Forecasting Covariances using High-Frequency Data and Positive Semi-Definite Matrix Multiplicative Error Models

Kevin Sheppard*
Department of Economics
University of Oxford

First Version: May 2007
This version: August 5, 2008

Abstract

Positive semi-definite matrices arise in a number of interesting contexts, most notably in covariance modeling and forecasting but also in conditional duration and hazard models. Most existing covariance models are constructed as Multivariate GARCH models which use the distribution of returns to estimate any unknown parameters in the conditional covariance. This paper introduced the positive semi-definite matrix multiplicative error model as a generalization of MV-GARCH models to allow the modeling of positive semi-definite matrices which do not have a representation using returns. The paper defines the models and examines some choices for likelihood-based estimation. The paper concludes with an application to modeling the conditional covariance of ETFs spanning the Dow Jones 30, the NASDAQ 100 and mid-cap Spiders. The application is extended to show how novel models can be produced in the PSD-MEM framework by decomposing the covariance into an over-night component, which is only measured from the close to the following open and an intra-day covariance which is measured using realized covariance.

JEL Classification Codes: C32, C51, C52, C53

Keywords: Covariance Forecasting, Epps Effect, Multiplicative Error Models, Multivariate GARCH, Realized Volatility, Wishart

*Mailing address: Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford, OX1 3UQ, UK, email: kevin.sheppard@economics.ox.ac.uk.

1 Introduction

Following the introduction of realized variance and covariance to econometrics by Andersen & Bollerslev (1998) and Andersen, Bollerslev, Diebold & Labys (2003), and Barndorff-Nielsen & Shephard (2004), the measurement and modeling of conditional variance and correlation have been fundamentally altered. The standard practice of using variance estimates based on daily returns using ARCH models (Engle 1982) has changed to accommodate the simple, model-free estimates of realized measures. This paper provides a model-based approach to forecasting positive semi-definite covariance matrices that can accommodate both precise high-frequency covariance measures as well as lower frequency proxies.

Recent research in this field has focused primarily on providing improved accuracy in the measurement of conditional variance in the presence of market microstructure noise. Initial efforts were concerned with improving the estimation of realized variance and include Zhang, Mykland & Aït-Sahalia (2004) and and Barndorff-Nielsen, Hansen, Lunde & Shephard (2006). The estimators proposed in these papers have clearly borne fruit and estimation of the quadratic variation over some period of time, even in the presence of MMN, can be considered a "solved" problem. Recent efforts have been made to extend these results to the measurement of covariance although significant hurdles, primarily driven by non-synchronous trading and possibly market segmentation remain (see (Lundin, Dacorogna & Muller 1999), (Zebedee 2001), Martens (2004), (Bandi & Russell 2005), Griffin & Oomen (2006), Sheppard (2006), Zhang (2006), Voev & Lunde (2007), and Corsi & Audrino (2007)). In light of the widespread empirical documentation of the *Epps effect*, it is difficult to treat the estimates of conditional covariance constructed using tick data as consistent. Despite the issues of measuring the covariance using tick data, there appears to be value even in relatively imprecise estimators (Fleming, Kirby & Ostdiek 2003).

The increased precision of realized volatility has allowed for improved precision for forecasts of future volatility and a better understanding of volatility dynamics. Forecast models for realized variance are usually constructed by taking a log-transform of the realized variance. This practice has been used in Andersen, Bollerslev, Diebold & Ebens (2001), Corsi (2004), and Lanne (2006*b*), among others, allowing models to be built using standard tools of time-series analysis, such as ARFIMA, HAR and VAR models. The forecasting performance of these models seems to be good and the residuals are approximately Gaussian. An alternative modeling approach that does not treat volatility as observed has been developed by Engle (2002*b*). These multiplicative error models parameterize a non-negative distribution that interacts multiplicatively with the conditional mean of a non-negative process. This class of models builds on the framework of ARCH and ACD (Engle & Russell 1998). Multiplicative error models have been further studied and extended in Engle & Gallo (2003), Lanne (2006*a*) and Cipollini, Engle & Gallo (2006), among others.

Simultaneously new parameterizations and increased data availability in the cross-section have led to a renaissance in the multivariate GARCH literature. The DCC model (Engle 2002*a*), the VC model (Tse & Tsui 2002), RSDCC (Pelletier 2003), the FLEXM Ledoit, Santa-Clara & Wolf (2003) and various Midas implementations (Ghysels, Santa-Clara & Valkanov 2007) have all produced tractable models for capturing interesting features of conditional covariance dynamics.[1] These models all parameterize a distribution for the conditional distribution of returns, usually normal, although occasionally another elliptical distribu-

[1] For a recent survey of multivariate GARCH models, see Laurent, Bauwens & Rombouts (2006).

tion, to facilitate parameter estimation. As a result, a disconnect exists between the MV-GARCH literature and the measurement of covariance using tick data. The few attempts at formally modeling realized covariance have typically treated both realized variance and realized correlation as observable and have modeled them using standard time-series tools such as ARFIMAs and HARs (see Andersen et al. (2001), Chiriac & Voev (2007), Corsi & Audrino (2007) and Voev (2007)), although there is a strand of the literature that have developed Wishart stochastic volatility models and processes, for example Philipov & Glickman (2004) and the Wishart Autoregressive process of Gourieroux (2006) and Gourieroux, Jasiak & Sufana (2005). The former class of models for realized covariance tends to ignore the structure of a covariance matrix while the latter class of models is very specific about the dynamics of conditional covariance and may be difficult to implement in practice (Chiriac 2006).

This paper bridge the gap between existing MV-GARCH models and higher-precision measures of conditional covariance by introducing the Positive Semi-Definite Multiplicative Error Model (PSD-MEM). The PSD-MEM builds on the existing literature in multiplicative error models and provides a unified framework for estimating time-series models for positive semi-definite processes.[2] With the increasing availability of tick data, there will be many opportunities to model conditional covariance, correlations and betas in order to better understand shock propagation mechanisms, persistence and forecasting. The PSD-MEM framework allows researchers to use off-the shelf models for positive semi-definite processes, most notable the wide range of MV-GARCH models, while retaining a key structure of the problem, the positivity of any covariance matrices.

The paper is organized as follows. Section 2 provides a motivation for PSD-MEMs based in realized covariance. Section 3 defines the model and one distribution suitable for estimating any unknown parameters. Section 4 examines alternative distributional assumptions. Section 5 examines three applications for PSD-MEMs that utilize both daily covariance proxies – ones which are only positive semi-definite – and realized covariances to study the effect of increased precision in the proxies, spill-over between market open hours and closed periods, and to examine whether the information contained in realized covariance is sufficient to crowd out the information in the outer-product of daily returns in a MV-GARCH-like application.

2 Motivation for Matrix MEM

Suppose that a k-vector of log prices, p_t, evolves according to a standard k-dimensional Brownian motion $\mathbf{B}_t$ with stochastic covariance,

$$d\mathbf{p}_t = \mathbf{\Omega}_t d\mathbf{B}_t \tag{1}$$

where $\Sigma_t = \Omega_t \Omega_t'$ is the instantaneous covariance. Define $RC^{(m)}$ to be a m-sample realized covariance computed from observed prices,

$$RV^{(m)} = \sum_{i=1}^{m} \left(\mathbf{p}_i - \mathbf{p}_{i-1}\right)\left(\mathbf{p}_i - \mathbf{p}_{i-1}\right)'. \tag{2}$$

[2] Without loss of generality the results of this paper can be directly applied to negative definite and negative semi-definite matrices by multiplying by -1.

Further assuming that the covariance is piece-wise constant, so that $\mathbf{\Omega}_t = \bar{\mathbf{\Omega}}_i$ for $t \in (i, i-1)$, the realized covariance estimator can be re-formulated as

$$RV^{(m)} = m^{-1} \sum_{i=1}^{m} \bar{\mathbf{\Omega}}_i \mathbf{W}_i \bar{\mathbf{\Omega}}_i' \tag{3}$$

where $W_i \overset{\text{IID}}{\sim} \text{Wishart}_1(\mathbf{I}_k)$ are independent Wishart distributions with 1 degree of freedom, which follows from the normality of the Wiener process. In this simplification realized covariance follows a sum of Wishart distribution (see Tan & Gupta (1983) and Nel & van der Merwe (1986)). As the number of samples increases this distribution will approach that of in Barndorff-Nielsen & Shephard (2004), although in finite samples the BNS distribution may be very poor, particular when applied to functions of the realized covariance matrix. Unfortunately the sum of Wishart distribution is essentially untractable and is generally proxied by either a Wishart or a symmetric matrix normal. Despite this limitation, consider the distribution of the standardized realized covariance,

$$\left(m^{-1} \sum_{i=1}^{m} \bar{\mathbf{\Omega}}_i \bar{\mathbf{\Omega}}_i'\right)^{-\frac{1}{2}} \left(m^{-1} \sum_{i=1}^{m} \bar{\mathbf{\Omega}}_i \mathbf{W}_i \bar{\mathbf{\Omega}}_i'\right) \left(m^{-1} \sum_{i=1}^{m} \bar{\mathbf{\Omega}}_i \bar{\mathbf{\Omega}}_i'\right)^{-\frac{1}{2}}. \tag{4}$$

This random variable will have expectation equal to $\mathbf{I}_k$ and it seems natural to model the sum of the covariance and attempting to find a model which makes the standardized residuals as close to an identity matrix as possible.

3 Positive Semi-Definite Matrix Multiplicative Error Model

Let $\{\mathbf{X}_t\}_{t=1}^{T}$ be a time-series of random matrices with support on the set of positive semi-definite matrices, $\mathcal{M}^{+}$, or the set of positive definite matrices, $\mathcal{M}^{++}$. A positive semi-definite matrix multiplicative error model for a random, semi-definite is defined as

$$\mathbf{X}_t = \mathbf{H}_t^{-\frac{1}{2}} \mathbf{\Xi}_t \mathbf{H}_t^{-\frac{1}{2}} \tag{5}$$

where $\mathbf{\Xi} \overset{\text{IID}}{\sim} g(\mathbf{I}_k, \mathbf{\Upsilon})$ are independent random matrices with support on the set of positive-semi-definite or positive-definite matrices and $\mathbf{H}_t$ is the $\mathcal{F}_{t-1}$ measurable conditional expectation of $\mathbf{X}_t$. $\mathbf{\Upsilon}$ is a set parameters that govern higher-order moments of $\mathbf{\Xi}$. It is further assumed that $\mathbf{H}_t \in \mathcal{M}^{++}$ and $\mathbf{\Xi}_t \in \mathcal{M}^{+}$ or $\mathbf{\Xi}_t \in \mathcal{M}^{++}$ as appropriate for the data examined.

The actual model for $\mathbf{H}_t$ is left intentionally unspecified. The exact specification of $\mathbf{H}_t$ is problem dependent and may take many forms as long as $\mathbf{H}_t \in \mathcal{M}^{++}$ a.s. In the empirical application of this paper the base formulation of $\mathbf{H}_t$ follows a standard diagonal *vech*,

$$\mathbf{H}_t = \mathbf{C} + \mathbf{A} \odot \mathbf{X}_{t-1} + \mathbf{B} \odot \mathbf{H}_{t-1} \tag{6}$$

where $\mathbf{C}$, $\mathbf{A}$ and $\mathbf{B}$ are positive definite $k \times k$ matrices and so each element of $\mathbf{H}_t$ will have a GARCH(1,1)-like evolution (Bollerslev, Engle & Wooldridge 1988). However, this particular specification was been chosen for ease of estimation and directness of interpretation, and a wide variety of dynamics may be specified for the evolution of $\mathbf{H}_t$, including the evolution of most M-GARCH models. For example, symmetric DCC-like dynamics (Engle 2002*a*), BEKK-like dynamics (Engle & Kroner 1995) or ADC-like dynamics (Kroner & Ng 1998) could replace those in equation (23). By decomposing the $\mathbf{H}_t$ using either an eigenvalue decomposition, $\mathbf{H}_t = \mathbf{Q}_t \boldsymbol{\Lambda}_t \mathbf{Q}_t$ or a standard deviation-correlation decomposition, $\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$, factor and dynamic correlation models can be easily accommodated in this framework.

