Volatility Asymmetry in High Frequency Data

Julia Litvinova *

First draft: February 5, 2002
Last revised: February 13, 2003 (preliminary version)

Abstract

The paper examines the important stylized fact of asymmetric volatility in high frequency data. It demonstrates that one of the two main factors explaining this phenomenon, the leverage effect, is measured more precisely using higher frequency data. The paper also examines the lead-lag relation between volatility and returns in high frequency data. The impact of returns shocks on future squared returns is shown to decay geometrically and remain significant for at least 3 days. A temporal aggregation formula for the correlation between squared returns and lagged returns that permits more accurate measures of the daily correlation using high frequency data is also derived. Lastly, the paper examines the ability of specific stochastic volatility models to explain the pattern observed in high frequency data.

*graduate student, Department of Economics, Duke University, e-mail: jal@duke.edu

I would like to thank Tim Bollerslev, my supervisor, for his ideas, suggestions and constant support during this research. Also I am grateful to Bjorn Eraker, Pete Kyle, George Tauchen and all participants of the Duke Financial Econometrics group for their comments and suggestions.
1 Introduction

The relationship between stock returns and stock return volatility is at the center of derivative asset valuation and portfolio management. Empirical researchers have documented an asymmetry in the volatility of stock market prices. Asymmetry in this relation has important implications in valuations of options and volatility forecasting. Better understanding of volatility symmetry would permit economists to refine existing models by incorporating variables that account for the asymmetry in more efficient way.

The asymmetric nature in stock returns and volatility is well documented in the existing empirical literature. Researchers observe that when bad news reaches the stock market, future volatility generally increases. In other words, stock returns and stock return volatility are negatively correlated. This effect is more pronounced during stock market crashes. For example, the 22% drop on October 19, 1987 led to a huge increase in volatility. On the other hand, good news does not cause a sharp decrease in volatility. Thus, volatility is asymmetric with respect to good and bad news.

Early influential studies by Black (1976) and Christie (1982) attribute the asymmetric volatility to effects stemming from changes in financial leverage (debt-equity ratio). However, the measured effect of stock price changes on volatility is too large to be explained solely by financial leverage changes. Another explanation of asymmetry is the time-varying risk premium hypothesis proposed by French, Schwert and Stambaugh (1987): an anticipated increase in volatility raises the required return on equity, leading to an immediate stock price decline. Like the leverage-based explanation, the time-varying risk premium enjoys only partial explanatory success. Campbell and Hentschel (1992) suggest that two effects contribute to the asymmetric behavior of stock market volatility: the leverage effect and the volatility feedback effect (if expected stock return increases when volatility increases, and if expected dividends are unchanged, then stock prices should fall when volatility increases). However, their explanation also cannot fully account for the observed asymmetry. Comparing the magnitude of two the effects, Bekaert and Wu (2000) recently attributed most of the asymmetry to the volatility feedback effect. The name "volatility feedback" can be explained in the following way: "if volatility is priced, an anticipated increase in volatility raises the required return on equity, leading to an immediate stock price decline" (Wu (2001)).

All of the previous studies relied on "low" frequency data (one day was the highest frequency studied). Hence, it is natural to ask whether high frequency data (for example, every 5-minutes or even "tick-by-tick" observations) contain more information about both price and volatility. The inability of previous studies to find a consensus in determining the main components of volatility asymmetry can arise because of loss of information in low frequency data (daily, weekly or even monthly) and because of different features of leverage and volatility feedback effects in low and high frequency data. In this paper I demonstrate the value of using high frequency data. I show that the volatility-return relation is measured more accurately using higher frequency data. In addition,
I provide a new methodology for accurately aggregating the high frequency data. Based on these results, I can empirically evaluate the various theoretical explanations for the asymmetric volatility with far greater precision than the literature to date.

To show that the correlation between price and volatility processes can be measured more precisely and accurately using high frequency data, I use a continuous time stochastic volatility process. Diffusion processes are widely used for modeling market prices. There are a number of methods to estimate diffusion parameters. Most of them work with a difference equation based on Euler discretization (naive discretization) of the continuous time stochastic process. Those methods include quasi-maximum likelihood estimation (QMLE), moment based estimation: generalized method of moments (GMM), simulated method of moments (SMM, Duffie and Singleton (1993)), efficient method of moments (EMM, Gallant and Tauchen (1996)). Recently, the Markov Chain Monte Carlo (MCMC) approach initially proposed by Jacquier, Polson and Rossi (1994) has received a lot of attention; see e.g. Eraker (2001) and Eraker, Johannes and Polson (2001).

I focus first on the Heston (1993) square root stochastic volatility model widely used in term structure theory. I derive the closed form moments for a discretized version of this model. I estimate the correlation coefficient (the leverage effect coefficient) between the Brownian motions driving the price and volatility processes at different frequencies using GMM. My result that the standard error is smaller at the higher frequency data directly motivates the use of these data.

Both leverage and volatility feedback effects explain the relation between return and volatility. The important difference between these explanations lies in the causality. The leverage effect explains why a lower return leads to a higher volatility, while the volatility feedback shows how a higher volatility may reinforce a lower return. Hence, it is interesting to examine the lead-lag relation between volatility and return. Since volatility is unobserved, many researchers rely on squared return or the absolute value of return as measures of ex-post volatility. Since a squared daily return is a quite noisy estimate of the volatility (see, e.g. Andersen and Bollerslev (1998)), I obtained a very noisy picture based on the daily data. However, use of 5-minute data reveals a new effect of very substantial, prolonged (for four-five days) and slowly decaying impact of return shocks on the future volatility proxy (the leverage effect).

I study the correlation between square return and lagged return under temporal aggregation. The derived formula helps to examine this relation on a daily level using 5-minute data. As mentioned above, the squared return-lagged return correlation is very noisy on a daily level. Using the aggregation formula, I study this relation on a daily level using 5-minute data. The impact of return shocks on the future squared return (the leverage effect) is significant for 3 days. The use of high frequency data therefore demonstrates a much more important role of the leverage effect in explaining volatility asymmetry.

The observed slowly decreasing squared return-return correlation pattern in 5-minute data ex-
hibits a geometric rate of decay. Most of the simple models widely used in practice today fail to produce this pattern. For instance, the Heston square root model produces a geometrically decreasing pattern in high-frequency data but is unable to produce the pattern in daily data. According to the large amount of literature on this subject, one-factor stochastic volatility models are also typically rejected because of their inability to generate sufficient kurtosis. Possible extensions of these simple models include the affine and logarithmic multifactor stochastic volatility model studied by Chernov, Gallant, Ghysels and Tauchen (2002), and models with a jump component included in both return and volatility equations studied by Eraker, Johannes, Polson (2001). The presence of two volatility factors breaks the link between tail thickness and volatility persistence. The first factor is often referred to as a long memory component while the second one is capable of making rapid moves, which are prohibited by a single SV specification. This model can be estimated in high frequency data using EMM, the method proposed by Gallant and Tauchen (1996). The preliminary simulations suggest that the two-factor model would produce the observed patterns both in high and low frequency data.

Most of my work is highly computer intensive. In order to construct the evenly spaced 5-minute returns from the raw unevenly spaced tick-by-tick futures data, I wrote the code in SAS to linearly interpolate returns. SAS is especially good for a work with huge data sets like tick-by-tick data. I use MATLAB for my GMM estimates and GAUSS for the most of my data calculations like intraday pattern estimation, high frequency returns standardization, graphics etc. Although there are SNP and EMM FORTRAN codes written by Gallant and Tauchen and they are available free of charge, a lot of additional work should be done to convert the programs to my applications. My experience in fitting a relatively simple model to the data suggests the programs take weeks to debug and run. Finally, to calibrate the parameters of the Heston model at the 5-minute level I use SMM. There is no pre-packaged statistical software for this type of work because I am creating a new methodology and using the most recent estimation techniques. My SMM code relies on a fairly widely distributed optimization routine by Gill, Murray, Saunders and Wright (1986). NPSOL is a Fortran package designed to solve the nonlinear programming problem: the minimization of a smooth nonlinear function subject to a set of constraints on the variables. Creating and executing my SMM programs in FORTRAN is extremely time consuming.

The remainder of the article is organized as follows. Section 2 describes the Heston square root stochastic volatility model, its moments and provides motivation for examining high frequency data. Section 3 consists of two parts. Empirical findings of the lead-lag relation between squared return and return are described in Section 3.1; the decaying behavior of the correlation pattern is studied in Section 3.2. The temporal aggregation of the correlation between squared return and lagged return is considered in Section 4. Section 5 examines the ability of the Heston model to generate the correlation behavior observed in 5-minute and daily data. EMM estimates of the
daily returns and SMM estimates of the 5-minute returns of the Heston model are presented in Sections 5.1 and 5.2, while Section 5.3 contains the discussion related to the sampling frequency. The next two sections outline the possible extensions of my current work and directions for the future research. Specifically, Section 6 describes the two-factor SV model, which preliminary can generate the observed correlation patterns both in low and high frequency data; Section 7 discusses the time varying leverage hypothesis and possible ways to model it.

