A Realized RSDC Model

Aymard Kassi
North Carolina State University* and
Denis Pelletier
North Carolina State University†

Current version: Wednesday 3rd December, 2014

Abstract

This paper introduces a new multivariate conditional volatility model for returns that utilizes realized covariance matrices. The model decomposes the conditional and realized covariance matrices into standard deviations and correlations matrices. On a first level, the univariate variances are estimated by a modified Generalized Autoregressive Conditional Heteroskedasticity (GARCH) that exploits intraday information. On the second level the conditional correlation matrices follows a regime switching Markov process. The inference of the regimes and the regime-switching correlations exploits the information contained in the realized correlation. An empirical application shows the ease of estimation and a forecasting exercise shows superior predictive ability when the high-frequency information is incorporated.

* Mailing address: Department of Economics, Campus Box 8110, North Carolina State University, Raleigh, 27695-8110, USA. e-mail: ankassi@ncsu.edu.
† Mailing address: Department of Economics, Campus Box 8110, North Carolina State University, Raleigh, 27695-8110, USA. e-mail: denis_pelletier@ncsu.edu. Web site: http://www4.ncsu.edu/~dpellet.
1 Introduction

The easy access to intraday data in financial econometrics has created the need for models which can exploit efficiently large high-frequency information. The potential advantage of higher frequency information set in finance is a better understanding and a better forecasting of time-varying covariance matrices. Understanding and predicting covariances between assets or between markets has important implications in areas such as portfolio allocation, asset pricing and risk management. Since the important work of Engle (1982) and Bollerslev (1986), extensive research in the area of univariate and multivariate volatility models for the volatility using low-frequency (daily) returns has been undertaken.

In recent years however, starting with Andersen and Bollerslev (1998), the focus of volatility modeling has shifted from returns computed with low-frequency observations to high-frequency transactions data when available. Under some general assumptions, fairly precise measurements of daily volatility can be estimated by non-parametric methods. There exists a considerable amount of work on issues related to these estimators. The asymptotic properties [e.g., Barndorff-Nielsen and Shephard (2002)], the optimal sampling frequency [e.g. Bandi and Russell (2008)], the robustness to jumps [e.g. Barndorff-Nielsen and Sheppard (2004)] and the robustness to market microstructure noise [e.g. Hansen and Lunde (2006)] have all been extensively studied. However, the focus has mostly been on the measurement of volatility, less so on its modeling and forecasting.

Recently, on an univariate level, Shephard and Sheppard (2010) and Hansen et al. (2012) have introduced frameworks that take advantage of the information contained in intraday data to improve the modeling of daily variances. They have shown that the use of high frequency information in conditional volatility specification improves the reaction time of the forecasted variances to rapid change in the financial market. In both studies, the conditional variance is specified as a modified GARCH of Bollerslev (1986). The modification consists in replacing the lag square returns by the lag value of the realized variance. The improvement in the dynamic of the daily variance stems from the fact that the realized volatility is a better proxy for the true and unobserved conditional second moment of the returns. In both specifications, the evolution of realized variance is also modeled. The objective is to forecast the daily volatility at an horizon greater than one period ahead. Shephard and Sheppard (2010) uses the Multiplicative Error Model (MEM) of Engle (2002b) to describe the dynamic of the conditional mean of the realized measures. Hansen et al. (2012) supplement the conditional variance specification with a measurement equation linking linearly the realized and the conditional volatility.

High frequency information has also been applied in multivariate setting. Gourieroux et al. (2009) introduce the Wishart autoregressive (WAR) model in which the measures of realized covariance have a non-central Wishart distribution. The non-centrality matrix of the distribution accommodates temporal dependence in the scale matrix. Jin and Maheu (2012), however propose to jointly model the returns and realized covariances. They assume that the realized covariance matrices follow a Wishart distribution with time varying scale matrix. Noureldin and Sheppard (2012) extend their univariate model using multivariate MEM. They also assume that the conditional and realized variance are Wishart distributed.
Finally, Hansen et al. (2014) extend Hansen et al. (2012) to vector of assets returns. The marginal model for each asset is the same as in their univariate model. The correlation between assets is assumed to be determined exclusively by the correlation between each asset and an index factor (S&P500 for instance).

The goal in this paper is to introduce a new and parsimonious way to generalize to multivariate data the Realized GARCH framework of Hansen et al. (2012). The model is inspired by the decomposition of the covariance matrix adopted successfully for low frequency returns in the Constant Conditional Correlation (CCC) model of Bollerslev (1990) and the Dynamic Conditional Correlation (DCC) models of Engle (2002a) and Tse and Tsui (2002). As indicated by their names, the CCC assumes that the conditional correlations between assets are constant while the DCC describes a dynamic correlation matrix with an autoregressive process. The Regime Switching Dynamic Correlations (RSDC) model of Pelletier (2006) is an intermediary step because it assumes a Markov switching process with correlations that are different across regimes but constant within a particular regime. The transition probability matrix governs the persistence of the correlations. The model proposed in this article, the Realized RSDC (R-RSDC), builds on Pelletier (2006) by also assuming regime-dependent conditional correlations matrices. Furthermore, the realized covariance matrices follow a mixture of Wishart distributions.

There are three main benefits to our proposed model. First the information set for inference of the regime switching correlations is richer than in Pelletier (2006) potentially allowing for a better description of the correlations and covariances. Second, the method of estimation, the Expectation-Maximization algorithm introduced by Dempster et al. (1977), breaks the curse of dimensionality in that the correlation matrices can be estimated through simple weighted sums. Third, the idea of describing the market through different regimes allow to give economic interpretation to the overall levels of correlation. It is possible for instance to tie a general increase of correlations between assets to particular events in the market such as financial crisis.

The paper is organized as follows. Section 2 introduces the R-RSDC model. The estimation procedure is discussed in Section 3. An overview of the computation of realized measures is given in Section 4. Section 5 presents an empirical application with transaction-level data for seven stocks. Section 6 evaluates the forecasting ability of the model. Concluding remarks and future work are given in Section 6.

2 MODEL

Let \( Y_t \) be the \( n \times 1 \) vector of daily demeaned log-returns such that:

\[
\begin{align*}
Y_t &= H_t^{1/2} U_t, \\
E(U_t) &= 0, \\
\text{Var}(U_t) &= I_n,
\end{align*}
\]
where $H_t^{1/2}$ is a $n \times n$ positive definite matrix, $I_n$ is the identity matrix and $0$ is a $n \times 1$ vector of zeros. Conditional on $F_{t-1}$, the $t-1$ information set, the variance of $Y_t$ is:

\[ \text{Var}(Y_t | F_{t-1}) = Var_{t-1}(Y_t) = H_t. \] (3)

The conditional covariance matrix $H_t$ can be decomposed into standard deviations and correlations matrices:

\[ H_t = D_t \Gamma_t D_t, \] (4)

where $D_t = diag(\sqrt{h_{1,t}}, \sqrt{h_{2,t}}, ... , \sqrt{h_{n,t}})$ and $\Gamma_t$ is the correlation matrix. Therefore the sequence of daily conditional variances for the $i^{th}$ asset is \( \{h_{i,t}\}_{t=1}^{T} \), where $T$ is the time of the last observation.

Contrary to the basic GARCH model where the information set is composed exclusively of the daily log-returns, the R-RSDC also exploits the information from the corresponding daily realized covariance matrices $\{X_t\}_{t=1}^{T}$ computed with high-frequency intraday returns.

In what follows, the R-RSDC model proposes a dynamic for $\{H_t\}_{t=1}^{T}$ and $\{X_t\}_{t=1}^{T}$ assuming multivariate Gaussian distribution for the innovation, that is $U_t \sim i.i.d.N(0, I_n)$ and Wishart distribution for $X_t$.