3.1 The Distributions of Ξ

Once a class of models to describe the dynamics has been chosen, the final detail to implementing a PSD-MEM is the specification of the density of $\boldsymbol{\Xi}$. The natural choice is to specify the distribution as a standardized Wishart with a mean of $\nu^{-1}\mathbf{I}_k$ and a degree of freedom of ν, so that $\boldsymbol{\Xi}_t \sim W_\nu\left(\nu^{-1}\mathbf{I}_k\right)$ (see Gupta & Nagar (2000) for a comprehensive treatment of the Wishart). As discussed in section 2, the Wishart arises naturally from the outer-product of normals. The p.d.f. of a Wishart with mean $\mathbf{H}$ and degree of freedom ν is

$$W_\nu\left(\mathbf{X};\mathbf{H}\right) = \frac{\nu^{\frac{1}{2}\nu}\left|\mathbf{X}\right|^{\frac{1}{2}\nu-k-1}}{2^{\frac{1}{2}k\nu}\Gamma_k\left(\frac{\nu}{2}\right)\left|\mathbf{H}\right|^{\frac{1}{2}\nu}} \exp\left(-\frac{1}{2\nu}\operatorname{tr}\left(\mathbf{H}^{-1}\mathbf{X}\right)\right) \quad , \mathbf{X} \text{ non-singular} \tag{7}$$

where $\Gamma_k \nu/2 = \pi^{k(k-1)/4}\prod_{j=1}^{k}\Gamma\left((\nu-1-j)/2\right)$ is the multivariate gamma function. One imposition of the Wishart is that the degree of freedom parameter is greater than the dimension of the problem. This imposition arises from the requirement in equation (7) that $\hat{\Sigma}$ is positive definite. This will not always be the case when modeling actual data, for example when using the outer-products of returns in an M-GARCH-like application, and the likelihood will be 0 for any parameter values. The solution to this problem is to use the Pseudo-Wishart distribution which covers the case where the data are less than full rank. The p.d.f. of the Pseudo-Wishart is

$$PW_\nu\left(\mathbf{X};\mathbf{H}\right) = \frac{\nu^{\frac{1}{2}\nu}\left|\boldsymbol{\Lambda}\right|^{\frac{1}{2}\nu-k-1}}{2^{\frac{1}{2}k\nu}\Gamma_\nu\left(\frac{\nu}{2}\right)\left|\mathbf{H}\right|^{\frac{1}{2}\nu}} \exp\left(-\frac{1}{2\nu}\operatorname{tr}\left(\mathbf{H}^{-1}\mathbf{X}\right)\right) \quad , \mathbf{X} \text{ singular} \tag{8}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with the non-zero eigenvalues of $\mathbf{X}$ along its diagonal (Uhlig 1994).[3]

Despite the inconvenience of having likelihoods that depend on the rank of the data, the log-likelihoods share a common property. After a minor amount of manipulation, the log-likelihoods of both the Wishart and the Pseudo-Wishart can be expressed

$$LL(\mathbf{X};\mathbf{H},\nu) = c - \frac{1}{2}\left(\nu\left(\ln\left|\mathbf{H}\right| + \operatorname{tr}\left(\mathbf{H}^{-1}\mathbf{X}\right)\right)\right) \tag{9}$$

where c is a constant with respect to $\mathbf{H}$ – that is, c depends only on ν, $\mathbf{X}$, and, in the case of a singular data matrix, $\boldsymbol{\Lambda}$. This expression of the log-likelihood makes it obvious that the value of ν does not matter when estimating any unknown parameters in $\mathbf{H}$ since a solution to the score of the log-likelihood with respect to $\mathbf{H}$ (or $\mathbf{H}_t$) will be identical for any value of ν. This is a similar finding as was documented in Engle & Gallo

[3] Both the Wishart and Pseudo Wishart assume that the rank of $\mathbf{X}$ is known *a priori*.

(2003) in light of estimating a scalar multiplicative error model with a gamma error distribution. Additionally, the PSD-MEM when specified with a Wishart reduces to the model of Engle & Gallo when applied to a univariate process.

As a consequence of the form of the log-likelihood, consistent parameter estimates for the model parameters governing $\mathbf{H}_t$ can be computed by maximizing the quasi-log-likelihood,

$$QLL(\mathbf{X};\mathbf{H}) = \sum_{t=1}^{T} -\frac{1}{2}\left(\ln|\mathbf{H}_t| + \text{tr}\left(\mathbf{H}^{-1}\mathbf{X}_t\right)\right) \tag{10}$$

and the Hessian will be block diagonal between the parameters of $\mathbf{H}_t$ and the degree of freedom. The Wishart does impose a substantial amount of structure on the problem. If $\Xi \sim W_\nu\left(\nu^{-1}\mathbf{I}_k\right)$ then $\text{E}[\Xi] = \mathbf{I}_k$ and the covariance is given by

$$\text{V}\left[\text{vec}\left(\Xi\right)\right] = \nu^{-1}\left(\mathbf{I}_{k^2} + \mathbf{P}_{kk}\right) \tag{11}$$

where $\mathbf{P}_{kk}$ is the commutation matrix of order kk, defined as

$$\mathbf{P}_{kk} = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(\mathbf{C}_{ij} \otimes \mathbf{C}'_{ij}\right) \tag{12}$$

where $\mathbf{C}_{ij}$ is a $k \times k$ matrix of 0s with 1 in the i,j$^{\text{th}}$ position. In the simplest bivariate model,

$$\text{V}[[\text{vec}\left(\Xi\right)] = \nu^{-1}\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \tag{13}$$

which is obviously singular which follows from the repetition of ξ_{12} in the covariance of the *vec*. When scaled by $\mathbf{H}$, the mean becomes $\text{E}[\mathbf{H}_t^{\frac{1}{2}}\Xi\left(\mathbf{H}_t^{\frac{1}{2}}\right)'] = \mathbf{H}$ and the covariance is

$$\text{V}\left[\text{vec}\left(\mathbf{H}_t^{\frac{1}{2}}\Xi\left(\mathbf{H}_t^{\frac{1}{2}}\right)'\right)\right] = \nu^{-1}\left(\mathbf{I}_{k^2} + \mathbf{P}_{kk}\right)\left(\mathbf{H}_t \otimes \mathbf{H}_t\right) \tag{14}$$

The consequence of the choice of a Wishart is that all of the dependence of the shocks is contained in the mean. This may not be a desirable property. The next section explores alternative distributional assumptions for Ξ.

4 Alternative Distributions

While the standard Wishart leads to a simple, easy to interpret estimator, it is not obvious that such a simple distribution is ideal. The strong link between the covariance and the conditional expectation of the Wishart is particularly troubling. There may be cases, particularly if efficiency of parameter estimators is an overriding concern where a more flexible approach is warranted. This section examines some alternatives for the distribution of the unitary covariance shocks in a PSD-MEM.

4.1 Sums of Scaled Wisharts

As discussed in section 2, a natural candidate distribution for modeling realized covariance is a sum of Wishart distribution. This distribution has been extensively studied in the statistics literature as it is a crucial input for deriving the exact distribution when testing equality of multivariate means using a Behrens-Fisher statistic. Unfortunately deriving the distribution of sums of Wisharts when the number of Wisharts is greater than two is impractical, and even in the case of two, estimation is difficult as the likelihood involves a hypergeometric function with matrix arguments (Gupta & Nagar 2000). The common solution in the statistics literature to this problem is to approximate a sum of Wisharts with a single Wishart by matching moments. This typically involves setting the mean of the Wishart to the sum of the means of the component Wishart and then choosing the degree of freedom to match the variance as close as possible (see (Tan & Gupta 1983) and Nel & van der Merwe (1986), *inter alia*). This solution is based on a central limit argument where a sum of Wisharts converges to a symmetric matrix normal; that is a symmetric matrix whose lower triangular elements are a $k(k+1)/2$ multivariate normal. This corresponds to the result in Barndorff-Nielsen & Shephard (2004) in the case of a pure diffusion, and is of little value when considering alternative distributions for modeling positive semi-definite processes.

Noting that a Wishart with an integer degree of freedom can be derived as quadratic form of normal random variables with a covariance matrix of $\mathbf{\Sigma}$, a number of authors have studies a generalization of the form $\mathbf{X}'\mathbf{X}$ where $\mathbf{X}_{k,n} \sim N(\mathbf{\Sigma} \otimes \mathbf{\Psi})$ is a k by n matrix of normals. Each column of $\mathbf{X}$ has a covariance proportional to $\mathbf{\Sigma}$ and $\mathbf{\Psi}$ controls the covariance between columns. The distribution of quadratic forms of this type has been explicitly derived in de Waal (1979) (An up-to-date overview can be found in Gupta & Nagar (2000)). However, in the case of modeling realized covariance no-arbitrage would restrict $\mathbf{\Psi}$ to be diagonal. An extended distribution for the innovations to the PSD-MEM could be parameterized by specifying $\mathbf{\Psi}$ to be diagonal with positive elements that sum to one. This is a fairly minor improvement over the standard Wishart and is not pursued further here.

4.2 Matrix Log-normal

An alternative to using random matrices derived by quadratic forms would be to use a matrix valued log-normal. Log-normals have often been found to provide a good description of the distribution of realized variance (Andersen et al. 2001). While raw realized variances have long-tailed errors, log transforms produce errors that are approximately normal. This had led many authors to model realized variance as an observable process, in logs, as an ARMA or ARFIMA with Gaussian errors (Corsi 2004). This transformation can be extended into conditional covariance space using a matrix logarithm. For a real, symmetric positive-definite matrix $\mathbf{X}$, the matrix logarithm is defined as

$$\text{logm}(\mathbf{X}) = \mathbf{Q} \ln(\mathbf{\Lambda}) \mathbf{Q}' \tag{15}$$

where $\mathbf{X} \equiv \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$, $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_k$ are orthonormal matrices and $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues of $\mathbf{X}$

along the diagonal. $\ln \mathbf{\Lambda}$ is a diagonal matrix with the natural log of the eigenvalues along its diagonal. The inverse operation, matrix exponentiation, is similarly defined for symmetric matrices,

$$\text{expm}\,(\mathbf{X}) = \mathbf{Q}\exp\left(\mathbf{\Lambda}\right)\mathbf{Q}'. \tag{16}$$

Together these two operators can be used to construct a bijection between the set of symmetric matrices, $\mathcal{S}$ and the set of positive definite matrices, $\mathcal{M}^{++}$. A matrix valued log-normal is constructed by taking the matrix exponential of a symmetric matrix normal distribution, which is itself constructed from a standard multivariate normal distribution. Let $\mathbf{X}$ be a $k(k+1)/2$ variable standard multivariate normal with mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Omega}$. A symmetric matrix normal is constructed as $\mathbf{Y} = \text{vech}^{-1}(\mathbf{X})$. The mean and covariance can be directly derived using the vech^{-1} operator and the commutation matrix. A matrix variate log-normal is constructed by taking the matrix exponential of $\mathbf{Y}$, $\mathbf{Z} = \text{expm}\,(\mathbf{Y})$.

A shock for a PSD-MEM can be constructed by defining $\mathbf{\Xi} \overset{\text{IID}}{\sim} MLN(\mathbf{I}_k, \mathbf{\Upsilon})$ where $\mathbf{I}_k$ is the mean of $\mathbf{\Xi}$ and $\mathbf{\Upsilon}$ are shape parameters inherited from the underlying normal. Using this type of error, computing the likelihood is easy since

$$\mathbf{X} = \mathbf{H}^{\frac{1}{2}}\mathbf{\Xi}\left(\mathbf{H}^{\frac{1}{2}}\right)' \tag{17}$$

and so

$$\mathbf{e} \equiv \text{vech}\left[\text{logm}\left(\mathbf{H}^{-\frac{1}{2}}\mathbf{X}\left(\mathbf{H}^{-\frac{1}{2}}\right)'\right)\right] \overset{\text{IID}}{\sim} N(\boldsymbol{\mu}, \mathbf{\Omega}). \tag{18}$$

where $\mathbf{X}_t$ is replaced by one of the covariance measures augmented to include the close-to-open return. For example, in the model using 5-minute realized covariance, the dependent variable is $RC_t^{(72)} + RC_t^{(\text{CtO})}$. In order to facilitate identification, $\boldsymbol{\mu}$ was restricted to be $\mathbf{0}$. Matrix exponentials have appeared in the multivariate GARCH literature via the Multivariate Exponential GARCH model of Kawakatsu (2006). This model, however, parameterizes the matrix logarithm of the covariance matrix directly in the spirit of Nelson (1991). The literature on multivariate covariance modeling is otherwise silent on matrix logarithms and exponents to our knowledge.

