Beyond Merton's Utopia (I):
effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data

Xuezheng Bai, Jeffrey R. Russell, George C. Tiao*
University of Chicago, GSB

Abstract:
Volatility is central to financial theory. In practice volatility is not directly observed and must be estimated. Historically the sampling frequency was driven by data availability - first monthly, then weekly and daily. Now, we have at our disposal ultra high frequency data sets containing potentially 1000's of observations per day. Merton’s (1980) seminal work suggests that as the sampling interval approaches zero arbitrarily precise volatility estimates can be obtained. Realistically, however, the limiting case is not attainable since the sampling frequency cannot be any higher than transaction by transaction. We examine the precision of unconditional volatility estimates that use high-frequency data. We derive analytical expression for the precision of volatility estimates as a function of the prominent high frequency data characteristics including leptokurtosis, autocorrelation in the returns, deterministic patterns and volatility clustering in intra-day variance. Simulations are presented to explore the efficiency of some maximum likelihood estimates. Once these features are accounted for, we find that large amounts of high frequency data do not necessarily translate into very precise estimates. Our results provide a measure of the usefulness of high frequency data in estimating volatility.

Keywords: variance estimation, high frequency data, kurtosis, volatility clustering

* The authors would like to thank Torbin Andersen, Tim Bollerslev, John Cochrane, Rob Engle, Mark Jensen, Per Mykland, and participants at the 1999 NBER/NSF Time series meetings and the 2000 NBER forecasting and modeling in macro and finance seminars for very helpful comments.
Contents

1 Introduction 1

2 Notation 4

3 Features of High-Frequency Data 5

4 A Statistical Structure of Intra-Day Returns and Daily Variance Estimators 12
   4.1 A general statistical structure .............................. 13
   4.2 A fully specified model ................................... 14
   4.3 Daily volatility estimators using intra-day observations ................... 16

5 Effects of Non-normality and Dependence on the Precision of Variance Estimators 22
   5.1 Measuring the relative precision of variance estimators ...................... 22
   5.2 The effects of high frequency data characteristics on the sum of squares estimators 22
   5.3 Numerical Comparison of \( ER(\hat{V}_{\text{intra}}^{ss}, \hat{V}_{\text{day}}) \) and \( ER(\hat{V}_{\text{intra}}^{ssa}, \hat{V}_{\text{day}}) \) ............... 27
   5.4 The effect of high frequency data characteristics on the maximum likelihood estimators for fully specified model .......................... 29
   5.5 Empirical applications to the exchange rate series ......................... 32

6 Conclusion 35

Appendix. Variances for Sum of Squares Estimators \( \hat{V}_{\text{day}}, \hat{V}_{\text{intra}}^{ss} \) and \( \hat{V}_{\text{intra}}^{ssa} \) 37

Reference 43
List of Tables

1  Descriptive Statistics of Exchange Rate Returns ........................................ 7
2  Descriptive Statistics of Adjusted Exchange Rate Returns .......................... 12
3  MLE Estimates for the Adjusted Exchange Rate Return Series with Mixture Normal Innovations ................................................................. 17
4  Impacts of Autocorrelation and Kurtosis on $ER$ ............................................ 28
5  Impacts of Autocorrelation and Kurtosis on $ER$s (continued) ....................... 31
6  $ER(\hat{V}_{intra}^{ss}, \hat{V}_{day})$ for Exchange Rate Returns ............................... 34
7  $ER(\hat{V}_{intra}^{ssa}, \hat{V}_{day})$ for Exchange Rate Returns ............................... 34
8  $ER(\hat{V}_{intra}^{mle1}, \hat{V}_{day})$ and $ER(\hat{V}_{intra}^{mle2}, \hat{V}_{day})$ for Exchange Rate Returns ................................................................. 34

List of Figures

1  Intra-day Deterministic Pattern of Standard Deviation ................................ 8
2  QQplots of Adjusted Returns against Standard Normal Distribution .............. 9
3  Sample ACF of Adjusted Return Series ......................................................... 10
4  Sample ACF of Squared MA(1) innovations ................................................. 11
5  Diagnostic Plots for Adjusted Dollar/Deutsch Mark ..................................... 18
6  Diagnostic Plots for Adjusted Dollar/France Franc ....................................... 19
7  Diagnostic Plots for Adjusted Dollar/Japanese Yen ...................................... 20
1 Introduction

Volatility is central to financial theory and investment decisions. Basic mean variance analysis, for example, requires estimates of the variance for the assets under consideration. Sharpe ratios, often used to characterize risk adjusted returns, also require a variance estimate. Academics and practitioners have long recognized difficulty in obtaining accurate estimates of both means and variance associated with financial returns. Every time that a practitioner estimates a variance, a choice must be made about what frequency the data are to be sampled. Historically, the choice of the sampling frequency was driven by the data availability. In the early days of empirical finance weekly data was considered to be high frequency. Later, daily data became available. Now, we have intraday financial data giving us a picture of asset prices perhaps 1000’s of times within a single day. Despite the availability of intraday data it is still common today to use daily data to estimate the variance of an asset. What is the benefit of using intraday data to estimate the variance? The answer to this question will depend on the properties of the returns series.

It is somewhat surprising that given the importance of the variance in finance, very little research has examined the precision of variance estimates outside of the unrealistic iid normal assumption for asset returns. This is surprising since financial data, particularly when viewed at the weekly or higher frequency, exhibit strong departures from the iid normal assumption including large kurtosis and volatility clustering. This paper builds on work by Box (1953) and Bai et al. (2001b) to construct analytical representations for the precision of variance estimates that use intraday data. The precision of these estimates is then compared to the precision of estimates based on daily data. The ratio of the precision associated with the intraday and daily estimates can then be used to gauge the benefit of using the intraday data. Hence we provide both measures of the precision of variance estimates that use intraday data as well as a measure of the information content or benefit of using the high frequency data relative to the estimates based on the traditional daily data.

The idea that very high frequency data might be useful in estimating the variance is not new. Working with continuous time geometric Brownian motion, Merton (1980) demonstrated that although precise estimates of the mean return require long spans of data, precise estimates of the variance can be obtained over fixed time intervals provided that the sampling interval approaches zero. If the asset price follows a geometric Brownian motion and is sampled over time period
[0, T] at intervals $\Delta T$, straightforward calculations show that the variance of the return variance estimate is proportional to $\Delta T$, or the reciprocal of the number of observations over [0, T]. Recently, Nelson (1992), and Foster and Nelson (1996) extend the result to general continuous diffusion process and obtain similar result for conditional instantaneous variance.

In Merton’s spirit, we might think that with the intraday data now available we should expect that extremely precise estimates of volatility should be at hand. However, there are physical limitations associated with the data collection in financial market. In particular, the highest possible sampling rate is limited by the transaction frequency. Letting the sampling time interval shrink to zero is simply not possible. It is then natural to ask “how precisely can we estimate the volatility using high-frequency financial data?” Not surprisingly, the answer depends on the characteristics of the return series analyzed.

Financial data, particularly when viewed at a high-frequency, do not conform well with the geometric Brownian motion assumption. First, intra-day volatility clustering has been documented and analyzed by numerous studies. Early work includes Engle et al. (1990) for foreign exchange rate returns and Hamao et al. (1990) for equity index returns. Second, asset returns have consistently been documented to be leptokurtic, or fat tailed. The kurtosis tends to increase as the sampling interval shrinks. Andersen and Bollerslev (1998) report increasing kurtosis for the Dollar/Deutschmark exchange rate returns and for the S&P 500 index returns as the sampling frequency increases. For both series the kurtosis exceeds 20 when viewed at 5-minute intervals. Third, Harris (1986), Derity and Mulherin (1992), and Baillie and Bollerslev (1991) have examined the deterministic intra-day behavior of volatility. Among others, these studies suggest that higher volatility occurs near the open and close in the equity markets, or occurs when certain regions of the world are actively transacting in the continuously operated foreign exchange markets. Finally, asset returns tend to be serially correlated. Over time horizons of weeks or months the serial correlation may be quite small, but over short time horizons like daily or intra-day, the correlation can be very strong.

This paper derives expressions for the $MSE$ of variance estimates in the presence of volatility clustering, leptokurtosis, deterministic patterns in the volatility structure, and serial correlation of the returns. As is often the case in practice we do not wish to make any parametric assumptions regarding the dynamics of the return process when estimating the variance. We therefore use estimates based on the sum of squared returns which do not require parametric assumptions. The
precision of these estimates is then expressed as a function of the moments of the return process only. Since it is likely that making further parametric assumptions regarding the intra-day return dynamics would lead to better estimates of the variance we also conduct a simulation experiment under the assumption of a GARCH process with a mixture of normal, leptokurtic iid innovation. It should be made clear, however, that only the simulation results rely on the GARCH structure, not the results for the sum of squares estimators.

Our work is related to two other branches of literature. Andersen et al. (2001) use intraday data to estimate the variance of a particular day. They do not construct measures of precision for their estimates. Here, we focus on the more basic question of deriving expressions for the precision of the unconditional variance of daily returns. We consider the precision of ex-post estimates in a companion paper, Bai et al. (2001a). Another related branch is the Drost and Nijman (1993) GARCH aggregation results. They derive the relationship between high and low frequency GARCH parameters. Again, they do not consider the precision of the estimates as we do in this paper. Also, for the sum of squares type estimators we do not impose any model, but rather rely on moments of the return process. Hence, we believe this is the first paper to construct measures for the precision of variance estimates that use high frequency data. Our focus on the unconditional daily variance is purely for exposition. The theoretical results can be applied to estimates of the variance for any time period of interest such as weekly or monthly.

The outline of the paper is as follows. In the following section we introduce notation and discuss the use of high-frequency data in estimating the variance of daily returns. In Section 3, we document several features of high-frequency exchange rate returns. These features are related to findings in earlier empirical work and a set of characteristics of high-frequency data is established. Section 4 proposes a general statistical structure consistent with the established features, and discusses the daily volatility estimators that use the intra-day returns. In Section 5, theoretical results are presented relating the $MSE$ of each estimator to the characteristics of the high-frequency data discussed. Finally, Section 6 concludes.
2 Notation

Consider a sequence of prices observed at unit time intervals, and denote the logged sequence of prices by \( p_1, p_2, \ldots, p_N \) and the corresponding returns by \( r_n = p_n - p_{n-1} \). In the following, we take the unit of time to be one day although the results can be easily carried over to other time periods. We assume that within a day the price is observed on \( k \) evenly spaced occasions. We denote the \( k \) logged intra-day prices for day \( n \) by \( \{ p_{nt}, t = 1, \ldots, k \} \) and the \( t^{th} \) intra-day return by \( x_{nt} = p_{nt} - p_{n,t-1} \), where \( p_{n0} = p_{n-1,k} \). By transforming the double subscript \((n, t)\) into the single subscript \( s(n, t) = (n - 1) \cdot k + t \), the intra-day return series \( \{ x_{nt} : n = 1, 2, \ldots, N, t = 1, 2, \ldots, k \} \) becomes a univariate series \( \{ x_s : s = 1, \ldots, N \cdot k \} \). For convenience, both notations \( x_{nt} \) and \( x_s \) will be used interchangeably throughout this paper. To avoid confusion, we use \( x_{s(n,t)} \) to indicate the dependence of the single subscript \( s \) on the double subscript \((n, t)\) explicitly.

The intra-day returns \( \{ x_{nt} : t = 1, \ldots, k \} \) on day \( n \) can also identified as a \( k \)-dimensional vector \( x_n = (x_{n1}, x_{n2}, \ldots, x_{nk})^t \). The daily return \( r_n \) is simply the sum of the intra-day returns:

\[
\sum_{t=1}^k x_{nt} \tag{2.1}
\]

In this paper, we are interested in using the intra-day return to estimate the variance of daily return \( r_n \), which is assumed to follow a covariance stationary process. We denote the mean daily return by \( \mu \) and the variance by \( \sigma^2 \).

Let \( x_n \) have joint distribution \( g(x) \) with mean vector and variance-covariance matrix given by,

\[
\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}, \quad \Omega = \begin{bmatrix} \sigma^2_1 & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma^2_2 & & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \cdots & \sigma_{k,k-1} & \sigma^2_k \end{bmatrix} \tag{2.2}
\]

If \( x'_n \)'s have the same joint distribution, we have the following identities:

\[
\mu = \mu_1 + \mu_2 + \cdots + \mu_k \tag{2.3}
\]
\[
\sigma^2 = \sum_{t=1}^k \sigma^2_t + 2 \sum_{t<s} \sigma_{st} \tag{2.4}
\]

If \( \Omega \) is diagonal, \( \sigma^2 \) simply reduces to the sum of the \( k \) intra-day variances \( \sigma^2_t \). In general, \( \sigma^2 \) is the sum of the \( k \) intra-day variances \( \sigma^2_t \) and terms \( \sigma_{st} \) that account for the correlations among the intra-day returns.
In general, the intra-day returns \( x_{nt} \) are correlated between days, but such relations don’t affect the calculation of variance \( \sigma^2 \) of daily returns. For simplicity, we assume that \( \mu = 0 \) in the following.

The financial data at daily frequency or higher are not collected consecutively in calendar time, either because the data at weekends are not available, or because the data at weekends are very different from the data at weekdays and thus treated differently. We use the same notation \( x_{nt} \) or \( x_{s(n,t)} \) for both consecutive or non-consecutive data which should be clear from the context. If necessary, we will use the subscript \( n^* \) instead of \( n \) to indicate that the data may be collected non-consecutively in calendar time.