2 Motivation for looking at HFD

My primary objective is to show that the correlation between price and volatility processes can be measured more precisely and accurately using high frequency data. Since volatility is unobserved, many researchers rely on the time series of squared returns or absolute returns as measures of ex-post volatility. I consider a simple one factor stochastic volatility model also known as a Heston square root model. Assume that the stock log-price process is generated by a continuous time stochastic volatility model:

\[
\begin{align*}
dp_t &= (\mu + cV_t)dt + \sqrt{V_t} \cdot dW_{p,t} \\
dV_t &= (\alpha + \beta V_t)dt + \sigma \sqrt{V_t} \cdot dW_{\sigma,t}
\end{align*}
\]  

(1)

where \( W_{p,t}, W_{\sigma,t} \) are correlated with the correlation coefficient \( \rho \). The parameter \(-\beta\) captures the speed of mean reversion, \(-\alpha/\beta\) determines the unconditional long-run mean of volatility, and \( \sigma \) determines the size of the volatility of volatility. Stationarity of volatility requires that: \( \sigma^2 \leq 2\alpha \).

The second equation in the system above is a special case of the well known constant elasticity of variance (CEV) model widely used in finance. In general, it has the form:

\[
dY_t = (\alpha + \beta Y_t)dt + \sigma Y_t \nu dW_t
\]  

(2)

Here, \( \nu = 0 \) corresponds to the Ornstein-Ulenbek process; \( \nu = 1/2 \) gives the "square root" process used by Cox, Ingerdsoll and Ross (1985). Clearly, \( \nu = 1/2 \) in (2) produces the second equation in system (1). (1) is then the Heston’s (1993) square root model - widely used because of the availability of the closed form solution. When \( \nu = 1 \), (1) is the continuous time model for the volatility process given by the diffusion limit (weak convergence) of the EGARCH(1,1) as shown by Nelson (1990).

Let \( p_{(t+1)\Delta} - p_t = \log(P_{(t+1)\Delta}) - \log(P_{t\Delta}) = r_{(t+1)\Delta} \) be the discretely observed time series process of return with \( \Delta \) observations per day. The discretized version of system (1) is given by:

\[
\begin{align*}
r_{t+1,\Delta} &= (\mu + c\sigma_{t,\Delta}^2)\Delta + \sigma_{t,\Delta} \sqrt{\Delta} \epsilon_{1,t+1} \\
\sigma_{t+1,\Delta}^2 &= \alpha \Delta + (1 + \beta \Delta)\sigma_{t,\Delta}^2 + \sigma \sigma_{t,\Delta} \sqrt{\Delta} \epsilon_{2,t+1}
\end{align*}
\]
where $\epsilon_{1,t+1}$ and $\epsilon_{2,t+1}$ are standard normal variables with correlation $\rho$ and $\sigma_t^2$ is standard notation for volatility in discrete models. Assume $c = 0$, $\Delta = 1$ and denote $\eta = 1 + \beta$ for simplicity. I show that the correlation between the return and the volatility process decreases geometrically with lags:

$$\text{corr}(r_{t+1}, \sigma_{t+1+k}^2) = \begin{cases} \rho \sqrt{(1 - \eta^2)} \eta^k & \text{for } k \geq 0 \\ 0 & \text{for } k < 0 \end{cases}$$

The covariance between the return and squared return is:

$$\text{cov}(r_{t+1}, r_{t+1+k}^2) = \begin{cases} \frac{\alpha}{1-\eta} \rho \sigma \eta^{k-1} & \text{for } k > 0 \\ \frac{2\alpha \mu}{1-\eta} & \text{for } k = 0 \\ 0 & \text{for } k < 0 \end{cases}$$

As one can see, $\text{cov}(r_t, r_t^2) > 0$ as long as the unconditional mean of returns is positive. This contradicts the numerous empirical findings that the unconditional correlation between return and squared return is negative.

Assume $\mu = c = 0$ and $\Delta \neq 1$. I compute explicitly the following moments of a discrete stochastic volatility model: $E(r_t^2)$, $E(r_t^4)$, $E(r_t r_{t+k}^2)$:

$$m(r_t, r_{t+k}) = \begin{pmatrix} r_t^2 + \frac{\alpha}{\beta} \Delta \\ r_t^4 - 3 \frac{\alpha^2 \Delta^2 - \frac{\alpha}{\beta} \sigma^2 \Delta - 2 \frac{\alpha}{\beta} \alpha (1 + \beta \Delta) \Delta^2}{1 - (1 + \beta \Delta)^{-1}} \\ r_t r_{t+k}^2 + \frac{\alpha^2}{\beta^2} \Delta^2 \rho \sigma (1 + \beta \Delta)^{k-1} \end{pmatrix}$$

Table 1 shows the GMM estimates of a discrete model for daily and half-daily simulated data based on 6 moments: $E(r_t^2)$, $E(r_t^4)$, $E(r_t r_{t+k}^2)$ for $k = 1, \ldots, 4$. The parameters are taken from Bollerslev and Zhou (2001). They correspond to the scenario with a moderately persistent variance process and not too high variance-of-variance parameter. $\beta$, a volatility persistence parameter, is negative and close to zero. The model was simulated at a 5-minute level ($\Delta = 1/80$) and then aggregated to daily and half-daily levels. Half-daily data provide the estimates with the smaller biases and smaller standard errors. In particular, the standard error of the correlation coefficient is around 3 times less for half-daily data than for daily in Table 1.

The next Table 2 shows the Monte Carlo experiment based on 500 simulations of 4000 days. RMSE is in parenthesis and the average GMM standard errors are in square brackets. Around 10% of all simulations are so-called ”crashes” when the estimated $\beta$ falls outside the $[-1, 0]$ region and $\rho$ falls outside the $[-1, 1]$ interval. The first thing to notice is that all daily estimates are insignificant. The reason can lie in the insufficient number of days (the sample size of 4000 days may be relatively

\[\text{I use Mike Cliff’s GMM code is written in MATLAB.}\]
Table 1: GMM estimation of the simulated discrete Heston square root model: $r_{t+1,\Delta} = \sigma_{t,\Delta}\sqrt{\Delta} \epsilon_{1,t+1}$ and $\sigma_{t+1,\Delta}^2 = \alpha \Delta + (1 + \beta \Delta) \sigma_{t,\Delta}^2 + \sigma_{t,\Delta} \sqrt{\Delta} \epsilon_{2,t+1}$ based on six moments $E(r_t^2), E(r_t^4), E(r_t r_{t+k}^2)$ for $k = 1, \ldots, 4$. GMM standard errors are in brackets. Based on 1 simulation of 100000 days.

<table>
<thead>
<tr>
<th>Data generating process</th>
<th>Half-daily estimates</th>
<th>Daily estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(80 five-minute periods per day)</td>
<td>200000 observations</td>
<td>100000 observations</td>
</tr>
<tr>
<td>$\Delta = \frac{1}{2}$</td>
<td>$\Delta = 1$</td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.025$</td>
<td>0.026</td>
<td>0.028</td>
</tr>
<tr>
<td>[0.003]</td>
<td>[0.008]</td>
<td></td>
</tr>
<tr>
<td>$\beta = -0.1$</td>
<td>-0.101</td>
<td>-0.109</td>
</tr>
<tr>
<td>[0.011]</td>
<td>[0.032]</td>
<td></td>
</tr>
<tr>
<td>$\sigma = 0.1$</td>
<td>0.104</td>
<td>0.105</td>
</tr>
<tr>
<td>[0.008]</td>
<td>[0.032]</td>
<td></td>
</tr>
<tr>
<td>$\rho = -0.5$</td>
<td>-0.509</td>
<td>-0.438</td>
</tr>
<tr>
<td>[0.038]</td>
<td>[0.118]</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$ (2)</td>
<td>1.162</td>
<td>1.371</td>
</tr>
<tr>
<td>$\text{Prob}[\chi^2(2) &gt; J]$</td>
<td>0.559</td>
<td>0.504</td>
</tr>
</tbody>
</table>

small) and the small number of moments. Andersen and Sørensen (1996) show that for a satisfactory GMM estimator performance of a simple discrete model with a constant volatility of volatility, one needs at least 4000 observations. My model has a stochastic volatility of volatility term, and its estimation via GMM is more complicated. The 5-minute estimates are significant, although the GMM standard errors are still high. The RMSE of $\rho$, surprisingly, is higher at the 5-minute level. The 5-minute estimates are much more reliable than the daily ones based on the GMM standard errors.

To summarize, the gain in accuracy of the measured correlation coefficient is substantial. This serves as good motivation for my current research.