While the matrix log-normal has some strong points in its favor, namely that log realized covariance is normal, this choice is not without drawbacks. First, unlike the Wishart distribution which is viable using either positive-definite or positive semi-definite data, a log-normal distribution can only be applied to positive definite data. Second, the exact conditions for $\mathbf{\Xi}$ to have expectation of an identity matrix are difficult in a general setting. Finally, estimation requires the matrix logarithm to be computed for every data point for ever iteration of an optimizer. This is a computationally expensive operation and it makes modeling moderately large dimension matrices difficult.

5 Empirical Application

Three applications, all involving realized covariance (or other realized covariance-like structures), are provided to illustrate the PSD-MEM. The first is a standard application to modeling the time-series dynamics of realized covariance using a diagonal *vech* inspired model. The model is fit, and comparisons are made between, the Wishart and the Log-normal densities. The second decomposes the open-to-close and the close-to-open covariation to examine whether there is value in explicitly modeling each of the components, and whether the information contained in one is nested by the other. The third application examines the fit of a standard MV-GARCH as estimated as a PSD-MEM against a model based on realized covariance, again testing whether there is any information contained in daily returns needed for modeling the covariance of daily returns when confronted with realized covariance.

All applications make use of a time-series of realized covariances constructed using Diamonds (DIA), a DJIA tracking exchange traded fund, Midcap Spiders, MDY, a mid-cap tracking ETF, and PowerShares QQQ, a NASDAQ 100 tracking ETF. Data was available from January 2, 2001 until December 31, 2006, a total of 1507 days. Prices were constructed from mid-quotes and data was pre-filtered by the following rules:

- Quotes originating from the NYSE, NASDAQ and AMEX
- Valid quotes, that is quotes where the ask greater than the bid, positive bid and ask depth, and no 0 values in these 4 fields
- Quotes with a spread less than the 99$^{\text{th}}$ percentile on the same day
- Quotes with the bid less than the daily high and the ask greater than the daily low where the high and low were taken from CRSP

The opening quote was assumed to occur at 9:45 AM and the closing quote was assumed to occur at 15:45. While the actual opening and closing were always before 9:45 and after 15:45, respectively, this choice was made to mitigate any opening delays or early closings. The opening price on day t, denoted $\mathbf{p}_t^{\mathrm{O}}$, is the last prior to 9:45:00 and the closing price on day t is the last price available before 15:45:00 and is denoted $\mathbf{p}_t^{\mathrm{C}}$.

5.1 Constructing Realized Covariance

Prices were sampled every minute between 9:45 and 15:45. Realized covariance was constructed using 1, 5, 10, 15 and 30 minute windows. While mid-quotes are at worst moderately affected by bid-ask bounce-like microstructure noise, covariance sampled from tick data suffer from the widely documented Epps effect (Epps 1979), and tend to exhibit substantial bias when sampled very frequently. While recent research has made some progress on sampling tick data to compute realized covariance – see Hayashi & Yoshida (2005), Griffin & Oomen (2006), Zhang (2006), Bandi & Russell (2005) and Sheppard (2006) – there is little work on how to optimally sample more than 2 series. As a result, this paper uses simple calendar time-sampling with last price interpolation.

The sub-sample realized covariance is computed as

$$RC_t^{(m)} = \kappa \sum_{i=1}^{300/m} \sum_{j=0}^{m-1} \left(\mathbf{p}_{im+j,t} - \mathbf{p}_{(i-1)m+j,t}\right)\left(\mathbf{p}_{im+j,t} - \mathbf{p}_{(i-1)m+j,t}\right)' \tag{19}$$

where m is the number of samples used to compute the standard version of realized covariance and $\kappa = m(301 - m)/300$ is a scaling factor to ensure that the correct number of samples are used, on average. Realized covariance was computed using 1, 5, 10 15 and 30 minutes sampling schemes. The sub-sampling used in this application is a naïve version that is only intended to reduce the importance of choice of the initial observations and to provide a mild reduction in variance for infrequently sampled estimators. This estimator, at least for modeling and forecasting, has been found to have good properties in Andersen, Bollerslev & Meddahi (2006). The results of the paper are robust to using the usual realized covariance estimator sampled at the same frequency. In addition, two other "realized covariances" are used, although these are both computed using the outer-product of a single return. The outer-product of the close-to-close return is denoted

$$RC_t^{\text{CtC}} = \left(\mathbf{p}_t^{\text{C}} - \mathbf{p}_{t-1}^{\text{C}}\right)\left(\mathbf{p}_t^{\text{C}} - \mathbf{p}_{t-1}^{\text{C}}\right)' \tag{20}$$

and the outer-product of the close-to-open return is denoted

$$RC_t^{\text{CtO}} = \left(\mathbf{p}_t^{\text{O}} - \mathbf{p}_{t-1}^{\text{C}}\right)\left(\mathbf{p}_t^{\text{O}} - \mathbf{p}_{t-1}^{\text{C}}\right)'. \tag{21}$$

The volatility signature plot (Andersen et al. 2003) for the different sampling frequencies is presented in figure 1. The average realized variances agree broadly with the exception of 1 minute sampling where MDY exhibits a large up-swing and QQQ exhibits a mild down-swing. A pseudo-correlation signature plot, constructed by dividing the average realized covariance between two ETFs by the square root of their daily variance is presented in figure 2.[4] The correlations do exhibit substantial scaling bias, especially for either pair involving MDY. This discrepancy is difficult to explain with by a simple non-synchronization as DIA, MDY and QQQ had quote intensities such that the media time between quotes was 1.9, 3.4 and 1.9 seconds, respectively (and this is removing non-price informative quotes).[5] This non-observable non-synchronization has been described by Sheppard (2006) as "scrambling". As a result, model built using 1 and 5-minute returns should be treated with caution.

5.2 Baseline Modeling

The description of the PSD-MEM has been intentionally vague on the dynamics for the conditional expectation of the time-series being modeled. This was to facilitate application of the PSD-MEM framework to a wide variety of problems. However, in order to implement a PSD-MEM some stand must be taken on the dynamics of the random matrices being studied. For this application, a simple specification where each series evolves according to ARMA(1,1)-like dynamics, and where the evolution of the conditional mean of the PSD-MEM follows a diagonal *vech*-like specification, will be used. Thus,

$$\mathbf{H}_t = \mathbf{C} + \mathbf{A} \odot \mathbf{X}_{t-1} + \mathbf{B} \odot \mathbf{H}_{t-1} \tag{22}$$

[4] The average overnight covariance was included in the pseudo correlations of each high frequency measure to prevent and bias introduced by ignoring the over night return.

[5] Trading in the 3 is essentially continuous.

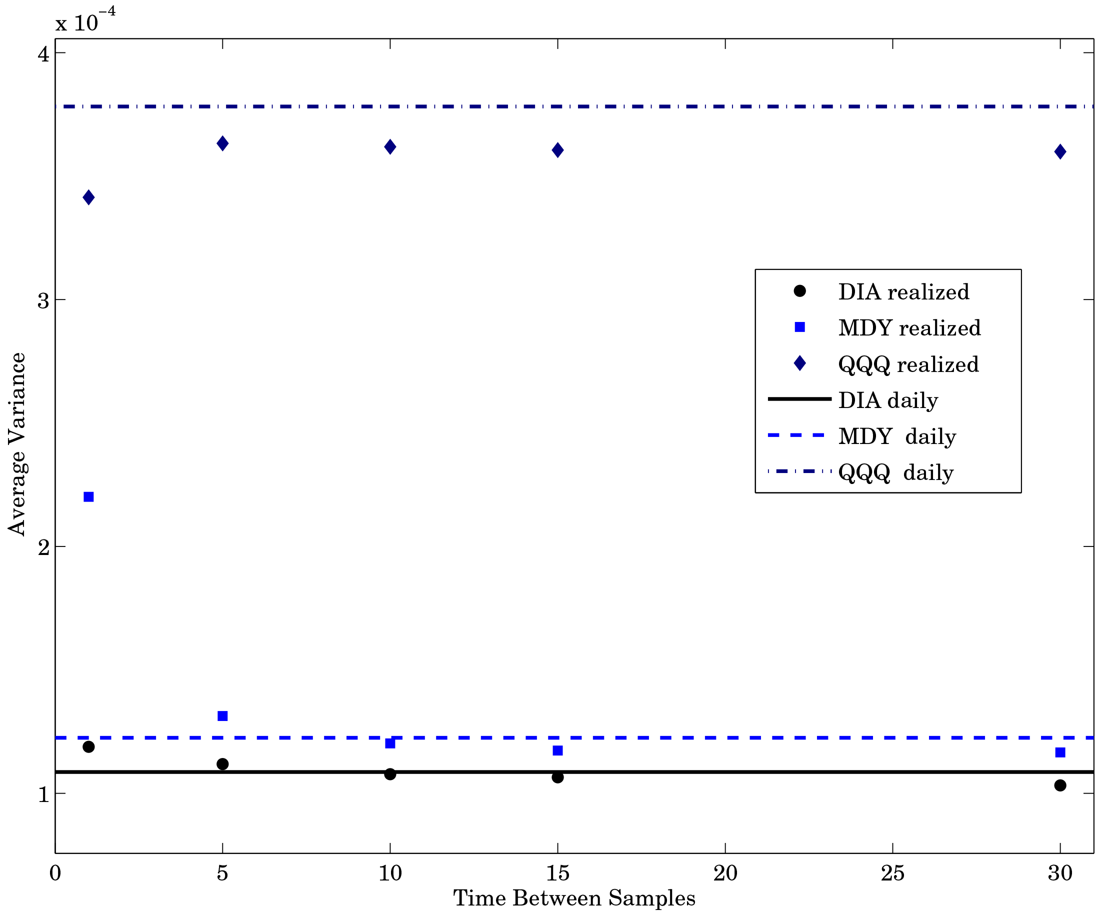

Figure 1: Volatility signature plot of the average realized variance for the three ETF used in the empirical application. With the exception of 1-minute sampling, the average variances appear to be unaffected by the sampling window. .