### 2.1 The Conditional Variances

With the set up above, the dependence of the elements of $D_t$ conditioning on the past information of the daily returns and realized variances is specified as the following modified GARCH:

\[ h_{i,t} = \omega_i + \beta_i h_{i,t-1} + \alpha_i x_{i,t-1}. \] (5)

As equation (5) shows, the lag realized variance replaces the lag square return from the usual GARCH specification. It has been shown in Lopez (2001) that although the squared return is an unbiased estimator for volatility it is very noisy. The objective of the above substitution is the improvement of the conditional variance dynamic. This can be seen by solving forward for $h_{i,t}$ in (5):

\[ h_{i,t} = \frac{\omega_i}{1 - \beta_i} + \sum_{j=0}^{\infty} [\alpha_i \beta_i^j] x_{i,t-1-j} \] (6)

The past values of the realized variance drive the dynamic of the conditional variance. The former, being a better proxy for the latent variance, will improve the dynamic of the latter. This is the underlying idea of the recent univariate models proposed by Shephard and Sheppard (2010) and Hansen et al. (2012). There is however an issue if one only specifies the dynamic of the conditional variance without explicitly modelling the evolution of the daily realized variances. That is, it is not possible to forecast at an horizon greater than one-period ahead. Therefore, as in Shephard and Sheppard (2010) and Hansen et al. (2012), we also propose a specification for the daily measures of realized volatility.
2.2 The Realized Variances

First, the realized covariances are assumed to follow a Wishart distribution with $\nu$ degree of freedom and time varying scale matrix $\Sigma_t$. The implication of that assumption is that $E(X_t|\mathcal{F}_{t-1}) = \Sigma_t$. Furthermore, the scale matrix can also be decomposed into standard deviations and correlations matrices. Finally, we extend the idea of the Regime Switching Dynamic Correlation of Pelletier (2006) to the correlation matrix implied by $\Sigma_t$.

More precisely we have the relations:

$$X_t \sim W_n(\nu, \Sigma_t/\nu)$$
$$\Sigma_t = \tilde{D}_t \Gamma_t \tilde{D}_t$$
$$\Gamma_t = \sum_{i=1}^{M} \Gamma^{(i)} \mathbb{1}_{\{s_t = i\}}, \quad \Gamma^{(i)} \neq \Gamma^{(j)} \quad \forall \quad i \neq j,$$

where $\tilde{D}_t$ is diagonal and holds the square root of the conditional mean of the realized variances. The correlation matrix $\Gamma_t$ follows a regime switching dynamic with $M$ regimes. The unobserved variable $s_t$ represents the state at time $t$ and the indicator function $\mathbb{1}_{\{s_t = i\}}$ is equal 1 when the condition $\{s_t = i\}$ is true or 0 otherwise. Let us denote by $\Pi$ the associated transition probability matrix and by $\pi_{i,j}$ the element on row $j$ and column $i$ of $\Pi$. Then, $\pi_{i,j}$ is the probability of transition from regime $i$ in period $t$ to regime $j$ in period $t+1$ and $\pi_i$ is the limiting probability of regime $i$. The standard assumptions on the Markov chain (aperiodicity, irreducibility and ergodicity) are made, see Ross (2006, Chapter 4).

Notice that the conditional correlation matrix indeed provides a link at time $t$ between the realized and conditional covariances. The conditional mean of the realized variance is modeled as a linear function of the conditional variance which is $\mathcal{F}_{t-1}$ measurable:

$$x_{i,t} = \hat{h}_{i,t} \epsilon_{i,t}$$
$$\hat{h}_{i,t} = \gamma_{i}^{(0)} + \gamma_{i}^{(1)} h_{i,t},$$

where $\hat{h}_{i,t}$ is the conditional mean of the realized variance. The multiplicative error term, $\epsilon_{i,t}$ is Gamma distributed and has a mean of 1. More precisely $\epsilon_{i,t}$ is $Ga(\nu/2, 2/\nu)$. This is a consequence of the assumption on the distribution of the realized covariance matrices. Given (8), let $\tilde{D}_t = diag(\sqrt{\hat{h}_{1,t}}, \sqrt{\hat{h}_{2,t}}, ..., \sqrt{\hat{h}_{n,t}})$. The following decomposition is obtained:

$$X_t = \tilde{D}_t \tilde{S}_t \tilde{D}_t,$$
$$\tilde{S}_t \sim W_n(\nu, \Gamma_t/\nu)$$

So by the properties of the Wishart distribution, the $i^{th}$ diagonal element of $X_t$ has a Gamma marginal distribution with time varying scale parameter equal to $\hat{h}_{i,t}$. The realized variance $x_{i,t}$ is equal to $\hat{h}_{i,t}$ times the element $(i, i)$ of a random matrix $\tilde{S}_t$. Finally, $E[\tilde{S}_t | \Gamma_t] = \Gamma_t$. 

Page 5


2.3 Comparison With Existing Models

To our knowledge, the R-RSDC is unique in the literature on joint modeling of returns and realized covariances because no other specifications model the correlations as regime dependent. However, we borrow from the multivariate stochastic volatility literature by assuming a Wishart distribution for the realized covariance matrices. The estimation method, explained below, is different because the Maximum Likelihood instead of the Bayesian methodology is implemented. The WAR of Gourieroux et al. (2009) also characterizes the covariance matrices by a Wishart distribution. However, the dynamic of the covariances are captured by a non-centrality parameter. the RCOV of Jin and Maheu (2012) also uses central and non central Wishart distributions but estimates their parameters by Bayesian method. Noureldin and Sheppard (2012) is close to our specification as they also assume the Wishart distribution for the realized covariances. However, the method of estimation and the dynamic is different. Hansen et al. (2014) extend their univariate models. To break the curse of dimensionality, they estimate the relationship between each asset in their sample to a factor (S&P 500) and assume that the correlations depends only on the relation between individual asset and the index.

On a univariate level, relationships between existing models and ours can be established. In the Realized GARCH model of Hansen et al. (2012), the contemporaneous link between the realized variance $x_{i,t}$ and the latent conditional variance $h_{i,t}$ of the daily return (neglecting a leverage effect term) is $x_{i,t} = \gamma_i^{(0)} + \gamma_i^{(1)} h_{i,t} + \epsilon_{i,t}$. Therefore as opposed to an additive error term the error is multiplicative in the R-RSDC. Even though Shephard and Sheppard (2010) also implement a MEM, our distributional assumption is different.

Compared to the regular GARCH specification, the univariate level of the R-RSDC adds the intraday information to the lower frequency (daily) information set. It also adds variation in the persistence of the conditional variance. To see this, equations (5) and (8) can be combined to find an alternative expression for the conditional volatility. Replacing $x_{i,t-1}$ in (5) by its expression yields:

$$h_{i,t} = [\omega_i + a_i \gamma_i^{(0)} \epsilon_{i,t-1}] + [\beta_i + a_i \gamma_i^{(1)} \epsilon_{i,t-1}] h_{i,t-1}$$

(11)

So the measure of persistence of the conditional variance in the R-RSDC is $\beta_i + a_i \gamma_i^{(1)}$. The latter changes with the innovations from the market conditions $\epsilon_{i,t-1}$. This allows the conditional variance to react faster to shock than in a typical GARCH. Given that $E(\epsilon_{i,t}) = 1$, the long run measure of the persistence is $\beta_i + a_i \gamma_i^{(1)}$ and the unconditional mean of the $i^{th}$ conditional and realized variances are:

$$E[h_i] = \frac{\omega_i + a_i \gamma_i^{(0)}}{1 - [\beta_i + a_i \gamma_i^{(1)}]}$$

(12)

$$E[x_i] = \gamma_i^{(0)} + \frac{\gamma_i^{(1)} (\omega_i + a_i \gamma_i^{(0)})}{1 - [\beta + a \gamma_1]}$$

(13)
3 Estimation

The estimation of the parameters by full Maximum Likelihood (ML) would require to evaluate the log-likelihood function given the observations \{Y_t, X_t\}_{t=1}^T:

\[
    \mathcal{L}(\theta; Y, X) = \sum_{t=1}^T \ln f(Y_t, X_t | \mathcal{F}_{t-1}),
\]

where \(Y = \{Y_t\}_{t=1}^T, X = \{X_t\}_{t=1}^T\) and \(\theta\) is the vector of parameters. When the covariance matrices to estimate are very large, maximum likelihood is not feasible. The estimation strategy instead takes advantage of the decomposition of the conditional and realized covariance matrices in expressions (4) and (9). The R-RSC model proposes to break the curse of dimensionality by decentralizing the estimation process. The method is of course less efficient than the full maximum likelihood estimation but yields consistent estimates.

