with jumps in the volatility, while the LV-with-jump model has jumps in the price.

Table 7: Stocks and codes

Large Stocks		Medium Stocks		Small Stocks	
Stock	Code	Stock	Code	Stock	Code
Exxon Mobil	XOM	TJX	TJX	Brown-Forman (B)	BFB
General Electric	GE	Tyco International	TYC	Constellation Energy Group	CEG
Procter & Gamble	PG	Viacom (B)	VIAB	Computer Sciences	CSC
John & Johnson	JNJ	Allergan	AGN	Jacobs Engineering Group	JEC
AT & T	T	Chesapeake Energy	CHK	American Int'l. Group	AIG
Chevron	CVX	Aon	AOC	Waters	WAT
JPMorgan Chase	JPM	Loews	CG	U.S. Steel	X
Wal Mart	WMT	Progress Energy	PGN	Moody's	MCO
IBM	IBM	Williams	WMB	McCormick	MKC
Pfizer	PFE	Baker Hughes	BHI	Comerica	CMA

Table 8: Sample period and data summary

Period	1	2	3
Begin date	2006-01-11	2007-03-13	2007-07-13
End date	2006-03-31	2007-06-04	2007-08-16
Begin index	1294.18	1377.95	1552.5
End index	1294.87	1539.18	1411.27
Number of days	56	58	25
Return in period	0.05%	11.70%	−9.1%
Annualized standard deviation of daily return	9.01%	9.85%	21.76%