### 3 Features of High-Frequency Data

When viewed over monthly or longer periods, many financial return series are approximately normal white noise. As the sampling period shrinks, however, temporal dependence and departures from normality become strikingly apparent. In fact, at the transaction-by-transaction frequency, the normal approximation is extremely poor since the irregular spacing between transactions, discreteness in transaction prices and dependence introduced by market micro-structure become dominant features of the data\(^1\). We consider financial data sampled at fixed intervals within the day. In particular, our empirical work focuses on half-hour foreign exchange rate data.

Several features are present in a broad class of intra-day financial data and are consistently observed in the literature at various intra-day sampling frequencies.

- First, volatility clustering is present in the intra-day data. Early work includes Engle, Ito and Lin (1990) for foreign exchange rate returns and Hamao, Masolis and Ng (1990) for equity index returns. Much applied work uses the ARCH type model of Engle (1982) and Bollerslev (1986) or the stochastic volatility models to account for this feature\(^2\).

- Second, intra-day volatility also appears to have a deterministic component. In the foreign exchange rate data, volatility varies as different regions of the world become active in the market. In stock data we tend to find the periods just after the open and just prior to the close are


\(^2\)See for example proceedings of the 2nd Olsen and Associates Conference on High-Frequency Data.
more volatile than the normal trading period. Harris (1986), Derity and Mulherin (1992), and Baillie and Bollerslev (1991) have examined the deterministic intra-day behavior of volatility. More recently, Andersen and Bollerslev (1998) document these patterns in volatility for equity and foreign exchange rate data.

• Third, returns are negatively autocorrelated. This correlation tends to become stronger as the sampling frequency increases.

• Fourth, as the sampling frequency increases, the kurtosis in the data can become exceptionally large. Andersen and Bollerslev (1998), for example, find a kurtosis of 18.5 for the 5-minute Dollar/Deutsch mark returns in their study. Chan and Karolyi (1991) find kurtosis ranging from 10 to over 25 for 5-minute S&P500 returns over several sub-samples ranging from 1984 to 1989. While it is well understood that time-varying volatility (stochastic or deterministic) increases the overall kurtosis, the common models used, such as the GARCH with conditional normal errors, cannot account for the exceedingly large kurtosis typically observed in the data, e.g., Bollerslev (1987) and Engle and Gonzalez-Rivera (1991). Hence even after accounting for the time-varying volatility, it is still unlikely that the intra-day returns have conditional normal distribution.

The data used in our empirical work are foreign exchange rate data obtained from Olsen and Associates. We analyze one complete year (1996) of half-hour returns on the Dollar/Deutsch mark (DD), Dollar/French Franc (DF) and Dollar/Japanese Yen (DY). It is well known that returns over the weekend, when many markets are inactive, are very different from returns during the week. Thus we only use the data from 12:00am Tuesday morning to 12:00 am Friday morning. This results in a sample of 156 days and 7488 half-hour returns. We now examine the half hour returns for the three series.

From descriptive statistics given in Table 1, all three return series have mean zero, are slightly skewed to the left and are extremely leptokurtic. The statistics presented in Table 1 are affected by the deterministic intra-day volatility pattern typically observed in such returns. From the intra-day returns series, we construct the sample variance $s_t^2$ by time of day. This is given by,

$$s_t^2 = \frac{1}{N} \sum_{n=1}^{N} x_{nt}^2, \ t = 1, 2, \cdots, k$$  (3.5)
Table 1: Descriptive Statistics of Exchange Rate Returns

<table>
<thead>
<tr>
<th>series</th>
<th>mean</th>
<th>std</th>
<th>skewness</th>
<th>kurtosis$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD</td>
<td>$0.0975 \times 10^{-4}$</td>
<td>$6.8270 \times 10^{-4}$</td>
<td>-0.3373</td>
<td>10.0316</td>
</tr>
<tr>
<td>DF</td>
<td>$0.0841 \times 10^{-4}$</td>
<td>$8.2022 \times 10^{-4}$</td>
<td>-0.2731</td>
<td>27.5292</td>
</tr>
<tr>
<td>DY</td>
<td>$0.0932 \times 10^{-4}$</td>
<td>$8.4762 \times 10^{-4}$</td>
<td>-0.2033</td>
<td>6.3823</td>
</tr>
</tbody>
</table>

This table provides the descriptive statistics of half hour returns for Dollar / Deutsch Mark (DD), Dollar / French Franc (DF) and Dollar / Japanese Yen (DY). Each series have 7488 observations.

Figure 1 shows plots of the sample standard deviations by time of day. The time of day on the horizontal axis is Greenwich Mean Time (GMT).\(^4\) Consistent with previous studies, we find that the most active period is in the afternoon GMT when both U.S. market and European market are active. Clearly, the returns cannot be regarded as homogeneous throughout a day. To partial out the time of day effect, we adjust the returns by dividing the intra-day returns by their corresponding intra-day sample standard deviations,

\[ \tilde{x}_{nt} = \frac{x_{nt}}{s_t} \]  

(3.6)

The adjusted series $\tilde{x}_{nt}$ are normalized to have unconditional variances close to 1, and should be free of deterministic volatility pattern. Throughout this paper, this is termed the adjusted return series. The descriptive statistics for the three adjusted return series are given in Table 2.

As expected, the adjusted return series have smaller overall kurtosis. However, the kurtosis is still quite large. Further examination of the QQplot in Figure 2 shows deviations in the tails of the distribution, indicative of leptokurtosis.

The sample autocorrelation function (ACF) for the three adjusted return series are presented in Figure 3. For each of the three series only the lag-one sample autocorrelation is significant, which is indicative of a moving average model of order one, MA(1), structure. Negative lag-one

$^4$Here, the kurtosis for a random variable $X$ is defined by $\gamma_2 = \frac{E[(X - E[X])^4]}{Var(X)^2} - 3$. The definition given here is sometimes referred to as excess kurtosis.

$^4$We first considered the deterministic patterns for the 3 days separately, and found that their patterns are similar. We thus treat them as the same.
Figure 1: Intra-day Deterministic Pattern of Standard Deviation

The horizontal axis is the time of day G.M.T. and the vertical axis is the intra-day standard derivation. Each series contains 7488 half-hour returns from 156 days. The sample standard deviation at each time of day is calculated from the 156 returns at that time.
The adjusted series, $\tilde{D}D$, $\tilde{D}F$ and $\tilde{D}Y$, are obtained from the original return series, DD, DF and DY, by factoring out their deterministic intra-day volatility pattern, respectively.
Figure 3: Sample ACF of Adjusted Return Series

The adjusted series are obtained from the original return series by factoring out the deterministic intra-day volatility pattern.
For each series of adjusted returns, the innovations are residuals obtained from the fitted normal MA(1) model.
Table 2: Descriptive Statistics of Adjusted Exchange Rate Returns

<table>
<thead>
<tr>
<th>series</th>
<th>mean</th>
<th>std</th>
<th>skewness</th>
<th>kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tilde{DD} )</td>
<td>0.0166</td>
<td>0.9999</td>
<td>-0.0802</td>
<td>8.1551</td>
</tr>
<tr>
<td>( \tilde{DF} )</td>
<td>0.0135</td>
<td>1.0000</td>
<td>-0.3233</td>
<td>13.4978</td>
</tr>
<tr>
<td>( \tilde{DY} )</td>
<td>0.0115</td>
<td>1.0000</td>
<td>-0.3874</td>
<td>5.2359</td>
</tr>
</tbody>
</table>

The adjusted series \( \tilde{DD}, \tilde{DF} \) and \( \tilde{DY} \) are obtained from the original return series DD, DF and DY by factoring out their deterministic intra-day volatility pattern, respectively.

Autocorrelation is commonly observed in intra-day returns. It is often considered the result of bid-ask bounce in equity markets and noisy quotes in foreign exchange markets.\(^5\)

Figure 4 presents the sample ACF of the squared residual series from the fitted normal MA(1) model for the adjusted return series. The large positive autocorrelation at lag one indicates strong correlation of the squared residuals at short lags, and the long span of positivity in the ACF plots is indicative of volatility clustering.

4 A Statistical Structure of Intra-Day Returns and Daily Variance Estimators

In this section we first propose a simple structure for the variance-covariance matrix of intra-day returns and their squares. This structure is consistent with the documented features, and sufficient to derive the \( MSE \) of the sum of squared estimators discussed later in this section. Next, we propose a fully specified model that is consistent with the structure proposed here. This parametric model will be used in studying the efficiency of the maximum likelihood estimator of the daily volatility. Finally, we discuss several daily volatility estimators that use the intra-day returns, including two commonly used sum of squared estimators and two maximum likelihood type estimators.

\(^5\) Hasbrouck (1988) discusses how market micro-structure noise can generate this autocorrelation structure.
4.1 A general statistical structure

The discussion on the features of high-frequency financial data in Section 3 suggests that the variance-covariance matrix of $x_n$ need not to be fully saturated as we find evidence of lag-one autocorrelation only. More specifically, the intra-day deterministic volatility pattern in Figure 1 and the lag-one autocorrelation of the adjusted returns in Figure 3 suggest the following stochastic structure for the intra-day returns, in terms of the univariate series $x_{s(n,t)}$

$$\frac{x_{s(n,t)}}{\delta_t} = \epsilon_{s(n,t)} + \theta \epsilon_{s(n,t)-1}$$ (4.1)

where $\epsilon_{s(n,t)}$ are uncorrelated with mean 0 and variance 1, and $\{\delta_t^2, t = 1, \cdots, k\}$ are deterministically time-varying. Equivalently, the model states that the intra-day returns $x_{nt}$ within a day have the following variance-covariance matrix:

$$\Omega = \text{cov}[x_n] = \begin{bmatrix} \delta_1 & \delta_2 & \cdots & \delta_k \\ \delta_2 & 1 + \theta^2 & \theta & \cdots & 0 \\ \vdots & \theta & 1 + \theta^2 & \theta & \cdots \\ \delta_k & 0 & \theta & \cdots & 1 + \theta^2 \end{bmatrix}$$ (4.2)

This model also relates the intra-day returns between the days. More explicitly, the first intra-day return on the next day is correlated with the last intra-day returns on the previous day, and is uncorrelated with other previous intra-day returns.

Figure 4 suggests that, while the innovation $\epsilon_{s(n,t)}$, or $\epsilon_{nt}$ in double-subscript, in (4.1) are uncorrelated, the squared innovations $\epsilon_{s(n,t)}^2$ are dynamically related. If we assume that the fourth moment of $\epsilon_{s(n,t)}$ exists and that $\epsilon_{s(n,t)}^2$ is covariance stationary, then the covariance between two squared innovations depends only on the time interval between them. We can therefore represent the auto-covariance of $\epsilon_{s(n,t)}^2$ as:

$$\text{cov}(\epsilon_{s(n,t)}^2, \epsilon_{s(n,t)-\ell}^2) = 2(1 + \frac{\gamma_2}{2}) \cdot \rho_\ell$$ (4.3)

where $\gamma_2$ is the kurtosis of $\epsilon_{s(n,t)}$. Since the variance of $\epsilon_{s(n,t)}$ is normalized to 1, it is easy to show that the variance of $\epsilon_{s(n,t)}^2$ is $2(1 + \frac{\gamma_2}{2})$. Note that if $\epsilon_{s(n,t)}$ is normal, then $\gamma_2 = 0$. The sample kurtosis of the adjusted series in Table 2 suggests a positive value of $\gamma_2$. The expression (4.3) shows explicitly the dependence of the variances and covariances of the $\epsilon_{s}^2$s on the kurtosis $\gamma_2$ of the $\epsilon_{s}'$s.
Autocorrelation of nearby squared innovations are typically positive in financial data, implying that $\rho_{\ell}$s are positive at least for small $\ell$. Furthermore, Figure 4 suggests the strong volatility persistence with positive autocorrelations for squared innovations even farther apart in time. Hence, most $\rho_{\ell}$s should be positive though they may decrease as $\ell$ increases.

A simple structure consistent with these characteristics is given by

$$\rho_j = \text{corr}(\epsilon_{s(n,t)}^2, \epsilon_{s(n,t)-j}^2) = \rho_1 \cdot \phi^{j-1}$$

with $0 < \rho_1 < 1$ and $0 < \phi < 1$. In other words, the squared innovations have an ARMA(1, 1) autocorrelation structure. In this structure, $\rho_1$ is the lag-one autocorrelation of $\epsilon_{s(n,t)}^2$ and determines the level of the volatility clustering, and $\phi$ is the rate of decay in the volatility clustering. As will be discussed in Section 4.2, the frequently used GARCH(1, 1) model and linear stochastic volatility AR(1) model are two parametric specification consistent with (4.4).

The structure given by (4.1), (4.3) and (4.4) specifies the variance-covariance structure for $x_{s(n,t)}$ and $x_{s(n,t)}^2$. This structure is consistent with the features documented in the previous section: deterministic intra-day volatility, serial correlation, fat tails and volatility clustering. With some further technical assumptions, this structure is sufficient to study the impact of documented market micro-structure effects on the precision estimation of sum of squared volatility estimates.

### 4.2 A fully specified model

The structure proposed in Section 4.1, however, does not provide a complete parametric specification that be required for MLE type estimators. This section proposes a model that will be used to study the MLE type volatility estimators.

Candidates for the parametric specifications of the statistical structure in (4.3) include GARCH and stochastic volatility models. Particularly, the GARCH(1, 1) and linear stochastic volatility AR(1) models can be regarded as special parametric representations of (4.3) and (4.4). In the following, we focus our discussion on the GARCH(1, 1) model. The discussion for the linear stochastic volatility AR(1) model is similar.