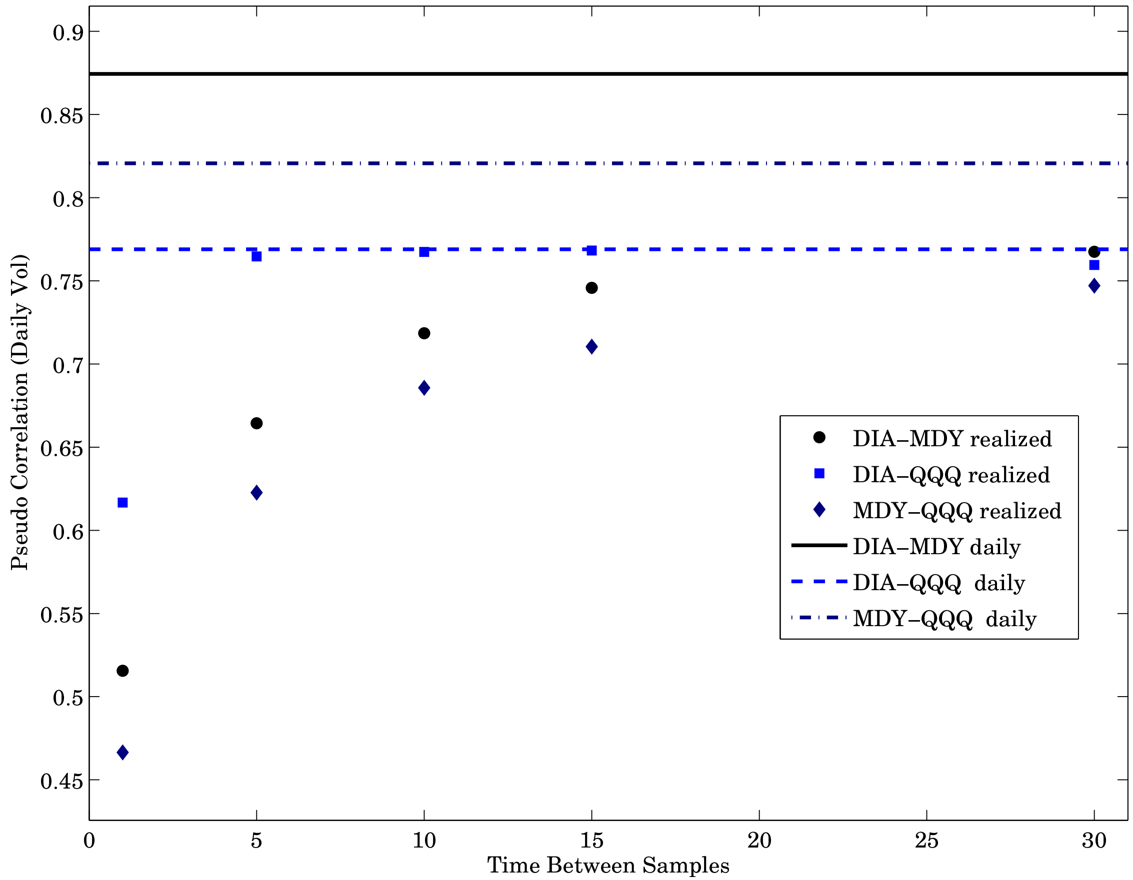

Figure 2: Pseudo correlation signature plots. Each point represents the average realized covariance between a pair of assets divided by the square root of the average daily variance of the pair. Unlike the volatility signature plot, there is a non-trivial bias in the pseudo correlations of either DIA or QQQ when matched against MDY. This may reflect sluggish adjustment of mid caps to market news.

which produces models for each element of the form

$$h_{ij,t} = c_{ij} + a_{ij} \odot x_{ij,t-1} + b_{ij} \odot h_{ij,t-1}. \tag{23}$$

The parameter matrices, **C**, **A** and **B** are restricted by re-parameterization using their Choleski decompositions to be positive semi-definite. This restriction, along with positive semi-definiteness of $\{\mathbf{X}_t\}$ and the initial condition $\mathbf{H}_0$, are sufficient to guarantee that $\mathbf{H}_t$ is positive definite (Ding & Engle 2001).

This baseline specification was fit to the 5-realized covariances and the outer-product of the close-to-close returns ($RC^{(\mathrm{CtC})}$) using both the Wishart and the Matrix Log-normal distributions.[6] Results of the models fit using the Wishart distribution are presented in table 1. A number of patterns are immediately evident in the results. First, all parameters are extremely significant with p-vals, computed using t-stats constructed using mis-specification robust standard errors, less than .01 in all cases.

The second obvious patter in the results is the effect of different frequencies on typical values in **A** and **B**. When sampled frequently, particularly for 1 and 5-minute realized covariance, the coefficients in **A** which measure the response to lagged realized covariance are on the order of 0.25 while the coefficients in the smoothing matrix are between .75 and .8. When compared to the model for the close-to-close returns, which is a standard diagonal *vech*, the numerical differences are not severe but the implications for the amount of smoothing needed to provide an adequate fit are dramatic. Using the values in **B**, the weights on the 10^{th} lags of realized covariance would be close to 0.02. For the close-to-close model the weight on the 10^{th} lag would be approximately 0.03. When contrasted with the 1^{st} lag weights of 0.20 and 0.05 for the 5-minute and close-to-close models, respectively, it is clear that a model based on daily returns requires substantially more smoothing to provide an ideal fit.

The final results from this set of models for realized covariance is the measured persistence. The average persistence is monotonically decreasing as the sampling frequency is reduced from 1-minute to 30-minutes and then the close-to-close model. The indicated persistence for the 1-minute realized covariance are very close to 1, and in the case of variance of MDY, the sum of the two coefficients is slightly greater than 1.[7] As the sampling frequency increases the persistence declines, although 30-minutes realized covariances still had coefficients that summed to 0.97. This behavior may be a consequence of the systematic bias in realized covariances computed using frequent samples, although the precise cause is beyond the scope of this study.

Table 2 presents the results for the baseline model using the matrix log-normal distribution for the 5 time-series of realized covariance based on intra-day prices. The results are qualitatively similar although some differences are noticeable. The primary difference is in the implied persistence for shocks which is uniformly lower when using the distribution. The other difference between the two is the response to recent news. The log-normal models are less responsive to recent news, exhibit somewhat lower persistence, but have larger coefficients in their smoothing term (**B**). A graphical representation of the difference in the fit variance is presented in figure 3 and for the differences in covariance is presented in figure 4. The slight difference in dynamics is noticeable with the log-normal model returning to its long-run level faster than the Wishart model. A more stark difference between the two is the difference in the level. The log-normal model

[6]The outer product of the close-to-close return cannot be modeled using the matrix log-normal as it is singular.
[7]No stationarity constraints were imposed on the model.

	DIA	MDY	QQQ	DIA	MDY	QQQ	LL
DIA	0.240 (0.00)			0.752 (0.00)			-3950.4
MDY	0.244 (0.00)	0.250 (0.00)		0.751 (0.00)	0.753 (0.00)		
QQQ	0.229 (0.00)	0.233 (0.00)	0.219 (0.00)	0.765 (0.00)	0.765 (0.00)	0.778 (0.00)	

$RC^{(72)}$

	A			**B**			LL
DIA	0.210 (0.00)			0.778 (0.00)			-2199.8
MDY	0.204 (0.00)	0.202 (0.00)		0.781 (0.00)	0.785 (0.00)		
QQQ	0.197 (0.00)	0.194 (0.00)	0.188 (0.00)	0.791 (0.00)	0.794 (0.00)	0.804 (0.00)	

$RC^{(36)}$

	A			**B**			LL
DIA	0.191 (0.00)			0.793 (0.00)			-1658.3
MDY	0.186 (0.00)	0.186 (0.00)		0.793 (0.00)	0.794 (0.00)		
QQQ	0.178 (0.00)	0.177 (0.00)	0.169 (0.00)	0.806 (0.00)	0.807 (0.00)	0.820 (0.00)	

$RC^{(24)}$

	A			**B**			LL
DIA	0.175 (0.00)			0.807 (0.00)			-1428.1
MDY	0.170 (0.00)	0.168 (0.00)		0.808 (0.00)	0.809 (0.00)		
QQQ	0.163 (0.00)	0.161 (0.00)	0.156 (0.00)	0.819 (0.00)	0.820 (0.00)	0.830 (0.00)	

$RC^{(12)}$

	A			**B**			LL
DIA	0.144 (0.00)			0.835 (0.00)			-1136.9
MDY	0.139 (0.00)	0.138 (0.00)		0.835 (0.00)	0.835 (0.00)		
QQQ	0.133 (0.00)	0.131 (0.00)	0.127 (0.00)	0.846 (0.00)	0.847 (0.00)	0.858 (0.00)	

$RC^{(CtC)}$ (Diagonal *vech* MV-GARCH(1,1))

	A			**B**			LL
DIA	0.047 (0.01)			0.938 (0.00)			-1150.6
MDY	0.046 (0.00)	0.048 (0.00)		0.935 (0.00)	0.932 (0.00)		
QQQ	0.038 (0.03)	0.040 (0.00)	0.033 (0.00)	0.947 (0.00)	0.944 (0.00)	0.956 (0.00)	

Table 1: This table contains parameter estimates from the baseline model using a Wishart distribution for the standardized shocks. The baseline model evolves according to $\mathbf{H}_t = \mathbf{C} + \mathbf{A} \odot \mathbf{X}_{t-1} + \mathbf{B} \odot \mathbf{H}_{t-1}$ where **C**, **A** and **B** are positive semi-definite matrices (**C** is not reported). Three patterns are evident in the table: 1. All parameters are highly significant. 2. The relative importance of recent news is increasing in the frequency of sampling. 3. The persistence is increasing in the sampling frequency.

indicates a lower long-run level than Wishart model which may reflect that the choice of the identification restriction, that the mean of the matrix log of the standard residuals is 0, was restrictive.

Figures 5 through 8 and table 3 contains results based on the estimated "shocks" $\hat{\boldsymbol{\Xi}}_t$,

$$\hat{\boldsymbol{\Xi}}_t = \mathbf{H}_t^{-1/2}\mathbf{X}_t\mathbf{H}_t^{-1/2}. \tag{24}$$

If the Wishart were an adequate description of the standardized residuals these figures and tables should indicate:

- Mean 1 for $\bar{\hat{\Xi}}_{ii}$ and mean 0 for $\bar{\hat{\Xi}}_{ij}$, $i \neq j$.
- Variance 2 for $\bar{\hat{\Xi}}_{ii}$ and variance 1 for $\bar{\hat{\Xi}}_{ij}$, $i \neq j$.
- No correlation among any of the elements
- No obvious dependence

Only the first, which is probably the most important, property appears to be satisfied. The mean for the diagonal elements are very close to 1 while the means for the off diagonal elements are close to 0. There is obviously excess dependence in the data, and expecting the Wishart to provide an adequate description of the dependence in the data, in addition to fitting the mean, is difficult in light of the results of Barndorff-Nielsen & Shephard (2004) which indicate the covariance of the realized covariance changes every day and is a function of the intra-daily volatility of volatility and covariance. The figures confirm this, particularly for the diagonal elements which have obvious dependence. If the Wishart were an adequate description of the data, these should appear to be scatter plots of independent (standardized to have mean 1) χ^2 random variables. Instead the figures indicate positive dependence in the fit "errors".

5.3 Modeling The close-to-open and the open-to-close

One of the advantages of the PSD-MEM framework over existing models and estimation strategies for realized covariance is the flexibility to seamlessly handle both positive definite and positive semi-definite data using the same likelihood – modulo the degree of freedom parameter, which would likely not be of interest. To illustrate this, consider the modeling of the total covariance within a 24 hour period by decomposing it into the over-night covariance and the intra-day covariance. The natural measures of these two quantities are the realized covariance computed using within day returns and the outer-product of the close-to-open returns, $RC^{(\mathrm{CtO})}$. We considered a system model which nested the case where each element evolved independently of the other element and allowed for interaction between the two. The complete model can be described by

$$\begin{bmatrix} \mathbf{H}_t^{OtC} \\ \mathbf{H}_t^{CtO} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} RC_{t-1}^{OtC} \\ RC_{t-1}^{CtO} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix} \begin{bmatrix} \mathbf{H}_{t-1}^{OtC} \\ \mathbf{H}_{t-1}^{CtO} \end{bmatrix} \tag{25}$$

When the off-diagonal blocks of $\mathbf{A}_{ij}$ are set to zero the model becomes to processes without any interactions. However, by allowing these blocks to be non-zero, a model can be developed where interactions and