3 Empirical findings

3.1 The squared return-lagged return correlation pattern

To test the hypothesis of prolonged leverage and volatility feedback effects empirically, I examine the correlation between squared returns and lagged returns for positive and negative lags; i.e. I look at the lead-lag relation between the square return and return in high frequency data.
Table 2: GMM estimation of the simulated discrete Heston square root model: $r_{t+1,\Delta} = \sigma_{t,\Delta}\sqrt{\Delta}\epsilon_{1,t+1}$ and $\sigma_{t+1,\Delta}^2 = \alpha\Delta + (1 + \beta\Delta)\sigma_t^2 + \sigma_t\Delta\sqrt{\Delta}\epsilon_{2,t+1}$ based on six moments $E(r_t^2), E(r_t^4), E(r_t r_{t+k}^2)$ for $k = 1, \ldots, 4$. Monte Carlo standard errors (RMSE) are in parenthesis; average GMM standard errors are in brackets. Based on 500 simulations of 4000 days.

<table>
<thead>
<tr>
<th>Data generating process</th>
<th>50 minute estimates</th>
<th>Daily estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(80 five-minute periods per day, $\Delta = 1/80$)</td>
<td>(8 fifty-minute periods per day, $\Delta = 1/8$)</td>
<td></td>
</tr>
<tr>
<td>$\alpha = 0.025$</td>
<td>0.034</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.012)</td>
</tr>
<tr>
<td></td>
<td>[0.013]</td>
<td>[0.133]</td>
</tr>
<tr>
<td>$\beta = -0.1$</td>
<td>-0.138</td>
<td>-0.102</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td>(0.049)</td>
</tr>
<tr>
<td></td>
<td>[0.054]</td>
<td>[0.551]</td>
</tr>
<tr>
<td>$\sigma = 0.1$</td>
<td>0.109</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.037)</td>
</tr>
<tr>
<td></td>
<td>[0.038]</td>
<td>[1.093]</td>
</tr>
<tr>
<td>$\rho = -0.5$</td>
<td>-0.464</td>
<td>-0.513</td>
</tr>
<tr>
<td></td>
<td>(0.213)</td>
<td>(0.184)</td>
</tr>
<tr>
<td></td>
<td>[0.223]</td>
<td>[2.496]</td>
</tr>
<tr>
<td>crashes</td>
<td>63*</td>
<td>68*</td>
</tr>
</tbody>
</table>

* Restriction $\beta \in [-1, 0]$, $\rho \in [-1, 1]$

two data sets: a Market Index constructed from the yearly average of 6340 stocks from January 5, 1993 to December 31, 1999 - 84 months, 1761 days, 139,119 five-minute intervals; and the S&P 500 tick-by-tick futures index from January 4, 1988 to March 9, 1999. I used futures because of their high trading frequency, small number of outliers and high purity in transactions recording. I interpolate five minute prices using the closest to the left tick price and construct five minute returns $r_{t,n} = \ln(P_{t,n}) - \ln(P_{t,n-1})$, where $t$ is the day number and $n$ is the intraday interval number. This way of interpolation can create a small number of additional jumps. Those jumps are negligible on the daily level but may be important on the high frequency level. I examined closely the interpolated

---

2I am thankful to Benjamin Zhang who provided me with the data. Compustat, CRSP, TAQ were the sources used to obtain the data.

3I wrote the code in SAS to linearly interpolate returns. SAS is especially good for a work with huge data sets like tick-by-tick data.

4I am thankful to George Tauchen for pointing that out to me. In general, it is better to take the log-difference of prices first, and then link the data together.
Table 3: Summary Statistics for five minute returns distributions

<table>
<thead>
<tr>
<th>Panel A: Market Index 5 minute Returns, 1993-1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>-2.24E-07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: S&amp;P 500 Futures Index 5 minute Returns, 1988-1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>2.16E-06</td>
</tr>
</tbody>
</table>

prices and found small jumps (around 1%) at the crossover dates. This, however, does not affect my series because I did not include the overnight. I also calculated daily returns ignoring the overnight (returns are calculated from open to close).

Summary statistics and plots of realized volatility and daily returns are given in Table 3 and Figures 1 and 2. The observed spikes correspond to the 1998 crisis. Both series do not include overnight returns. Moreover, the first return of the day is deleted because of very high volatility at the beginning of the day, leaving me with 77 five-minute returns per day. The figures described below are robust to the inclusion of the 78th return, although slightly less volatile. Because of the microstructure effects, the market index return series exhibits a strong positive first order autocorrelation, $\rho_1 = 0.19$. This can arise as a result of infrequent trading and the index averaging effect. One approach to deal with the high first order autocorrelation is to filter returns using a MA(1) model (see Section 5.3 for further discussion). The first order autocorrelation in the S&P 500 futures series is small and negative -0.028 and probably is a result of the market microstructure bounce effect.

The plots of the correlation between a proxy for the variance (square/absolute value of 5 minute return and realized volatility) and lagged 5 minute returns are given in Figure 3 and 4. Because of the fat tailed distribution of returns and the possible presence of outliers, a more robust proxy for volatility, namely the absolute value of returns, is also considered. The figures show a large negative impact of the return shock on volatility proxies, consistent with previous empirical studies. Since a squared daily return is quite a noisy estimate of the volatility, I obtained a very noisy picture for the daily data (see Figure 3). However, use of 5-minute data reveals a new effect of very substantial, prolonged (for around 350 five-minute lags or five days) and slowly decaying impact of the return shocks on future volatility’s proxy (the leverage effect). The decay of the leverage effect is immediate when using low frequency data. I also found an insignificant impact of volatility on future returns (the volatility feedback effect).

The observed slowly decreasing correlation between the volatility proxy and past returns is too significant to be solely explained by the leverage effect. This behavior might be partially explained by the empirically observed volatility behavior: volatility tends to rise quickly and sharply, but to
Table 4: GMM estimation of the \( \text{corr}(r_t^2, r_{t-j}) \), \( j = -5, \cdots, 5 \) for daily returns. 10 lags are included in the estimated weighting matrix.

<table>
<thead>
<tr>
<th>Lag</th>
<th>Coeff</th>
<th>SE</th>
<th>Coeff</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>lag-5</td>
<td>0.0421</td>
<td>0.0202</td>
<td>0.0128</td>
<td>0.0093</td>
</tr>
<tr>
<td>lag-4</td>
<td>0.0114</td>
<td>0.0214</td>
<td>0.0091</td>
<td>0.0076</td>
</tr>
<tr>
<td>lag-3</td>
<td>-0.0021</td>
<td>0.0213</td>
<td>-0.0073</td>
<td>0.0088</td>
</tr>
<tr>
<td>lag-2</td>
<td>0.0079</td>
<td>0.0212</td>
<td>0.0099</td>
<td>0.0090</td>
</tr>
<tr>
<td>lag-1</td>
<td>0.0923</td>
<td>0.0616</td>
<td>0.0131</td>
<td>0.0098</td>
</tr>
<tr>
<td>lag0</td>
<td>-0.0464</td>
<td>0.1352</td>
<td>-0.1072</td>
<td>0.0543</td>
</tr>
<tr>
<td>lag1</td>
<td>-0.0999</td>
<td>0.0485</td>
<td>-0.0384</td>
<td>0.0112</td>
</tr>
<tr>
<td>lag2</td>
<td>-0.0661</td>
<td>0.0321</td>
<td>-0.0223</td>
<td>0.0093</td>
</tr>
<tr>
<td>lag3</td>
<td>-0.0143</td>
<td>0.0109</td>
<td>-0.0195</td>
<td>0.0080</td>
</tr>
<tr>
<td>lag4</td>
<td>0.0181</td>
<td>0.0193</td>
<td>-0.0022</td>
<td>0.0077</td>
</tr>
<tr>
<td>lag5</td>
<td>-0.0016</td>
<td>0.0147</td>
<td>-0.0066</td>
<td>0.0100</td>
</tr>
</tbody>
</table>

decrease slowly and gradually.

The \( 2/\sqrt{T} \) standard errors band, shown on the figures, is a poor approximation of the standard error of a heterogeneous series. I use GMM to check the standard errors for correlations. I consider the cross-correlation with +/- 5 lags and include 10 lags in the information matrix. Reduction of the number of lags in the information matrix to 5 only slightly reduces the standard errors leaving all of the results mainly unchanged. The left columns in Table 4 show that all correlations except lag 1 and 2 computed from the daily data directly are insignificant. Correlations at lag 1 and 2 are significant at 5%. The right columns compute the same correlation using 5-minute returns and the aggregation technique discussed in Section 4. Now the daily correlations are significant up to three days.