Consider the GARCH(1, 1) model for the innovation $\epsilon_{s(n,t)}$ in (4.1),

$$\epsilon_{s(n,t)} = \sqrt{h_{s(n,t)}} z_{s(n,t)}$$

$$h_{s(n,t)} = \omega + \alpha \epsilon_{s(n,t)-1}^2 + \beta h_{s(n,t)-1}$$

with $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $0 < \lambda < 1$. The parameters $\omega$, $\alpha$, $\beta$, $\lambda$ are estimated by maximum likelihood. The precision of these estimates is affected by the market micro-structure effects.
where $h_{s(n,t)}$ is the conditional variance of $\epsilon_{s(n,t)}$, and $z_{s(n,t)}$ are iid with mean 0 and variance 1. It is well known that the GARCH(1, 1) formulation implies that $\epsilon_{s(n,t)}^2$ follow the ARMA(1, 1) covariance structure provided that the 4th moment of $\epsilon_{s(n,t)}$ exists, and thus it is consistent with our general structure.

While (4.6) provides a good representation for volatility clustering, the GARCH(1, 1) with conditional normal innovations that matches the volatility clustering pattern in the data is not capable of capturing the exceedingly large sample kurtosis. For exchange rate returns analyzed in this paper, we find that the kurtosis induced by the conditional normal GARCH(1, 1) model is far smaller than the kurtosis observed in the sample. Similar findings have been reported in the literature. To capture the large kurtosis, Bollerslev (1987) used t-distributions for the iid innovations $z_{s(n,t)}$ and Engle and Gonzalez-Rivera (1991) propose a non-parametric approach.

Here we propose using a mixture of two normals with common zero mean but different variances to capture the large kurtosis. This model can generate kurtosis of any positive value. Specifically, we assume that

$$z_{s(n,t)} \sim \begin{cases} N(0, \sigma_0^2) & \text{with probability } 1 - \eta \\ N(0, \lambda \sigma_0^2) & \text{with probability } \eta \end{cases}$$

(4.7)

where the three parameters $(\sigma_0^2, \eta, \lambda)$ are constrained by $(1 - \eta + \lambda \cdot \eta)\sigma_0^2 = 1$ since $z_{s(n,t)}$ is normalized to have unit variance. One can easily find the kurtosis of $z_{s(n,t)}$ is given by $\gamma^z_2 = 3(\lambda - 1)^2(\xi + \lambda)^{-2}$, where $\xi = (1 - \eta)\eta^{-1}$ is the odds ratio.

It is interesting to explicitly relate the parametric GARCH(1, 1) model (4.6) with the normal mixture (4.7) to the general structure in (4.3) and (4.4). For volatility clustering, it is easy to show that $\rho_1 = \frac{\alpha(1 - \alpha + \beta)\beta}{1 - (2\alpha + \beta)\beta}$ and $\phi = \alpha + \beta$. For the kurtosis $\gamma_2$ of $\epsilon_{s(n,t)}$, Bai, Russell and Tiao (2001c) have obtained the following simple relation among the kurtosis:

$$\gamma_2 = \frac{\gamma^z_2 + \gamma^g_2 + \frac{5}{6}\gamma^z_2 \gamma^g_2}{1 - \frac{5}{6}\gamma^z_2 \gamma^g_2}$$

(4.8)

where $\gamma_2$ is the kurtosis of $\epsilon_{s(n,t)}$, $\gamma^z_2$ is the kurtosis of $z_{s(n,t)}$ and $\gamma^g_2$ is the kurtosis of the conditional normal GARCH process. The conditional normal GARCH(1, 1) process has kurtosis $\gamma^g_2 = \frac{6\alpha^2}{1 - (\alpha + \beta)^2 - 2\alpha^2}$. It is clear from (4.8) that the mixture normal innovations gives rise to arbitrary large overall kurtosis.

In the following, we will call $\gamma^z_2$ the iid kurtosis and $\gamma_2$ the overall kurtosis. Obviously, if
there is no volatility clustering, the overall kurtosis equals the iid kurtosis, and we will simply call it the kurtosis.

The fully parametrized model given by (4.1), (4.6) and (4.7) will be used in Section 5 to study the precision of maximum likelihood volatility estimates. Now, we estimate this model and perform model diagnostics to demonstrate that it provides a good representation of the data. Table 3 gives the maximum likelihood estimates for the three adjusted series and Figures 5-7 present the ACF plots of the residual, the ACF plots of the squared standardized residuals and the QQplots. We can conclude that our model fits the three adjusted return series well and captures the documented features in these series. In addition, the implied kurtosis for the adjusted Dollar/Deutsch Mark and Dollar/Japanese Yen series are close to the sample kurtosis in Table 2. This suggests that the model is adequate in capturing the fat-tail property. On the other hand, the implied kurtosis for the adjusted Dollar/France Franc is $\infty$. This may be due to possible mis-specification and is left to further investigation.

4.3 Daily volatility estimators using intra-day observations

We are interested in estimating $\sigma^2$ or the variance of daily returns. First, consider estimating $\sigma^2$ from daily returns. Since for many financial assets the daily returns are essentially uncorrelated, it is common to consider the sum of squares estimator for the daily variance,

$$\hat{V}_{\text{day}} = \frac{1}{N} \sum_{n^*=1}^{N} r_n^2$$

(4.9)

where $N$ is the number of days. Note here the subscript $n^*$ indicates that the summation is taken over all available squared daily returns, which may not be consecutive in calendar time. If the daily returns are iid normal, this estimator will be efficient. In general, it is unbiased.

Several estimators of low frequency variances that use high-frequency data have been proposed and used in the literature. Perhaps the simplest and most frequently used estimator is the one obtained by summing the squared high frequency returns. This type of estimator has been used in

---

6The deterministic pattern $\delta_t$ could be estimated for the fully specified model. For simplicity, however, we assume that $(\delta_1, \cdots, \delta_k)$ are known up to a constant. More explicitly, we assume that $(\delta_1, \cdots, \delta_k)$ is proportional to the sample intra-day standard deviation pattern $s_1, s_2, \cdots, s_k$. 

16
Table 3: MLE Estimates for the Adjusted Exchange Rate Return Series with Mixture Normal Innovations

<table>
<thead>
<tr>
<th>Series</th>
<th>( \theta )</th>
<th>( \omega )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \eta )</th>
<th>( \lambda )</th>
<th>( \gamma_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tilde{DD} )</td>
<td>-0.1415 (0.0115)</td>
<td>0.1328 (0.0157)</td>
<td>0.1296 (0.0125)</td>
<td>0.7372 (0.0230)</td>
<td>0.1163 (0.0103)</td>
<td>6.9950 (0.3159)</td>
<td>8.3161</td>
</tr>
<tr>
<td>( \tilde{DF} )</td>
<td>-0.2139 (0.0115)</td>
<td>0.1301 (0.0102)</td>
<td>0.2419 (0.0182)</td>
<td>0.6484 (0.0184)</td>
<td>0.1795 (0.0109)</td>
<td>7.7964 (0.2734)</td>
<td>( \infty )</td>
</tr>
<tr>
<td>( \tilde{DY} )</td>
<td>-0.1765 (0.0113)</td>
<td>0.0807 (0.0100)</td>
<td>0.1085 (0.0103)</td>
<td>0.8122 (0.0165)</td>
<td>0.1539 (0.0148)</td>
<td>5.7367 (0.2753)</td>
<td>6.5861</td>
</tr>
</tbody>
</table>

This table provides the maximum likelihood estimates and their estimated standard errors (in parentheses) for the model

\[
x_{s(n,t)} = \epsilon_{s(n,t)} + \theta \epsilon_{s(n,t)-1}, \quad h_{s(n,t)} = \omega + \alpha \cdot \epsilon_{s(n,t)-1}^2 + \beta \cdot h_{s(n,t)-1}
\]

where \( h_{s(n,t)} = \text{var}(\epsilon_{s(n,t)}|\epsilon_{s(n,t)-1}, \cdots) \), and

\[
h_{s(n,t)}^{-1/2} \epsilon_{s(n,t)} \sim \begin{cases} N(0, \sigma_0^2) & \text{with probability } 1 - \eta \\ N(0, \lambda \sigma_0^2) & \text{with probability } \eta \end{cases}
\]

The \( \gamma_2 \) in the last column is the kurtosis of \( \epsilon_{s(n,t)} \) implied by the model, and is calculated from (4.8).

To derive the above estimates, we assumed that adjusted returns series is free of intra-day deterministic pattern but with unknown unconditional variance. Equivalently, we assumed the deterministic patterns are proportional to the sample intra-day standard deviations. In the estimation, we treat the whole series as consecutive.
Plot (a) gives the sample ACF of the fitted residuals $\hat{z}_{s(n,t)}$, and Plot (b) gives the sample ACF of the squared fitted residuals $\hat{z}^2_{s(n,t)}$.

Plot (c) is the QQplot of adjusted Dollar/Deutsch Mark returns versus the distribution implied by the fitted model. In other words, we use the estimated model for the adjusted Dollar/Deutsch Mark returns to simulate a series, and present the QQplot for the data versus this simulated series.
Plot (a) gives the sample ACF of the fitted residuals $\hat{z}_{s(n,t)}$, and Plot (b) gives the sample ACF of the squared fitted residuals $\hat{z}_{s(n,t)}^2$.

Plot (c) is the QQplot of adjusted Dollar/French Franc returns versus the distribution implied by the fitted model. In other words, we use the estimated model for the adjusted Dollar/French Franc returns to simulate a series, and present the QQplot for the data versus this simulated series.
Plot (a) gives the sample ACF of the fitted residuals $\hat{z}_{s(n,t)}$, and Plot (b) gives the sample ACF of the squared fitted residuals $\hat{z}_{s(n,t)}^2$.

Plot (c) is the QQplot of adjusted Dollar/Japanese Yen returns versus the distribution implied by the fitted model. In other words, we use the estimated model for the adjusted Dollar/Japanese Yen returns to simulate a series, and present the QQplot for the data versus this simulated series.
Poterba and Summers (1986), French, Schwert and Stambaugh (1987), and Schwert (1989; 1990). In these studies the ex-post estimate for the variance of a given month was estimated by sampling at daily frequency. Here we focus on estimation of the unconditional variance but this previous research suggests possible estimate that uses intra-day returns to consider. More formally we define the simple sum of squares estimator by

\[
\hat{V}_{\text{intra}}^{ss} = \frac{1}{N} \sum_{n^*=1}^{N} \sum_{t=1}^{k} x_{s(n^*,t)}^2
\]

(4.10)

If the returns are uncorrelated this estimator will be unbiased. Otherwise, this estimator is biased and the magnitude of bias depends on the off-diagonal terms in the variance covariance matrix \(\Omega\) in (2.2). Another simple yet more robust estimator that allows for some serial dependence is given by

\[
\hat{V}_{\text{intra}}^{ssa} = \frac{1}{N} \sum_{n^*=1}^{N} \left( \sum_{t=1}^{k} x_{s(n^*,t)}^2 + 2 \sum_{t=1}^{k-1} x_{s(n^*,t)} \cdot x_{s(n^*,t)+1} \right)
\]

(4.11)

This estimator is simply \(\hat{V}_{\text{intra}}^{ss}\) plus terms that account for the lag-one autocorrelation, and will be unbiased if at most lag-one autocorrelation exists in the returns. This estimator has been used by French, Schwert and Stambaugh (1987).

It is important to note that for these studies, only the daily returns within a given month are used to estimate the volatility of the particular month. That is, \(N = 1\). Here we consider using intra-day returns across many days to estimate unconditional variance of daily returns.

The above two intra-day return based estimators, \(\hat{V}_{\text{intra}}^{ss}\) and \(\hat{V}_{\text{intra}}^{ssa}\), are constructed without any explicit use of the statistical structure proposed in Section 4.1 and do not require parametric assumption on the distribution of \(x_n\). We refer to these as sum of squared estimators. We now turn to some maximum likelihood estimators. Such estimators will make parametric assumptions but they may also be more efficient and are therefore interesting to consider. It is easy to show that under the structure of (4.1), (4.3) and (4.4), the unconditional variance of the daily return takes the following form

\[
\sigma^2 = (1 + \theta^2) \sum_{t=1}^{k} \delta_t^2 + 2\theta \sum_{t=1}^{k-1} \delta_t \delta_{t+1}
\]

(4.12)

Thus, for given estimators of the deterministic pattern \((\delta_1^2, \delta_2^2, \ldots, \delta_k^2)\) and MA(1) coefficient \(\theta\), the corresponding estimator for the daily volatility \(\sigma^2\) can be constructed. While the daily volatility estimates do not involve terms beyond the deterministic component and the lag one autocorrelation, the precision of these estimates will depend on the remaining two features, leptokurtosis and
volatility clustering, as outlined in section 3. In particular, an estimate of the daily volatility is obtained using the maximum likelihood estimates of the parameters on the right hand side of 4.12. We consider two maximum likelihood estimators. The first one, denoted by $\hat{V}^{\text{mle}}_{\text{intra},1}$, is maximum likelihood estimator under the assumptions (4.1), (4.6) and that $z_{s(n,t)}$ follows the standard normal distribution. If the normal specification for $z_{s(n,t)}$ is not correct the estimate $\hat{V}^{\text{mle}}_{\text{intra},1}$ will be a quasi maximum likelihood estimate and will still be consistent albeit inefficient. The second one, denoted by $\hat{V}^{\text{mle}}_{\text{intra},2}$, is the maximum likelihood estimator under the fully specified model of (4.1), (4.6) and (4.7).