The joint distribution of returns and conditional variances can be expressed as:

\[
    f(Y_t, X_t | \mathcal{F}_{t-1}) = f(Y_t | \mathcal{F}_{t-1}) f(X_t | Y_t, \mathcal{F}_{t-1})
\]

\[
    f(Y_t | \mathcal{F}_{t-1}) = (2\pi)^{-n/2} |H_t|^{-1/2} \exp\left(-\frac{1}{2} Y_t' H_t^{-1} Y_t\right)
\]

\[
    f(X_t | Y_t, \mathcal{F}_{t-1}) = \frac{|X_t|^{\nu/2}}{2^n \nu/2 |\Sigma_t|/\nu/2 \Gamma_n(\nu/2)} \exp\left(-\frac{1}{2} X_t' \Sigma_t^{-1} \right)
\]

\(\Sigma_n(\nu/2)\) is the multivariate gamma function evaluated at \(\nu/2\) and \(\text{etr}\) represents the exponential of the trace function. The probability distribution functions (pdf) in expressions (16) and (17) are used to derive the log-likelihood functions. With \(\theta_1 = \{\omega, \beta, \alpha, \gamma^{(0)}, \gamma^{(1)}\}, \theta_2 = \{\Gamma^{(1)}, \ldots, \Gamma^{(M)}, \pi_{1,1}, \pi_{1,2}, \ldots, \pi_{M,M}\}\) and after replacing the matrices \(X_t, H_t\) and \(\Sigma_t\) by their decompositions, the log-likelihood function can be broken up into two parts (proof in appendix A):

\[
    \mathcal{L}_{step1}(\theta_1) = \sum_{t=1}^T \left\{ -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln h_{i,t} - \frac{1}{2} \frac{y_{i,t}^2}{h_{i,t}} \right\}
    + \left\{ \frac{\nu}{2} \ln(\nu/2) - \ln \Gamma(\nu/2) - \frac{\nu}{2} \ln(\tilde{h}_{i,t}) + \frac{\nu-2}{2} \ln x_{i,t} - \frac{\nu x_{i,t}}{2 \tilde{h}_{i,t}} \right\}, \quad \forall i = 1, 2, ..., n
\]

(18)

and

\[
    \mathcal{L}_{step2}(\theta_2) = \sum_{t=1}^T \left\{ \frac{1}{2} \tilde{U}_t' \tilde{U}_t - \frac{1}{2} \ln |\Gamma_t| - \frac{1}{2} \tilde{U}_t' \Gamma_t^{-1} \tilde{U}_t \right\}
    + \left\{ \frac{1-n}{2} \ln |\tilde{V}_t| - \frac{\nu}{2} \ln |\Gamma_t| + n \ln \Gamma(\nu/2) - \ln \Gamma_n(\nu/2) + \frac{\nu}{2} \text{tr}(\tilde{S}_t[I_n - \Gamma_t^{-1}]) \right\}
\]

(19)
where, $y_{i,t}$ is the $i^{th}$ element of $Y_t$ in (2) and $\hat{U}_t = D_t^{-1/2} Y_t$. In a first step, the estimation of the parameters of the conditional and realized variances is achieved individually for each asset by maximizing (18). In a second step, and conditional on the first step estimates, we estimate the states, the transition probabilities $P_i$ and the correlation matrices $\Gamma^{(j)}, \forall j = 1, 2, \ldots, M$.

### 3.1 Univariate Models Estimation

The univariate parameters are estimated by maximizing (18). With the estimated parameters, the sequence of matrices and vectors $\{\tilde{S}_t\}_{t=1}^T$ and $\{\tilde{U}_t\}_{t=1}^T$ are constructed.

Notice that in the first step, $\nu$ has no effect on the estimation of the parameters because the univariate likelihood $(\mathcal{L}_{step1}(\theta_1))$ is a combination of Normal and Gamma distribution. The Gamma part, where $\nu$ appears, can be written as:

$$C_{x,\nu} \frac{-\nu}{2} \sum_{i=1}^T \left\{ \ln \hat{h}_{i,t} + \frac{x_{i,t}}{\hat{h}_{i,t}} \right\},$$

where $C_{x,\nu}$ is function of $\nu$ and $\{x_{i,t}\}_{t=1}^T$ only. So the first order conditions involved for the parameters of $\hat{h}_{i,t}$ do not depend on $\nu$. This is, however, not the case in the second step as we see it below.

### 3.2 Regime Switching Model Estimation

In large system, direct maximization of (19) may not be feasible. For instance, if we assume a two-state Markov process for $n = 10$ assets, not only do we need to infer the unobserved states at each time $t$ but also the model implies that there are $2^{n(n-1)} = 90$ unique correlation coefficients to estimate. A solution to this problem is the Expectation-Maximization (EM) algorithm introduced by Dempster et al. (1977). In the following, we show that by using the EM framework to maximize the likelihood we can break the curse of dimensionality in that the correlation matrices $\Gamma^{(1)}, \ldots, \Gamma^{(M)}$ can be estimated through simple weighted sums.

The central concept of the EM algorithm is the auxiliary function $Q(•|•)$ called the intermediate quantity of EM. It is the expected value of the log-likelihood with respect to the unknown states given the data and the current guess of the parameters value. The algorithm consists in building iteratively a sequence $\{\theta_2^{(k)}\}_{k \geq 1}$ of parameter estimates given an initial guess $\theta_2^{(0)}$. For the $k^{th}$ iteration, the steps of the EM algorithm are:

- **Expectation:** Compute $Q(\theta|\theta_2^{(k)})$.
- **Maximization:** Find $\theta_2^{(k+1)}$ such that $Q(\theta|\theta_2^{(k)})$ is maximized.

Given regularity assumptions (Cappé et al. (2005)), the convergence of the iteration achieves local maximization of the the log-likelihood.
Applied to the problem at hand, the EM algorithm consists of the a set of simple updating equations (proof in appendix B). For states, $i, j = 1, ..., M$ there are:

\[
\hat{\pi}_{ij} = \frac{\sum_{t=1}^{T} P(s_t = j, s_{t-1} = i|\tilde{\mathbf{U}}, \tilde{\mathbf{S}}, \theta_2^{(k)})}{\sum_{t=1}^{T} P(s_{t-1} = i|\mathbf{U}, \tilde{\mathbf{S}}, \theta_2^{(k)})},
\]

\[
\hat{f}^{(j)} = \frac{\sum_{t=1}^{T} C_t P(s_t = j|\tilde{\mathbf{U}}, \tilde{\mathbf{S}}, \theta_2^{(k)})}{\sum_{t=1}^{T} P(s_t = j|\mathbf{U}, \tilde{\mathbf{S}}, \theta_2^{(k)})}.
\]

Finally, if the only available data is a time series of realized covariances we have the following updating equations:

\[
\hat{\pi}_{ij} = \frac{\sum_{t=1}^{T} P(s_t = j, s_{t-1} = i|\tilde{\mathbf{S}}, \theta_2^{(k)})}{\sum_{t=1}^{T} P(s_{t-1} = i|\tilde{\mathbf{S}}, \theta_2^{(k)})},
\]

where an appropriate model for the conditional mean of the realized variance is estimated in a first step to construct the sequence of $\tilde{S}$ matrices.