### 3.2 Exponential vs Hyperbolic decay

This section examines the slowly decreasing squared return-return correlation pattern that is observed in high frequency data. Exponential and hyperbolic decays can be described by the following exponential and power functions\(^5\):

\[ f(j) = -k h^{-j}, \text{ where } k = a, h = e^{1/b} \text{ are two constants.} \]
Figure 1: Realized volatility and daily returns for the Market Index 1993-99.
Figure 2: Daily returns for the S&P 500 futures 1988-99.
Figure 3: Correlations between squared returns/realized volatility and lagged returns on the Market Index 1993-99.
Figure 4: Correlations between absolute value of returns and lagged returns on the S&P 500 Futures Index 1988-99.
\[ \text{corr}(r_t^2, r_{t-j}) = -a \exp\left(-\frac{j}{b}\right) \]
\[ \text{corr}(r_t^2, r_{t-j}) = -kj^{-h} \]

where \(a, b, k, h\) are constants. GMM estimates of the exponential function based on 200 lags computed with a step of 5 lags (total 41 moments) are \(a = 0.021, b = 53.08\) with the standard errors 0.004, 11.92 correspondingly. The J-statistic is 21.79 with probability \(P[\chi^2(39) > J] = 0.988\). Both the amplitude \(a\) and the time of decay \(b\) are significant. The lags \(j\) are five-minute, hence, the time of decay is approximately between 30 and 77 five-minute periods, which is at most one day. Bouchaud, Matacz and Potters (2001) found the time of decay to be around 9.3 days for the average of the 7 major stock indices.

GMM estimates of the hyperbolic function based on 200 lags computed with a step of 5 lags (total 41 moments) are \(k = 0.036, h = 0.365\) with the standard errors 0.006, 0.650 correspondingly. The J-statistic is 32.84 with probability \(P[\chi^2(39) > J] = 0.746\). The amplitude \(k\) is significant but the time of decay \(h\) is not. The decay behavior is better described by the exponential function based on 200 lags and the step 5, as one can judge from the J-statistic. The same conclusion is robust to the number of lags equal to 400 and the step 5. Figure 5 shows the decay pattern with the exponential and hyperbolic functions.

4 Aggregation

4.1 Single variable case

I study the correlation between the squared return and lagged return under temporal aggregation. The derivation of an aggregation formula on the daily level for the correlation between the squared return and lagged return calculated from 5-minute returns is presented in Appendix A in equation (8). The derived formula helps to examine this relation on a daily level using 5-minute data.

\[
\tau \neq 0 \quad \text{cov}(r_{t,1}^2; r_{t-\tau,1}) = \sum_{i=-n}^{n} (n - |i|) \text{cov}(r_{t-i\Delta,\Delta}^2; r_{t-\tau,\Delta}) \\
\tau = 0 \quad \text{cov}(r_{t,1}^2; r_{t,1}) = 3 \sum_{i=-n}^{n} (n - |i|) \text{cov}(r_{t-i\Delta,\Delta}^2; r_{t,\Delta}) - 2n \text{cov}(r_{t,\Delta}^2, r_{t,\Delta})
\]

where \(r_{t,1} = \sum_{i=1}^{n} r_{t-i\Delta,\Delta}\) denotes daily return, \(r_{t,\Delta}\) denotes high-frequency (5 minute) return and \(n, T\) are number of intraday intervals and days respectively.

As mentioned above, the squared return-lagged return correlation is very noisy on the daily level. Using my temporal aggregation formula, I study this relation on a daily level using 5-minute data.
Figure 5: The exponential fit to the return-squared return correlation in 5-minute S&P 500 futures data. Three colored lines show the exponential decay estimated with 100 lags and the step 10 lags, 200 lags and the step 5 lags, 400 lags and the step 5 lags.

As discussed in Section 3.1, daily correlations computed directly from the daily data are all insignificant. The same correlations computed using the aggregation technique are significant up to three days (right columns in Table 4). Figures 6 and 7 show the daily correlations computed from daily data and using the aggregation formula for unstandardized and standardized returns on S&P 500 futures and Market Index data. Different standardization techniques are described in the Appendix, Section B. Evidently high frequency data contains more information about the process, which can be lost in low frequency data. The use of high frequency data demonstrates a much more important role of the leverage effect in explaining volatility asymmetry.

4.2 Bivariate case

An interesting generalization of the aggregation formula is the bivariate case. Similar to the one variable case, it can be shown that

\[
\tau \neq 0 \quad \text{cov}(x_{t,1}^2, y_{t-\tau,1}) = \sum_{i=-n}^{n} (n - |i|) \text{cov}(x_{t-i\Delta,\Delta}^2 ; y_{t-\tau,\Delta})
\]

\[
\tau = 0 \quad \text{cov}(x_{t,1}^2, y_{t,1}) = \sum_{i=-n}^{n} (n - |i|) \left( 2\text{cov}(x_{t-i\Delta,\Delta} x_{t,\Delta}; y_{t,\Delta}) + \text{cov}(x_{t-i\Delta,\Delta}^2 ; y_{t,\Delta}) \right) - 2n \text{cov}(x_{t,\Delta}^2, y_{t,\Delta})
\]
Figure 6: Daily squared return-lagged return correlation computed from daily (red) data and using the aggregation formula for 5-minute standardized (yellow) and unstandardized (black) returns on S&P 500 futures.

where $x_{t,\Delta}, y_{t,\Delta}$ are high frequency returns on two different stocks or indexes, or $x_{t,\Delta}$ is the market return and $y_{t,\Delta}$ is the firm return. The later case is a generalization of the aggregation formula to the firm-market level. An alternative measure of volatility in this case is a product of market and firm returns. Only when this product is positive can the volatility feedback effect be observed at the firm level. The formulas are transformed to

\[
\begin{align*}
\tau \neq 0 & \quad \text{cov}(x_{t,1}y_{t,1}; y_{t-\tau,1}) = \sum_{i=-n}^{n} (n - |i|) \text{cov}(x_{t-i,\Delta}y_{t-1-i,\Delta}; y_{t-\tau,\Delta}) \\
\tau = 0 & \quad \text{cov}(x_{t,1}y_{t,1}; y_{t,1}) = \sum_{i=-n}^{n} (n - |i|) \left( \text{cov}(x_{t-i,\Delta}y_{t,\Delta}; y_{t,\Delta}) + \text{cov}(x_{t,\Delta}y_{t-1-i,\Delta}; y_{t,\Delta}) \right) \\
& \quad + \text{cov}(x_{t-1-i,\Delta}y_{t-i,\Delta}; y_{t,\Delta})) - 2n \, \text{cov}(x_{t,\Delta}y_{t,\Delta}; y_{t,\Delta})
\end{align*}
\]

The bivariate formula would help to study the behavior of pairs of indexes and the interaction between them. The mutual effects of different indexes on different financial world markets would help to measure markets integration and better explain the mechanisms of world crises. The idea is to test whether a crisis in one country (a large index price decline) has power to predict crises in another country (a high level of volatility and hence, a very low price level).
Figure 7: Daily squared return-lagged return correlation computed from daily data (red) and using the aggregation formula for 5-minute standardized (yellow) and unstandardized (black) returns on MI data.

5 Estimated Heston model

The observed, slowly decreasing squared return-lagged return correlation pattern in 5-minute data exhibits a geometric rate of decay. The daily correlation decays substantially quicker. Most of the simple models widely used in practice today fail to produce this pattern on a daily level. The Heston square root model produces the geometrically decreasing pattern (see Section 2), which matches the pattern observed in high-frequency data, but is unable to produce the pattern in daily data.

To show this I first estimate the Heston model on a daily level using the EMM, the method proposed by Gallant and Tauchen (1996)\(^6\), and demonstrate that the squared return-lagged return correlation patterns obtained from the daily data and from the simulated Heston model are different. Second, I calibrate the parameters of the Heston model that generate the pattern observed in 5 minute data. I match the sample correlations, \(\text{corr}(r_t^2, r_{t-j})\), computed on 5 minute data to the population correlations computed using a sequence of simulated observations generated from the Heston model. Basically, I "estimate" the Heston model on the 5 minute level using the simulated method of moments proposed by Duffie and Singleton (1993). As the set of moments, I use only the

\(^6\)SNP and EMM FORTRAN codes are written by Gallant and Tauchen and are available free of charge at UNC/Duke server.
correlations between squared returns and lagged returns. The Heston square root model cannot be really estimated on the 5 minute level since it is unable to capture all aspects of the high frequency data like extremely high kurtosis (around 70). By using the restricted set of moments of interest, I can nevertheless calibrate the parameters that produce the observed pattern.