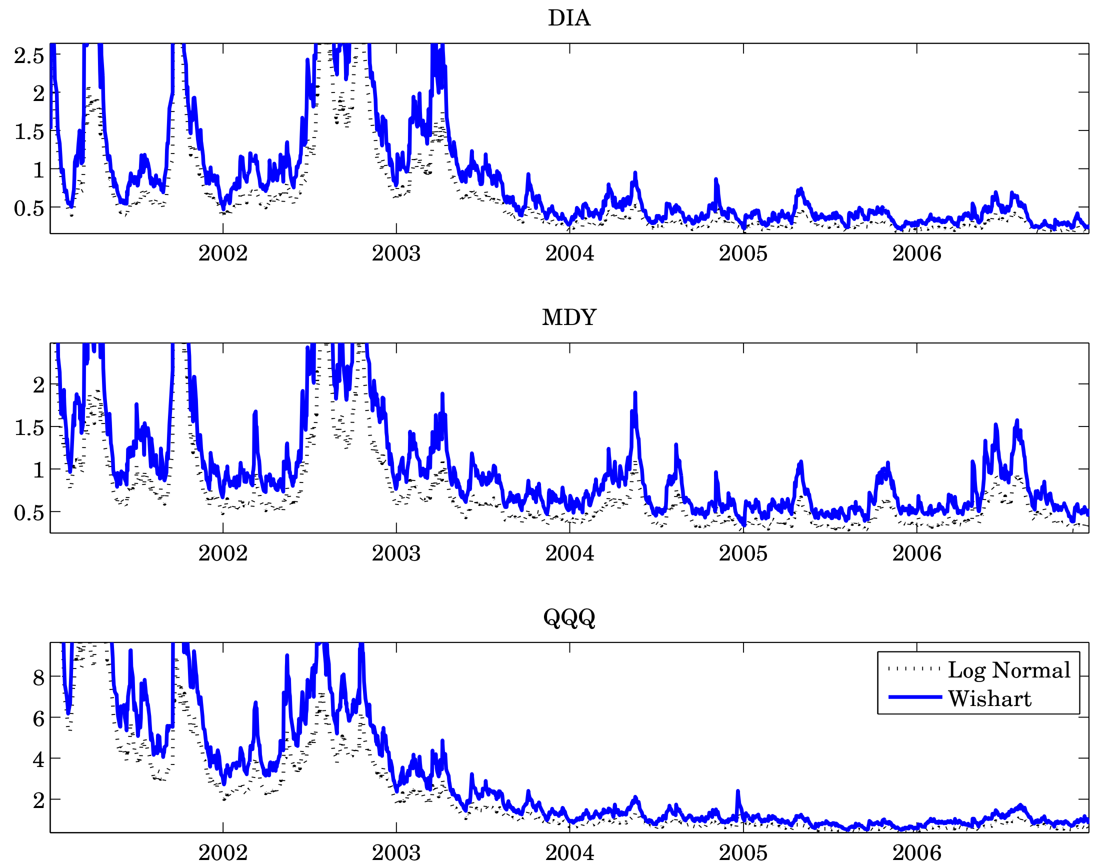

Figure 3: These three figures are typical of the fits of the Wishart PSD-MEM and the log-normal PSD-MEM. The fit values from the Wishart models are typically larger than those from the log-normal model, are more responsive to news and exhibit higher persistence.

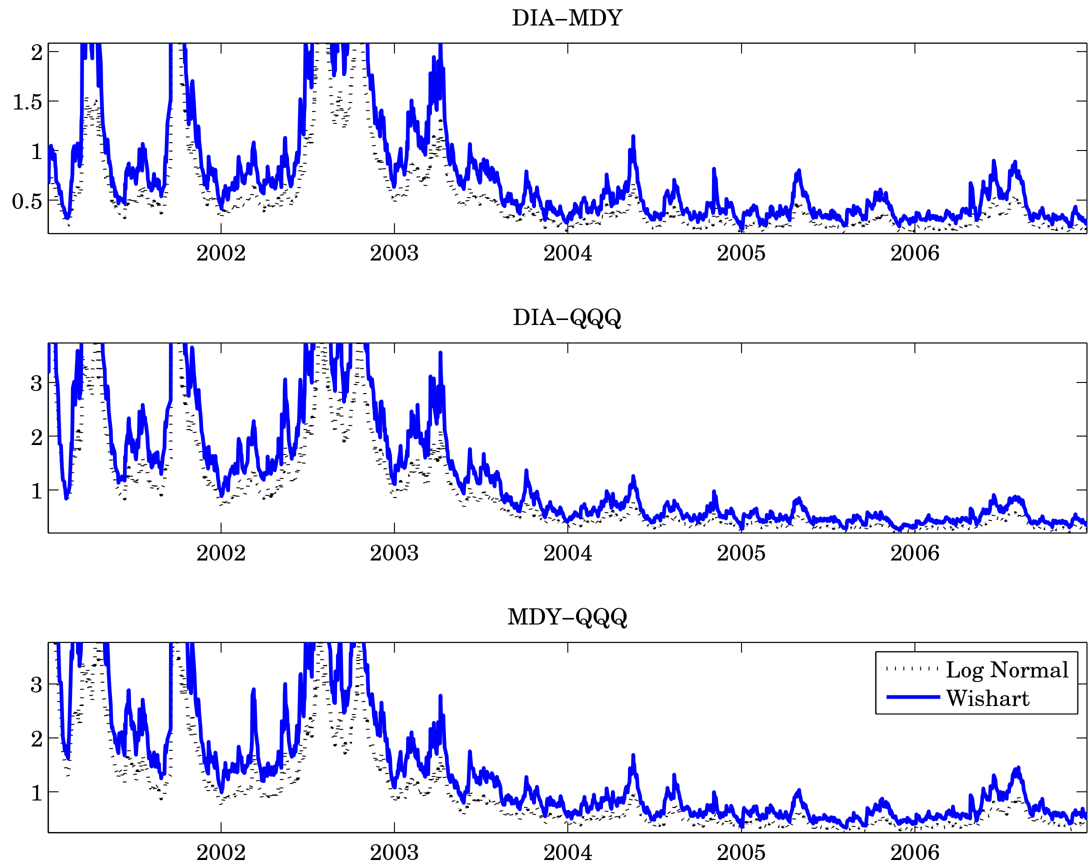

Figure 4: These three figures contain the fit covariance from the Wishart and the log-normal PSD-MEM when fit to the 30-minute data. Like the variances, the log-normal fit models have a lower level, less persistence and are less responsive.

Results using the Symmetric Matrix Log-Normal Log-likelihood

$RC^{(360)}$

	A			B			
	DIA	MDY	QQQ	DIA	MDY	QQQ	LL
DIA	0.136			0.816			-5569.5
	(0.00)			(0.00)			
MDY	0.139	0.141		0.782	0.749		
	(0.00)	(0.00)		(0.00)	(0.00)		
QQQ	0.140	0.143	0.155	0.799	0.766	0.785	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	

$RC^{(72)}$

	A			B			LL
DIA	0.103			0.848			-4778.7
	(0.00)			(0.00)			
MDY	0.101	0.103		0.837	0.826		
	(0.00)	(0.00)		(0.00)	(0.00)		
QQQ	0.100	0.101	0.102	0.850	0.839	0.853	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	

$RC^{(36)}$

	A			B			LL
DIA	0.090			0.860			-3927.0
	(0.00)			(0.00)			
MDY	0.087	0.088		0.856	0.851		
	(0.00)	(0.00)		(0.00)	(0.00)		
QQQ	0.086	0.086	0.087	0.865	0.860	0.870	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	

$RC^{(24)}$

	A			B			LL
DIA	0.086			0.864			-3260.6
	(0.00)			(0.00)			
MDY	0.083	0.083		0.860	0.856		
	(0.00)	(0.00)		(0.00)	(0.00)		
QQQ	0.082	0.081	0.081	0.870	0.866	0.877	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	

$RC^{(12)}$

	A			B			LL
DIA	0.065			0.883			-1574.6
	(0.00)			(0.00)			
MDY	0.064	0.065		0.880	0.877		
	(0.00)	(0.00)		(0.00)	(0.00)		
QQQ	0.062	0.063	0.061	0.889	0.886	0.896	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	

Table 2: This table contains parameter estimates from the baseline model using a matrix Log-normal distribution for the standardized shocks. The baseline model evolves according to $\mathbf{H}_t = \mathbf{C} + \mathbf{A} \odot \mathbf{X}_{t-1} + \mathbf{B} \odot \mathbf{H}_{t-1}$ where **C**, **A** and **B** are positive semi-definite matrices (**C** is not reported). The results are qualitatively similar to those using he Wishart distribution, although models fit using the log-normal distribution indicate less persistence and are somewhat less responsive to recent news.

Properties of Standardized Residuals for $RC^{(72)}$

Mean

Ξ_{11}	Ξ_{22}	Ξ_{33}	Ξ_{12}	Ξ_{13}	Ξ_{23}
1.004	1.010	0.997	-0.009	0.016	-0.004

Variance

Ξ_{11}	Ξ_{22}	Ξ_{33}	Ξ_{12}	Ξ_{13}	Ξ_{23}
0.618	1.079	0.948	0.169	0.280	0.214

Correlation

	Ξ_{11}	Ξ_{22}	Ξ_{33}	Ξ_{12}	Ξ_{13}	Ξ_{23}
Ξ_{11}		0.349	0.423	0.362	-0.117	-0.219
Ξ_{22}			0.240	-0.087	0.012	-0.277
Ξ_{33}				0.302	-0.572	-0.462
Ξ_{12}					-0.217	-0.134
Ξ_{13}						0.500
Ξ_{23}						

Properties of Standardized Residuals for $RC^{(12)}$

Mean

Ξ_{11}	Ξ_{22}	Ξ_{33}	Ξ_{12}	Ξ_{13}	Ξ_{23}
1.005	1.008	1.004	0.004	0.027	0.016

Variance

Ξ_{11}	Ξ_{22}	Ξ_{33}	Ξ_{12}	Ξ_{13}	Ξ_{23}
0.940	0.829	1.565	0.307	0.510	0.372

Correlation

	Ξ_{11}	Ξ_{22}	Ξ_{33}	Ξ_{12}	Ξ_{13}	Ξ_{23}
Ξ_{11}		0.448	0.373	0.349	-0.049	-0.172
Ξ_{22}			0.310	0.083	-0.002	-0.143
Ξ_{33}				0.322	-0.608	-0.587
Ξ_{12}					-0.284	-0.284
Ξ_{13}						0.545
Ξ_{23}						

Table 3: These two tables present the mean, variance and correlation of the standardized "shocks" to the PSD-MEM estimated using the Wishart distribution. The top correspond to using 5-minute returns while the bottom corresponds to using 30-minute returns (72 and 12 samples per day, respectively). If the assumption of a Wishart were correct, the means would be 1's for each of the diagonal elements, 0 for the off diagonal elements, the variances would be 2 for the on diagonal elements and 1 for the off diagonal elements and the correlation matrix would be diagonal. These results indicate that the model performs well on average but that the strong link in the Wishart between the location and the scale is not correct, even on average.

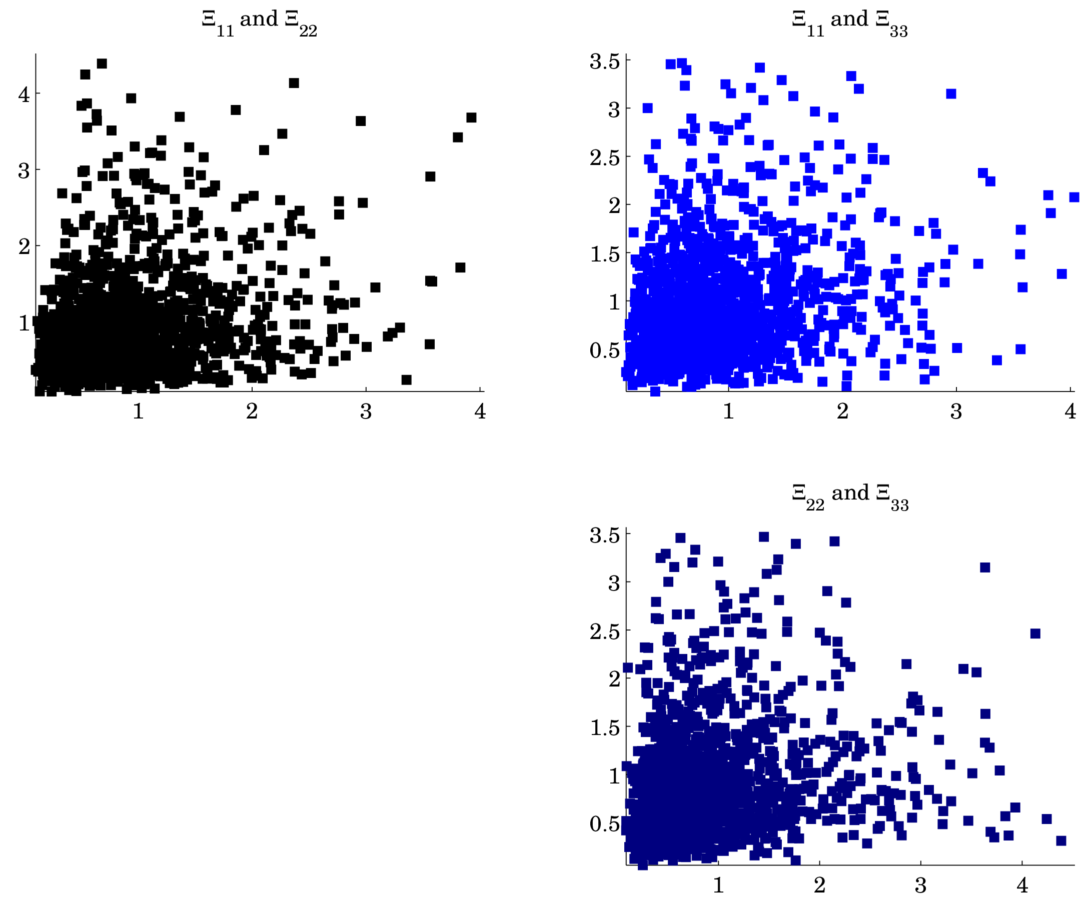

Figure 5: These three plots contain scatter plots of the diagonal elements of the estimated "shocks" to the PSD-MEM process, $\hat{\boldsymbol{\Xi}}_t = \hat{\mathbf{H}}_t^{-1/2}\hat{\Sigma}_t\hat{\mathbf{H}}_t^{-1/2}$. If the assumption of a Wishart were correct, these scatter plots should and contain the result of scattering two independent χ^2 random variables against one another. However, there is clear positive dependence evident in the figures.