The smoothed, joint and conditional probabilities required in the updating equations are computed by forward-backward inference algorithm as explained in Hamilton (1990) or Kim (1994) and pioneered by Baum et al. (1970) (see appendix C for the details).

The impact of the degree of freedom $\nu$ on the estimation when the number of assets increases can be seen in the updating expression (21). $C_t$ is a weighted average of $\tilde{U}_t$ and $\tilde{S}_t$ with the weights depending on $\nu$. Given that the non-singular Wishart distribution requires that $\nu > n - 1$, on the one hand, the contribution of the standardized returns in the estimation will decrease as the ratio $\frac{1}{\nu + 1}$ get smaller. On the other hand, most of the multivariate information would be provided by the scaled realized correlation $\tilde{S}_t$ as $\frac{\nu}{\nu + 1} \approx 1$.

### 3.3 Asymptotic Distribution of the Two-Step MLE

Murphy and Topel (2002) derive the asymptotic distribution of the Two-step MLE. Let us
3 ESTIMATION

denote by \( \frac{1}{T} \gamma_1 \) the asymptotic covariance matrix of the first step estimators. Furthermore, let \( \frac{1}{T} \gamma_2 \) be the asymptotic covariance matrix of the second step estimators without accounting for the first step. \( \frac{1}{T} \gamma_2 \) must be corrected for the fact that an estimate of \( \theta_1 \) is used in the estimation of \( \theta_2 \) instead of the true \( \theta_1 \). The corrected form is \( \frac{1}{T} \gamma_2^* \) where \( \gamma_2^* \) is defined as:

\[
\gamma_2^* = \gamma_2 + \gamma_2 (E \gamma_1 - E \gamma_1 R') \gamma_2
\]

where

\[
E = E \left( \begin{array}{cc}
\frac{\partial L_{\text{step2}}}{\partial \theta_2} & \frac{\partial L_{\text{step1}}}{\partial \theta'_1}
\end{array} \right)
\]

In finite sample, the estimates of \( \gamma_1, \gamma_2, E \) and \( R \) are:

\[
\hat{\gamma}_1 = \left[ \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\partial L_{\text{step1}}}{\partial \theta_1} \right) \left( \frac{\partial L_{\text{step1}}}{\partial \theta'_1} \right) \right]^{-1}
\]

\[
\hat{\gamma}_2 = \left[ \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\partial L_{\text{step2}}}{\partial \theta_2} \right) \left( \frac{\partial L_{\text{step2}}}{\partial \theta'_2} \right) \right]^{-1}
\]

\[
\hat{E} = \sum_{t=1}^{T} \left( \frac{\partial L_{\text{step2}}}{\partial \theta_2} \right) \left( \frac{\partial L_{\text{step1}}}{\partial \theta'_1} \right)
\]

\[
\hat{R} = \sum_{t=1}^{T} \left( \frac{\partial L_{\text{step2}}}{\partial \theta_2} \right) \left( \frac{\partial L_{\text{step1}}}{\partial \theta'_1} \right)
\]

3.4 Multistep Ahead Forecast

Unfortunately, as in the DCC of Engle (2002a), the non-linearity of our specification precludes from the derivation of a closed form expression for the multistep ahead forecast. However, the assumption of Markov regime switching allows to forecast the correlation matrices independently of the variances. We also can easily forecast the individual conditional and realized variances. So the following forecasting scheme is proposed:

\[
E_T[H_{T+L}] = H_{T+L|T} + D_{T+L|T} \Gamma_{T+L|T} D_{T+L|T}, \quad \forall \quad L \geq 1
\]

where \( D_{T|T+L} \) holds the square root of the \( L \)-step ahead forecast of the conditional variances and \( \Gamma_{T|T+L} \) is the \( L \)-step ahead forecast of the correlation matrix.

3.4.1 Conditional and Realized Variances Forecast

Given the specification (5) and (8), the \( L \)-step ahead forecast of the conditional variance is (See appendix 4 for proof):

\[
E_T[h_{i,T+L}] = h_{i,T+L|T} = \sum_{j=0}^{L-1} \omega_j \beta_j^1 + \beta_0^1 h_{i,T} + \sum_{k=0}^{L-1} \alpha_l \beta^k_i x_{i,T+L-1-k|T}, \quad \forall \quad L \geq 1
\]

where \( x_{i,T+L-1-k|T} = E_T x_{i,T+L-1-k} = \gamma_0 + \gamma_1 E_T h_{i,T+L-1-k|T} \). The forecast is therefore recursive.

The \( L \)-period forecast of the conditional variance given information at \( T \) will be non trivially dependent on the forecast of the realized variance. Typically, it is expected that \( \alpha_i \beta_l < 1 \)
so the effect of past forecast of the realized variance should decrease with the number of periods to forecast. Similarly, \( \beta_1 < 1 \) and the longer the horizon is, the smaller the effect of the conditional variance at time \( T \) will be. It is also possible to generate \( E_T[\sqrt{h_{i,T+L}}] \) by simulation to avoid the Jensen’s Inequality \( E_T[\sqrt{h_{i,T+L}}] \neq \sqrt{E_T[h_{i,T+L}]} \).

### 3.4.2 Correlations Forecast

Given the properties of the Markov chain and using the notation introduced in appendix (D), the \( L \)-step ahead forecast of the correlation matrix is:

\[
E_T \Gamma_{T|T+L} = \Gamma_{T|T+L} = \sum_{s_{T+L}=l}^{M} \hat{\xi}^{(i)}_{T+L}/T \Gamma_{j}^{(i)}_{T+L}, \quad \forall i = 1, ..., M
\]  

(30)

where \( \hat{\xi}^{(i)}_{T+L}/T \) is the \( i^{th} \) row of:

\[
\hat{\xi}^{(i)}_{T+L}/T = \Pi^L \hat{\xi}^{(i)}_{T}/T
\]  

(31)

### 4 Multivariate Realized Kernel Computation

As it is common in the literature, we assume that the vector of intraday efficient log prices is modeled as a Brownian semimartingale defined on a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t), P)\):

\[
P(t) = \int_0^t a(s) ds + \int_0^t \sigma(s) dB(s)
\]  

(32)

Where \( a(t) \) and \( \sigma(t) \) are the predictable processes for the drift and the instantaneous volatility matrices. \( B(t) \) is a vector of independent Brownian motions. The integrated covariance process is then

\[
[P](t) = \int_0^t \sigma(s)\sigma(s)' ds
\]  

(33)

Consider the quadratic covariation, \( \nu_n(t) = \sum_{i=0}^{n-1}(P(t_{i+1}^n) - P(t_{i+1}^n))(P(t_{i+1}^n) - P(t_{i+1}^n))' \) for a partition \( \{t_i^n\} \) of \([0, t]\). It can be shown that \( \nu_n(t) \) converges in probability to \([P](t)\). See for example Klebaner (2005, Section 8.5) and Protter (2004, Chapter 2, Section 6). Therefore by increasing the sampling frequency over a day worth of data we can approximate the integrated covariance by the quadratic covariation. However, microstructure noise renders imprecise the estimation of the realized (co)volatilities when the sampling frequency increases (Barndorff-Nielsen et al. (2011)). The market microstructure effects cause the observed price to be the noisy measurement of the efficient price. Examples of microstructure noise
include bid-ask bounces and discreteness of the price. If we denote the noise as $MN$ then the observed price at a given $\tau_i$ is:

$$\tilde{P}(\tau_i) = P(\tau_i) + MN(\tau_i)$$

(34)

In the equation above, $\tilde{P}(\tau_i)$ is the observed price. The semi-martingale properties of the returns is not satisfied for the observed high frequency prices because of the microstructure noise. The quadratic covariation is no longer a consistent estimator for the latent Integrated covariances. Hansen and Lunde (2006) have found that the market microstructure noise is:

1. Correlated with the efficient price
2. Time-dependent
3. Small for DJIA stocks

In a multivariate framework, the measurement of realized covariance is complicated further because the data is irregularly spaced and non synchronous causing a bias toward zero of the covariances, the so called Epps effect [see Epps (1979)].