5.1 EMM Estimates of Daily Returns

The well-known problem of the non-availability of a closed form for the a transition density makes a lot of standard estimation techniques infeasible. However, the expectation of a function \( g \) can be computed, for a given \( \rho \), by a simulation \( \hat{y}_t \) from the generic system as long as the number of simulations, \( N \), is large enough:

\[
\varepsilon_\rho(g) = \int \ldots \int g(y_{-L}, \ldots, y_0)p(y_{-L}, \ldots, y_0|\rho)dy_{-L} \ldots dy_0
\]

\[
\varepsilon_\rho(g) = \frac{1}{N} \sum_{t=1}^{N} g(\hat{y}_{t-L}, \ldots, \hat{y}_t)
\]

where \( y_t \) is a discrete stationary time series, \( p(y_{-L}, \ldots, y_0|\rho) \) is the density, \( \rho \) is a vector of unknown parameters. Gallant and Tauchen (1996) propose a minimum chi-square estimator for the parameter of interest termed as the efficient method of moments (EMM). The moments of the minimum chi-square estimator are obtained from the score vector \( \frac{\partial}{\partial \theta} \log f(y_t|x_{t-1}, \theta) \) of an auxiliary model (score generator) \( f(y_t|x_{t-1}, \theta) \) where \( x_{t-1} \) is a lagged state vector. It is shown that if the SNP density \( f_K(y|x, \theta_K) \) described below is taken as the score generator, then the efficiency of the EMM can be made as close to the maximum likelihood as desired by taking \( K \) large enough. Briefly the steps of EMM are: use the score generator \( f(y_t|x_{t-1}, \theta) \) to get the data \( \{\hat{y}_t, \hat{x}_{t-1}\}_{t=1}^n \) and compute the QMLE:

\[
\hat{\theta}_n = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \log[f(\hat{y}_t|\hat{x}_{t-1}, \theta)]
\]

with the information matrix estimated as

\[
\hat{\Upsilon}_n = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} \log f(\hat{y}_t|\hat{x}_{t-1}, \hat{\theta}_n))\left[\frac{\partial}{\partial \theta} \log f(\hat{y}_t|\hat{x}_{t-1}, \hat{\theta}_n)\right]'
\]

Note that the terms are serially uncorrelated if the score generator is a good approximation to the transition density of the data. The moment \( m(\rho, \theta) = \varepsilon_\rho(\frac{\partial}{\partial \theta} \log f(y_0|x_{-1}, \theta)) \) is computed by averaging over a long simulation

\[
m(\rho, \theta) = \frac{1}{N} \sum_{t=1}^{N} \frac{\partial}{\partial \theta} \log f(\hat{y}_t|\hat{x}_{t-1}, \theta)]
\]

\[7\]There is no pre-packaged statistical software for this type of work. My SMM code relies on a fairly widely distributed optimization routine by Gill, Murray, Saunders and Wright (1986). NPSOL is a Fortran package designed to solve the nonlinear programming problem: the minimization of a smooth nonlinear function subject to a set of constraints on the variables.
Then the EMM estimator is

\[ \hat{\rho}_n = \arg \min_{\rho \in \mathbb{R}^p} m'(\rho, \tilde{\theta}_n)(\tilde{Y}_n)^{-1}m(\rho, \tilde{\theta}_n) \]

The SNP is the best score generator for the EMM. The method is based on the Hermite series functional expansion of the conditional density of a multivariate process. It is shown that the score of the auxiliary model that closely approximates the density of the data is the best choice of a moment function to implement EMM.

Let \( y_t \) denote a multivariate state vector of length \( M \); \( x_{t-j} \) is the lagged state vector with lag \( j \). The stationary density \( p(y_{-L}, \cdots, y_0 | \rho) \) has the transition density:

\[ p(y_t | x, \rho) = \frac{p(x, y_t | \rho)}{\int p(x, y | \rho) \, dx} \]

As noted above, one has to expand the square root of the density in a Hermite series. Then it is shown that the transition density of the truncated expansion \( f_K(y_t | x_{t-1}) \) has a location-scale transformation:

\[ y_t = R z_t + \mu_{x_{t-1}} \]  

The density function of the innovation \( z_t \) is

\[ h_K(z|x) = \frac{[P(z, x)]^2 \phi(z)}{\int[P(u, x)]^2 \phi(u) \, du} \]

where \( P(z, x) \) is a polynomial in \((z, x)\) of degree \( K = K_z + K_x \) and \( \phi(z) \) denotes the multivariate standard normal density. The polynomial is of degree \( K_z \) in \( z \) whose coefficients are, in turn, polynomials of degree \( K_x \) in \( x \):

\[ P(z, x) = \sum_{\alpha=0}^{K_z} \left( \sum_{\beta=0}^{K_x} a_{\alpha \beta} x^\beta \right) z^\alpha \]  

To ensure a unique representation, the constant term in (4), \( a_{00} \), is set equal to 1. With this normalization, the density function of innovation is equal to the normal plus some higher order terms.

The location function in (3) is linear \( \mu_x = b_0 + B x_{t-1} \). The scaling function is allowed to depend on \( x, y = R x z + \mu_x \). In general, \( R_{x_{t-1}} \) has the ARCH/GARCH specification

\[ vech(R_{x_{t-1}}) = \rho_0 + \sum_{i=1}^{L_u} P_{(i)} |y_{t-1-L_u+i} - \mu_{x_{t-2-L_u+i}}|^2 + \sum_{i=1}^{L_g} diag(G_{(i)}) R_{x_{t-2-L_g+i}} \]

The model \((L_u, L_g, L_r, L_p, K_z, I_z, K_x, I_x) = (1, 1, 1, 8, 0, 0, 0, 0, 0)\) is the preferred model under BIC for the S&P 500 futures daily data. This SNP model can be characterized as a GARCH(1,1) with a nonparametric error density represented as an eight degree Hermite expansion. Most of the
Figure 8: Daily squared return-lagged return correlation computed from the daily S&P 500 data (black), from 5-minute data using the aggregation formula (yellow) and from the estimated Heston model using simulated observations (pink).

\( t \)-statistics of even coefficients (odd power), \( A_{2k} = a_{0,2k-1} \), are insignificant, as one can see from Table 5.

The EMM estimates of daily S&P 500 futures are in Table 6. To prevent the problem of the discrete volatility being negative, I simulate the square root volatility process, \( Z_t = \sqrt{V_t} \), using the Ito lemma. Alternatively, one can simulate the log-transform. The square root process, however, has a constant diffusion term which makes the simulation more stable.

\[
\begin{align*}
dZ_t &= -\frac{\beta}{2Z_t} \left( -\frac{\alpha}{\beta} + \frac{\sigma^2}{2\beta} - Z_t^2 \right) dt + \frac{1}{2}\sigma dW_t \\
\end{align*}
\]

Figure 8 shows the daily correlations based on daily data, the aggregation formula and the estimated Heston model. The simulated correlations (computed using a sequence of simulated observations from the estimated model) are very different from the observed daily correlations. This fact together with a high J-statistic (\( \chi^2(8) = 29.95 \)) confirms a well-known result that the Heston model fits the data poorly.
Table 5: t-ratio diagnostics.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location Function:</td>
<td></td>
</tr>
<tr>
<td>$b_0$</td>
<td>psi(1)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>psi(2)</td>
</tr>
<tr>
<td>Scale Function:</td>
<td></td>
</tr>
<tr>
<td>$\tau_0$</td>
<td>tau(1)</td>
</tr>
<tr>
<td>$\tau_{gx}$</td>
<td>tau(2)</td>
</tr>
<tr>
<td>$\tau_{gx}$</td>
<td>tau(3)</td>
</tr>
<tr>
<td>Hermite Polynomial:</td>
<td></td>
</tr>
<tr>
<td>$a_{0,1}$ A(2)</td>
<td>-2.044</td>
</tr>
<tr>
<td>$a_{0,2}$ A(3)</td>
<td>-17.999</td>
</tr>
<tr>
<td>$a_{0,3}$ A(4)</td>
<td>1.231</td>
</tr>
<tr>
<td>$a_{0,4}$ A(5)</td>
<td>9.943</td>
</tr>
<tr>
<td>$a_{0,5}$ A(6)</td>
<td>-0.928</td>
</tr>
<tr>
<td>$a_{0,6}$ A(7)</td>
<td>-7.266</td>
</tr>
<tr>
<td>$a_{0,7}$ A(8)</td>
<td>0.487</td>
</tr>
<tr>
<td>$a_{0,8}$ A(9)</td>
<td>6.342</td>
</tr>
</tbody>
</table>

5.2 SMM Estimates of Five-Minute Returns

Second, the results in Section 2 show that the Heston model can produce the slowly decreasing correlation pattern observed in high frequency data. This model cannot fit the data at the 5 minute level, but I calibrate the parameters that produce the observed high frequency data pattern using the simulated method of moments (SMM), as proposed by Duffie and Singleton (1993). The idea is to match the sample moments of the correlation between squared returns and lagged returns to the population moments calculated from the model, i.e. moments computed using a sequence of simulated observations generated from the assumed model. Briefly, the SMM problem can be described as follows. Denote the sample correlation between the squared return and lagged return as $c_i = \text{corr}(r_{t,\Delta}^2; r_{t-i,\Delta})$, where the notation is consistent with that used in Appendix A. Denote $c_i(\xi) = \text{corr}(r_{t,\Delta}^2(\xi); r_{t-i,\Delta}(\xi))$ - the correlation computed using a sequence of simulated returns generated from model (1). Then the moment conditions are:
Table 6: EMM estimation for the Heston square root model fitted to daily S&P 500 futures 1988-99: 
\[ dp_t = \mu dt + \sqrt{V_t} dW_{t,1}, \text{ and } dV_t = (\alpha - \beta V_t) dt + \sigma \sqrt{V_t} dW_{t,2}. \] Conventional Wald-type standard errors determined by numerical differentiation are the third column.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu)</td>
<td>0.0174</td>
<td>0.0131</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>0.0152</td>
<td>0.0056</td>
</tr>
<tr>
<td>(\beta)</td>
<td>0.0259</td>
<td>0.0099</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>0.1160</td>
<td>0.0259</td>
</tr>
<tr>
<td>(\rho)</td>
<td>-0.3738</td>
<td>0.1278</td>
</tr>
<tr>
<td>(\chi^2)</td>
<td>29.9475</td>
<td></td>
</tr>
</tbody>
</table>

\[ m_t(\xi) = \begin{pmatrix}
    c_{-i}(\xi) - c_{-i} \\
c_{-i+1}(\xi) - c_{-i+1} \\
    \ldots \\
c_{-1}(\xi) - c_{-1} \\
c_0(\xi) - c_0 \\
c_1(\xi) - c_1 \\
    \ldots \\
c_i(\xi) - c_i
\end{pmatrix} \]

where \(\xi = (\mu, \kappa, \theta, \sigma, \rho)\) is a vector of parameters, \(i\) is the number of lags. \(^8\) Since \(E[m_t(\xi)] = 0\) by construction, the GMM/SMM estimator is the one that solves the problem:

\[ \min_{\xi} \hat{m}(\xi)' W \hat{m}(\xi) \]

where \(\hat{m}(\xi) = \frac{1}{T} \sum_{t=1}^{T} m_t(\xi)\). Because of the structure of the moment conditions, I use the heteroskedasticity and autocorrelation robust Newey-West covariance matrix estimator.