5 Effects of Non-normality and Dependence on the Precision of Variance Estimators

The statistical structure presented in the preceding section is consistent with the established features of the high-frequency data, and thus provides the framework to determine the impacts of various features on the precision of daily variance estimators. In this section, we first investigate the impact of the various features of the data on the precision of the commonly used intra-day sum of squares estimators given in (4.10) and (4.11) using just the moment structure of (4.1), (4.3) and (4.4). We then examine the precision of maximum likelihood type volatility estimators implied by the fully specified model discussed in Section 4.2. In each case the $MSE$ of the estimate is expressed as functions of the data characteristics. Additionally we study the relative gain in efficiency by comparing the $MSE$ of estimate based on the high-frequency data to those based on low frequency data. Finally, we discuss some empirical implications of the results. We begin with a discussion of a measure of relative efficiency.

5.1 Measuring the relative precision of variance estimators

To examine the usefulness of the high frequency data in estimating the variance of daily returns, we use the efficiency ratio ($ER$) given by

$$ER(\hat{V}_{\text{intra}}, \hat{V}_{\text{day}}) = \frac{MSE(\hat{V}_{\text{intra}})}{MSE(\hat{V}_{\text{day}})}$$

(5.1)

where $\hat{V}_{\text{intra}}$ is an estimator of the variance of daily returns derived from intra-day returns and $\hat{V}_{\text{day}}$ is the estimator using only the daily returns as given in (4.9). This quantity measures the
MSE of the $\hat{V}_{\text{intra}}$ estimator as a fraction of the $\text{MSE}$ of $\hat{V}_{\text{day}}$. If the daily returns follow a normal distribution, $\text{MSE}(\hat{V}_{\text{day}}) = \frac{1}{N} \cdot 2\sigma^4$ where $\sigma^2$ is the daily variance and $N$ is the number of days.

A good estimator $\hat{V}_{\text{intra}}$ using intra-day data would satisfy $ER(\hat{V}_{\text{intra}}, \hat{V}_{\text{day}}) \leq 1$, and we can view $ER$ as a measure of the usefulness of the intra-day data relative to the daily data in estimating the daily variance.

### 5.2 The effects of high frequency data characteristics on the sum of squares estimators

Recall that $\hat{V}_{\text{ss} \text{ intra}}$ in (4.10) and $\hat{V}_{\text{ssa} \text{ intra}}$ in (4.11) are two sum of squares estimators of the daily variance based on intra-day returns. To examine their relative efficiency with respect to the estimator $\hat{V}_{\text{day}}$ based on daily returns, we employ the structure in (4.1), (4.3) and (4.4), and make the following further assumption

(A.1) $E[\epsilon^3_{ni}\epsilon_{nj}] = 0, E[\epsilon^2_{ni}\epsilon_{nj}\epsilon_{nk}] = 0, \text{ if } i \neq j \neq k \neq l.$

(A.2) $A_{i,j,\ell} = O(A)$, where $A_{i,j,\ell} = \frac{1}{k} \sum_{t=1}^{k} (\delta_{i,t} - \bar{\delta_{i}})(\delta_{j,t} - \bar{\delta_{j}})^2/((i+j)/2)^2, \bar{\delta_{i}} = \frac{1}{k} \sum_{t=1}^{k} \delta_{i,t}$ and $A = \frac{1}{k} \sum_{t=1}^{k} (\delta_{i,t} - \bar{\delta_{i}})^2/(\bar{\delta_{i}})^2$,

Together with (4.3) and (4.4), Assumption (A.1) completely specify the fourth moments of $\epsilon_{s(n,t)}$. It rules out the leverage effect discussed in the GARCH literature.

In Assumption (A.2), $\bar{\delta_{i}} = \frac{1}{k} \sum_{t=1}^{k} \delta_{i,t}$ is the average of $\delta_{i,t}$. Notice that $A$ is the squared coefficient of variation for $\{\delta_{1}^2, \cdots, \delta_{k}^2\}$. If the deterministic pattern is absent, i.e., $\delta_{1}^2 = \delta_{2}^2 = \cdots = \delta_{k}^2$, all the terms $A$ and $A_{i,j,\ell}$ will vanish. They are all quite small in practice. This assumption is unnecessary for the result presented below but can greatly simplify the discussion.

Since we only consider the intra-day returns of the weekdays, we denote $w_b$ the beginning of the weekday and $w_e$ the end of the weekday used.

In the Appendix, we prove the following result for the simple sum of squared estimator defined by equation (4.10):

**Proposition 5.1** If $x_{s(n,t)}$ satisfies (4.1), (4.3), (4.4), (A.1), and (A.2), we have
a. \( \text{MSE}(\hat{V}_{\text{intra}}^{ss}) = \frac{2\sigma_{\theta}^2}{Nk} \left(1 + 4\theta + 6\theta^2\right) \left(1 + \frac{2\rho_1}{1-\phi} + f_1(\theta) - b_1 \cdot \frac{1}{k}\right) + 2N \cdot k \left(1 - \frac{2}{k}\right) \theta^2 + o(k^{-2}, \theta^2, A) \)

b. \( \text{ER}(\hat{V}_{\text{intra}}^{ss}, \hat{V}_{\text{day}}) = \frac{1}{k} (1 + b_0 \cdot \frac{1}{k})^{-1} \left[ \left(1 + \frac{2\rho_1}{1-\phi}\right) \left(1 + A + \frac{2\rho_1}{1-\phi}\right) + f_1(\theta) - b_1 \cdot \frac{1}{k} + 2N \cdot \left(k - 2\right) \theta^2 + o(k^{-2}, \theta^2, A) \right] \)

\( f_1(\theta) = a_{11} \cdot \theta + a_{12} \cdot \theta^2 + o(\theta^2) \)

\( a_{11} = -4 \left(1 + \frac{2\rho_1}{2}\right) \left(1 + A + \frac{2\rho_1}{1-\phi}\right) \)

\( a_{12} = 2 \left[1 + \left(1 + \frac{2\rho_1}{2}\right) (6 + 2\rho_1 + \frac{12\rho_1}{1-\phi})\right] \)

\( b_0 = \frac{2\rho_1}{2} + 6 \left(1 + \frac{2\rho_1}{2}\right) \frac{\rho_1}{1-\phi} \)

\( b_1 = \frac{1}{w_{\text{intra}} - w_{\text{day}}} \left(1 + \frac{2\rho_1}{2}\right) \frac{2\rho_1}{(1-\phi)^2} \)

In the special case when the intra-day returns are iid normal, \( \text{ER}(\hat{V}_{\text{intra}}^{ss}, \hat{V}_{\text{day}}) = \frac{1}{k} \), corresponding to the case when all the features are absent from the data. This is the result obtained by Merton (1980) under the geometric Brownian motion framework.

Proposition 5.1 quantifies the efficiency loss of the simple sum of squares estimator \( \hat{V}_{\text{intra}}^{ss} \) induced by departures from geometric Brownian motion, where the departures include the autocorrelation in the returns and their squares, intra-day deterministic pattern, and leptokurtosis. For simplicity, we focus our discussion on the expression of \( \text{ER}(\hat{V}_{\text{intra}}^{ss}, \hat{V}_{\text{day}}) \) to order \( o(k^{-1}, \theta^2, A) \).

First, consider the case where the intra-day returns follow covariance stationary white noise, that is, \( \theta = 0 \). The efficiency ratio is given by \( \frac{1}{k} \left(1 + \frac{2\rho_1}{2}\right) \left(1 + A + \frac{2\rho_1}{1-\phi}\right) \). In this product, the first factor \( (1 + \frac{2\rho_1}{2}) \) quantifies the efficiency loss due to leptokurtosis of the intra-day returns. This is potentially important since the kurtosis of many high-frequency series exceeds 10, which would result in an efficiency loss by a factor of at least 6. In the second factor, the term \( A \), which is the squared coefficient of variation of the \( \delta^2 \)'s, quantifies the efficiency loss due to the deterministic intra-day volatility pattern. The more heterogeneous the intra-day returns are, the larger the value \( A \) is and the greater is the efficiency loss. The other term in the second factor, \( \frac{\rho_1}{1-\phi} = \rho_1 \left(1 + \phi + \phi^2 + \phi^3 + \cdots\right) \) is the sum of all the autocorrelations in the squared innovations, and thus quantifies the efficiency loss due to the stochastically changing intra-day volatility. The stronger the volatility persistence is, the greater this term is and the less information is contained in the high-frequency data. With strong volatility clustering typically observed in high-frequency data, this term may also greatly reduce the benefits of using intra-day returns to estimate daily volatility. Overall the second factor \( (1 + A \)
\[ \frac{2 \rho}{1 - \phi} \] captures the impact of the changing, deterministic or stochastic, intra-day volatility. Notice that kurtosis and changing volatility affect the \( ER \) in a multiplicative manner, which exacerbates the efficiency loss.

Second, consider the impact of the correlation structure of the intra-day returns. The parameter \( \theta \), representing the lag-one correlation of intra-day returns, measures the deviation of intra-day returns from covariance stationary white noise, and affects \( ER(\hat{V}_{\text{intra}}^{ss}, \hat{V}_{\text{day}}) \) via the function \( \frac{1}{k} f_1(\theta) \) and \( 2N(1 - \frac{2 + b_0}{k})\theta^2 \). Notice that the coefficient of \( \theta \) in \( f_1(\theta) \) is \( a_{11} = -4(1 + \frac{\gamma}{2})(1 + A + \frac{2 \rho}{1 - \phi}) \), so the efficiency ratio is augmented by a factor \( (1 - 4\theta) \) in the first order expansion. Interestingly, negative autocorrelation results in greater efficiency loss while positive autocorrelation results in less efficiency loss. Next, note that the term \( 2N(1 - \frac{2 + b_0}{k})\theta^2 \) is approximately \( 2N\theta^2 \) when the number of intra-day observations \( k \) is large, and explodes as the number of days \( N \) gets large. This is due to the bias of \( \hat{V}_{\text{intra}}^{ss} \). As a direct consequence, we find whenever \( |\theta| > \sqrt{\frac{1}{2N}} \), it would be worse to use \( \hat{V}_{\text{intra}}^{ss} \) to estimate the daily variance than \( \hat{V}_{\text{day}} \), even though the former uses intra-day returns. For the special case when \( N = 1 \), which corresponds to using one day’s intra-day returns to estimates the volatility of that day, this \( 2N\theta^2 \) term alone sets a lower bound of \( ER \) at 0.02 for \( \theta < -0.1 \) as typically observed in our high frequency data.

In summary, the efficiency ratio is approximately given by \( ER(\hat{V}_{\text{intra}}^{ss}, \hat{V}_{\text{day}}) \approx \frac{1}{k} \cdot (1 + \frac{\gamma}{2})(1 - 4\theta)(1 + A + \frac{2 \rho}{1 - \phi}) \) for very small \( \theta \) and large \( k \), where the first factor \( (1 + \frac{\gamma}{2}) \) represents the impact of kurtosis, the second factor \( (1 - 4\theta) \) represents the impact of autocorrelation structure, and the last factor \( (1 + A + \frac{2 \rho}{1 - \phi}) \) represents the impact of changing volatility. If the intra-day prices are sampled from a geometric Brownian motion, all three factors reduce to 1 and \( ER(\hat{V}_{\text{intra}}^{ss}, \hat{V}_{\text{day}}) \) reaches the ideal efficiency ratio \( \frac{1}{k} \). Here, small \( \theta \) is relative to \( \sqrt{\frac{1}{2N}} \). If \( \theta \) and \( \sqrt{\frac{1}{2N}} \) are comparable, the bias has a big effect on the estimator \( \hat{V}_{\text{intra}}^{ss} \), and \( \hat{V}_{\text{intra}}^{ss} \) may be even worse than \( \hat{V}_{\text{day}} \).

We now add some remarks on the number of intra-day observations \( k \). The above discussion focused on the impact of different features when \( k \) is large enough so that terms of order higher than \( k^{-1} \) can be ignored. For moderate size of \( k \), terms of order \( k^{-2} \) will be needed to measure the additional impact of large leptokurtosis and strong volatility clustering. This includes \( k = 48 \) in our empirical study. The first formula of \( ER \) in Proposition 5.1 takes it into account and gives a more accurate approximation.

We next establish results for \( \hat{V}_{\text{intra}}^{ssa} \) in (4.11), the sum of squared estimator adjusted for the
Proposition 5.2 If \( x_{s(n,t)} \) satisfies (4.1), (4.3), (4.4), (A.1), and (A.2), we have

a. \[ \text{MSE}(\hat{V}_{\text{intra}}^{\text{ssa}}) = \frac{2\sigma^4}{Nk} (1 + 4\theta + 6\theta^2)[(1 + \frac{\gamma_2}{2})(1 + A + 2\rho_1) + \{2 + 4(1 + \frac{\gamma_2}{2})\rho_1\} + f_2(\theta) - b_1 \cdot \frac{1}{k}] + o(k^{-2}, \theta^2) \]

b. \[ \text{ER}(\hat{V}_{\text{intra}}^{\text{ssa}}, \hat{V}_{\text{day}}) = \frac{1}{k}(1 + b_0 \cdot \frac{1}{k})^{-1}[\{2 + 4(1 + \frac{\gamma_2}{2})\rho_1\} + \{2 + 4\rho_1(1 + \frac{\gamma_2}{2})\} + f_2(\theta) - b_1 \cdot \frac{1}{k}] + o(k^{-2}, \theta^2) \]

where \( A, b_0 \) and \( b_1 \) are given in Proposition 5.1, \( f_2(\theta) = a_{21}\theta + a_{22}\theta^2 \) and

\[ a_{21} = -4[1 + (1 + \frac{\gamma_2}{2})(A + 2\rho_1)] \]
\[ a_{22} = 2(6 + (1 + \frac{\gamma_2}{2})(10\rho_1 + 2\rho_1\phi)) \]

Comparing \( \text{ER}(\hat{V}_{\text{intra}}^{\text{ssa}}, \hat{V}_{\text{day}}) \) with \( \text{ER}(\hat{V}_{\text{intra}}^{\text{ss}}, \hat{V}_{\text{day}}) \), we have an additional term \( \{2 + 4\rho_1(1 + \frac{\gamma_2}{2})\} \) in \( \text{ER}(\hat{V}_{\text{intra}}^{\text{ssa}}, \hat{V}_{\text{day}}) \). When \( \gamma_2 \) and \( \rho_1 \) are non-negative, this term is always positive and hence, \( \hat{V}_{\text{intra}}^{\text{ssa}} \) is less efficient than \( \hat{V}_{\text{intra}}^{\text{ss}} \) for small \( \theta \). This additional term in \( \hat{V}_{\text{intra}}^{\text{ssa}} \) is induced by estimating \( \theta \), representing the lag-one autocorrelation. Regardless of the features of the intra-day returns, the value 2 quantifies the fixed effect of estimating \( \theta \). Hence even if the returns are sampled from a geometric Brownian motion, the benefit of using intra-day returns is effectively reduced by a factor of 3 for the estimator \( \hat{V}_{\text{intra}}^{\text{ssa}} \). The other additional term \( 4\rho_1(1 + \frac{\gamma_2}{2}) \) partially quantifies the impact of volatility clustering compounded by the fat tails. Together, \( \{2 + 4\rho_1(1 + \frac{\gamma_2}{2})\} \) represents the “cost” incurred by having estimated \( \theta \).