We use the procedure proposed by Barndorff-Nielsen et al. (2011), to construct the matrices of realized covariances that correct for non synchronous transactions and presence of noise. The multivariate realized kernel are computed as:

$$X_t = \gamma_0 + \sum_{h=1}^{Q} K\left(\frac{q}{Q}\right) (\gamma_q + \gamma'_q)$$

(35)

Where, $K(q/Q)$ is the Parzen kernel function$^1$, $Q$ is the bandwidth and $\gamma_q$ the auto covariance at $q$ lags. The data cleaning process, the synchronization of prices and the bandwidth selection for each day in the sample are performed following the prescriptions of Barndorff-Nielsen et al. (2009) and Barndorff-Nielsen et al. (2011).

5 Empirical Illustration

The model is applied to high-frequency prices for seven assets from the financial sector of the S&P 100 $^2$. The main focus of the present section is to illustrate the ability of the R-RSDC to model the correlations and the conditional variances. The sample period runs from January 3, 2006 to March 30, 2012 corresponding to a sample size of 1,593. However, days for which the daily transaction period is less than the normal 6.5 hours are removed$^3$. The final sample size is 1,581 days.

Often high-frequency data contains a considerable number of errors in recording, therefore data cleaning is an important preliminary step of a volatility estimation based on intraday data.

$^1K(x) = 1 - 6x^2 + 6x^3$ if $0 \leq x \leq 0.5$, $= 2(1 - x)^3$ if $0.5 \leq x \leq 1$, $= 0$ if $x > 1$.

$^2$The ticker symbols of the assets are AXP, BAC, COF, GS, JPM, MS and USB

$^3$Independence Day, the days before Thanksgiving and Christmas
5.1 Procedure for Cleaning the High frequency Data

The data cleaning process follows the recommendations of Barndorff-Nielsen et al. (2009). The analysis is restricted only to the regular trading hours, that is from 9:30am to 4:00pm EST. Furthermore, entries with corrected trades are deleted. The transactions with a sale condition equal to Z are also excluded because they are reported to the tape at a later time than their occurrences. Multiple transactions with the same time stamp are aggregated and the median price is used as the aggregated price.

5.2 Data description

5.2.1 Realized Volatility Description

The computation of the multivariate realized kernel starts by the synchronization of the high-frequency prices for each day in the sample. The sampling method introduced by Barndorff-Nielsen et al. (2011) is called the “refresh sampling” and define the time clock for the \( i \)th trade as the time at which the last asset in the sample is traded for that \( i \)th trade. More precisely, if \( \tau_i \) is the refresh time then:

\[
\tau_1 = \max(t_{1}^{(1)}, t_{1}^{(2)}, \ldots, t_{1}^{(n)})
\]

and subsequent refresh times are

\[
\tau_{j+1} = \max(t_{N_j+1}^{(1)}, t_{N_j+1}^{(2)}, \ldots, t_{N_j+1}^{(n)})
\]

where \( N_j \) is a count of the transactions up to refresh time \( \tau_j \). A measure of the data retention by the refresh time sampling method is

\[
p = \frac{N}{\sum_{j=1}^{n} n_j}
\]

where \( n \) is the number of assets, \( n_j \) the initial number of transactions for the \( j \)th asset and \( N \) is the common number of transactions selected after synchronization. Figure 1 presents the evolution of the retention percentage for the sample period. On average 36.98% of the daily data is retained. The minimum retention is 22.48% while the maximum is 79.93%. It is worthwhile to notice that the maximum retention percentage occurred on September 18, 2008\(^4\), when all the stocks in the sample were actively traded. Table 1 shows the statistics for the percentage of retention per asset in the sample. The percentages are computed as the ratio of the number of daily transactions selected over the total of daily transactions for a particular asset. The sampling percentage can be as low as 9.66\(^5\) and as high as 88%.

An important input of the multivariate realized kernel estimation is the bandwidth. As recommended by Barndorff-Nielsen et al. (2009) and Barndorff-Nielsen et al. (2011), we apply a univariate optimal mean square error bandwidth selection to each asset. On a single day, the bandwidth selected for the kernel estimation is the median of the bandwidths estimated for each asset. The minimum, maximum and mean of these unique daily bandwidths are respectively 29, 141 and 60. The bandwidth also corresponds to the number of autocovariances included in the computation of the realized matrices because the Parzen kernel is the weight function.

The summary statistics for the annualized realized volatilities are reported in the second part of Table 2. The annualized realized volatility is the square root of 252 times the real-

---

\(^4\)That day, in all the major financial markets, swap lines were expanded by $180 Billion enabling central banks to lend dollars into their domestic banking system.

\(^5\)There were 17,790 total transactions for BAC on 12/27/2010, only 1,7186 transactions were selected. The reason may be the cyber attack targeted at Bank of America on that day.
ized kernel estimates. Morgan Stanley (MS) is on average the most volatile of all the assets with the highest average (42.38%) and the highest standard deviation (43.64%). Figures 2 and 3 show the behavior of the annualized realized volatility for the entire sample period. The financial crisis of 2008-2009 is noticeable midway in the sample as there is increased volatility. The period starting from September 1, 2008 to June 1, 2009 is depicted in Figures 4 and 5. Notice that the maximum values of the annualized realized volatility occur in this specific time frame.

The averages and standard deviations of the realized correlations of the pairs of assets are presented in Table 3. The mean of the correlations oscillates between 0.46 and 0.62 while the standard deviations are between 0.12 and 0.19. The last two columns of Table 3 present the maximum and minimum realized correlation along with the day they occur. The majority of the lowest correlations dates (17 out of 21) are before the financial crises of 2008. While, except from the correlation between AXP and GS, the highest values of realized correlations are recorded between 2008 and 2011.

5.2.2 Daily Log-Return Description

The lower frequency information is the daily log close-to-close returns. The statistics are presented in the first part of Table 2. As expected, the averages are in the neighbourhood of 0% except for BAC. It appears also that MS is the most volatile stock in the sample even at the daily frequency. The range of the return is large. For instance the minimum daily return for BAC is $-33.79\%$ and the maximum is 31.66%.

5.3 Estimation Results

This subsection presents the results of the estimation as introduced in section 3. For illustration purpose, it is assumed that there are two regimes ($\mathbb{M} = 2$) with respectively high and low correlations. With a convergence criteria for the second step set at $10^{-8}$, the local maximum is attained after 33 iterations. The individual estimation of the univariate Realized GARCH models, which are based on a combination of a Gaussian and a Gamma term, are presented in Tables 4 and 5. The estimates of the unique coefficients of correlation for the different regimes and the transition probabilities are presented in Tables 6, 7, 8 and 9.

5.3.1 Step 1: Univariate Results

The autoregressive parameters in the conditional variance specification ($\beta_i$) is considerably less than it would be under a regular GARCH model. The estimated coefficients are in the range of 0.36 to 0.50. The typical value in a GARCH specification is between 0.95 and 0.99 which is indicative of the strong persistence of the conditional variance. However a direct comparison of $\beta_i$ with the GARCH parameter is not correct as shown in Section 2.3. The appropriate R-RSDC counterpart is $\beta_i + \alpha_i g_i^{(1)}$ which can be found in Table 5. Notice that the actual measure of persistence in our model is time changing because of the innovation term appearing in the expression $\beta_i + \alpha_i g_i^{(1)} \epsilon_{i,t-1}$. Table 5 is showing only the averages. We can see that the values are very similar to what a GARCH would predict.
The effect of the lag realized volatility is in the range of 0.65 to 0.92. Hansen et al. (2012) have found a similar value for the SPY index (0.87). The estimates of \( \gamma^{(1)} \) are less than one, suggesting that the variance measured while the market is open corresponds to between 56% and 78% of the daily variance.