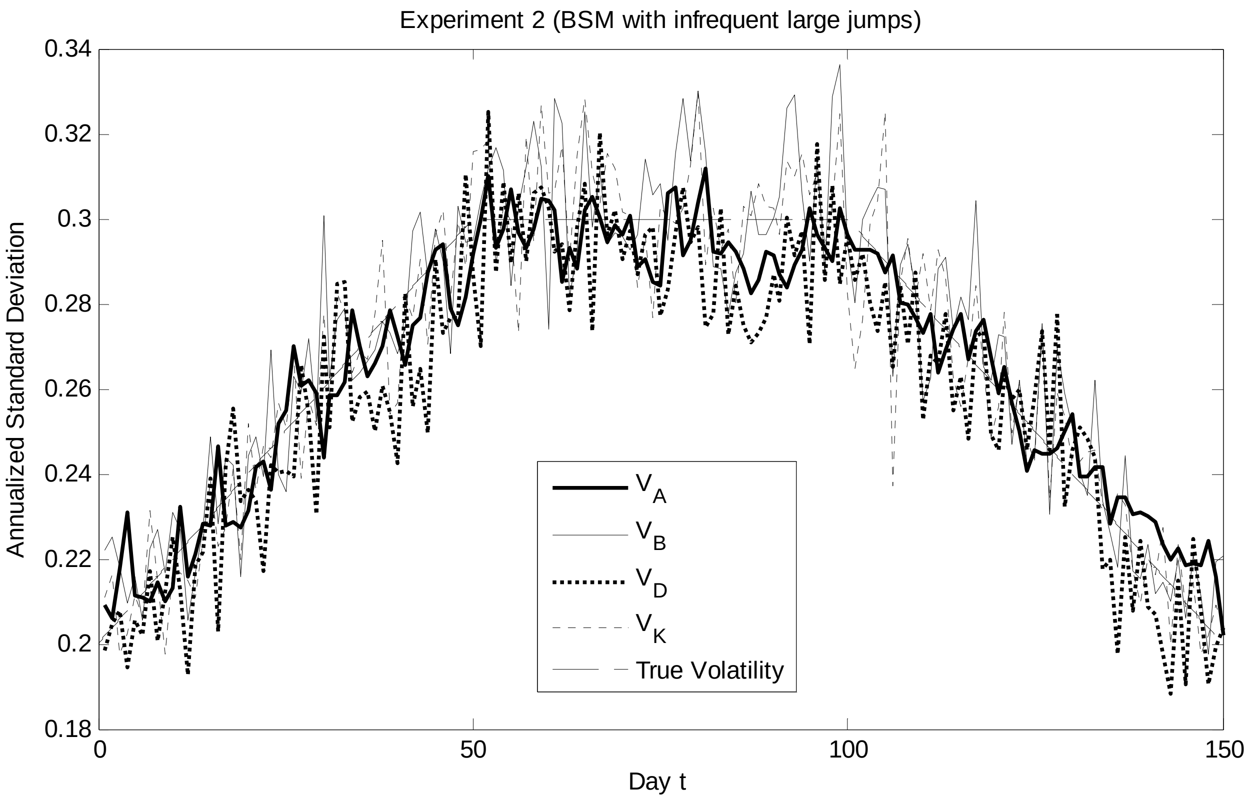


Figure 1: Estimates of Deterministic Volatility Model

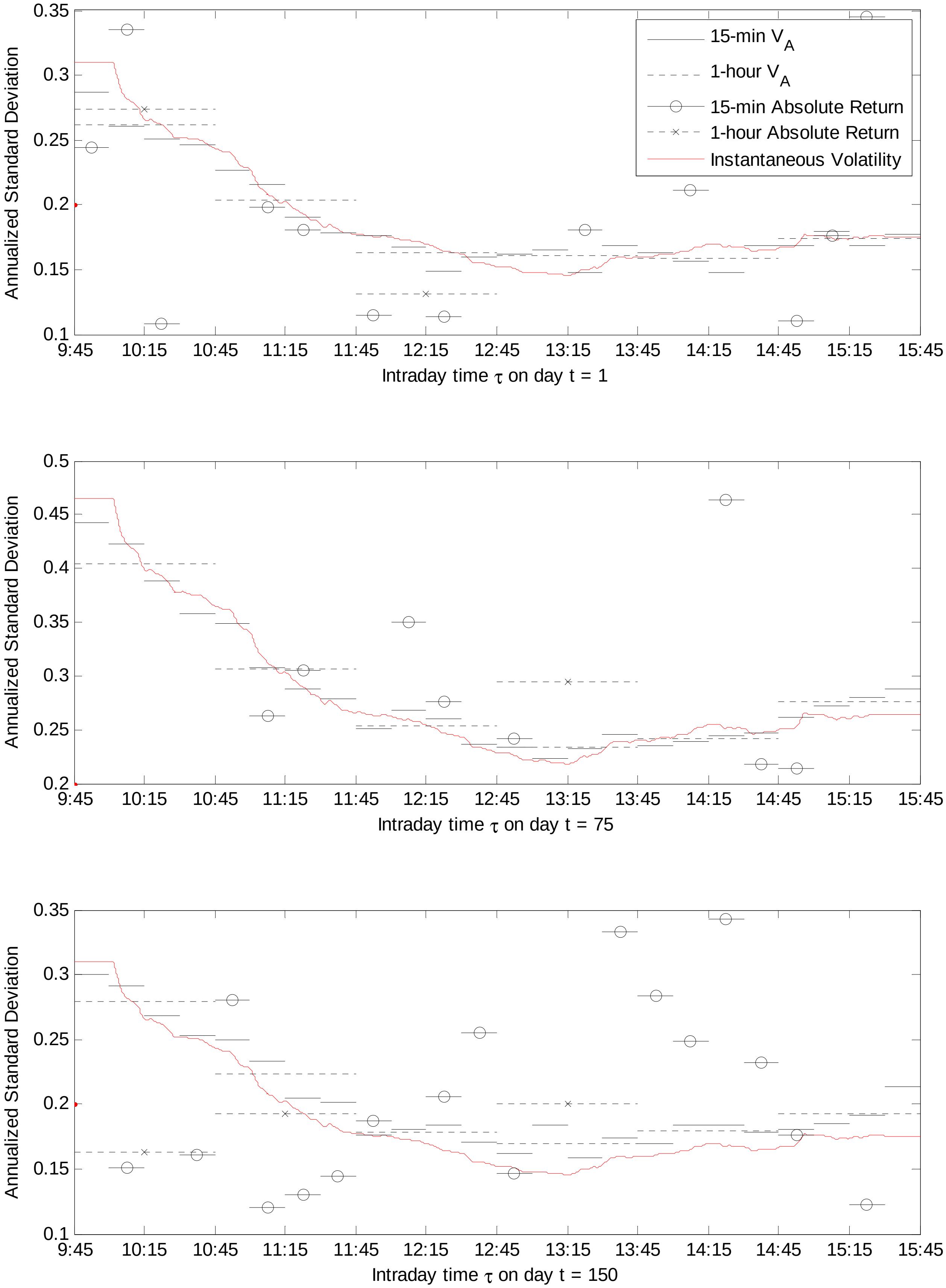
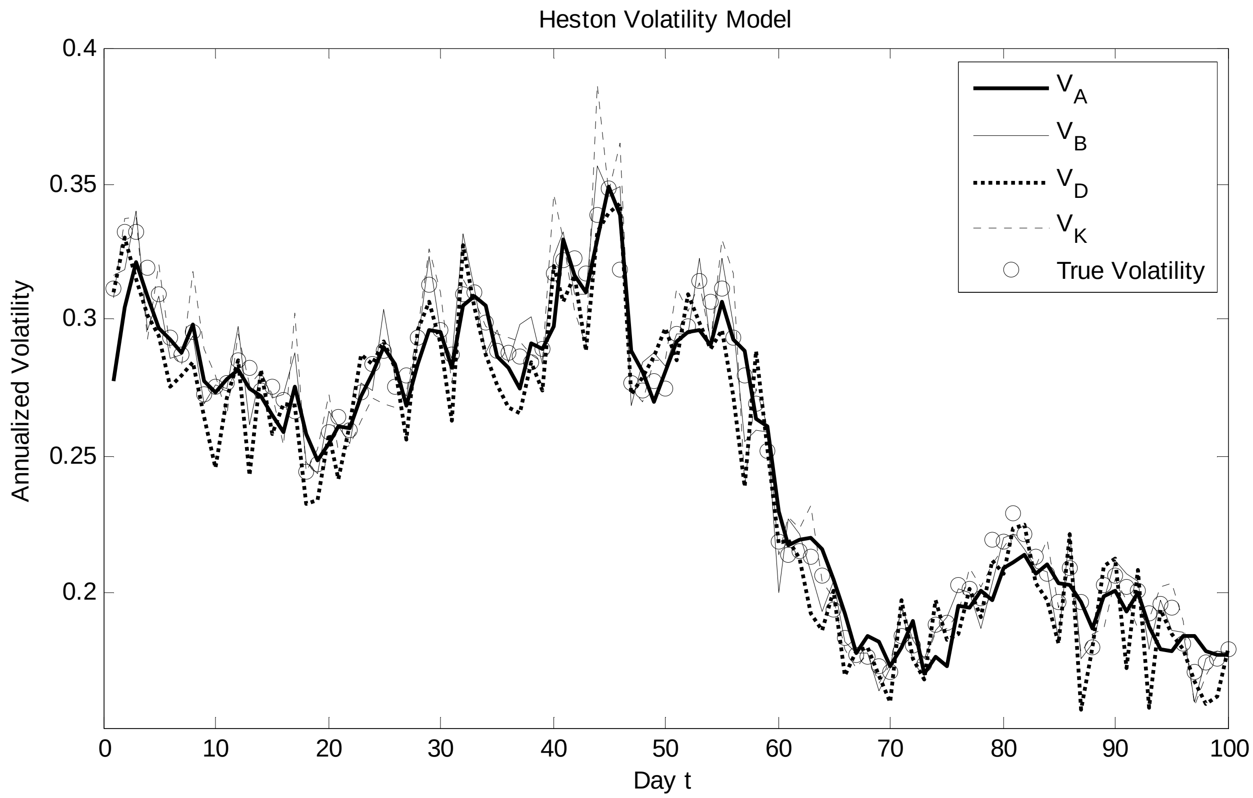


Figure 2: Intraday V_A Estimates, Experiment 2 (BSM with infrequent large jumps)