When \( \theta < 0 \), there is also “benefit” of estimating \( \theta \) by reducing the bias effect induced by lag-one autocorrelation. In the linear terms of \( f_1(\theta) \) and \( f_2(\theta) \), \( a_{21} \) can be written as \( a_{21} = a_{11} + 4[(1 + \frac{\gamma_2}{2})\phi\rho_1] \), so that for non-negative \( (\gamma_2, \rho_1, \phi) \), \( f_2(\theta) < f_1(\theta) \) in the first order approximation. This is part of the benefit for estimating \( \theta \). For large \( k \) or large \( N \), a substantial reduction is achieved by removing from \( \hat{V}_{\text{intra}}^{\text{ssa}} \) the bias effect, i.e., the term \( 2N(1 - \frac{2 + b_0}{k})\theta^2 \) in \( \hat{V}_{\text{intra}}^{\text{ssa}} \). So accounting for lag-one autocorrelation in \( \hat{V}_{\text{intra}}^{\text{ssa}} \) effectively removing the bias effect, and can substantially reduce the impact of the autocorrelation in returns on the precision of the variance estimation.

Thus, even for \( N = 1 \), there is a trade-off between the cost of estimating \( \theta \) and removing bias, which can be regarded as the benefit of estimating \( \theta \). For fixed \( N, k \) and small \( \theta \), the bias effect is relatively small and it may be better-off to use \( \hat{V}_{\text{intra}}^{\text{ssa}} \) by ignoring the lag-one autocorrelation. On
the other hand, even for a very small $\theta$, the bias effect becomes important for large $N$ or $k$, and $\hat{V}_{sxa}^{ss\text{ intra}}$ will be more efficient eventually. Finally, for moderate $k$, a second order expansion in $\frac{1}{k}$ would give more accurate approximation and is given in the first expression of $\text{ER}(\hat{V}_{sxa}^{ss\text{ intra}}, \hat{V}_{\text{day}})$ in Proposition 5.2.

5.3 Numerical Comparison of $\text{ER}(\hat{V}_{sxa}^{ss\text{ intra}}, \hat{V}_{\text{day}})$ and $\text{ER}(\hat{V}_{sxa}^{ss\text{ intra}}, \hat{V}_{\text{day}})$

Propositions 5.1 and 5.2 can be used to study the benefit of using intra-day returns to estimate daily volatility when the intra-day returns deviate from iid normal. In the following, we will further investigate impacts of autocorrelation and fat-tails using numerical examples.

Table 4 presents the $\text{ERs}$ for $N = 156$, $k = 48$, $w_b = 2$, $w_e = 4$ when $A = 0$, and $\rho_1 = \phi = 0.0$, focusing on the impact of autocorrelation and kurtosis. That is, we use one year half-hour returns to estimates the daily volatility when the deterministic pattern and volatility clustering are absent from the intra-day return series.

The $\text{ERs}$ in each column are computed based on the same kurtosis but different autocorrelations. For example, Column 3 corresponds to the case when the intra-day returns are normally distributed, i.e. $\gamma_2 = 0$, but have different lag-one autocorrelations. The following observations can be made along the column direction, and they conform to the discussions in Section 5.2:

- $\text{ER}(\hat{V}_{sxa}^{ss\text{ intra}}, \hat{V}_{\text{day}}) > 1$ whenever $\theta \neq 0$, i.e., $\hat{V}_{sxa}^{ss\text{ intra}}$ is less efficient than $\hat{V}_{\text{day}}$ when the lag-one autocorrelation is present in the intra-day returns; this suggests that when estimating aggregated quantities, the disaggregated data or the high frequency data only help when the model are specified correctly and hurt a lot otherwise.

- At $\theta = 0.0$, $\hat{V}_{sxa}^{ss\text{ intra}}$ performs better than $\hat{V}_{sxa}^{ss\text{ intra}}$;

- $\text{ER}(\hat{V}_{sxa}^{ss\text{ intra}}, \hat{V}_{\text{day}})$ increases as $\theta$ becomes more negative, that is, negative autocorrelation structure reduces the benefit of using intra-day data;

- $\text{ER}$ for $\hat{V}_{sxa}^{ss\text{ intra}}$ are asymmetric with respect to $\theta = 0.0$, because positive and negative $\theta$ have different effect on the daily returns which are the sum of intra-day returns. Consider the extreme cases when $\theta = -1.0$, the daily return is $\epsilon_{nk} - \epsilon_{n0}$, the difference between last and first error terms. On the other hand, when $\theta = 1.0$, the daily return is $\epsilon_{n0} + 2\epsilon_{n1} + \cdots + 2\epsilon_{n,n-1} + \epsilon_{nk}$, almost twice of the sum of all the intra-day errors.

The impact of kurtosis can be analyzed similarly. Along the row direction, the $\text{ERs}$ correspond
Table 4: Impacts of Autocorrelation and Kurtosis on $ER$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\theta$</th>
<th>0</th>
<th>3</th>
<th>6</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{V}^{ss}_{intra}$</td>
<td>-0.3</td>
<td>106.7</td>
<td>103.6</td>
<td>100.6</td>
<td>97.87</td>
</tr>
<tr>
<td>$\hat{V}^{ssa}_{intra}$</td>
<td>-0.3</td>
<td>0.136</td>
<td>0.163</td>
<td>0.188</td>
<td>0.211</td>
</tr>
<tr>
<td>$\hat{V}^{ss}_{intra}$</td>
<td>-0.2</td>
<td>28.52</td>
<td>27.73</td>
<td>26.99</td>
<td>26.29</td>
</tr>
<tr>
<td>$\hat{V}^{ssa}_{intra}$</td>
<td>-0.2</td>
<td>0.093</td>
<td>0.121</td>
<td>0.147</td>
<td>0.171</td>
</tr>
<tr>
<td>$\hat{V}^{ss}_{intra}$</td>
<td>-0.1</td>
<td>4.545</td>
<td>4.454</td>
<td>4.368</td>
<td>4.287</td>
</tr>
<tr>
<td>$\hat{V}^{ssa}_{intra}$</td>
<td>-0.1</td>
<td>0.072</td>
<td>0.101</td>
<td>0.127</td>
<td>0.152</td>
</tr>
<tr>
<td>$\hat{V}^{ss}_{intra}$</td>
<td>0.0</td>
<td>0.021</td>
<td>0.051</td>
<td>0.078</td>
<td>0.105</td>
</tr>
<tr>
<td>$\hat{V}^{ssa}_{intra}$</td>
<td>0.0</td>
<td>0.062</td>
<td>0.090</td>
<td>0.117</td>
<td>0.142</td>
</tr>
<tr>
<td>$\hat{V}^{ss}_{intra}$</td>
<td>0.1</td>
<td>2.072</td>
<td>2.031</td>
<td>1.992</td>
<td>1.955</td>
</tr>
<tr>
<td>$\hat{V}^{ssa}_{intra}$</td>
<td>0.1</td>
<td>0.056</td>
<td>0.084</td>
<td>0.111</td>
<td>0.137</td>
</tr>
</tbody>
</table>

to the same lag-one autocorrelation but different kurtosis. We have the following observations:

- $ER(\hat{V}^{ssa}_{intra}, \hat{V}_{day})$ increases with $\gamma_2$. In other words, the benefit of using intra-day returns diminishes as the observations become more leptokurtically distributed;

- At $\theta = 0$, $ER(\hat{V}^{ss}_{intra}, \hat{V}_{day})$ also increases with $\gamma_2$.

Based on the above observations for Table 4, we conclude that the benefit of using intra-day returns to estimate daily volatility decreases substantially when lag-one autocorrelation becomes negative and/or the leptokurtosis becomes stronger.

Propositions 5.1 and 5.2 can be readily extended to more general autocorrelation structure that that given in (4.1), volatility covariance structure specified by equation (4.3) and distributions with leptokurtic property. In particular, more complicated forms of volatility dependence than the special case (4.6) can be considered. We remark here that that the results presented in Sections 5.2 and 5.3 only depend on the statistical structure given in (4.1), (4.3) and (4.4).
5.4 The effect of high frequency data characteristics on the maximum likelihood estimators for fully specified model

In the previous two sections, we found that estimation accuracy can be greatly reduced for the estimators $\hat{V}_{\text{ss}}^{\text{intra}}$ and $\hat{V}_{\text{ssa}}^{\text{intra}}$ if the intra-day returns have negative serial correlation, fat tails or volatility clustering. A natural question then arises: can we improve the estimation accuracy by making further parametric assumptions and using maximum likelihood estimation? In this section, we examine the ER's of the maximum likelihood estimators based on the fully specified model in Section 4.2.

Consider the model specified in (4.1), (4.6) and (4.7). Under this model, we study how modeling volatility clustering and non-normality can impact the precision of daily variance estimation. For simplicity, we assume that the model is free of the deterministic volatility pattern, i.e., $\delta_1 = \cdots = \delta_k$.

Specifically, the fully specified model is a GARCH(1, 1) model with MA(1) conditional mean structure,

\[
\begin{align*}
    x_{s(n,t)} &= \epsilon_{s(n,t)} + \theta \cdot \epsilon_{s(n,t)-1}, \quad t = 1, 2, \cdots, k, \\
    h_{s(n,t)} &= \omega + \alpha \cdot h_{s(n,t)-1} \cdot z_{s(n,t)-1}^2 + \beta h_{s(n,t)-1}
\end{align*}
\]

where $h_{s(n,t)}$ is the conditional variance, and $z_{s(n,t)} = \frac{\epsilon_{s(n,t)}}{\sqrt{h_{s(n,t)}}}$ follows the mixture of two normals with parameters $(\sigma_0^2, \eta, \lambda)$.

Since this fully specified model is a special case of the structure (4.1), (4.3) and (4.4), we will have the same ER for the estimators $\hat{V}_{\text{ss}}^{\text{intra}}$ and $\hat{V}_{\text{ssa}}^{\text{intra}}$. The further parametric assumptions should provide superior estimates compared with $\hat{V}_{\text{ss}}^{\text{intra}}$ and $\hat{V}_{\text{ssa}}^{\text{intra}}$. We consider two MLE type estimators: $\hat{V}_{\text{intra}}^{\text{mle1}}$ obtained under the normal assumption for $z_{s(n,t)}$, and $\hat{V}_{\text{intra}}^{\text{mle2}}$ obtained under the mixture normal assumption for $z_{s(n,t)}$. Both estimators are of interest. The latter is asymptotically efficient and therefore its efficiency loss is only subject to the features in the model. The former is easily implemented with available software.

Unfortunately closed-form analytical representation for $\hat{V}_{\text{intra}}^{\text{mle1}}$ and $\hat{V}_{\text{intra}}^{\text{mle2}}$ are not available, so we cannot analytically quantify the effects of lag-one autocorrelation, volatility clustering and non-normality as we have done for the other three estimators. Nevertheless, numerical results can be obtained by simulation. Using the estimated GARCH parameters ($\alpha = 0.1296, \beta = 0.7372$) for the Dollar/Deutsch Mark exchange rate returns, we have constructed Table 5. In this table, we only
consider the $\hat{V}_{ssa\text{ intra}}$, $\hat{V}_{mle1\text{ intra}}$ and $\hat{V}_{mle2\text{ intra}}$, since $\hat{V}_{ssa\text{ intra}}$ is less efficient than $\hat{V}_{day}$ unless $\theta = 0$ and thus is of less interest.

In Table 5, we have chosen iid kurtosis for the four values 0, 3, 6, and 9. As a reminder, the iid kurtosis refers to the kurtosis of the iid innovation $z_{s(n,t)}$ and the overall kurtosis refers to the kurtosis of the process $\epsilon_{s(n,t)}$. Back in Table 4 where volatility clustering is absent, the overall kurtosis coincides with the iid kurtosis. However, the overall kurtosis no longer equals to the iid kurtosis when the volatility clustering is present as in Table 5. Corresponding to the four iid kurtosis values 0, 3, 6, and 9, the overall kurtosis are 0.47, 6.06, 16.58 and 43.70 by formula (4.8). We label the row direction in both Tables by the overall kurtosis.