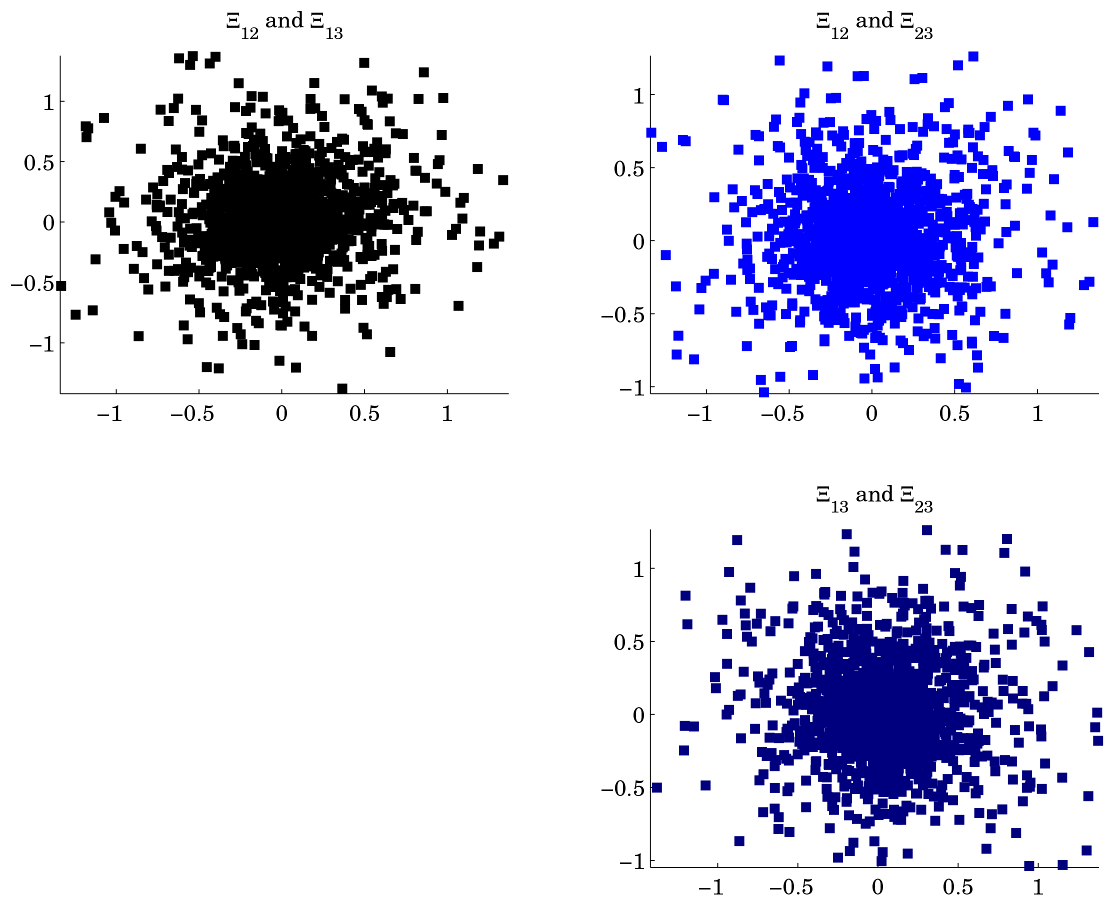

Figure 6: These three plots contain scatter plots of the off-diagonal elements of the estimated "shocks" to the PSD-MEM process, $\hat{\boldsymbol{\Xi}}_t = \hat{\mathbf{H}}_t^{-1/2}\hat{\Sigma}_t\hat{\mathbf{H}}_t^{-1/2}$. These should be independent is the assumption of the Wishart were valid, and while they superficially appear independent, they all exhibit statistically significant correlations.

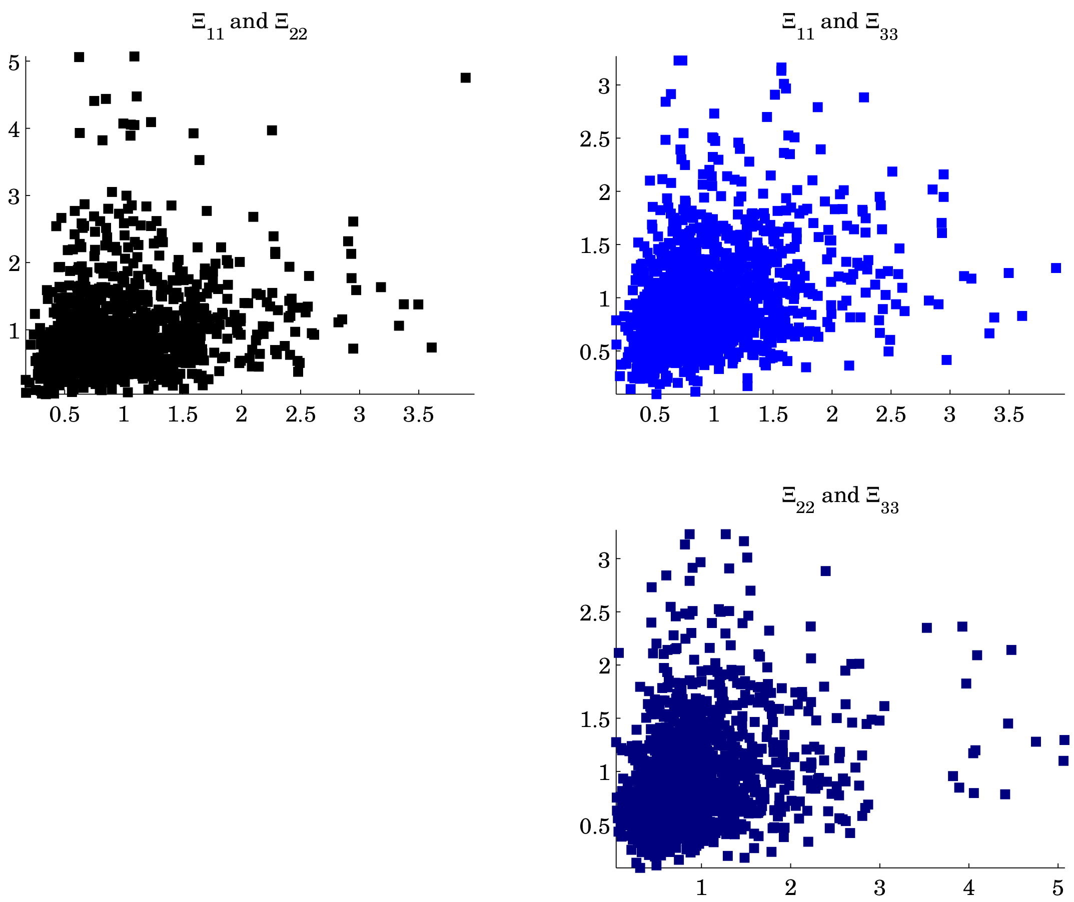

Figure 7: These three scatter plots show that like the 30 minute standardized shocks, the diagonal elements of the 5 minute standardized shocks all exhibit more dependence that a Wishart would imply for the data.

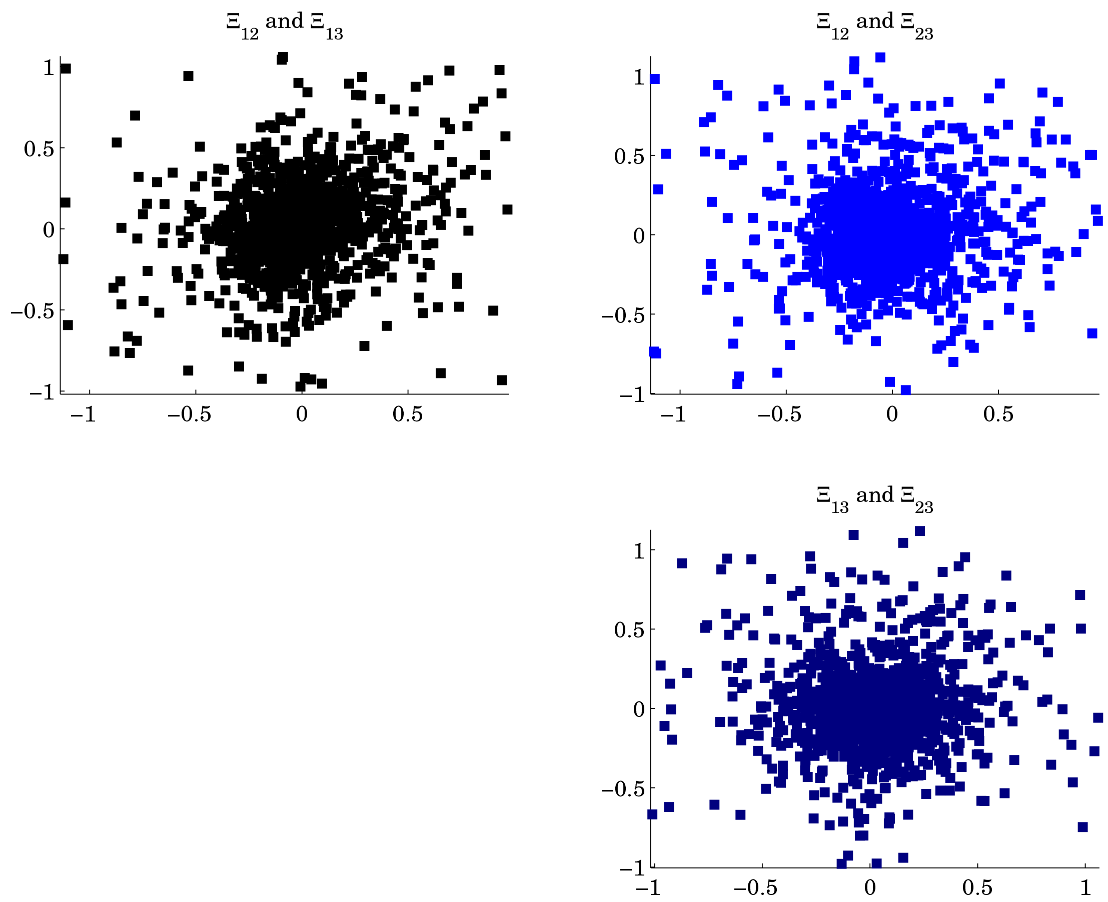

Figure 8: These three scatter plots show that like the 30 minute standardized shocks, the off-diagonal elements of the 5 minute standardized shocks all exhibit little dependence although they all have statistically significant correlations, primarily due to non-pictured outliers.

own lagged effects of each component are allowed. Each block was restricted to be positive semi-definite by Choleski re-parameterization which is sufficient to ensure that fit values are positive definite.