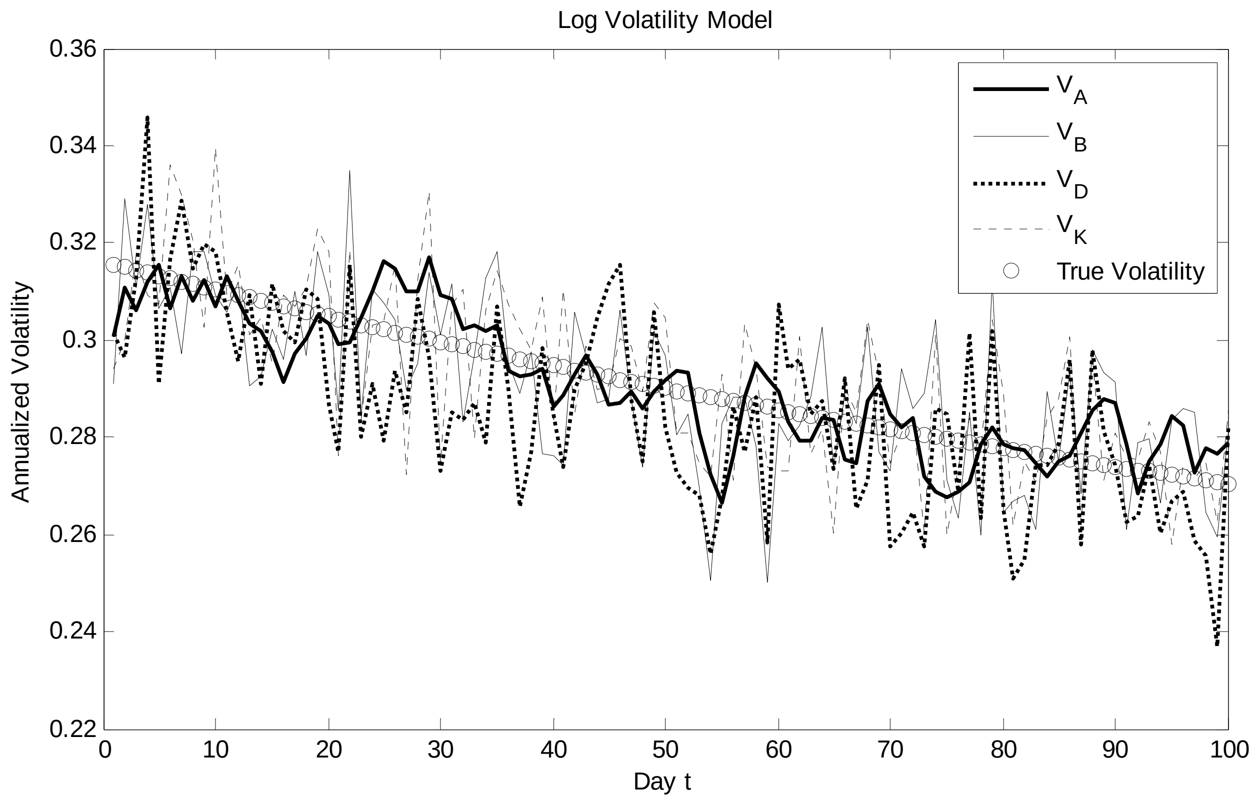


Figure 3: Estimates of Stochastic Volatility Model

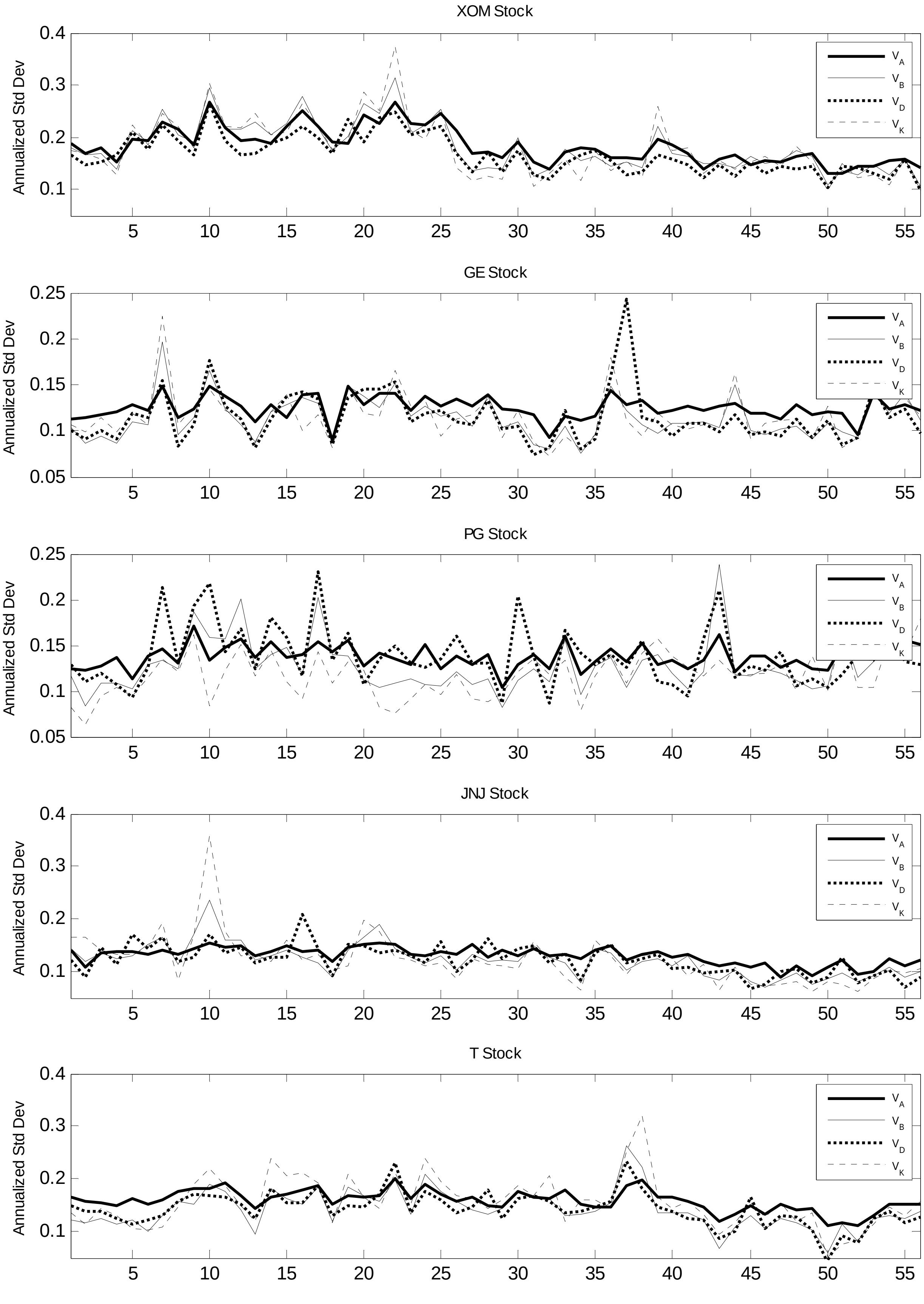


Figure 4: Volatility of Large Stocks, Period 1: Jan 11, 2006 -- March 31, 2006

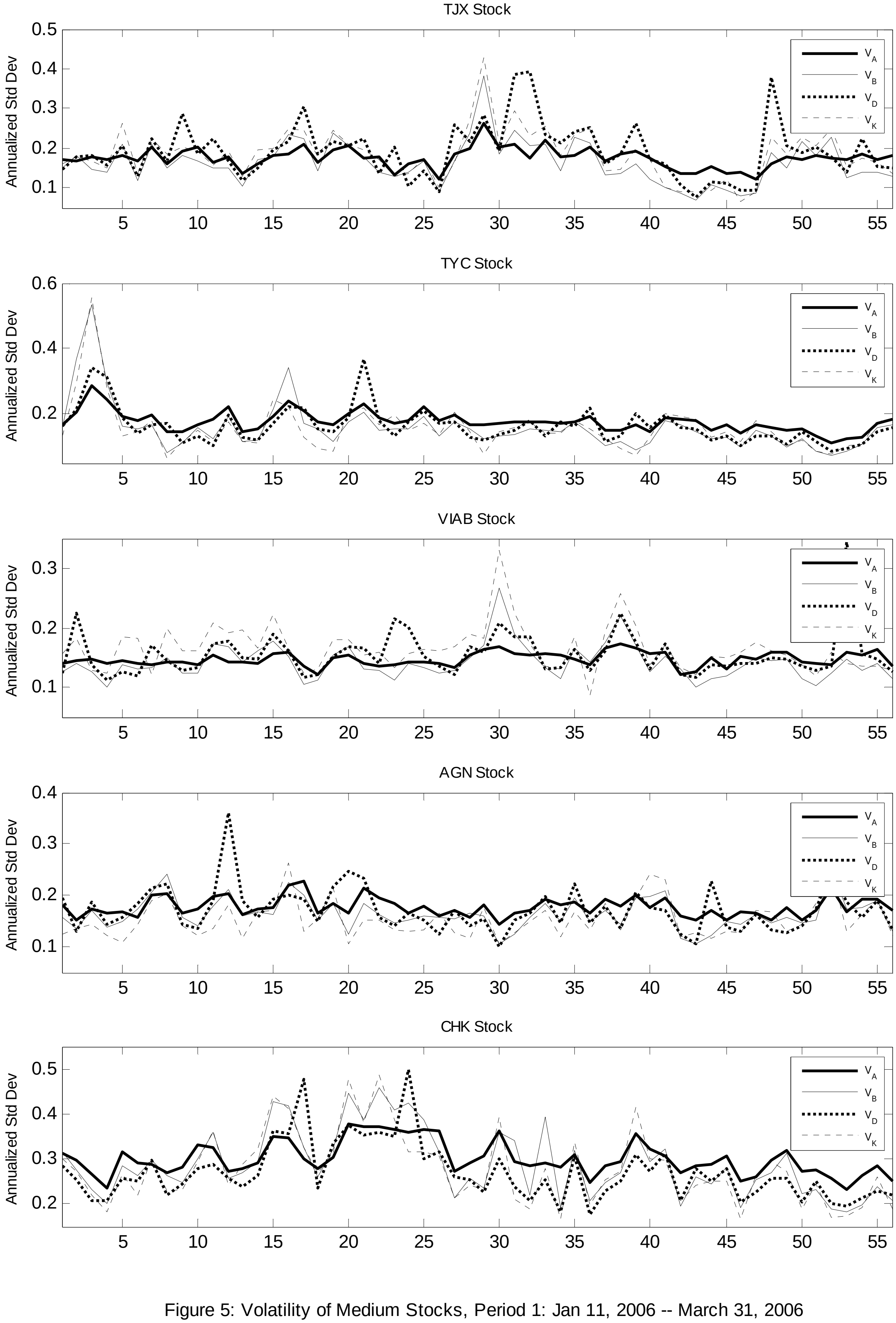