Corresponding to the same autocorrelation structure and the same iid kurtosis, $ER(\hat{V}_{ssa\text{ intra}}, \hat{V}_{day})$ in Table 5 are much larger than their count-parts in Table 4. The differences gives the impact of volatility clustering on estimating daily volatility. Therefore, volatility clustering in the intra-day data would also substantially reduce the benefit of using intra-day data to estimate daily volatility.

From Table 5, we can also study, under the volatility clustering, how lag-one autocorrelation and non-normality would affect the efficiency in estimating the daily volatility. We have the following observations:

- The $ERs$ for estimator $\hat{V}_{ssa\text{ intra}}$ have similar patterns as in Table 4. In other words, under a given volatility clustering structure, the benefit of using intra-day returns to estimate daily variance would be reduced as the lag-one autocorrelation becomes more negative or the kurtosis increases.

- When applicable, $\hat{V}_{mle2\text{ intra}}$ is more efficient than $\hat{V}_{mle1\text{ intra}}$, and $\hat{V}_{mle1\text{ intra}}$ is more efficient than $\hat{V}_{ssa\text{ intra}}$. This ordering is of course not surprising. While all three estimators adjust for autocorrelation structure in the intra-day returns, $\hat{V}_{ssa\text{ intra}}$ only uses a crude measure for the lag-one correlation and thus is the least efficient one. $\hat{V}_{mle1\text{ intra}}$ uses the maximum likelihood method to estimate lag-one correlation and GARCH parameter, and thus improves over $\hat{V}_{ssa\text{ intra}}$. This estimator benefits from using the correct model but is penalized by using the wrong distribution. $\hat{V}_{mle2\text{ intra}}$ uses all the available information and is thus the most efficient one among the three.

- The $ERs$ for $\hat{V}_{mle1\text{ intra}}$ and $\hat{V}_{mle2\text{ intra}}$ increase with strong negative lag-one autocorrelation. In other words, the presence of autocorrelation reduces the benefit of using intra-day data. The impact
Table 5: Impacts of Autocorrelation and Kurtosis on ERs (continued)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\theta$</th>
<th>$\gamma_2$</th>
<th>0.468</th>
<th>6.06</th>
<th>16.58</th>
<th>43.70</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{V}_{ss\text{ intra}}$</td>
<td>0.933</td>
<td>1.236</td>
<td>1.508</td>
<td>1.756</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{ssa\text{ intra}}$</td>
<td>0.210</td>
<td>0.337</td>
<td>0.452</td>
<td>0.556</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle1\text{ intra}}$</td>
<td>-0.3</td>
<td>0.273</td>
<td>0.378</td>
<td>0.473</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle2\text{ intra}}$</td>
<td>0.152</td>
<td>0.226</td>
<td>0.233</td>
<td>0.173</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle3\text{ intra}}$</td>
<td>NA</td>
<td>0.199</td>
<td>0.185</td>
<td>0.128</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\theta$</th>
<th>$\gamma_2$</th>
<th>0.468</th>
<th>6.06</th>
<th>16.58</th>
<th>43.70</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{V}_{ss\text{ intra}}$</td>
<td>0.342</td>
<td>0.550</td>
<td>0.736</td>
<td>0.906</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{ssa\text{ intra}}$</td>
<td>0.158</td>
<td>0.274</td>
<td>0.378</td>
<td>0.473</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle1\text{ intra}}$</td>
<td>-0.2</td>
<td>0.255</td>
<td>0.356</td>
<td>0.448</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle2\text{ intra}}$</td>
<td>0.137</td>
<td>0.210</td>
<td>0.220</td>
<td>0.164</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle3\text{ intra}}$</td>
<td>NA</td>
<td>0.186</td>
<td>0.176</td>
<td>0.123</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\theta$</th>
<th>$\gamma_2$</th>
<th>0.468</th>
<th>6.06</th>
<th>16.58</th>
<th>43.70</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{V}_{ss\text{ intra}}$</td>
<td>0.134</td>
<td>0.276</td>
<td>0.404</td>
<td>0.519</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{ssa\text{ intra}}$</td>
<td>0.133</td>
<td>0.244</td>
<td>0.343</td>
<td>0.432</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle1\text{ intra}}$</td>
<td>-0.1</td>
<td>0.240</td>
<td>0.339</td>
<td>0.428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle2\text{ intra}}$</td>
<td>0.125</td>
<td>0.199</td>
<td>0.210</td>
<td>0.158</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle3\text{ intra}}$</td>
<td>NA</td>
<td>0.177</td>
<td>0.169</td>
<td>0.119</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\theta$</th>
<th>$\gamma_2$</th>
<th>0.468</th>
<th>6.06</th>
<th>16.58</th>
<th>43.70</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{V}_{ss\text{ intra}}$</td>
<td>0.071</td>
<td>0.167</td>
<td>0.253</td>
<td>0.331</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{ssa\text{ intra}}$</td>
<td>0.120</td>
<td>0.228</td>
<td>0.324</td>
<td>0.411</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle1\text{ intra}}$</td>
<td>0.0</td>
<td>0.228</td>
<td>0.324</td>
<td>0.411</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle2\text{ intra}}$</td>
<td>0.116</td>
<td>0.190</td>
<td>0.203</td>
<td>0.153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle3\text{ intra}}$</td>
<td>NA</td>
<td>0.170</td>
<td>0.164</td>
<td>0.116</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\theta$</th>
<th>$\gamma_2$</th>
<th>0.468</th>
<th>6.06</th>
<th>16.58</th>
<th>43.70</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{V}_{ss\text{ intra}}$</td>
<td>0.061</td>
<td>0.126</td>
<td>0.183</td>
<td>0.235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{ssa\text{ intra}}$</td>
<td>0.113</td>
<td>0.219</td>
<td>0.314</td>
<td>0.399</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle1\text{ intra}}$</td>
<td>0.1</td>
<td>0.217</td>
<td>0.312</td>
<td>0.397</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle2\text{ intra}}$</td>
<td>0.108</td>
<td>0.183</td>
<td>0.197</td>
<td>0.150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{V}_{mle3\text{ intra}}$</td>
<td>NA</td>
<td>0.164</td>
<td>0.160</td>
<td>0.113</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results in this table are based on the GARCH(1, 1) with parameter ($\alpha = 0.1296, \beta = 0.7372$). The overall kurtosis of $\epsilon_{nt}$ in the row direction correspond to standard normal, mixture normal(0.10, 6.00), (0.07, 10) and (0.058, 14) innovation $z_{nt}$. The ERs for $\hat{V}_{ss\text{ intra}}$, $\hat{V}_{ssa\text{ intra}}$ and $\hat{V}_{mle1\text{ intra}}$ are exact, while ERs for $\hat{V}_{mle2\text{ intra}}$ and $\hat{V}_{mle3\text{ intra}}$ are calculated based on 10,000 simulation. Finally, NA indicated non-existence of ER because of over-parameterization in the model.
of leptokurtosis on ERs, however, is mixed and requires further study.

- Even for the most efficient estimator $\hat{V}_{\text{mle}}^{\text{intra}}$, its ERs are much larger than the ideal situation (0.02, corresponding to iid normal). Therefore, once dependence, volatility clustering and non-normality are present in the intra-day returns, the benefit of using intra-day data to estimate daily volatility decreases substantially.

5.5 Empirical applications to the exchange rate series

Propositions 5.1 and 5.2 can be used not only to study the overall benefit of using intra-day returns, but also to assess the relative importance of different features on the daily volatility estimators. In order to better understand the impact of each individual features, we apply the results in Propositions 5.1 and 5.2 to the three half-hour exchange rate returns series analyzed in Section 3. Using parameter estimation results on all the features in Table 3, we have constructed Tables 6 and 7.

Tables 6 - 7 quantify non-normality, the deterministic pattern, and volatility clustering in the three columns under “$1 + \frac{\gamma^2}{2}$”, “$A$”, and “$1 + \frac{2\rho_1}{1-\phi}$”, respectively. The combined impact of these features is approximately the product $(1 + \frac{\gamma^2}{2})(1 + A + \frac{2\rho_1}{1-\phi})$, given in the column “$a_0$”. Clearly, both the non-normality and the volatility clustering can reduce the benefit of using intra-day returns dramatically. For the Dollar/Deutschmark series, the kurtosis alone and volatility clustering alone reduces the benefit of using intra-day observations by a factor of 5.2 and 3.65, respectively. Similar observations can also be made for the other two series. Surprisingly, although all three series display noticeable deterministic intra-day volatility pattern, the deterministic pattern only plays a minor role as represented by $A$.

The next columns “$c = 2 + 4\rho_1(1 + \frac{\gamma^2}{2})$” represent the cost of estimating $\theta$, the columns “$a_j\theta$”, and “$a_j\theta^2$”, $j = 1, 2$, represent the first and second order effects of lag-one autocorrelation, and the columns “bias” represent the bias effect $2N(1 + \frac{2+k\phi_0}{k})\theta^2$. Since estimating $\theta$ is not performed on $\hat{V}_{\text{intra}}^{\text{ss}}$, “c” is 0.0 for this estimator. The efficiency loss associated with MA(1) coefficient $\theta$, given in columns “$a_j\theta$” and “$a_j\theta^2$”, are much higher for $\hat{V}_{\text{intra}}^{\text{ss}}$ than for the estimator $\hat{V}_{\text{intra}}^{\text{ss}}$. For the Dollar/Deutschmark series, the contrasts are 11.28 versus 2.23 for $a_j\theta$, and 4.88 versus 0.92 for $a_j\theta^2$. Finally, the efficiency loss associated with bias, given in the columns “bias”, is extremely big for the estimator $\hat{V}_{\text{intra}}^{\text{ss}}$ given 156 days observations and the MA(1) coefficient is around -0.1 to -0.2. For the Dollar/Deutschmark series, this effect is 287.4 while “bias” effect on $\hat{V}_{\text{intra}}^{\text{ss}}$ is 0. The
effects of $\theta$ and bias indicate that the benefit of estimating $\theta$ outweighs the fixed cost associated with the estimation even for $\theta$ as small as -0.1 to -0.2.

Tables 6 - 7 also give the overall benefit in columns “$ER(ss)$” and “$ER(ssa)$” obtained from results in the Appendix. The corresponding approximations, to order $o(k^{-2}, \theta^2, A)$ given in Propositions 5.1 - 5.2, to these $ER$ are presented in columns “$ER^a(ss)$” and “$ER^a(ssa)$”, respectively. Clearly, the exact values and approximations are close for estimator $\hat{V}_{intra}^{ssa}$. The large difference between the exact values and approximations for the estimator $\hat{V}_{intra}^{ss}$ come from the fact that the number of observations per day, $k = 48$, is small comparing with the number of days, $N = 156$, used.
Table 6: \( ER(\hat{V}_{intra}^{ss}, \hat{V}_{day}) \) for Exchange Rate Returns

<table>
<thead>
<tr>
<th>Series</th>
<th>( 1 + \frac{\gamma^2}{2} )</th>
<th>( A )</th>
<th>( 1 + \frac{2\pi}{1-\phi} )</th>
<th>( a_0 )</th>
<th>( c )</th>
<th>( a_{11}\theta )</th>
<th>( a_{22}\theta^2 )</th>
<th>bias</th>
<th>( ER^a(ss) )</th>
<th>( ER(ss) )</th>
<th>( ER(N) )</th>
<th>equi. #</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD</td>
<td>5.20</td>
<td>0.22</td>
<td>3.65</td>
<td>19.9</td>
<td>0.00</td>
<td>11.28</td>
<td>4.88</td>
<td>287.4</td>
<td>3.42</td>
<td>5.84</td>
<td>0.02</td>
<td>—</td>
</tr>
<tr>
<td>DF</td>
<td>7.75(*)</td>
<td>0.37</td>
<td>8.01</td>
<td>64.93</td>
<td>0.00</td>
<td>55.55</td>
<td>36.26</td>
<td>656.7</td>
<td>3.65</td>
<td>8.39</td>
<td>0.02</td>
<td>—</td>
</tr>
<tr>
<td>DY</td>
<td>4.29</td>
<td>0.14</td>
<td>5.21</td>
<td>22.94</td>
<td>0.00</td>
<td>16.20</td>
<td>8.71</td>
<td>441.7</td>
<td>4.63</td>
<td>10.41</td>
<td>0.02</td>
<td>—</td>
</tr>
</tbody>
</table>

The tables 6 - 7 are constructed using the estimation results in Table 3. The explicit expression for \( A, a_0, a_{j1}, \) and \( a_{j2}, j = 1, 2 \), can be found in Proposition (5.1 - 5.2). \( ER(ss) \) and \( ER(ssa) \) exact based on Appendix, and \( ER^a's \) are the approximations of order \( o(k^{-2}, \theta^2) \) given by Proposition (5.1 - 5.2).

(*): The sample kurtosis is used to calculate the relevant \( ER \), since implied one does not exist.