This model was fit using the Wishart log-likelihood, which is the only appropriate choice since RC_t^{OtC} is a rank 1 matrix, and results are presented in table 4. The open-to-close was modeled by the realized covariance computed using 30-minute samples. When estimated separately the parameter estimates are consistent with was seen in the previous application. Both models are very persistent, although each is less than what was documented modeling the entire day's covariance using the sum of these two elements. The model for the intra-day realized covariance shows a larger response to news and requires less smoothing, while the parameters of the close-to-open covariance are similar to those of a standard MV-GARCH.

When estimated as a system all of the **A** parameters are reduced in magnitude although there is widespread evidence of significance in the coefficients on both proxies, although some caution is needed when interpreting the magnitude. While the close-to-open proxy is significant in the model for the open-to-close realized covariance, the coefficients are relatively small as is the typical magnitude of the over night return, where only 25% or so of the total covariance is expressed. As a results, and pictured in figure 9, the economic significant of the close-to-open covariance is very small. The is not true when modeling the close-to-open using the open-to-close. These parameters are jointly significant and the log-likelihood of the model, when treated as a system, indicates a much better fit than when estimated separately. Including the open-to-close realized covariance in the model for the close-to-open covariance dramatically reduced the required level of smoothing as indicated by **B** and appears to be economically interesting as illustrated in the bottom panel of figure 9. The fit of the combined model is meaningfully different than the fit of the model that used close-to-open exclusively.

5.4 Comparing MV-GARCH and PSD-MEMs using Realized Covariance

The final application of the PSD-MEM framework is to the question of whether the information content in realized covariance is sufficient to crowd-out the information content in the outer-product of daily returns. In this example the natural model is a standard diagonal *vech* MV-GARCH(1,1) that uses close-to-close returns to proxy for the latent covariance. This specification is extended to allow either or both lagged realized covariance or lagged outer-product of close-to-close returns.

$$\mathbf{H}_t^{\text{CtC}} = \mathbf{C} + \mathbf{A}_1 \odot RC_{t-1}^{(\text{CtC})} + \mathbf{A}_2 \odot \left(RC_{t-1}^{(12)} + RC_{t-1}^{(\text{CtO})} \right) + \mathbf{B} \odot \mathbf{H}_{t-1}^{\text{CtC}} \tag{26}$$

where the "exogenous" regressor is the sum of the within day covariance, as measured by the 30-minute sampled realized covariance and the close-to-open realized covariance as measured by the outer-product of the close-to-open return. The inclusion of the close-to-open covariance was to ensure that the alterative proxy had sufficient information to proxy for the entire day's covariance.

The results of three estimates of the model in equation (26) are presented in table 5. The three specifications correspond to a standard MV-GARCH(1,1), which sets $\mathbf{A}_2 = \mathbf{0}$, a model drive exclusively by the 30-minute realized covariance term, $\mathbf{A}_1 = \mathbf{0}$, and a model which allows both terms to be non-zero. The parameter estimates indicate that while the higher-precision proxy is useful, as indicated by an improvement of 50-log-likelihood points using a model with the same number of parameters, it is insufficient to crowd

Modeling the open-to-close and the close-to-open Covariance

$RC^{(12)}$ and $RC^{(\mathrm{CtO})}$ modeled separately

	$\mathbf{A}_{11}$			$\mathbf{A}_{12}$			$\mathbf{B}_1$			
	DIA	MDY	QQQ	DIA	MDY	QQQ	DIA	MDY	QQQ	LL
DIA	0.177 (0.00)						0.799 (0.00)			4756.2
MDY	0.166 (0.00)	0.160 (0.00)					0.803 (0.00)	0.807 (0.00)		
QQQ	0.165 (0.00)	0.155 (0.00)	0.154 (0.00)				0.808 (0.00)	0.814 (0.00)	0.825 (0.00)	
	$\mathbf{A}_{21}$			$\mathbf{A}_{22}$			$\mathbf{B}_2$			
DIA				0.074 (0.00)			0.903 (0.00)			
MDY				0.063 (0.00)	0.055 (0.00)		0.912 (0.00)	0.922 (0.00)		
QQQ				0.061 (0.00)	0.053 (0.00)	0.052 (0.00)	0.919 (0.00)	0.929 (0.00)	0.936 (0.00)	

$RC^{(12)}$ and $RC^{(\mathrm{CtO})}$ modeled as a system

	$\mathbf{A}_{11}$			$\mathbf{A}_{12}$			$\mathbf{B}_1$			
DIA	0.165 (0.00)			0.028 (0.00)			0.800 (0.00)			4969.5
MDY	0.153 (0.00)	0.147 (0.00)		0.031 (0.00)	0.034 (0.00)		0.801 (0.00)	0.803 (0.00)		
QQQ	0.153 (0.00)	0.142 (0.00)	0.142 (0.00)	0.036 (0.00)	0.040 (0.00)	0.047 (0.00)	0.804 (0.00)	0.807 (0.00)	0.814 (0.00)	
	$\mathbf{A}_{21}$			$\mathbf{A}_{22}$			$\mathbf{B}_2$			LL
DIA	0.052 (0.19)			0.051 (0.26)			0.789 (0.00)			
MDY	0.049 (0.28)	0.045 (0.53)		0.050 (0.36)	0.050 (0.44)		0.803 (0.00)	0.817 (0.00)		
QQQ	0.052 (0.21)	0.048 (0.30)	0.052 (0.21)	0.040 (0.31)	0.041 (0.45)	0.033 (0.40)	0.819 (0.00)	0.833 (0.00)	0.850 (0.00)	

Table 4: This table contains the parameter estimates (p-values in parenthesis) and log-likelihoods of for the model of the open-to-close and the close-to-open estimated separately (top) and as a system (bottom). The system estimation provides a large improvement in the fits of the model, particularly in the close-to-open period. The model used in each panel is $\mathbf{H}_t = \mathbf{C}_j + \mathbf{A}_{j1} RC^{(12)} + \mathbf{A}_{j1} RC^{(\mathrm{CtO})} + \mathbf{B}\mathbf{H}_{t-1}$ where $j = 1$ corresponds to the model using $RC^{(12)}$ as the dependent variable and $j = 2$ corresponds to the model with $RC^{(\mathrm{CtO})}$ as the dependent variable.

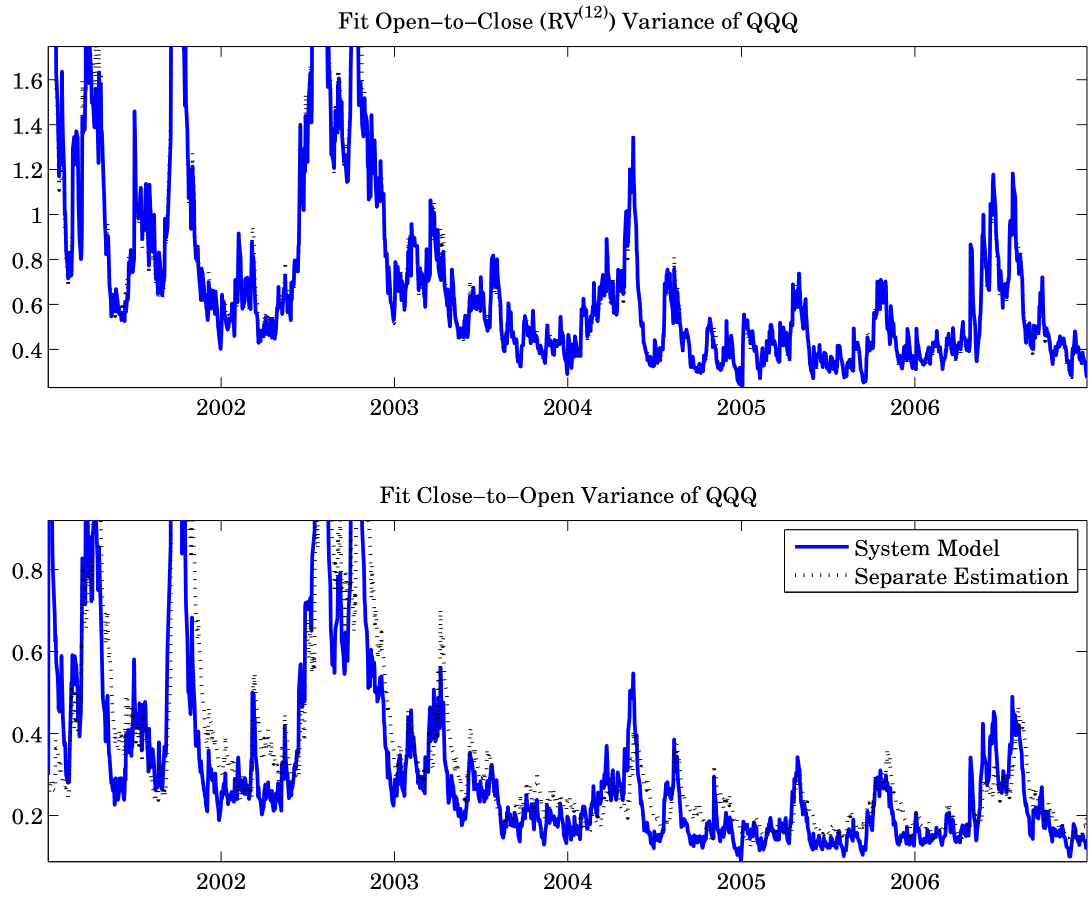

Figure 9: These two figures show the fit of the open-to-close, as measured by the realized covariance between 9:45 and 15:45, and the close-to-open, as measured by the return spanning 15:45 until 9:45 the next day. The top shows that while the coefficients on the over night proxy were significant, the economic significance from the over-night is minor. The bottom shows that the over night variance of QQQ is typically 40% of the open-to-close and is affected by the open-to-close volatility.

Valuing Realized for MV-GARCH Estimation

$RC^{(\text{CtC})}$ (Diagonal *vech* MV-GARCH(1,1))

	$\mathbf{A}_1$			$\mathbf{A}_2$			$\mathbf{B}$			
	DIA	MDY	QQQ	DIA	MDY	QQQ	DIA	MDY	QQQ	LL
DIA	0.047 (0.01)						0.938 (0.00)			-1150.6
MDY	0.046 (0.00)	0.048 (0.00)					0.935 (0.00)	0.932 (0.00)		
QQQ	0.038 (0.02)	0.039 (0.00)	0.033 (0.00)				0.947 (0.00)	0.944 (0.00)	0.956 (0.00)	

$RC^{(12)}$ as regressor

	$\mathbf{A}_1$			$\mathbf{A}_2$			$\mathbf{B}$			LL
DIA				0.120 (0.06)			0.838 (0.00)			-1094.3
MDY				0.111 (0.07)	0.105 (0.02)		0.849 (0.00)	0.860 (0.00)		
QQQ				0.090 (0.17)	0.087 (0.10)	0.074 (0.12)	0.870 (0.00)	0.882 (0.00)	0.904 (0.00)	

Both as regressors

	$\mathbf{A}_1$			$\mathbf{A}_2$			$\mathbf{B}$			LL
DIA	0.027 (0.12)			0.081 (0.20)			0.859 (0.00)			-1075.1
MDY	0.028 (0.01)	0.030 (0.03)		0.071 (0.37)	0.063 (0.42)		0.867 (0.00)	0.876 (0.00)		
QQQ	0.024 (0.12)	0.025 (0.02)	0.021 (0.05)	0.057 (0.35)	0.053 (0.43)	0.045 (0.36)	0.886 (0.00)	0.895 (0.00)	0.914 (0.00)	

Table 5: This table presents the results comparing a standard MV-GARCH, a special case of the Wishart PSD-MEM, against a model where the dependent variables, the outer-product of close-to-close returns is driven by realized covariance computed using 30 minutes returns, $RC^{(12)}$, or both 30-minute returns and close-to-close returns. The complete specification is $\mathbf{H}_t = \mathbf{C} + \mathbf{A}_1 \odot RC^{(\text{CtC})} + \mathbf{A}_2 \odot RC^{(12)} + \mathbf{B} \odot \mathbf{H}_{t-1}$. As measured by fit, the more precise proxy provides a much better fit than close-to-close returns and substantially improves the fit of the model when both, although neither proxy is sufficient to crowd out the other.

out the lagged outer-product. This contrasts with the often found ability of realized variance to crowd-out daily returns squared in augmented GARCH models. There are two reasonable explanations for this finding. First, it may be the case that even 30-minute returns are being sampled too frequently to capture all of the information in realized covariance. The pseudo-correlation plot indicates there may be some support for this hypothesis although it is not completely clear. An alternative hypothesis is that the realized covariance is not being sampled fast enough, and that despite being a higher quality proxy, the added precision is not enough to completely crows out the outer-product of daily returns, at least in finite samples. It is clear, however, that more work is needed in finding both precise and unbiased proxies for realized covariance in-order to distinguish between these competing hypotheses.