The calibrated parameters are in reported Table 7. The standard errors are very high, which can be explained by the attempt to fit the misspecified model to high frequency data. The point estimates are correct, since the simulated and observed patterns are very close to each other as one can see in Figure 9. Finally, the aggregated calibrated Heston model produces the pattern which closely matches the one obtained from the aggregation of 5-minute data as shown in Figure 10.

\(^8\)The efficiency of the constructed estimator depends on the particular choice of instruments. Here I propose to use simply lagged moments as instruments. The number of moment conditions is flexible and is established during the estimation procedure.
An interesting approach is described in Bollerslev and Zhou (2001). Their idea is to match sample moments of the realized volatility to the population moments of integrated volatility implied by a particular continuous time stochastic volatility model. They consider the Heston square root model with the leverage effect and jumps, and obtain closed form expressions for the population moments of integrated volatility and estimate the model via GMM. The closed form expression for the correlation between realized volatility and log-price allows them to estimate the leverage effect. Their method can be applied to solve my problem when one considers the moments generated by the correlation between the realized volatility and lagged return instead of between the squared return and lagged return.

5.3 Discussion of sampling frequency

The optimal frequency of returns that can be used in an empirical analysis is still an open question. Ideally, it should be high enough to produce a good estimate of the integrated volatility and low enough to be free of market microstructure bias. Some recent studies, for example Andersen, Bollerslev, Diebold and Labys (2000), show that in order to have an unbiased estimator with FX data, one should use returns with frequencies not higher than 30 minutes - 2 hours. Another possibility for dealing with market microstructure effects is to perform MA(1) filtering. The filtered
Figure 10: Correlations between squared returns and lagged returns observed in daily data (black), and obtained from aggregation of the calibrated Heston model (pink) and of 5-minute data (yellow), S&P 500.

data will be free of negative first order autocorrelation, which can reach up to 40% for tick-by-tick data, and which is usually a sign of the microstructure effect. The last approach is reasonable for an individual stock but might be too simplistic for a stock index. A non-trivial lead-lag structure between different stocks in the index induces a more complex autocorrelation structure in the index. A Monte Carlo study of an efficient estimation of volatility using high frequency data is recently considered by Zumbach, Corsi and Trapletti (2002).

The optimal sampling frequency should be determined during the estimation procedure. A good starting point would be to aggregate the data to the hourly level. Market micro structure noise would be partially washed out after the aggregation procedure, and the remaining noise would be small enough to get an unbiased estimate according to the empirical studies mentioned above. Another interesting possibility is to estimate the data on the 5 minute or even higher frequency levels in order to have the market micro structure noise present. The time varying leverage effect might arise as a result of the processes on the market micro level as well. Currently I am working on estimating the Heston model and the two factor model that is discussed below in the hourly data.
Table 7: SMM estimation of the Heston square root model "fitted" to 5 minute S&P 500 futures
1988-99: \( dp_t = (\mu + cV_t)dt + \sqrt{V_t}dW_{t,1}, \) and \( dV_t = (\alpha - \beta V_t)dt + \sigma \sqrt{V_t}dW_{t,2}. \) All numbers except the correlation coefficient \( \rho \) are multiplied by 10.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SMM: 11 lags</th>
<th>SMM: 135 lags</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>0.044</td>
<td>-0.038</td>
</tr>
<tr>
<td></td>
<td>(0.165)</td>
<td>(1.081)</td>
</tr>
<tr>
<td>( c )</td>
<td>-1.1E-05</td>
<td>7.9E-05</td>
</tr>
<tr>
<td></td>
<td>(0.167)</td>
<td>(1.189)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.159</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.318)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.185</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.361)</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>1.499</td>
<td>0.890</td>
</tr>
<tr>
<td></td>
<td>(0.904)</td>
<td>(2.528)</td>
</tr>
<tr>
<td>( \rho )</td>
<td>-0.308</td>
<td>-0.325</td>
</tr>
<tr>
<td></td>
<td>(0.161)</td>
<td>(0.921)</td>
</tr>
<tr>
<td>( \chi^2 - \text{square} )</td>
<td>14.602</td>
<td>0.812</td>
</tr>
<tr>
<td>( df )</td>
<td>53</td>
<td>53</td>
</tr>
</tbody>
</table>

\( n2: E(r_t), E(r_t^2) \)
\( c57: E(|r_t| r_{t-j}), j = -80...200, \) step 5

Standard Errors should be multiplied by \( \sqrt{1 + \tau} = 1.77, \) \( \tau = n/N = 2.12 \)

6 Log Linear Two Factor Stochastic Volatility Model

An extension of the one factor Heston model is a model involving two stochastic volatility factors. The presence of two volatility factors breaks the link between tail thickness and volatility persistence (Chernov et al. (2002)). The first factor in logarithmic specification captures the long-memory persistent features of volatility, while the second one is responsible for the very short-lived but erratic and extreme event behavior.

The preliminary simulations suggest that the two-factor model can produce the observed patterns both in high and low frequency data. I simulated a two factor SV model to check whether the correlation pattern produced by this model matches the pattern observed in real data better than the Heston one factor SV model. I took the parameters from Chernov, et al. (2002), who estimated a two factor model on Dow Jones Industrial Average 1953-99 returns. They found that the log linear
model has to be extended to two factors with feedback in the mean reverting factor (LL2VI). Thus, their preferred logarithmic model is:

\[
\frac{dP_t}{P_t} = (\alpha_{10} + \alpha_{12} U_{2t}) dt + \exp(\beta_{10} + \beta_{13} U_{3t} + \beta_{14} U_{4t}) \left( \sqrt{1 - \psi_{13}^2 - \psi_{14}^2} dW_{1t} + \psi_{13} dW_{3t} + \psi_{14} dW_{4t} \right)
\]

\[
dU_{2t} = \alpha_{22} U_{2t} dt + dW_{2t}
\]

\[
dU_{3t} = \alpha_{33} U_{3t} dt + dW_{3t}
\]

\[
dU_{4t} = \alpha_{44} U_{4t} dt + (1 + \beta_{44} U_{4t}) dW_{4t}
\]

The simulated model (although it was estimated initially on DJNS returns) produces the correlation pattern close to that observed in S&P 500 futures data on the daily level. The pattern is similar to that observed in the high-frequency data, but the pattern generated by the simulated model decays slightly slower. Currently I am working on estimating LL2VI model on S&P 500 futures data on a daily and hourly level.

7 Time varying leverage effect

The slowly decreasing, and different for different periods correlation between return and future volatility together with the study of the intraday pattern of this correlation for different lags suggests that the leverage effect is time varying. Empirical evidence based on options data supports this fact. Garsia, Luger and Renault (2001) considered a discrete time model with a time varying leverage effect. Another way to incorporate a time varying leverage effect is to consider the aforementioned multifactor stochastic model for the variance process where the Brownian motion associated with the volatility’s factors are correlated to the stock’s Brownian motions. Then the leverage effect would be different for the various components of volatility (Chernov, Gallant, Ghysels and Tauchen (2002)).