Figure 5: Volatility of Medium Stocks, Period 1: Jan 11, 2006 -- March 31, 2006

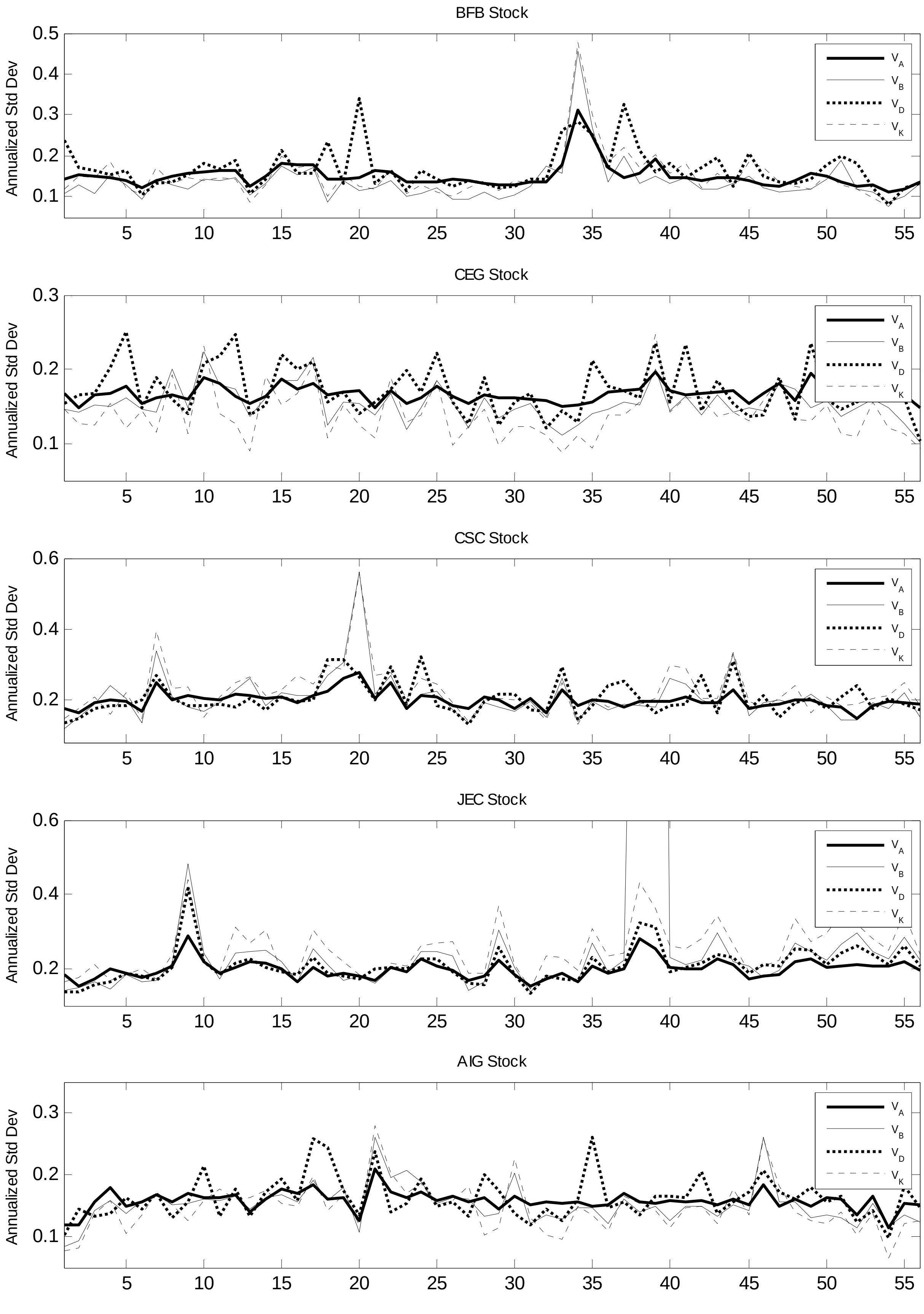


Figure 6: Volatility of Small Stocks, Period 1: Jan 11, 2006 -- March 31, 2006

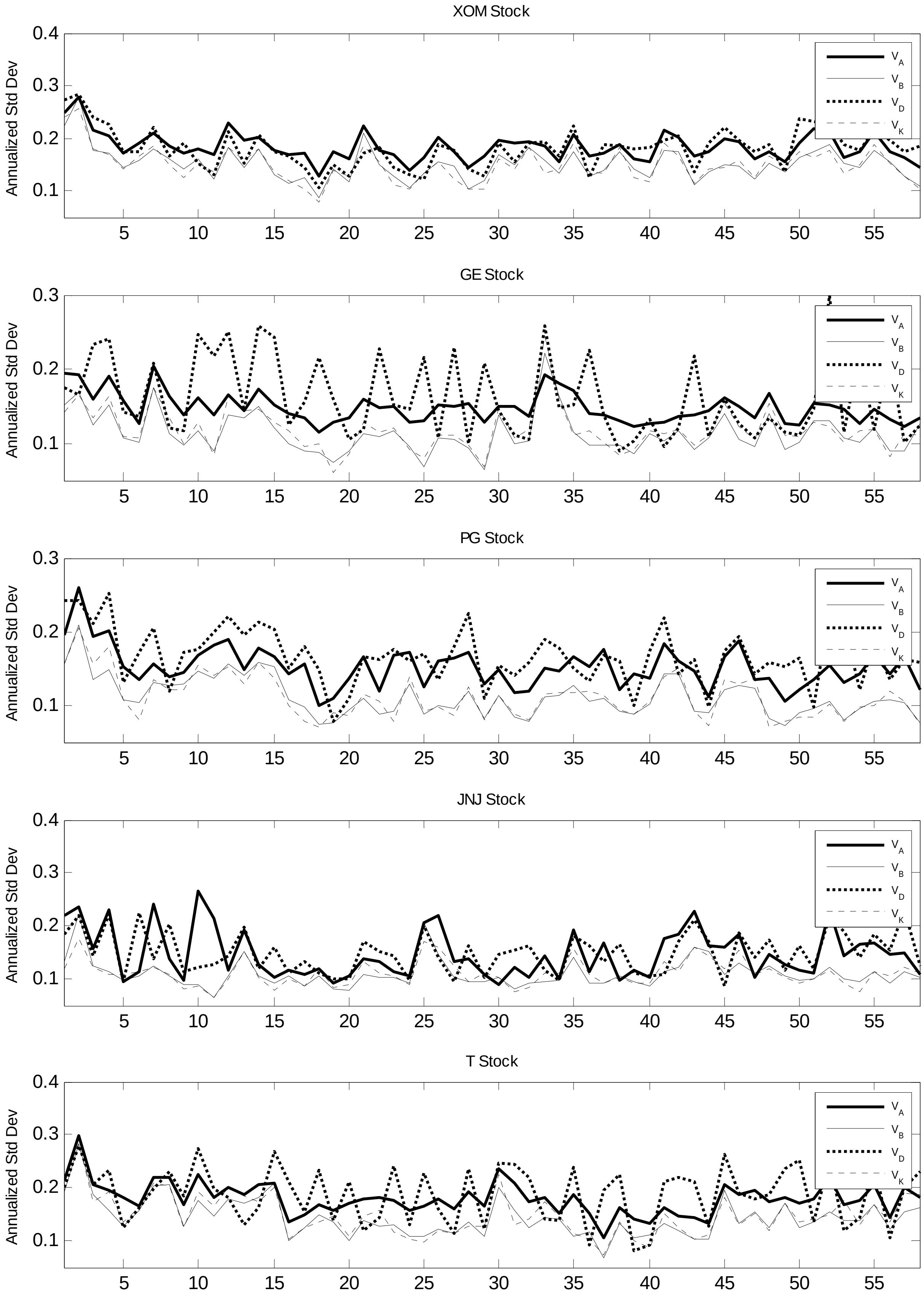