6 Conclusions

This paper has introduced the positive semi-definite multiplicative error model for modeling conditional covariances. The PSD-MEM provides an important bridge between the standard framework of multivariate

GARCH modeling, which specifies a distribution for returns in order to estimate parameters of the conditional covariance, and the emerging standard of realized covariance which employs ultra-high frequency data to improve the measurement of volatility and correlation. By treating realized covariance as a positive semi-definite process for random matrices, the link between the conditional distribution of returns and inference on the covariance can be broken.

The main specification for the PSD-MEM is intentionally vague over the dynamics for the positive semi-definite process. This "omission" was made to allow researchers to specify a model appropriate for the data being studied and any specific requirements of needed. As a simple baseline, most existing MV-GARCH models can be specified as PSD-MEM with little change. This would allow the modeling of conditional asymmetries in covariance or for the modeling of the realized variance and the realized correlation to be decomposed in a DCC-like model (Engle 2002*a*), using univariate MEMs for each conditional variance and then modeling the transformed quantity using a PSD-MEM with the restriction that the expected value of the diagonal elements is unity. This should allow for improved estimation of models of volatility transmission and contagion, where volatility is allowed to spill-over into future correlation.

Two likelihoods have been investigated to estimate the parameters governing the dynamics in a PSD-MEM. The first is the Wishart. The Wishart has a number of desirable properties including an easy to interpret score and a rich tradition in the statistic literature as a 'loss function" for covariance risk. More importantly, it was shown that the value of the degree of freedom parameter in the Wishart is not needed, and is in fact asymptotically irrelevant, for estimation of the dynamics in a PSD-MEM, a very desirable property. In the rare case where the degree of freedom is needed, it can be efficiently estimated on the transformed shocks. This said, there are some drawbacks to using the Wishart, primarily the link between the dependence between the elements and the conditional expectation. While finding a flexible specification to overcome this restriction is an interesting topic for future research, it is unclear what the gain would be. If standardized shocks were conditionally homoskedastic, but not uncorrelated, it could lead to efficient estimation in an analogous manner to Cipollini et al. (2006). However, the asymptotic theory of Barndorff-Nielsen & Shephard (2004) clearly shows that this would not normally be the case and 1st order efficient would require an auxiliary model for the quadratic covariance, something that is notoriously difficult to estimate.

The second likelihood investigated was that of a matrix valued log normal. This distribution arises naturally out of applying a matrix exponential function to a symmetric matrix normal. The advantage of this choice for the likelihood is that the model provides a substantial amount of flexibility in the standardized shock. However, there are some substantial drawbacks to using the log-normal including difficulty in establishing non-binding but required identification restrictions and the requirement to compute the matrix logarithm at for every data point for each iteration of a parameter estimation routine. The differences between the models fit with the Wishart and the log-normal were minor and the Wishart is simpler and more direct to use. As a result it is difficult to see how the log-normal would be preferred to the Wishart except in circumstances where the normality of the log matrix might provide improved accuracy in forecasting.

The paper concludes with three example application of the PSD-MEM. The first examines the effect of using different proxies for the realized covariance. This application documented that the precision decreases the need for smoothing and seemed to increase the persistence as indicated by the model. The model was estimated using both the Wishart likelihood and the log-normal, and only minor differences in the estimated parameters were evident, aside from a level shift. The second application examined the decomposition of a day's worth of covariance into the over night component and the within trading hours portion. This allowed the model to be estimated separately, where each component was driven by itself and as a system where feedback between each was allowed. The model indicates that feedback between the close-to-open and the open-to-close improved the fit of the model. The final application compared the results of a standard MV-GARCH(1,1) to an augmented MV-GARCH where an intra-daily proxy was allowed to drive the process. It was found that while higher quality proxies are important for modeling the conditional covariance of daily returns, they do not contain sufficient information (or precision, depending on the interpretation) to crowd out the outer-product of daily returns.

While this paper has introduced the PSD-MEM, there are still many interesting questions remaining to be answered. First, can a more flexible distribution be constructed that will allow for improved estimation of the conditional covariance, and if so, what types of gains should be expected. Another remaining challenge, one which is pervasive in the MV-GARCH literature, is to study when a PSD-MEM may be consistent for the true parameters of the process when the distribution is mis-specified. The structure of the first order condition of a Wishart is such that this may be a possibility, but further investigation is warranted. Another interesting application of PSD-MEMs would be to examine the properties of a jump-free measure of realized covariance in relationship to a measure of the quadratic covariation, as was done in this study. This has been done for univariate time-series by Lanne (2006*b*) and an extension to multivariate systems would be interesting. We leave these as topics for future research.

References

Andersen, T. G. & Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**(4), 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The Distribution of Stock Return Volatility', *Journal of Financial Economics* **61**, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and Forecasting Realized Volatility', *Econometrica* **71**(1), 3–29.

Andersen, T. G., Bollerslev, T. & Meddahi, N. (2006), Realized volatility forecasting and market microstructure noise, Technical report, Kellogg School of Management, Northwestern University.

Bandi, F. & Russell, J. (2005), Realized covariation, realized beta, and microstructure noise. University of Chicago.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2006), Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Oxford Financial Research Centre Economics Series Ref: 2006-FE-05.

Barndorff-Nielsen, O. E. & Shephard, N. (2004), 'Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics', *Econometrica* **73**(4), 885–925.

Bollerslev, T., Engle, R. F. & Wooldridge, J. M. (1988), 'A capital asset pricing model with time-varying covariances', *Journal of Political Economy* **96**(1), 116–131.

Chiriac, R. (2006), Estimating realized volatility wishart autoregressive model, Technical report, University of Konstanz.

Chiriac, R. & Voev, V. (2007), Long memory modelling of realized covariance matrices, Technical report, University of Konstanz.

Cipollini, F., Engle, R. F. & Gallo, G. M. (2006), Vector multiplicative error models: Representation and inference, Technical report, NYU Stern School of Business Working Paper.

Corsi, F. (2004), Simple long memory models of realized volatility, Technical report, University of Lugano.

Corsi, F. & Audrino, F. (2007), Realized correlation tick-by-tick, Technical report, Univeristy of St. Galen. Discussion Paper 2007-02.

de Waal, D. J. (1979), 'General theorem on quadratic forms of matrix normal variables', *South African Statistics* **13**, 103–122.

Ding, Z. & Engle, R. (2001), 'Large scale conditional matrix modeling, estimation and testing', *Academia Economic Papers* **29**(2), 157–184.

Engle, R. F. (1982), 'Autoregressive conditional heteroskedasticity with esimtates of the variance of U.K. inflation', *Econometrica* **50**(4), 987–1008.

Engle, R. F. (2002*a*), 'Dynamic conditional correlation - a simple class of multivariate GARCH models', *Journal of Business and Economic Statistics* **20**(3), 339–350.

Engle, R. F. (2002*b*), 'New frontiers for arch models', *Journal of Applied Econometrics* **17**, 425–446.

Engle, R. F. & Gallo, G. M. (2003), A multiple indicators model for volatility using intra-daily data. NBER Working Papers 10117.

Engle, R. F. & Kroner, K. F. (1995), 'Multivariate simultaneous generalized ARCH', *Econometric Theory* **11**(1), 122–150.

Engle, R. F. & Russell, J. R. (1998), 'Forecasting transaction rates: the autoregressive conditional duration model', *Econometrica* **66**, 1127–1162.

Epps, T. W. (1979), 'Comovements in stock prices in the very short run', *Journal of the American Statistical Society* **74**, 291–296.

Fleming, J., Kirby, C. & Ostdiek, B. (2003), 'The economic value of volatility timing using "realized" volatility', *Journal of Financial Economics* **67**(3), 473–509.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2007), 'Predicting volatility: How to get most out of returns data sampled at different frequencies', *Journal of Econometrics* . Forthcoming.

Gourieroux, C., Jasiak, J. & Sufana, R. (2005), The wishart autoregressive process of multivariate stochastic volatility, Technical report, Univeristy of Toronto.

Gourieroux, C. S. (2006), 'Continuous time wishart process for stochastic risk', *Econometric Reviews* **25**, 177–217.

Griffin, J. E. & Oomen, R. C. (2006), Covariance measurement in the presence of non-synchronous trading and market microstructure noise. Mimeo.

Gupta, A. K. & Nagar, D. K. (2000), *Matrix Variate Distributions*, Vol. 104 of *Monographs and Surveys in Pure and Applied Mathematics*, Chapman and Hall/CRC, Boca Raton.

Hayashi, T. & Yoshida, N. (2005), 'On covariance estimation of non-synchronously observed diffusion processes', *Bernoulli* **11**, 359–379.

Kawakatsu, H. (2006), 'Matrix exponential garch', *Journal of Econometrics* **134**(1), 95–128.

Kroner, K. E. & Ng, V. K. (1998), 'Modeling asymmetric comovements of asset returns', *Review of Financial Studies* **11**(4), 817–844.

Lanne, M. (2006*a*), Forecasting realized volatility by decomposition, Technical report, European Univeristy Institute. EUI Working Paper ECO No. 2006/20.

Lanne, M. (2006*b*), A mixture multiplicative error model for realized volatility, Technical report, European Univeristy Institute. ECO No. 2006/3.

Laurent, S., Bauwens, L. & Rombouts, J. V. K. (2006), 'Multivariate garch models: a survey', *Journal of Applied Econometrics* **21**(1), 79–109.

Ledoit, O., Santa-Clara, P. & Wolf, M. (2003), 'Flexible multivariate GARCH modeling with an application to international stock markets', *Review of Economics and Statistics* **85**(3), 735–747.

Lundin, M., Dacorogna, M. M. & Muller, U. A. (1999), Correlation of high frequency financial time series, *in* P. Lequeux, ed., 'Financial Markets Tick by Tick', John Wiley and Sons.

Martens, M. (2004), Estimating unbiased and precise realized covariances. Erasmus University Rotterdam.

Nel, D. G. & van der Merwe, C. A. (1986), 'A solution to the multivariate behrens-fisher problem', *Communication in Statistics: Theory and Methods* **15**(12), 3719–3735.

Nelson, D. (1991), 'Conditional heteroskedasticity in asset returns: a new approach', *Econometrica* **59**(2), 347–370.

Pelletier, D. (2003), Regime switching for dynamic correlations. CIRANO.

Philipov, A. & Glickman, M. E. (2004), Multivariate stochastic volatility via wishart processes, Technical report, Department of Health Services, Boston University.

Sheppard, K. (2006), Realized covariance and scrambling. Univeristy of Oxford.

Tan, W. Y. & Gupta, R. P. (1983), 'On approximating a linear combination of central wishart matrices with positive coefficients', *Communication in Statistics: Thoery and Methods* **12**(22), 2589–2600.

Tse, Y. K. & Tsui, A. K. (2002), 'A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations', *Journal of Business & Economic Statistics* **20**(3), 351–362.

Uhlig, H. (1994), 'On singular wishart and singular multivariate beta distributions', *The Annals of Statistics* **22**(1), 395–405.

Voev, V. (2007), Dynamic modeling of large dimensional covariance matrices, Technical report, University of Konstanz.

Voev, V. & Lunde, A. (2007), 'Integrated covariance estimation using high-frequency data in the presence of noise', *Journal of Financial Econometrics* **5**(1), 68–104.

Zebedee, A. A. (2001), A closer look at comovements among stock returns. Department of Finance, San Diego State University.

Zhang, L. (2006), Estimating Covariation: Epps Effect, Microstructure Noise. University of Illinois at Chicago.

Zhang, L., Mykland, P. & Aït-Sahalia, Y. (2004), A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Forthcoming *Journal of the American Statistical Association.*