Another approach described by Meddahi (2001) is to model the correlation coefficient simply as a two stage discrete Markov chain. One stage will correspond to a positive shock (good news) and another to a negative shock (bad news). An alternative approach outlined by Meddahi (2001) is based on modeling the instantaneous correlation coefficient as a stochastic process by itself, e.g. a Jacobi Diffusion process, Karlin and Taylor (1981):

\[
d\rho_t = \frac{1}{2}[(\beta + 1)(1 - \rho_t) - (\alpha + 1)(1 + \rho_t)] dt + \sqrt{1 - \rho_t^2} dW_{\rho t} (6)
\]

This diffusion process is defined on a finite closed interval \([-1, 1]\). The volatility term \(\sigma^2(\rho_t) = (1 - \rho_t^2)^2\) is continuous and positive over the whole interval. For a positive \(\alpha\) and \(\beta\), (6) is a well defined process with mean reverting correlation. The unconditional mean of the process is \((\beta - \alpha)/(\alpha + \beta + 2)\).
The process $\rho_t$ is defined on the $[-1, 1]$ interval. However, the discretized process does not always stay inside the interval. The difficulty of its simulation can be overcome by taking the logistic transform of the $\rho_t$ process, $Y_t = \ln \left( \frac{1 + \rho_t}{1 - \rho_t} \right)$, which transforms the $[-1, 1]$ interval into a $(-\infty, +\infty)$ interval. The transformed process follows the following stochastic process according to the Ito lemma:

$$dY_t = \left[ \frac{1}{2Y_t} \left( 1 - \frac{1}{2Y_t} e^{Y_t} + 2(\beta + 1) - 2e^{Y_t}(\alpha + 1) \right) \right] dt + \frac{\sqrt{e^{Y_t}}}{Y_t} dW_{\rho,t} \quad (7)$$

First, I propose to estimate the Heston square root model (1) with a third equation of time-varying leverage effect (6). Once the model is successfully estimated via EMM, it will be possible to add other features, like multifactor volatility. Finally, I can check whether the two effects, the leverage and the volatility feedback, can fully explain volatility asymmetry in high frequency data.

8 Conclusion

Empirical study of the lead-lag relation between volatility proxies and returns in high frequency data shows that the leverage effect has a very prolonged exponentially decreasing pattern significant for around 3 days.

I derive the temporal aggregation formula of the correlation between squared returns and lagged returns. This formula helps to measure the daily correlation using high frequency data more accurately than using daily data directly. One possible and very interesting extension of the aggregation formula is to the bivariate case. The bivariate formula would help to study the behavior of pairs of indices and the interaction between them. The mutual effects of different indices on different financial world markets would help to measure markets integration and better explain the mechanisms of world crises. The idea is to test whether a crisis in one country (a large index price decline) has power to predict crises in another country (a high level of volatility and hence, a very low price level).

I shown that the simple one-factor continuous time stochastic volatility model, the Heston square root model, is unable to capture the observed features in high and low frequency data. The extension to a two-factor SV model can preliminary produce the pattern observed in both high and low frequency data.

The slowly decreasing, and different for different periods squared return-lagged return correlation together with the study of the intraday pattern of this correlation for different lags suggests that the leverage effect is time varying. Empirical evidence based on options data support this fact. Garcia, Luger and Renault (2001) considered a discrete time model with a time varying leverage effect. Another way to incorporate a time varying leverage effect is to consider the aforementioned multifactor stochastic model for the variance process where the Brownian motion
associated with the volatility factors are correlated with the Brownian motion for the price. The leverage effect would be then different for the various components of volatility (Chernov, Gallant, Ghysels and Tauchen (2000)). Another approach is to model the correlation coefficient simply as a two stage discrete Markov chain. One stage will correspond to a positive shock (good news) and another to a negative shock (bad news). An alternative approach is based on modeling the instantaneous correlation coefficient as a stochastic process by itself (e.g. a Jacobi Diffusion process). I propose to estimate the Heston model with the stochastic correlation coefficient to explore whether the relaxing of the constant correlation assumption can improve the fit to the data. To be completed
References


30


[29] Karatzas, I. and Shreve, S., 1988, "Brownian Motion and Stochastic Calculus", *Springer-Verlag*


A Aggregation formula derivation

Let me use the following notations: \( r_{t,h} \equiv \log P_t - \log P_{t-h} \), \( r_{t,1} \equiv \sum_{i=1}^{n} r_{t-i\Delta,\Delta} \) denotes daily return, \( r_{t,\Delta} \) denotes high-frequency (5 minute) return and \( n, T \) are number of intraday intervals and days respectively. Two assumptions are necessary to derive the aggregation formula:

1. \( r_t \) is covariance stationary:

\[
\text{cov}(r_{t-(i-h)\Delta,\Delta}; r_{t-\tau+h\Delta,\Delta}) = \text{cov}(r_{t-i\Delta,\Delta}; r_{t-\tau,\Delta})
\]

This assumption is likely to hold for de-seasonalized high-frequency returns.

\[
r_{t,h} = \frac{\tilde{r}_{t,h}}{s_h}, \quad s_h = \frac{1}{T} \sum_{t=1}^{T} r_{t,h}^2, \quad h = 1...n,
\]

2. Expected value of product of returns equal to zero, unless two span the same time interval:

\[
E(r_{t-i\Delta,\Delta}r_{\tau-j\Delta,\Delta}r_{s-h\Delta,\Delta}) \begin{cases} 
\neq 0 & \text{if } t = \tau, i = j \text{ or } t = s, i = h \\
\text{or } s = \tau, h = j & \\
= 0 & \text{otherwise}
\end{cases}
\]

Lemma 1

\[
\tau \neq 0 \quad \text{cov}(r_{t,1}^2; r_{t-\tau,1}) = \sum_{i=-n}^{n} (n - |i|) \text{cov}(r_{t-i\Delta,\Delta}^2; r_{t-\tau,\Delta}) \tag{8}
\]

\[
\tau = 0 \quad \text{cov}(r_{t,1}^2; r_{t,1}) = 3 \sum_{i=-n}^{n} (n - |i|) \text{cov}(r_{t-i\Delta,\Delta}^2; r_{t,\Delta}) - 2n \text{ cov}(r_{t,\Delta}^2, r_{t,\Delta})
\]

\[
\text{corr}(r_{t,1}^2; \tilde{r}_{t-\tau,1}) \approx \sigma^3(r_{t,\Delta}) \frac{\text{cov}(r_{t,1}^2; r_{t-\tau,1})}{\sigma(r_{t,1})^2 \sigma(\tilde{r}_{t,1})}
\]

Proof:

Assume \( \tau \neq 0 \):

\[
\text{cov}(r_{t,1}^2; r_{t-\tau,1}) = \text{cov} \left( \left( \sum_{i=1}^{n} r_{t-i\Delta,\Delta} \right)^2; \sum_{j=1}^{n} r_{t-\tau-j\Delta,\Delta} \right)
\]

\[
= \text{cov} \left( \sum_{i=1}^{n} r_{t-i\Delta,\Delta}^2; \sum_{j=1}^{n} r_{t-\tau-j\Delta,\Delta} \right) + 2 \text{cov} \left( \sum_{i=1}^{n} r_{t-i\Delta,\Delta} r_{t-k\Delta,\Delta}; \sum_{j=1}^{n} r_{t-\tau-j\Delta,\Delta} \right)
\]

\[
\approx \sum_{i=1}^{n} \sum_{j=1}^{n} \text{cov} (r_{t-i\Delta,\Delta}^2; r_{t-\tau-j\Delta,\Delta}) \approx \sum_{i=1}^{n} \sum_{j=1}^{n} \text{cov} (r_{t-(i-j)\Delta,\Delta}^2; r_{t-\tau,\Delta})
\]

\[
= \sum_{i=1}^{n} \sum_{k=1-n}^{i-1} \text{cov} (r_{t-k\Delta,\Delta}^2; r_{t-\tau,\Delta}) = \sum_{i=-n}^{n} (n - |i|) \text{cov}(r_{t-i\Delta,\Delta}^2; r_{t-\tau,\Delta})
\]

33
where equality * is made since the last term is equal to zero by assumption 2, ** is by assumption 1, *** - substitute \( k = i - j \).

Now assume that \( \tau = 0 \):

\[
\text{cov}(r_{t,1}^2; r_{t,1}) = \text{cov}\left( \sum_{i=1}^{n} r_{t-i,\Delta} \sum_{j=1}^{n} r_{t-j,\Delta} \right)
\]

\[
= \text{cov}\left( \sum_{i=1}^{n} r_{t-i,\Delta}^2 \sum_{j=1}^{n} r_{t-j,\Delta} \right) + 2 \text{cov}\left( \sum_{i<k} r_{t-i,\Delta} r_{t-k,\Delta} \sum_{j=1}^{n} r_{t-j,\Delta} \right)
\]