Figure 7: Volatility of Large Stocks, Period 2: Mar 13, 2007 -- Jun 4, 2007

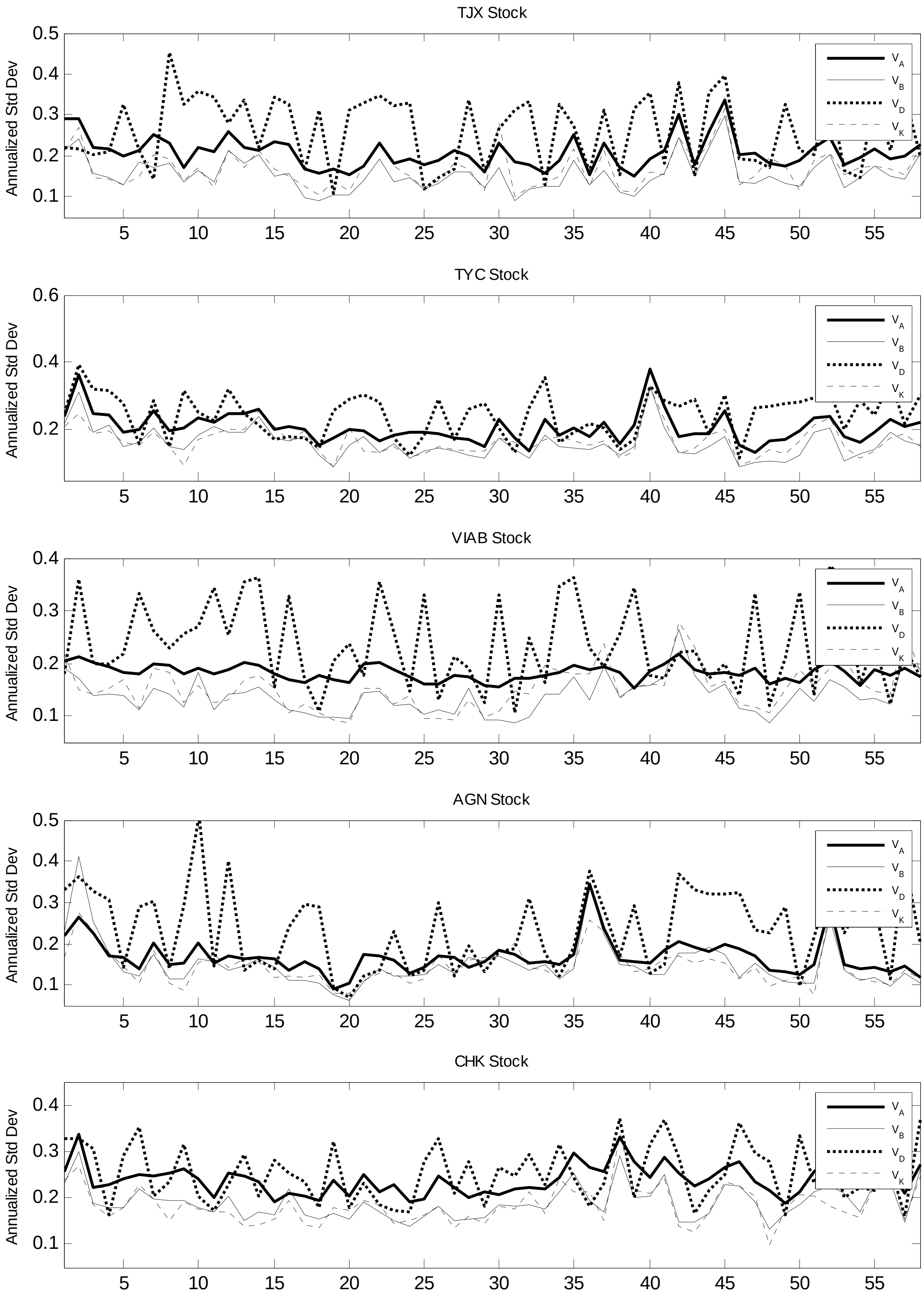


Figure 8: Volatility of Medium Stocks, Period 2: Mar 13, 2007 -- Jun 4, 2007

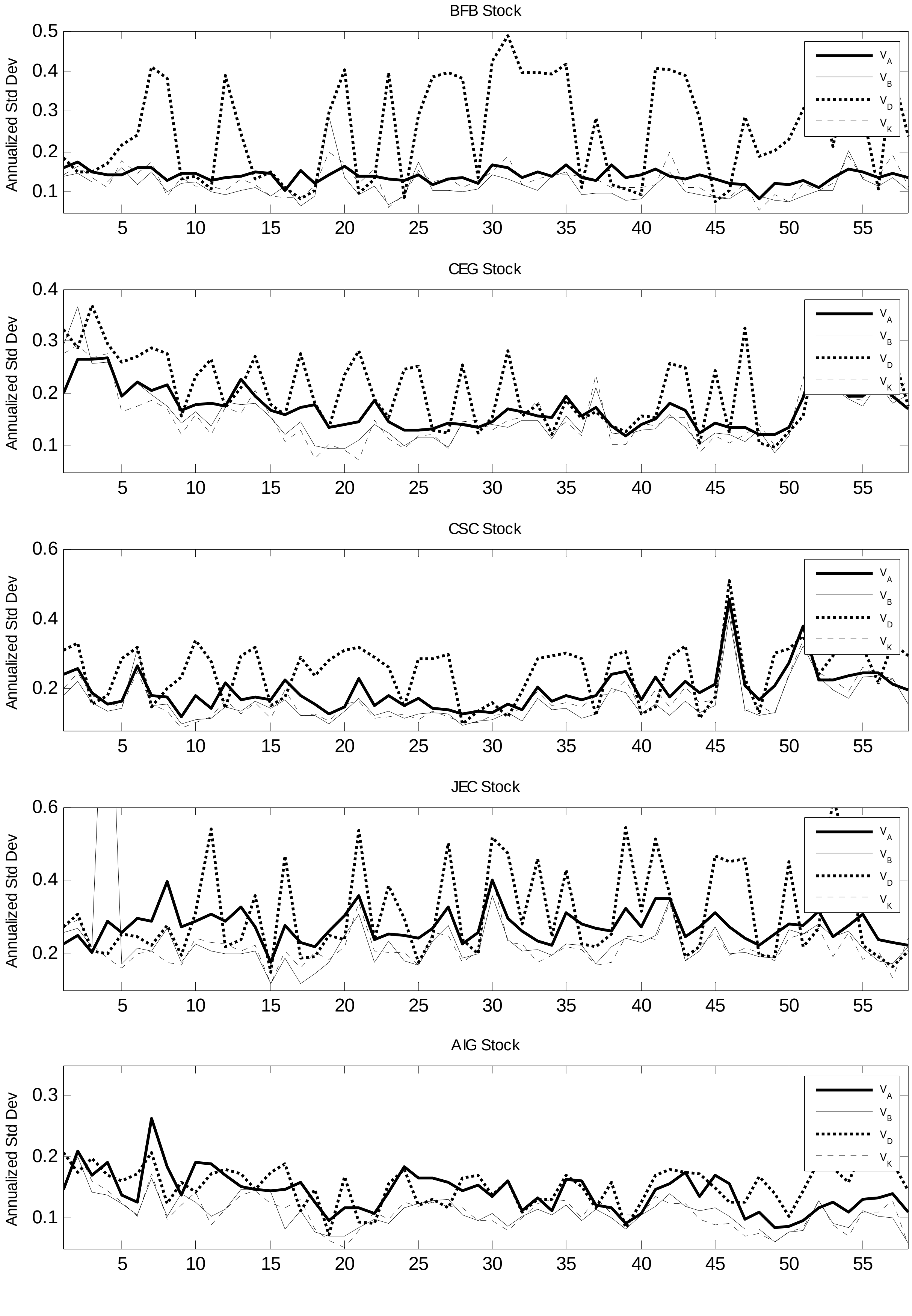