As for the estimates of the shape parameter \( \nu \), they vary between 6.35 and 8.87 with a median value of 8.3955. A non singular Wishart distribution requires that the degree of freedom be greater than the number of assets minus one. In our case, the degree of freedom must be greater than 6. For the second step therefore, the unique shape parameter is set to 8.3955, the median first step estimates.

5.3.2 Step 2: Multivariate Results

The sequence of scaled covariance matrices \( \{\tilde{S}_t\} \) and the standardized return vectors \( \{\hat{U}\} \) are constructed given the first step parameters estimates. As evidenced in Table 8, the difference in correlation between the two regimes is between 15.85% and 44.13%. Table 9 shows that the regime of high correlation is slightly more persistent. Figures 6 and 7 show the evolution of the probabilities (smooth and filtered) of the occurrence of regime 1 and 2 for the entire sample period. Except for brief periods, the regime constantly switches between regime 1 and 2. One of these exceptions is the financial crisis of 2008-2009 where the level of correlation appears higher (regime 2). Figures 8 and 9 reproduce the probability dynamic but only for the period of September 1, 2008 to June 1, 2009. We can see that large portions in the month of October 2008 to April 2009 belong to the regime of high correlation. This suggests that during period of high volatility, the financial assets in the sample tend to behave uniformly. For instance, the evolution of the annualized realized volatility along with the realized, conditional correlations and covariances for BAC and JPM are presented in Figures 10 and 11. It is noticeable that in Figure 10, the period of financial crisis is characterized by high volatilities and covariances. The picture is clear when we zoom in the period in Figure 11. The behavior of JPM and BAC realized measures of volatility behave almost identically from September to December 2008. The conditional covariance implied by the R-RSDC specifications follows closely the erratic realized covariance.

5.3.3 Model Fit

To assess the fit of the model, the multivariate Ljung-Box portmanteau (or HM) test of Hosking (1980) is performed on the standardized residuals. The test is used to detect any residual ARCH effects. It has the form:

\[
HM(W) = T^2 \sum_{j=1}^{W} (T - j)^{-1} \text{tr}\{A_{\hat{z}_t}^{-1}(0)A_{\hat{z}_t}(j)A_{\hat{z}_t}^{-1}(0)A_{\hat{z}_t}(j)\}
\]

where \( \hat{z}_t = \text{vec} h(D^{-1/2} Y_t Y_t' D^{-1/2}) \) and \( A_{\hat{z}_t}(j) \) is the sample auto covariance matrix of order \( j \). Under the null hypothesis of no ARCH effects, the test is distributed asymptotically as Chi-square with degree of freedom \( W \times n^2 \).
With $W = 20$ lags, the HM statistics is 998.3220. Given that the 5% critical value is 1053.9, there is no sign of ARCH effects in the residual\(^6\).

## 6 Out-of-sample Model Evaluation

The goal of the analysis is to show the advantage of the high-frequency information in the forecasting of conditional covariance matrices. Therefore, our model is compared to a multivariate specification that uses only daily information. Notice that a benchmark specification that uses high frequency information as well would be the HEAVY of Noureldin and Sheppard (2012). To conduct the forecast analysis, the sample is split into two parts. The in-sample for the initial estimation runs from January 4, 2006 to April 29, 2011 and has a size of 1329. With a size of 251 (one year), the out-of-sample spans from May 2, 2011 to April 30, 2012.

The R-RSDC model is closed in its design to DCC model of Engle (2002) because it splits the covariances into standard deviations and correlations. Therefore to assess the gain from the proposed model, the DCC is the competing model. The parameters of both models are updated daily and the horizon considered for the predictions are 1, 5, 10 and 20. These correspond respectively to one day, one week, two weeks and one month ahead forecast.

The pair-wise comparison of the volatility forecast uses the Diebold and Mariano (2002) and West (1996) test (DMW) given a loss function. The quasi-likelihood (QLIK) is the chosen loss function:

$$L_{t,L}(H^0_{t+L}, H^i_{t+L|t}) = \ln |H^i_{t+L|t}| + \text{tr}\left(\left(H^i_{t+L|t}\right)^{-1} H^0_{t+L} \right), \quad \forall \ i = 1, 2, \quad (36)$$

where the index $i = 1, 2$ represents respectively the DCC and R-RSDC. $H^0_{t+L}$ is the true and unobserved covariance matrix while $H^i_{t+L|t}$ is the $L$-step-ahead forecast generated by the $i^{th}$ model. The QLIK function is robust in the sense of Patton and Sheppard (2009) and Laurent et al. (2013) because it preserves statistical ordering regardless of the proxy of the latent covariance matrix.

Given that $H^0_{t+L}$ is unobserved a proxy is constructed by scaling the multivariate realized kernel. The scaling is required because the realized kernel measures the volatility while the market is open (6.5 hours) and the forecast objective is the daily variance and covariance. For each univariate forecast of the conditional variance, the scale factor is:

$$\zeta_i = \frac{\sum_{t=1}^{T} y_{i,t}^2}{\sum_{t=1}^{T} x_{i,t}}, \quad \forall \ i = 1, \ldots, n$$

where $\tilde{T}$ is the size of the estimation sample. One effect of the scaling is that the sum of the scaled daily realized variance is equal to the sum of the square daily returns. The overall realized covariance matrix is constructed with the scaled realized variance and the realized

---

\(^6\)A 20 lags HM test on $Y_t$ was initially performed and the Test statistics was 3,240.3
correlation. The realized correlation are not transformed because it is assumed that the overnight effect on the correlation is negligible.

With a proxy for the latent covariance matrix and the forecasted conditional covariance matrices, the loss is calculated for each model and the loss differential is computed as:

\[ d_{t,L} = L_{t,L}(H_{t+L}^{DCC}, H_{t+L|t}^{R-RSDC}) - L_{t,L}(H_{t+L}^{R-RSDC}) \]

The direction of the difference implies that a positive number of \( d_{t,L} \) is the sign that the R-RSDC performs better than the DCC in term of the loss function. With the time series of \( d_{t,L} \), we test for all \( L \):

- \( H_0 : E[L_{t,L}(H_{t+L}^{DCC}, H_{t+L|t}^{R-RSDC})] = E[L_{t,L}(H_{t+L}^{R-RSDC})] \)
- \( H_1 : E[L_{t,L}(H_{t+L}^{DCC}, H_{t+L|t}^{R-RSDC})] > E[L_{t,L}(H_{t+L}^{R-RSDC})] \)
- \( H_2 : E[L_{t,L}(H_{t+L}^{DCC}, H_{t+L|t}^{R-RSDC})] < E[L_{t,L}(H_{t+L}^{R-RSDC})] \)

Therefore the DMW tests the null of equal predictive accuracy against the alternative that one of the model performs better. The test is computed using a standard t-test after accounting for serial correlation. Under the null, Diebold and Mariano (2002) show that the test is asymptotically Normally distributed. Practically, the differential is regressed on a constant and a t-test is constructed with heteroskedasticity and autocorrelation consistent estimate of the standard errors.

Comparison of forecast in a multivariate context has the additional issue of interpretation. The reason is the difficulty to separate the effect of the forecasted variances and covariances on the predictive ability of the model. Noureldin and Sheppard (2012) propose to decompose the QLIK in (36) into marginal and copula-style contribution. The DMV test is performed for each asset in the sample and the copula-like term is the contribution of the correlation forecast on the predictive ability of the model. We perform such a test also and the results are reported in Table 10.