Table 7: \( ER(\hat{V}_{intra}^{ssa}, \hat{V}_{day}) \) for Exchange Rate Returns

<table>
<thead>
<tr>
<th>Series</th>
<th>( 1 + \frac{\gamma^2}{2} )</th>
<th>( A )</th>
<th>( 1 + \frac{2\pi}{1-\phi} )</th>
<th>( a_0 )</th>
<th>( c )</th>
<th>( a_{21}\theta )</th>
<th>( a_{22}\theta^2 )</th>
<th>bias</th>
<th>( ER^a(ssa) )</th>
<th>( ER(ssa) )</th>
<th>( ER(N) )</th>
<th>equi. #</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD</td>
<td>5.20</td>
<td>0.22</td>
<td>3.65</td>
<td>19.92</td>
<td>5.64</td>
<td>2.23</td>
<td>0.92</td>
<td>0.000</td>
<td>0.28</td>
<td>0.31</td>
<td>0.02</td>
<td>3.24</td>
</tr>
<tr>
<td>DF</td>
<td>7.75(*)</td>
<td>0.37</td>
<td>8.01</td>
<td>64.93</td>
<td>13.92</td>
<td>8.39</td>
<td>5.30</td>
<td>0.000</td>
<td>0.37</td>
<td>0.44</td>
<td>0.02</td>
<td>2.29</td>
</tr>
<tr>
<td>DY</td>
<td>4.29</td>
<td>0.14</td>
<td>5.21</td>
<td>22.94</td>
<td>4.86</td>
<td>2.14</td>
<td>1.08</td>
<td>0.000</td>
<td>0.25</td>
<td>0.29</td>
<td>0.02</td>
<td>3.45</td>
</tr>
</tbody>
</table>

Table 8: \( ER(\hat{V}_{intra}^{mle1}, \hat{V}_{day}) \) and \( ER(\hat{V}_{intra}^{mle2}, \hat{V}_{day}) \) for Exchange Rate Returns

<table>
<thead>
<tr>
<th>( \hat{V}_{intra}^{mle1} )</th>
<th>( \hat{V}_{intra}^{mle2} )</th>
<th>series</th>
<th>( ER(mle1) )</th>
<th>equi. #</th>
<th>( ER(mle2) )</th>
<th>equi. #</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D\hat{D} )</td>
<td>0.22</td>
<td>4.64</td>
<td>0.18</td>
<td>5.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( D\hat{F} )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( D\hat{Y} )</td>
<td>0.24</td>
<td>4.21</td>
<td>0.21</td>
<td>4.74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The \( ER \) in this table are based on 1000 simulations, and the parameters used in the simulation are taken from Table 3.
Comparing the actual \( ER \) with 0.02 (= \( \frac{1}{48} \)) under \( ER(N) \) corresponding to iid normal intra-day observations, the \( ER \) are very large, indicating that the benefit of using the intra-day data to estimate daily volatility is substantially reduced by the features presented in these data. Most noticeably, we find that using \( \hat{V}_{ss}^{intra} \) to estimate the daily volatility can be even much worse than using the estimator \( \hat{V}_{day} \), which only use the daily observations. Primarily, this is because \( \hat{V}_{ss}^{intra} \) is a biased estimator. Alternatively, the reciprocal of the \( ER \), given in the last columns under “equi. #”, is the equivalent number of iid normal intra-day returns which would gain the same efficiency over the daily returns, and can also be used to measure the benefit of using intra-day returns. The “equi. #” are not listed for \( \hat{V}_{ss}^{intra} \), since they are less than 1. Corresponding to the actual 48 observations, only 1.2 - 4 iid normal intra-day observations are needed to give the same relative efficiency for estimator \( \hat{V}_{ss}^{intra} \). Thus, the benefit of using intra-day returns to estimate the daily volatility may be much smaller than that under the iid normal scenario, once the deterministic pattern, autocorrelation structure, volatility clustering and non-normality typically observed in the intra-day data are accounted for.

Finally, \( ER(\hat{V}_{mle1}^{intra}, \hat{V}_{day}) \) and \( ER(\hat{V}_{mle2}^{intra}, \hat{V}_{day}) \) for the three exchange rate returns series are constructed in Table 8 in analogue to Table 6 - 7. These \( ER \) are determined by simulation. Compared with Table 6 - 7, not surprisingly, \( \hat{V}_{mle2}^{intra} \) is the most efficient one among all the four estimators. However, it should be noticed that the benefit of using intra-day returns is still not comparable with the case of iid normal intra-day observations as shown in the columns “equi. #” in Table 8.

6 Conclusion

Merton’s (1980) seminal work suggested that volatility estimates can be made arbitrarily precise provided that the sampling interval be allowed to shrink to zero. With the availability of high-frequency financial data sets that record asset prices sometimes 1000’s of times within a single day it appears that arbitrarily precise estimators may be at hand. While many studies have proceeded along this line, the precision of the estimates has not been considered outside of the unrealistic geometric Brownian motion setting. We identify fat tails, dependence in the returns and in their squared returns, and deterministic patterns in the variance, as departures from this simple continuous time setting. Using a discrete time model that is consistent with the observed
features of the high-frequency data, we derive the mean squared error of several estimators ranging from a simple sum of squares estimator to the efficient MLE.

Our results suggest that, after accounting for these features of the high-frequency data, the precision of the estimators can be greatly reduced. Although the deterministic pattern is a noticeable feature observed in the exchange rate data considered, it has a relatively small impact on the precision of the variance estimates. Most notably, as soon as the returns are autocorrelated, the estimator $\hat{V}_{\text{intra}}^{ss}$ can be much less efficient than $\hat{V}_{\text{day}}$ depending the number of days used in the estimation. Also, the efficiency ratio $ER(\hat{V}_{\text{intra}}^{ss}, \hat{V}_{\text{day}})$ for $\hat{V}_{\text{intra}}^{ss}$ triples and may be even larger depending on the magnitude and persistence of the autocorrelation. The kurtosis in the high-frequency data is typically quite large and greatly reduces the precision of the volatility estimate by a factor of half the kurtosis. Furthermore, time persistent varying volatility can also greatly reduce the precision of our volatility estimates. When all the high frequency data characteristics are accounted for, the benefit of using high frequency data to estimate daily volatility is much smaller than the one under ideal situation, namely, the asset price follows the geometric Brownian motion. For example, 48 intra-day returns with the leptokurtic property and autocorrelation in both mean and variance are only equivalent to 2 - 4 intra-day normal iid normal returns in terms of $MSE$. One direct implication of the results is that the benefit of using high-frequency data to estimate daily volatility may be quite different depending on the features of the return series under consideration.

Of the four estimators considered, the impacts of the high frequency data features on their $ER$s and $MSE$s are different. Although the benefits for all of them are far less than the one under the ideal situation, the results suggest that the more information is incorporated in the estimation the more efficient will the estimator be.

An interesting feature of high-frequency data is that often the kurtosis and the autocorrelation increase as the sampling interval decreases. Hence there is a trade off between using higher frequency data to obtain larger sample sizes while exacerbating the influence of the kurtosis and autocorrelation. This raises the interesting possibility of finding an optimal sampling frequency for a given data set. This appears to be an interesting problem for future research.
Appendix

Variances for Sum of Squares Estimators $\hat{V}_{\text{day}}$, $\hat{V}_{\text{intra}}^{\text{ss}}$ and $\hat{V}_{\text{intra}}^{\text{ssa}}$

Suppose that the intra-day returns $x_{nt}$, or $x_{s(n,t)}$, satisfy the structure specified by (4.1) and (4.3), and the assumption (A.1). In this appendix, we will derive the MSEs for the sum of squared estimators $\hat{V}_{\text{day}}$, $\hat{V}_{\text{day}}^{\text{ss},1}$ and $\hat{V}_{\text{day}}^{\text{ss},2}$ and the ERs for $\hat{V}_{\text{day}}^{\text{ss},1}$ and $\hat{V}_{\text{day}}^{\text{ss},2}$. As will be seen, the exact representations can be obtained. To better understand the impacts of different features discussed in Section 3, we also find their expansion at $\theta = 0$.

Since we only use the intra-day returns from the weekdays, denote $n(w,d) = 7 \cdot (w - 1) + d$ the $d$-th day in the $w$-th week. Let $M$, $b_1$ and $b_2$ denote the number of weeks, the begining and the end of the weekdays used in estimation, respectively. Unless stated explicitly, we assume $b_2 - b_1 + 1 < 7$. Also, we assume $N = M \cdot (b_2 - b_1 + 1)$ which would simplify the discussion.

\[
V_{\text{day}} = \frac{1}{N} \sum_{n^* = 1}^{N} r_{n^*}^2 = \frac{1}{N} \sum_{w=1}^{M} \sum_{d=b_1}^{b_2} r_{n(w,d)}^2
\]

\[
= \frac{1}{N} \sum_{w=1}^{M} \sum_{d=b_1}^{b_2} \left( \sum_{t=1}^{k} \delta_t (\epsilon_{s(n(w,d),t)} + \theta \epsilon_{s(n(w,d),t-1)})^2 \right)
\]

\[
= \frac{1}{N} \sum_{w=1}^{M} \sum_{d=b_1}^{b_2} \left( \theta \delta_1 \epsilon_{s(n(w,d),1)} - 1 + \delta_k \epsilon_{s(n(w,d),k)} + \sum_{t=1}^{k-1} (\theta \delta_{t+1} + \theta \delta_{t+1}) \epsilon_{s(n(w,d),t)} \right)^2
\]

\[
= I_1 + I_2
\]

where

\[
I_1 = \frac{1}{N} \sum_{w=1}^{M} \sum_{d=b_1}^{b_2} I_{1wd}
\]

\[
I_{1wd} = \sum_{t=0}^{k} v_t^2 \cdot \epsilon_{s(n(w,d),1) - 1 + t}^2
\]

\[
I_2 = \frac{2}{N} \sum_{w=1}^{M} \sum_{d=b_1}^{b_2} I_{2wd}
\]

\[
I_{2wd} = \sum_{t=0}^{k} \sum_{u=t+1}^{k} v_t v_{1,u} \epsilon_{s(n(w,d),1) - 1 + t + u} \epsilon_{s(n(w,d),1) - 1 + u}
\]

\[
v_{(1)t} = \begin{cases} 
\theta \delta_1 & t = 0 \\
\delta_t + \theta \delta_t & t = 1, \ldots, k - 1 \\
\delta_k & t = k
\end{cases}
\]

37
From (A.1) and the structural equation (4.3), we have

\[
\text{var}(I_{1wd}) = (2 + \gamma_2)\left(\sum_{t=0}^{k} v_{(1)t}^4 + 2 \sum_{i=1}^{k} \rho_i \sum_{j=0}^{k-i} v_{(1)j} v_{(1), i+j}^2\right)
\]

\[
\text{var}(I_{2wd}) = \sum_{t=0}^{k} \sum_{u=t+1}^{k} v_{(1)t}^2 v_{(1)u}^2 + (2 + \gamma_2)\sum_{i=1}^{k} \rho_i \sum_{j=0}^{k-i} v_{(1)j} v_{(1), i+j}^2
\]

\[
\text{cov}(I_{1wd}, I_{1w'd'}) = \left(2 + \gamma_2\right)\sum_{t=0}^{k} \sum_{u=t+1}^{k} v_{(1)t}^2 v_{(1)u}^2 \rho(n(w', d') - n(w, d) - k + t - u)
\]

\[
= \left\{
\begin{array}{ll}
\phi(n(w', d') - n(w, d) - 1)k \cdot F\left(\{v_{(1)t}^2\}\right) & \text{if } n(w', d') - n(w, d) > 1 \\
F\left(\{v_{(1)t}^2\}\right) + (2 + \gamma_2) v_{(1)0}^2 v_{(1)k}^2 & \text{if } n(w', d') - n(w, d) = 1
\end{array}
\right.
\]

\[
\text{cov}(I_{1wd}, I_{2w'd'}) = 0 \quad \forall \quad (m, d) \neq (m', d')
\]

\[
F\left(\{v_{vt}^2\}\right) = (2 + \gamma_2)\sum_{0 \leq t, u \leq k} v_{tu}^2 v_{u-t}^2 \rho_{k+u-t} - (2 + \gamma_2) v_{(1)0}^2 v_{(1)k}^2
\]

\[
= (2 + \gamma_2)\sum_{i=1}^{k} \rho_i \sum_{j=k-i}^{k} v_{j}^2 \cdot v_{i+j-k}^2 + (2 + \gamma_2) \phi^k \sum_{i=1}^{k} \rho_i \sum_{j=0}^{k-i} v_{j}^2 \cdot v_{i+j}^2
\]

In deriving \(\text{cov}(I_{1w}, I_{1w'})\), we use the fact that \(\rho_\ell = \rho_1 \cdot \phi^{\ell-1}\). We have

\[
\text{var}(I_{1wd}) + 4 \cdot \text{var}(I_{2wd})
\]

\[
= 2\left(\sum_{t=0}^{k} v_{(1)t}^2\right)^2 + 2\left(\sum_{t=0}^{k} v_{(1)t}^4 + 6(\gamma_2 + 2) \sum_{0 \leq i < j \leq k} v_{(1)i} v_{(1)j}^2 \rho_{j-i}\right)
\]

\[
= 2k^2(\overline{\theta}^2)^2((1 + 2(1 + \sqrt{A}A_{1,1}) \cdot \theta + \theta^2)^2 + \frac{1}{k}\left(\frac{\gamma_2}{2}(1 + A) + 6(1 + \frac{\gamma_2}{2}) \frac{\rho_1}{1 - \phi}\right) + o(\theta^2, k^{-1}))
\]

\[
= 2k^2(\overline{\theta}^2)^2(1 + 4\theta + 6\theta^2 + \frac{1}{k}\left(\frac{\gamma_2}{2}(1 + A) + 6(1 + \frac{\gamma_2}{2}) \frac{\rho_1}{1 - \phi}\right) + o(k^{-1}, \theta^2, A))
\]

\[
\text{mse}(\hat{V}_{\text{day}}) = \text{var}(\hat{V}_{\text{day}})
\]

\[
= \frac{1}{N} (\text{var}(I_{1wd}) + 4 \cdot \text{var}(I_{2wd})) + \frac{1}{N^2} \sum_{(w, d)\neq (w', d')} \text{cov}(I_{1wd}, I_{1w'd'})
\]

\[
= \frac{2}{N} k^2(\overline{\theta}^2)^2(1 + 4\theta + 6\theta^2 + \frac{1}{k}\left(\frac{\gamma_2}{2}(1 + A) + 6(1 + \frac{\gamma_2}{2}) \frac{\rho_1}{1 - \phi}\right) + o(k^{-1}, \theta^2, A))
\]