Figure 9: Volatility of Small Stocks, Period 2: Mar 13, 2007 -- Jun 4, 2007

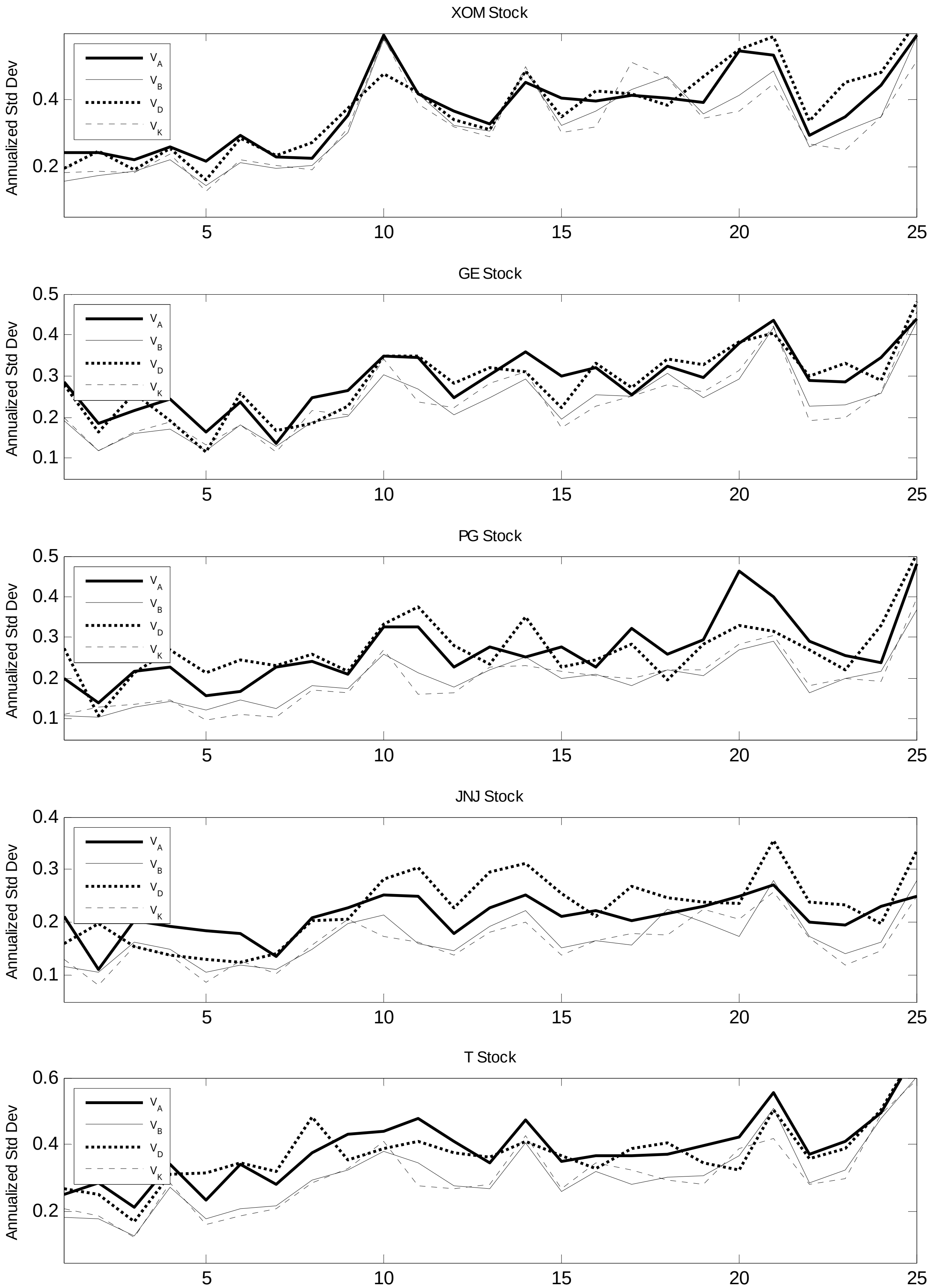


Figure 10: Volatility of Large Stocks, Period 3: Jul 13, 2007 -- Aug 16, 2007

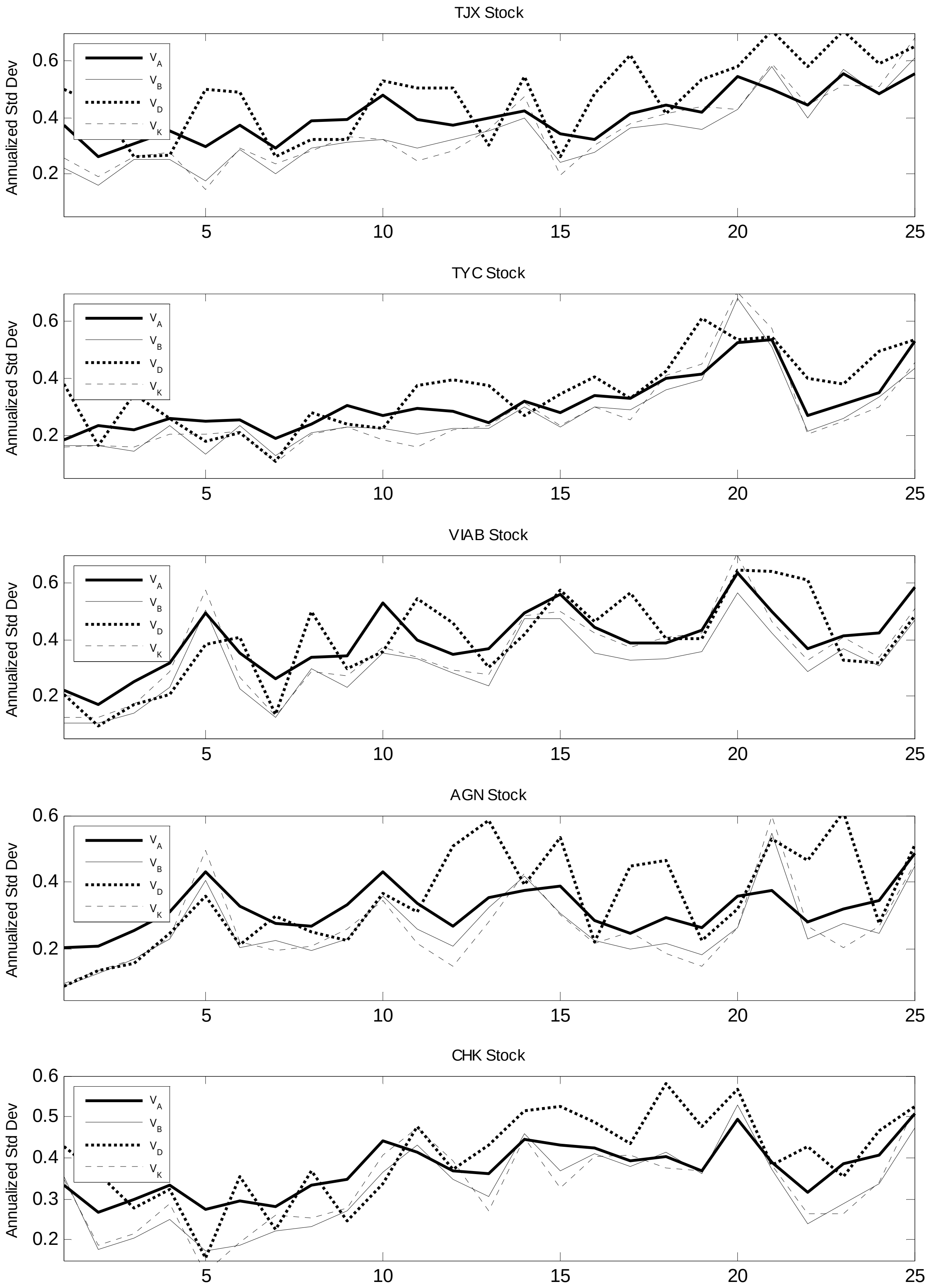