The results show that the R-RSDC outperforms the DCC. The univariate variance model has clear superior predictive ability, especially in shorter horizon. Except for BAC, all the tests in the one-step ahead case reveal that the R-RSDC is more accurate than the DCC with a 95% confidence level or more. The DMV test for the correlation forecasts shows a clear advantage over the DCC even at longer horizon. As a results, with a confidence level of 99%, the multivariate forecast is more accurate in all the horizons, even though on the margin the gain from the univariate model is mixed in longer horizon. This suggests that the high frequency information incorporated in the model increases the predictive ability of the conditional correlation.
7 Conclusion

The paper introduced the Realized Regime Switching for Dynamic Correlations (R-RSDC) model for vectors of daily returns and corresponding realized measures. This model builds on the univariate Realized GARCH of Hansen et al. (2012), the DCC of Engle (2002a) and the RSDC of Pelletier (2006).

The Realized GARCH framework allows to specify a modified regressive equation for the conditional variance. The decomposition of the realized and conditional covariance matrices permits to specify a univariate measurement equation for each asset. The inference of regime switching correlation matrices exploits the daily information and the high frequency realized correlation matrices.

The curse of dimensionality is dealt with by the DCC-like decomposition of the matrices of covariance. The realized covariance matrices considered are the multivariate realized kernel of Barndorff-Nielsen et al. (2011). The estimation is broken down into two steps where the first one consists of estimating the parameters of univariate models. The second step uses the EM algorithm introduced by Dempster et al. (1977).

In a illustrative empirical application, we estimated the model for a dataset of seven assets. The results show that we are able to infer with accuracy the state of the market as there is increased in correlation during period of crisis. In an out-of-sample comparison with the DCC, the proposed model has a higher level of accuracy because of the additional high frequency information. The marginal (asset by asset) accuracy level deteriorates slightly with the increase of the forecast horizon. However, the correlation forecast contribution is considerable even at our largest horizon of prediction (twenty-two days).

The model as introduced does not account for asymmetry in the univariate specification. The R-RSDC also assumes multivariate normal distribution. For future research, the model can be extended to include asymmetry and fat tailed distribution.

Finally, the assumed distribution for the realized covariance matrix implies that the degree of freedom must be greater than the number of asset minus one. The likelihood in our first step estimation also implies that the estimation of the degree of freedom has no effect on the estimation of the univariate parameters. However the expressions for the updating equations for the EM algorithm suggest that the degree of freedom will affect considerably the contribution of the lower frequency information. We postpone future work on the exact effect of the degree of freedom on the R-RSDC.
Table 1: Summary Statistics for refresh sampling: January, 03 2006 to March, 30 2012

<table>
<thead>
<tr>
<th>Asset</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXP</td>
<td>46.25</td>
<td>14.63</td>
<td>87.94</td>
</tr>
<tr>
<td>BAC</td>
<td>26.42</td>
<td>9.66</td>
<td>76.03</td>
</tr>
<tr>
<td>COF</td>
<td>54.20</td>
<td>18.42</td>
<td>88.06</td>
</tr>
<tr>
<td>GS</td>
<td>35.76</td>
<td>14.83</td>
<td>75.64</td>
</tr>
<tr>
<td>JPM</td>
<td>30.33</td>
<td>15.72</td>
<td>77.05</td>
</tr>
<tr>
<td>MS</td>
<td>38.99</td>
<td>18.31</td>
<td>75.68</td>
</tr>
<tr>
<td>USB</td>
<td>48.16</td>
<td>23.58</td>
<td>82.95</td>
</tr>
</tbody>
</table>

The numbers are the statistics for the percentages of retention after synchronization.

Table 2: Summary Statistics for the Daily Log Returns and the Annualized Realized Volatilities: January, 03 2006 to March, 30 2012

<table>
<thead>
<tr>
<th>Stocks</th>
<th>Log returns</th>
<th>Realized volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>St. dev</td>
</tr>
<tr>
<td>AXP</td>
<td>0.01</td>
<td>2.93</td>
</tr>
<tr>
<td>BAC</td>
<td>-0.11</td>
<td>4.32</td>
</tr>
<tr>
<td>COF</td>
<td>-0.03</td>
<td>3.88</td>
</tr>
<tr>
<td>GS</td>
<td>-0.01</td>
<td>2.95</td>
</tr>
<tr>
<td>JPM</td>
<td>0.00</td>
<td>3.21</td>
</tr>
<tr>
<td>MS</td>
<td>-0.08</td>
<td>4.34</td>
</tr>
<tr>
<td>USB</td>
<td>0.00</td>
<td>2.91</td>
</tr>
</tbody>
</table>

The log return are in percentage point. The Annualized realized volatility is the square root of 252 times the realized kernel.
Table 3: Summary Statistics for the realized correlation coefficients: January, 03 2006 to March, 30 2012

<table>
<thead>
<tr>
<th>Pairs</th>
<th>Mean</th>
<th>St.dev</th>
<th>Max</th>
<th>Date Max</th>
<th>Min</th>
<th>Date Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXP-BAC</td>
<td>0.49</td>
<td>0.16</td>
<td>0.88</td>
<td>14-Mar-2008</td>
<td>-0.07</td>
<td>13-Mar-2006</td>
</tr>
<tr>
<td>AXP-COF</td>
<td>0.50</td>
<td>0.19</td>
<td>0.91</td>
<td>08-Jun-2010</td>
<td>-0.17</td>
<td>23-Jan-2007</td>
</tr>
<tr>
<td>AXP-GS</td>
<td>0.47</td>
<td>0.16</td>
<td>0.84</td>
<td>30-Mar-2007</td>
<td>-0.13</td>
<td>21-Dec-2010</td>
</tr>
<tr>
<td>AXP-JPM</td>
<td>0.51</td>
<td>0.17</td>
<td>0.89</td>
<td>12-Aug-2011</td>
<td>-0.14</td>
<td>02-Nov-2010</td>
</tr>
<tr>
<td>AXP-MS</td>
<td>0.47</td>
<td>0.17</td>
<td>0.84</td>
<td>14-Sep-2011</td>
<td>-0.27</td>
<td>21-Apr-2011</td>
</tr>
<tr>
<td>AXP-USB</td>
<td>0.48</td>
<td>0.19</td>
<td>0.91</td>
<td>14-Mar-2008</td>
<td>-0.40</td>
<td>13-Jul-2007</td>
</tr>
<tr>
<td>BAC-COF</td>
<td>0.46</td>
<td>0.17</td>
<td>0.87</td>
<td>25-Feb-2008</td>
<td>-0.17</td>
<td>19-Jan-2007</td>
</tr>
<tr>
<td>BAC-GS</td>
<td>0.49</td>
<td>0.15</td>
<td>0.85</td>
<td>01-Nov-2011</td>
<td>-0.11</td>
<td>19-Sep-2006</td>
</tr>
<tr>
<td>BAC-JPM</td>
<td>0.59</td>
<td>0.15</td>
<td>0.91</td>
<td>13-Mar-2008</td>
<td>-0.18</td>
<td>25-Jan-2007</td>
</tr>
<tr>
<td>BAC-MS</td>
<td>0.50</td>
<td>0.15</td>
<td>0.86</td>
<td>04-Mar-2008</td>
<td>-0.20</td>
<td>25-Apr-2011</td>
</tr>
<tr>
<td>BAC-USB</td>
<td>0.54</td>
<td>0.17</td>
<td>0.93</td>
<td>25-Feb-2008</td>
<td>-0.11</td>
<td>17-Oct-2006</td>
</tr>
<tr>
<td>COF-GS</td>
<td>0.44</td>
<td>0.17</td>
<td>0.81</td>
<td>28-Sep-2011</td>
<td>-0.12</td>
<td>13-Mar-2006</td>
</tr>
<tr>
<td>COF-JPM</td>
<td>0.48</td>
<td>0.18</td>
<td>0.90</td>
<td>26-Aug-2011</td>
<td>-0.20</td>
<td>17-Mar-2006</td>
</tr>
<tr>
<td>COF-MS</td>
<td>0.44</td>
<td>0.18</td>
<td>0.85</td>
<td>01-Sep-2011</td>
<td>-0.31</td>
<td>19-Mar-2007</td>
</tr>
<tr>
<td>COF-USB</td>
<td>0.46</td>
<td>0.19</td>
<td>0.88</td>
<td>26-Aug-2011</td>
<td>-0.34</td>
<td>16-Feb-2007</td>
</tr>
<tr>
<td>GS-JPM</td>
<td>0.54</td>
<td>0.15</td>
<td>0.85</td>
<td>12-Aug-2011</td>
<td>-0.05</td>
<td>09-May-2006</td>
</tr>
<tr>
<td>GS-MS</td>
<td>0.62</td>
<td>0.12</td>
<td>0.87</td>
<td>01-Jul-2008</td>
<td>0.10</td>
<td>15-Mar-2006</td>
</tr>
<tr>
<td>GS-USB</td>
<td>0.46</td>
<td>0.17</td>
<td>0.83</td>
<td>13-Mar-2008</td>
<td>-0.22</td>
<td>27-Dec-2006</td>
</tr>
<tr>
<td>JPM-MS</td>
<td>0.54</td>
<td>0.15</td>
<td>0.87</td>
<td>12-Jul-2010</td>
<td>0.01</td>
<td>17-Mar-2006</td>
</tr>
<tr>
<td>JPM-USB</td>
<td>0.56</td>
<td>0.18</td>
<td>0.92</td>
<td>12-Aug-2011</td>
<td>-0.35</td>
<td>20-Apr-2007</td>
</tr>
<tr>
<td>MS-USB</td>
<td>0.47</td>
<td>0.18</td>
<td>0.89</td>
<td>26-Aug-2011</td>
<td>-0.32</td>
<td>27-Dec-2006</td>
</tr>
</tbody>
</table>