Similarly, we have
\[ V_{\text{intra}}^{ss} = \frac{1}{N} \sum_{w=1}^{M} \sum_{d=b_1+1}^{b_2} \sum_{t=1}^{k} \delta_t^2 (\epsilon_s(n(w,d),t) + \theta \epsilon_s(n(w,d),t-1))^2 \]

\[ = II_1 + II_2 \]

\[ II_1 = \frac{1}{N} \sum_{w=1}^{M} \sum_{d=b_1+1}^{b_2} II_{1wd} \]

\[ II_{1wd} = \sum_{t=0}^{k} v_{(2)t}^2 \delta_t^2 \epsilon_s(n(w,d),1) + t - 1 \]

\[ II_2 = \frac{2\theta}{N} \sum_{w=1}^{M} \sum_{d=b_1+1}^{b_2} II_{2wd} \]

\[ II_{2wd} = \sum_{t=1}^{k} \delta_t^2 \epsilon_s(n(w,d),1) + t - 2 \cdot \epsilon_s(n(w,d),1) + t - 1 \]

\[ v_{(2)t}^2 = \begin{cases} 
\theta^2 \delta_1^2 & \text{if } t = 0 \\
\delta_t^2 + \theta^2 \delta_{t+1}^2 & \text{if } t = 1, 2, \ldots, k - 1 \\
\delta_k^2 & \text{if } t = k 
\end{cases} \]

\[ V_{\text{intra}}^{ssa} = III_1 + III_2 + III_3 \]

\[ III_1 = I_1 = \frac{1}{N} \sum_{w=1}^{M} \sum_{d=b_1+1}^{b_2} III_{1wd} \]

\[ III_{1wd} = III_{1wd} = \sum_{t=0}^{k} v_{(2)t}^2 \epsilon_s(n(w,d),1) + t - 1 \]

\[ III_2 = \frac{2\theta}{N} \sum_{w=1}^{M} \sum_{d=b_1+1}^{b_2} III_{2wd} \]

\[ III_{2wd} = \sum_{t=1}^{k} v_{(3)t}^2 \epsilon_s(n(w,d),1) + t - 2 \epsilon_s(n(w,d),1) + t - 1 \]

\[ III_3 = \frac{2\theta}{N} \sum_{w=1}^{M} \sum_{d=b_1+1}^{b_2} III_{3wd} \]

\[ III_{3wd} = \sum_{t=1}^{k-1} \delta_t \delta_{t+1} \epsilon_s(n(w,d),1) + t - 1 \epsilon_s(n(w,d),1) + t - 1 \]

\[ v_{(3)t} = \begin{cases} 
\theta \delta_t^2 + \theta^2 \delta_1 \delta_2 & \text{if } t = 1 \\
\delta_{t-1} \delta_t + \theta \delta_t^2 + \theta^2 \delta_t \delta_{t+1} & \text{if } t = 2, \ldots, k - 1 \\
\delta_{k-1} \delta_k + \theta \delta_k^2 & \text{if } t = k 
\end{cases} \]

and we have
\[
\begin{align*}
\text{var}(II_{1wd}) & = (2 + \gamma_2)\left(\sum_{t=0}^{k} v_{(2)t}^4 + 2 \sum_{i=1}^{k} \rho_i \sum_{j=0}^{k-i} v_{(2)j}^2 v_{(2)i+j}^2\right) \\
\text{var}(II_{2wd}) & = \sum_{t=1}^{k} \delta_t^4 + (2 + \gamma_2)\rho_1 \sum_{t=1}^{k} \delta_t^4 \\
\text{cov}(II_{1wd}, II_{1w'd'}) & = \begin{cases} \\
\phi(n(w',d') - n(w,d) - 1)k \cdot F\{\{v_{(2)t}\}\} & \text{if } n(w',d') - n(w,d) > 1 \\
F\{\{v_{(2)t}\}\} + (2 + \gamma_2)v_{(2)t}^2 v_{(2)k}^2 & \text{if } n(w',d') - n(w,d) = 1 \\
\end{cases}
\text{cov}(II_{1wd}, II_{2w'd'}) & = 0 \quad \forall \ w, w', d, d' \\
\text{cov}(II_{2wd}, II_{2w'd'}) & = 0 \quad \forall \ (m, d) \neq (m', d') \\
\text{var}(III_{1wd}) & = \text{var}(I_{1wd}) \\
\text{var}(III_{2wd}) & = (1 + (2 + \gamma_2)\rho_1) \sum_{t=1}^{k} v_{(3)t}^4 \\
\text{var}(III_{3wd}) & = (1 + (2 + \gamma_2)\rho_2) \sum_{t=1}^{k-1} \delta_t^2 \delta_{t+1}^2 \\
\text{cov}(III_{1wd}, III_{1w'd'}) & = \text{cov}(I_{1wd}, I_{1w'd'}) \\
\text{cov}(III_{1wd}, III_{2w'd'}) & = 0 \quad \forall \ w, w', d, d' \\
\text{cov}(III_{1wd}, III_{3w'd'}) & = 0 \quad \forall \ w, w', d, d' \\
\text{cov}(III_{2wd}, III_{2w'd'}) & = 0 \quad \forall \ (m, d) \neq (m', d') \\
\text{cov}(III_{2wd}, III_{3w'd'}) & = 0 \quad \forall \ w, w', d, d' \\
\text{cov}(III_{3wd}, III_{3w'd'}) & = 0 \quad \forall \ (m, d) \neq (m', d') \\
\text{cov}(III_{3wd}, III_{3w'd'}) & = 0 \quad \forall \ (m, d) \neq (m', d')
\end{align*}
\]

So,

\[
\text{var}(\hat{V}_{\text{intra}}^{ss}) = \frac{1}{N}(\text{var}(II_{1wd}) + 4\theta^2 \cdot \text{var}(II_{2wd})) + \frac{1}{N^2} \sum_{(w,d) \neq (w',d')} \text{cov}(II_{1wd}, II_{1w'd'})
\]

\[
= \frac{2}{N} k(\delta^2)^2 \left(1 + \frac{\gamma_2}{2}\right)\left(1 + A + \frac{2\rho_1}{1 - \phi}\right)
\]

\[
+ 2(1 + (1 + \frac{\gamma_2}{2})(1 + 2\rho_1 + \frac{2\rho_1}{1 - \phi}))\theta^2
\]

\[
- \frac{1}{k} b_2 - b_1 + 1 \left(1 + \frac{\gamma_2}{2}\right)\left(\frac{2\rho}{(1 - \phi)^2}\right) + o(k^{-1}, \theta^2, A))
\]
\[ \text{bias}^2 (\hat{V}_{\text{intra}}^{ss}) = 4 \theta^2 \sum_{t=1}^{k-1} \delta_t \delta_{t+1}^2 = 2k (\delta^2)^2 \cdot ((2k - 4) \theta^2 + o(k^{-1}, \theta^2, A)) \]

\[ \text{mse} (\hat{V}_{\text{intra}}^{ssa}) = \text{var} (\hat{V}_{\text{intra}}^{ssa}) = 2k (\delta^2)^2 ((1 + \frac{2\gamma}{2}) (A + (1 + 4 \theta + 6 \theta^2)(1 + 2 \frac{\rho_1}{1 - \phi})) + 2\theta^2 (1 + 2(1 + \frac{\gamma}{2}) \rho_1 \cdot \phi) + 2(1 + 2(1 + \frac{\gamma}{2}) \rho_1)(1 + 2 \theta + 3 \theta^2) - \frac{1}{k} \frac{1}{b_2 - b_1 + 1} (1 + \frac{2\gamma}{2}) \frac{2\rho}{(1 - \phi)^2} + o(k^{-1}, \theta^2, A)) \]

We thus have the following theorem, which gives Proposition (5.1 - 5.2).

**Theorem A.1** Suppose that \( \{x_{nt} : t = 1, 2, \cdots, n, n = 1, 2, \cdots, N\} \) satisfy the statistical structure (4.1), (4.3) and (4.4), and the assumption \((A.1), (A.2)\), then the MSE of the estimators \( \hat{V}_{\text{day}}, \hat{V}_{\text{intra}}^{ss} \) and \( \hat{V}_{\text{intra}}^{ssa} \) are given by the following expansion at \( \theta = 0 \), \( A = \frac{1}{k} \sum_{t=1}^{k} (\frac{\delta^2}{\delta t})^2 = 0 \) and large \( k \):

\[ \text{mse} (\hat{V}_{\text{day}}) = \frac{2}{N} k (\frac{2\gamma}{2}) (1 + 4 \theta + 6 \theta^2 + \frac{1}{k} (1 + \frac{2\gamma}{2}) (A + 6(1 + \frac{2\gamma}{2}) \frac{\rho_1}{1 - \phi} + o(k^{-1}, \theta^2, A)) \]

\[ \text{mse} (\hat{V}_{\text{intra}}^{ss}) = \frac{2}{N} k (\frac{2\gamma}{2}) (1 + A + \frac{2\rho_1}{1 - \phi}) + 2(1 + 2(1 + \frac{2\gamma}{2}) \rho_1 \cdot \phi) + 2(1 + 2(1 + \frac{2\gamma}{2}) \rho_1)(1 + 2 \theta + 3 \theta^2) - \frac{1}{k} \frac{1}{b_2 - b_1 + 1} (1 + \frac{2\gamma}{2}) \frac{2\rho}{(1 - \phi)^2} + o(k^{-1}, \theta^2, A)) \]

\[ \text{mse} (\hat{V}_{\text{intra}}^{ssa}) = 2k (\delta^2)^2 ((1 + \frac{2\gamma}{2}) (A + (1 + 4 \theta + 6 \theta^2)(1 + 2 \frac{\rho_1}{1 - \phi})) + 2\theta^2 (1 + 2(1 + \frac{\gamma}{2}) \rho_1 \cdot \phi) + 2(1 + 2(1 + \frac{\gamma}{2}) \rho_1)(1 + 2 \theta + 3 \theta^2) - \frac{1}{k} \frac{1}{b_2 - b_1 + 1} (1 + \frac{2\gamma}{2}) \frac{2\rho}{(1 - \phi)^2} + o(k^{-1}, \theta^2, A)) \]

The efficiency ratios \( \text{ER} (\hat{V}_{\text{intra}}^{ssa}, \hat{V}_{\text{day}}) \) and \( \text{ER} (\hat{V}_{\text{intra}}^{ssa}, \hat{V}_{\text{day}}) \) have the following expansion at \( \theta = 0 \), \( A = \frac{1}{k} \sum_{t=1}^{k} (\frac{\delta^2}{\delta t})^2 = 0 \), and large \( k \):

\[ \text{ER} (\hat{V}_{\text{day}}, \hat{V}_{\text{day}}) = \frac{1}{k} (1 + b_0 \cdot \frac{1}{k})^{-1} [(1 + \frac{2\gamma}{2})(1 + A + \frac{\rho_1}{1 - \phi}) + f_1 (\theta) - b_1 \cdot \frac{1}{k} + N \cdot (2k - 4) \cdot \theta^2 + o(k^{-1}, \theta^2)] \]

\[ \text{ER} (\hat{V}_{\text{day}}, \hat{V}_{\text{day}}) = \frac{1}{k} (1 + b_0 \cdot \frac{1}{k})^{-1} [(1 + \frac{2\gamma}{2})(1 + A + \frac{\rho_1}{1 - \phi}) + 2(1 + \frac{2\gamma}{2}) \rho_1 + f_2 (\theta) - b_1 \cdot \frac{1}{k} + o(k^{-1}, \theta^2)] \]

where

\[ f_1 (\theta) = a_{11} \cdot \theta + a_{12} \cdot \theta^2 + o(\theta^2), f_2 (\theta) = a_{21} \theta + a_{22} \theta^2 + o(\theta^2) \]
and \( \delta^2 = \frac{1}{k} \sum_{i=1}^{k} \delta_i^2 \)

\[
\begin{align*}
a_{11} &= -4(1 + \frac{\gamma^2}{2})(1 + A + 2\rho_1) \\
a_{12} &= 2[1 + (1 + \frac{\gamma^2}{2})(6 + 5A + 2\rho_1 + \frac{12\rho_1}{1 - \phi})] \\
a_{21} &= -4[1 + (1 + \frac{\gamma^2}{2})(A + 2\rho_1)] \\
a_{22} &= 2[6 + (1 + \frac{\gamma^2}{2})(10\rho_1 + 2\rho_1\phi)] \\
b_0 &= \frac{\gamma^2}{2}(1 + A) + 6(1 + \frac{\gamma^2}{2}) \frac{\rho_1}{1 - \phi} \\
b_1 &= \frac{1}{\rho_2 - b_1 + 1}(1 + \frac{\gamma^2}{2}) \frac{\rho_1}{(1 - \phi)^2}
\end{align*}
\]

In this theorem, we assumed that only the intra-day data collected in weekdays are used in estimation. If all the intra-day data, including the weekends, satisfy 4.1, 4.4 and assumptions (A.1), (A.2), then the coefficient of \( \frac{1}{k} \), \( b_1 \) for ERs’ representation, vanish. Also note that, the exact result can also be obtained as in the proof.

While the above result is based on the structure (4.1) and (4.4), it can be easily extended to general stochastic volatility given by (4.3) since the expressions on MSE don’t depends on (4.4). Similar results for relaxing (4.1) to general ARMA models are also obtained.
References