Figure 11: Volatility of Medium Stocks, Period 3: Jul 13, 2007 -- Aug 16, 2007

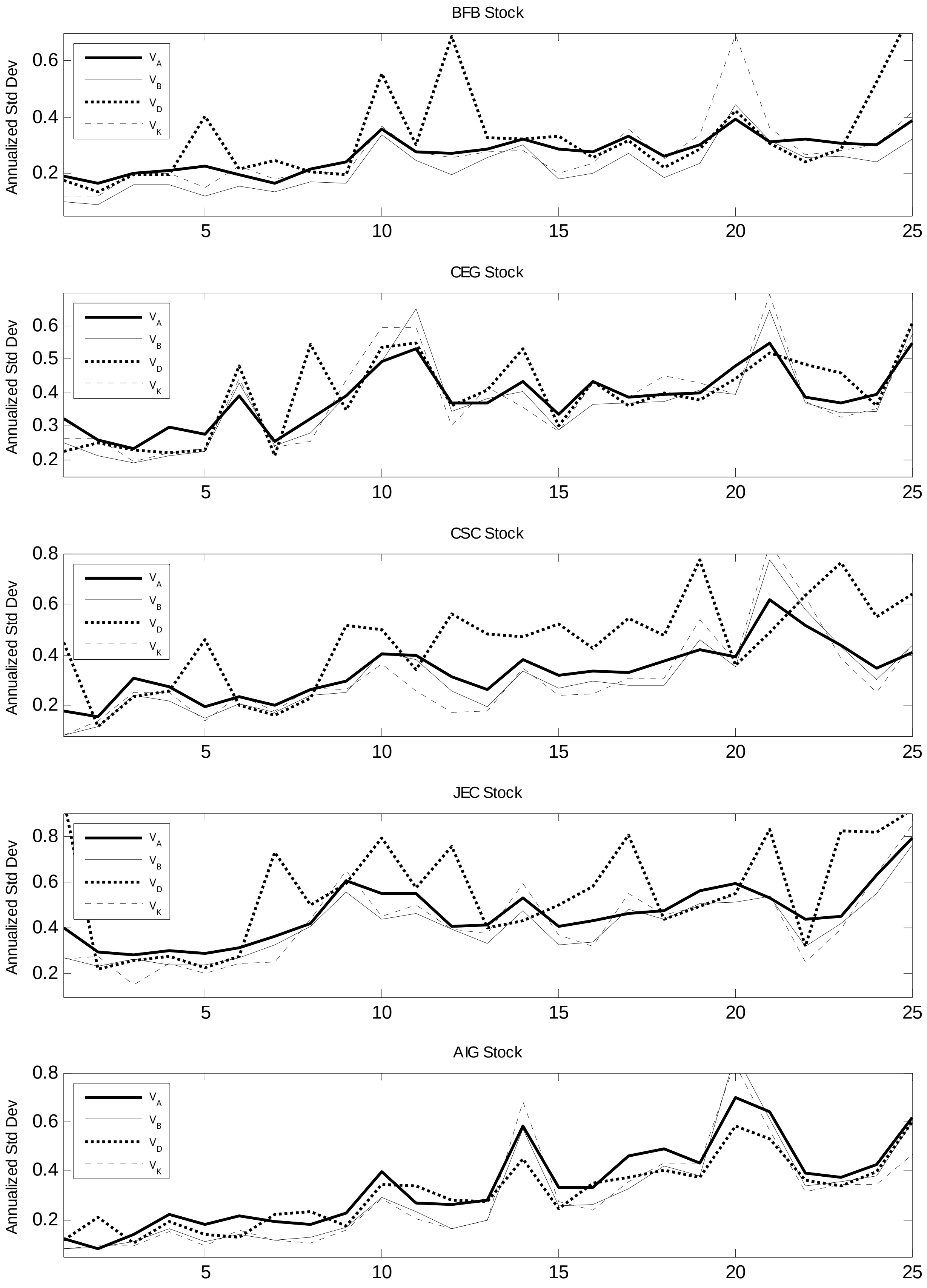


Figure 12: Volatility of Small Stocks, Period 3: Jul 13, 2007 -- Aug 16, 2007

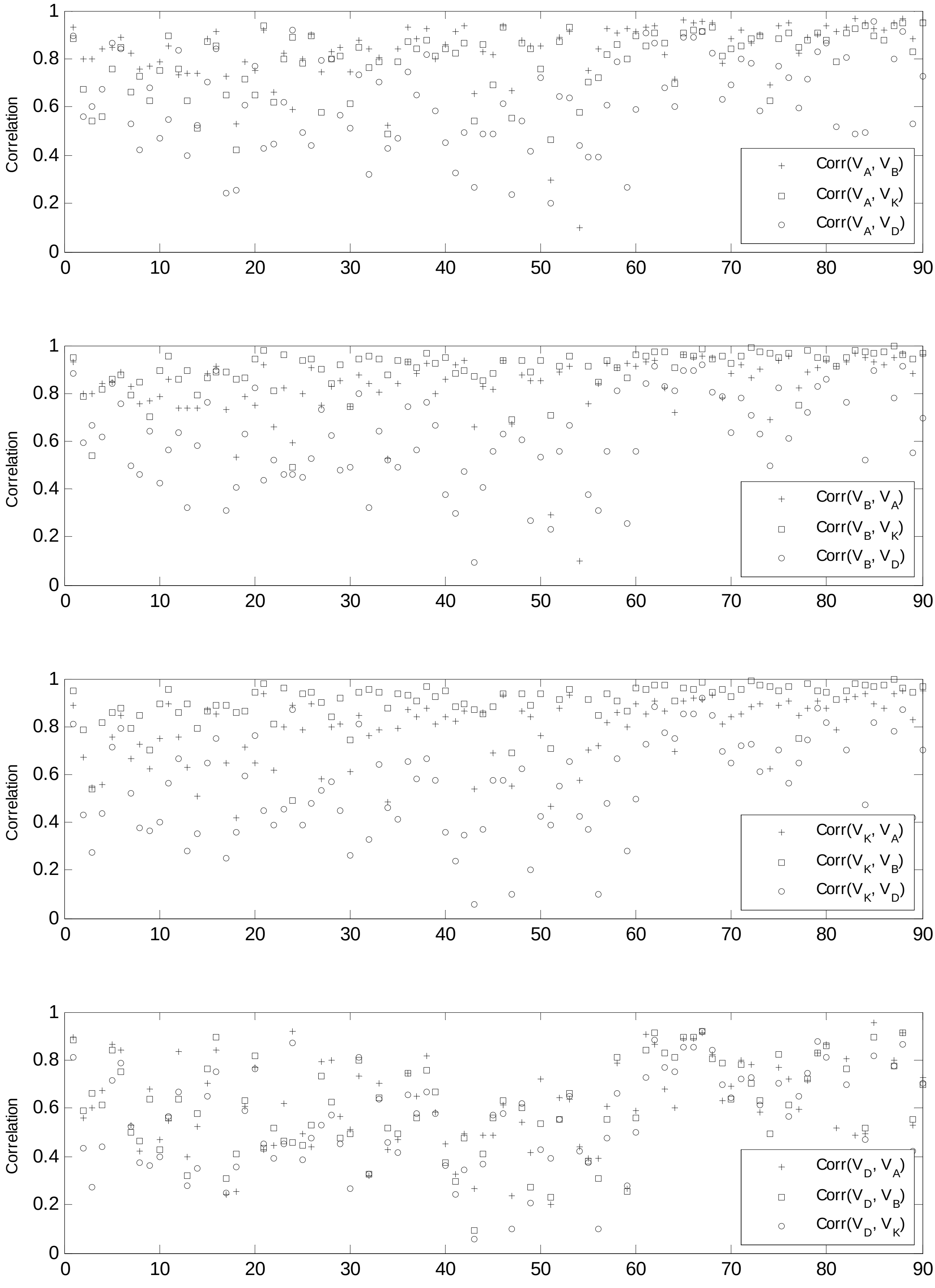