The last term is not equal to zero if and only if \( j = i \) or \( j = k \) (by assumption 2). Let’s rearrange the last term taking into account:

\[
\sum_{i<k} = \sum_{i=1}^{n-1} \sum_{k=i+1}^{n}
\]

\[
2 \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sum_{k=i+1}^{n} \text{cov}(r_{t-i,\Delta} r_{t-k,\Delta}; r_{t-j,\Delta})
\]

\[
= 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\text{cov}(r_{t-i,\Delta} r_{t-j,\Delta}; r_{t-i,\Delta}) + \text{cov}(r_{t-i,\Delta} r_{t-j,\Delta}; r_{t-j,\Delta}))
\]

\[
= 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{cov}(r_{t-i,\Delta} r_{t-(i-j),\Delta}; r_{t,\Delta}) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{m=-k}^{n} \text{cov}(r_{t-(i-j),\Delta} r_{t,\Delta}; r_{t,\Delta})
\]

\[
= 2 \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} \sum_{j=i+1}^{n} \text{cov}(r_{t-i,\Delta} r_{t-k,\Delta}; r_{t,\Delta}) + 2 \sum_{i=1}^{n-1} \sum_{m=1-n}^{n-1} \text{cov}(r_{t-m,\Delta} r_{t,\Delta}; r_{t,\Delta})
\]

\[
= 2 \sum_{i=1}^{n} (n-j) \left( \text{cov}(r_{t-i,\Delta} r_{t-j,\Delta}; r_{t,\Delta}) + \text{cov}(r_{t+j,\Delta} r_{t,\Delta}; r_{t,\Delta}) \right)
\]

\[
= 2 \sum_{j=-n}^{n} (n-|j|) \text{cov}(r_{t-i,\Delta} r_{t-i-j,\Delta}; r_{t,\Delta}) - 2n \text{cov}(r_{t,\Delta}^2, r_{t,\Delta})
\]

where equality * is made by assumption 1. Finally:

\[
\text{cov}\left( \sum_{i=1}^{n} r_{t-i,\Delta} \right)^2 \sum_{j=1}^{n} r_{t-j,\Delta} = \sum_{i=-n}^{n} (n-|i|) \text{cov}(r_{t-i,\Delta}^2; r_{t,\Delta})
\]

\[
+ 2 \sum_{i=-n}^{n} (n-|i|) \text{cov}(r_{t-i,\Delta} r_{t-i-j,\Delta}; r_{t,\Delta}) - 2n \text{cov}(r_{t,\Delta}^2, r_{t,\Delta})
\]

Note:

\[
\text{cov}(r_{t,\Delta} r_{t-i,\Delta}; r_{t,\Delta}) = E[r_{t,\Delta} r_{t-i,\Delta} r_{t,\Delta}] - E[r_{t,\Delta} r_{t-i,\Delta}] E[r_{t,\Delta}] = E[r_{t+i,\Delta}^2] - E[r_{t+i,\Delta}] E[r_{t,\Delta}] = \text{cov}(r_{t+i,\Delta}; r_{t,\Delta})
\]
B Modelling of Intraday Periodic Volatility Component

Consider the following decomposition for the intraday returns (Andersen and Bollerslev (1997)):

\[ R_{t,n} = E(R_{t,n}) + \sigma_t s_{t,n} z_{t,n} \sqrt{N} \]  

(9)

where \( R_{t,n} \) is intraday return, \( \sigma_t \) is conditional volatility factor for day \( t \), \( z_{t,n} \) is standard normal random variable, \( N \) is the number of return intervals per day (77 in my case), and \( s_{t,n} \) is an intraday periodic component. I will assume that \( s_{t,n} = s_n \), i.e. the periodic component is deterministic.

From equation (9), define:

\[ x_{t,n} \equiv 2 \log \left| R_{t,n} - E(R_{t,n}) \right| - \log \sigma_t^2 + \log N = \log s_{t,n}^2 + \log z_{t,n}^2 \]  

(10)

Now the problem is reduced to the following non-linear regression: \( x_{t,n} = f(\Theta; \sigma_t, n) + u_{t,n} \), where the error is \( u_{t,n} = \log z_{t,n}^2 - E(\log z_{t,n}^2) \). I assume that the function \( f(\Theta; \sigma_t, n) = f(\Theta, n) \), i.e. the intraday periodic component does not depend on the daily volatility. This assumption is consistent with the result in Andersen and Bollerslev (1997). Now I have a simple bivariate nonparametric regression which can be estimated using all available nonparametric techniques.

First, the \( x_{t,n} \) series is constructed from the intraday returns using formula (10). \( E(R_{t,n}) \) is replaced with the unconditional mean of the returns, \( \sigma_t \) with a daily volatility estimate (I used sum of squared five-minute returns). After the conditional expectation is nonparametrically estimated, one should convert the estimated \( \hat{f}_{t,n} \) to the intraday periodic component:

\[ \hat{s}_{t,n} = \frac{T \cdot \exp(\hat{f}_{t,n}/2)}{\sum_{t=1}^{T/N} \sum_{n=1}^{N} \exp(\hat{f}_{t,n}/2)} = \frac{N \cdot \exp(\hat{f}_n/2)}{\sum_{n=1}^{N} \exp(\hat{f}_n/2)} \]

where normalization \( T^{-1} \sum_{n=1}^{N} \sum_{t=1}^{[T/N]} s_{t,n} = 1 \) is assumed.

I used three different types of nonparametric estimators: the simple regressogram estimator, the local linear estimator with Gaussian kernel as a weighting function, and the cubic B-spline estimator. The final choice of the four nonparametric estimators (the regressogram, local linear, cubic B-spline and FFF estimators) is not random. The FFF and B-spline estimators are series estimators. Hence, they belong to a different from local smoothers class of nonparametric estimators. The aim of this exercise is to compare the estimators not only within the local smoothers class, but also between classes: local smoothers vs series estimators.

Since the total number of 5 minute returns is more than 130 thousand, I decided to work with a restricted series of 10,000 observations while computing cross-validation.
Andersen and Bollerslev (1997) used flexible Fourier functional form proposed by Gallant (1981). The function $f(\Theta; \sigma_t, n)$ is modelled as a parametric expression:

$$f(\Theta; \sigma_t, n) = \sum_{j=0}^{J} \sigma_j^t \left[ \mu_{0j} + \mu_{1j} \frac{n}{N_1} + \mu_{2j} \frac{n^2}{N_2} + \sum_{i=1}^{D} \lambda_{ij} I_{n=d_i} + \sum_{p=1}^{P} (\gamma_{pj} \cos \frac{pm2\pi}{N} + \delta_{pj} \sin \frac{pm2\pi}{N}) \right]$$

where $N_1 \equiv N^{-1} \sum_{i=1,N} i = (N+1)/2$ and $N_2 \equiv N^{-1} \sum_{i=1,N} i^2 = (N+1)(2N+1)/6$ are normalizing constants. I also estimated FFF form assuming $J = 0, D = 0$ for simplicity, i.e. there is no interaction between the daily volatility level and the seasonal volatility pattern and there is no time specific dummies. The question of choosing $P$ (number of cos-sin terms in the Fourier expansion) is in some sense similar to the problem of choosing a bandwidth and degree of polynomial parameters. I plaid with $P = 1, 2, 3$ and concluded that $P = 3$ provides a reasonable fit to the data. The resulting estimates with the standard errors in parenthesis (non heteroscedasticity robust) are:

$$\hat{f}_n = \begin{bmatrix}
-0.424 & -2.391 \frac{n}{N_1} & +1.140 \frac{n^2}{N_2} \\
(0.188) & (0.550) & (0.364) \\
+0.328 \cos \frac{2n\pi}{N} & -0.331 \sin \frac{2n\pi}{N} & +0.042 \cos \frac{2n\pi}{N} \\
(0.109) & (0.020) & (0.029) \\
-0.115 \sin \frac{2n\pi}{N} & +0.014 \cos \frac{3n\pi}{N} & -0.072 \sin \frac{3n\pi}{N} \\
(0.013) & (0.015) & (0.011)
\end{bmatrix}$$

Figure 11 shows the fit of all four nonparametric estimators of the volatility intraday pattern into the data: the last graph is the average of absolute values of intraday returns (one can also consider the average squared returns, but the absolute value is more robust to outliers).

All four graphs are very close to each other inside the interval. However, there are obviously large differences at the ends of the interval. The reason is simple: number of points available for estimation is less close to the boundaries than inside the interval. Hence, the regressogram and the local linear estimators are different from the B-spline and FFF estimators close to the boundaries of the support. Smoothers, in general, do not perform well near the boundaries.

The knots for the spline estimator were set uniformly. One would expect this estimator to perform not very good, since, in general, it is necessary to set the knots in a nonuniform way according to the observed properties of the data. However, the B-spline estimator performs really good. The intraday pattern is a smooth and there is no need to concentrate knots artificially in one place and remove them from another.

The cubic B-spline and FFF-estimator are smoother than the regressogram and local linear, because they are the series estimators as discussed above and they are constructed using smooth functions. The cubic B-splines estimator performs well although it is of relatively low order polynomial. The shape of the intraday pattern is a "nice" function (no non-differentiability, no infinite slopes), and a relatively low order polynomial can well capture the data properties. The local linear
estimator was constructed using the Gaussian kernel as the weighting function. This provides some additional smoothness for the linear estimator, since the Gaussian kernel is defined on the whole support, and thus, all observations are taken into account.

One potential problem with the series estimators is that the polynomial/spline estimator can look very different from the Fourier estimator. The same is not a problem with the kernels, for instance, since the choice of the kernel has little impact on the shape of the function. The B-spline and FFF estimators are very close to each other in my case because of the nice and smooth regression function.
Figure 11: Four nonparametric estimators of the volatility intraday pattern, Market Index 1993-99: regressogram, local linear, B-spline, FFF and average absolute return.