Figure 13: Correlation Coefficient of Volatility Estimates

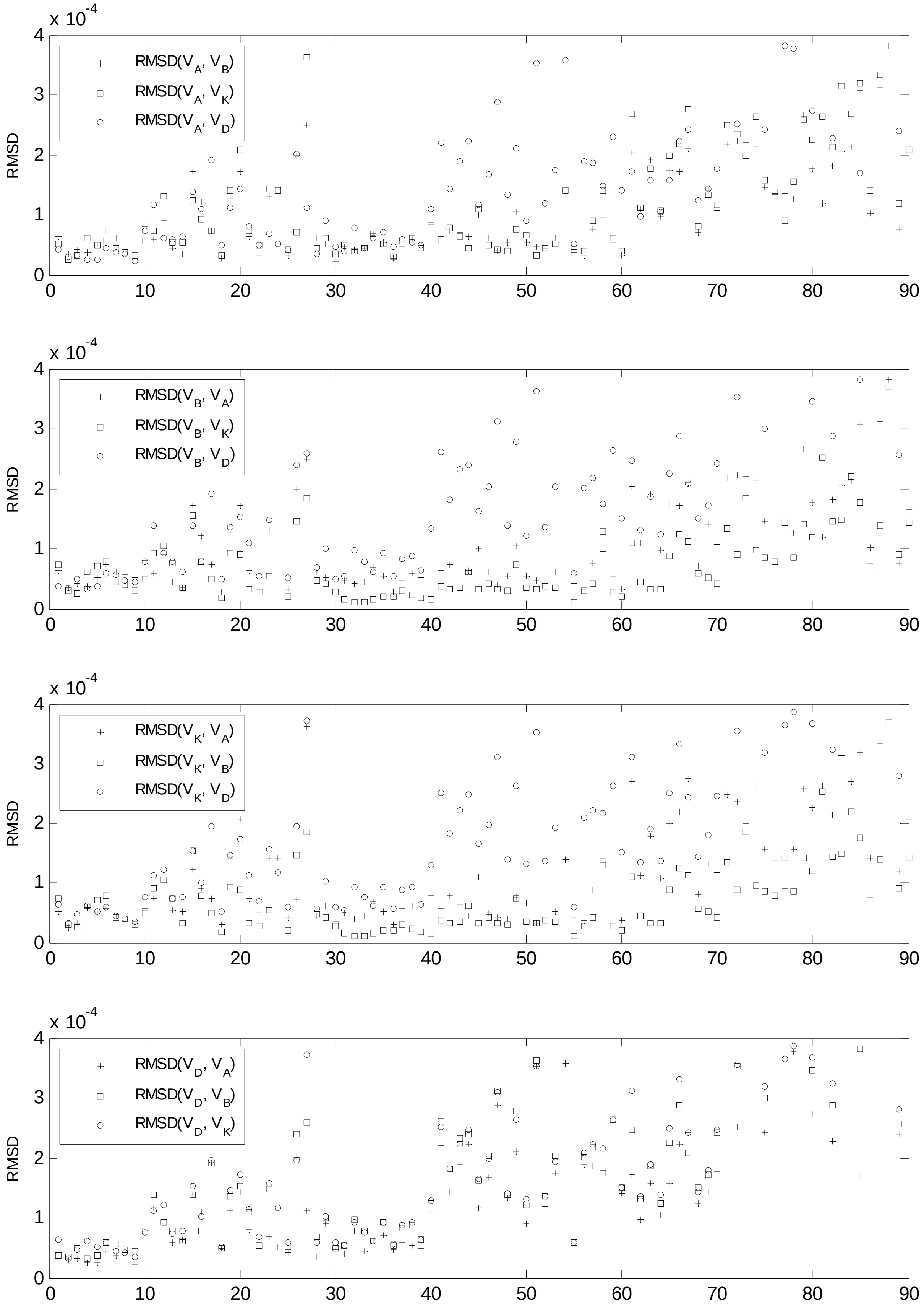


Figure 14: Root Mean-Squared Difference of Volatility Estimates

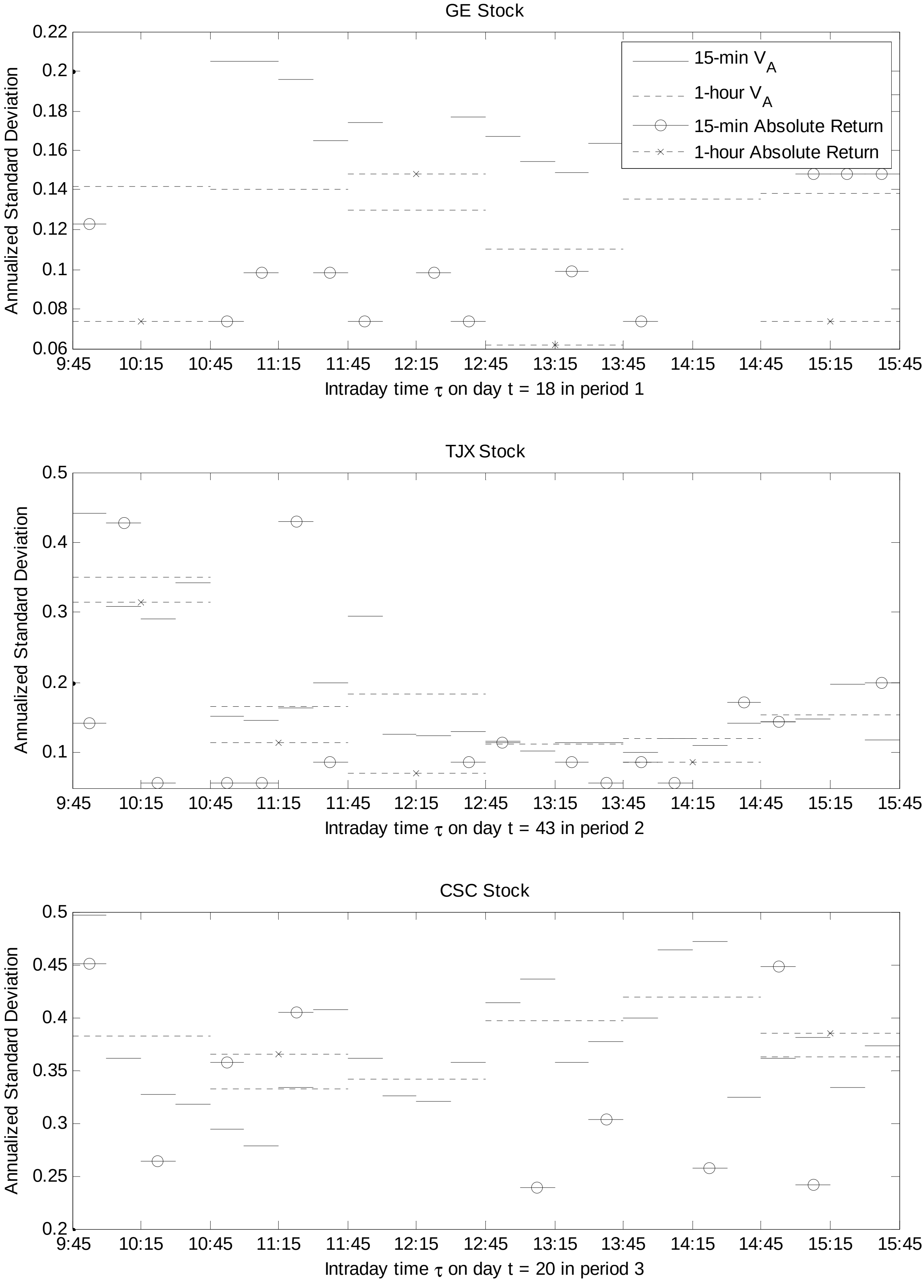


Figure 15: Intraday V_A Estimates of Three Selected NYSE Stocks