Table 4: Parameter estimates for the univariate Realized GARCH models.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>AXP</th>
<th>BAC</th>
<th>COF</th>
<th>GS</th>
<th>JPM</th>
<th>MS</th>
<th>USB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.3831</td>
<td>0.4345</td>
<td>0.3968</td>
<td>0.5090</td>
<td>0.4043</td>
<td>0.3637</td>
<td>0.3750</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.9093</td>
<td>0.9649</td>
<td>0.8584</td>
<td>0.6537</td>
<td>0.8720</td>
<td>0.9259</td>
<td>0.7784</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.6560</td>
<td>0.5686</td>
<td>0.6711</td>
<td>0.6613</td>
<td>0.6546</td>
<td>0.6314</td>
<td>0.7807</td>
</tr>
<tr>
<td>$\nu$</td>
<td>8.7902</td>
<td>6.1575</td>
<td>8.7315</td>
<td>6.3525</td>
<td>8.3948</td>
<td>8.3977</td>
<td>7.2103</td>
</tr>
</tbody>
</table>

The Standard errors are in parenthesis.
Table 5: Persistence of Conditional Variances Implied by the R-RSDC.

<table>
<thead>
<tr>
<th>$\beta_i + \alpha_i I_i^{(1)}$</th>
<th>AXP</th>
<th>BAC</th>
<th>COF</th>
<th>GS</th>
<th>JPM</th>
<th>MS</th>
<th>USB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9796</td>
<td>0.9831</td>
<td>0.9729</td>
<td>0.9413</td>
<td>0.9751</td>
<td>0.9484</td>
<td>0.9826</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Correlation Estimates Regime 1

<table>
<thead>
<tr>
<th></th>
<th>AXP</th>
<th>BAC</th>
<th>COF</th>
<th>GS</th>
<th>JPM</th>
<th>MS</th>
<th>USB</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAC</td>
<td>0.4307(0.0136)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COF</td>
<td>0.4284(0.0107)</td>
<td>0.4102 (0.0137)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS</td>
<td>0.4189(0.0160)</td>
<td>0.4501 (0.0126)</td>
<td>0.3771(0.0173)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JPM</td>
<td>0.4545(0.0131)</td>
<td>0.5680(0.0229)</td>
<td>0.4263(0.0211)</td>
<td>0.5068(0.0123)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.4181(0.0176)</td>
<td>0.4598(0.0207)</td>
<td>0.3789(0.0132)</td>
<td>0.5815(0.0118)</td>
<td>0.4988(0.0225)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USB</td>
<td>0.4124(0.0211)</td>
<td>0.4809(0.0140)</td>
<td>0.3978(0.0260)</td>
<td>0.4090(0.0202)</td>
<td>0.5118(0.0206)</td>
<td>0.4078(0.0224)</td>
<td></td>
</tr>
</tbody>
</table>

The Standard errors are in parenthesis
Table 7: Correlation Estimates Regime 2

<table>
<thead>
<tr>
<th></th>
<th>AXP</th>
<th>BAC</th>
<th>COF</th>
<th>GS</th>
<th>JPM</th>
<th>MS</th>
<th>USB</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAC</td>
<td>0.5796 (0.0162)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COF</td>
<td>0.6123 (0.0126)</td>
<td>0.5699 (0.0138)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS</td>
<td>0.5523 (0.0242)</td>
<td>0.5745 (0.0202)</td>
<td>0.5333 (0.0218)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JPM</td>
<td>0.6092 (0.0158)</td>
<td>0.6820 (0.0272)</td>
<td>0.5936 (0.0279)</td>
<td>0.6288 (0.0136)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.5632 (0.0233)</td>
<td>0.5928 (0.0272)</td>
<td>0.5451 (0.0225)</td>
<td>0.6736 (0.0219)</td>
<td>0.6348 (0.0269)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USB</td>
<td>0.5954 (0.0285)</td>
<td>0.6382 (0.0156)</td>
<td>0.5782 (0.0296)</td>
<td>0.5559 (0.0258)</td>
<td>0.6701 (0.0263)</td>
<td>0.5726 (0.0285)</td>
<td></td>
</tr>
</tbody>
</table>

The Standard errors are in parenthesis

Table 8: Percentage Change in Correlation Between Regime 1 and 2

<table>
<thead>
<tr>
<th></th>
<th>AXP</th>
<th>BAC</th>
<th>COF</th>
<th>GS</th>
<th>JPM</th>
<th>MS</th>
<th>USB</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAC</td>
<td>34.5496</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COF</td>
<td>42.9166</td>
<td>38.9228</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS</td>
<td>31.8445</td>
<td>27.6476</td>
<td>41.4282</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JPM</td>
<td>34.0596</td>
<td>20.0838</td>
<td>39.2275</td>
<td>24.0774</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>34.7222</td>
<td>28.9317</td>
<td>43.8695</td>
<td>15.8311</td>
<td>27.2773</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USB</td>
<td>44.3904</td>
<td>32.6932</td>
<td>45.3574</td>
<td>35.9282</td>
<td>30.9212</td>
<td>40.4127</td>
<td></td>
</tr>
</tbody>
</table>

Table 9: Transition Probabilities Estimates

<table>
<thead>
<tr>
<th></th>
<th>Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\pi}_{11} )</td>
<td>0.8514 (0.0180)</td>
</tr>
<tr>
<td>( \hat{\pi}_{22} )</td>
<td>0.8738 (0.0182)</td>
</tr>
</tbody>
</table>

The Standard errors are in parenthesis
Table 10: Statistical Test of Predictive Ability: Forecast Horizons = 1, 5, 10 and 22 days

<table>
<thead>
<tr>
<th>Margin 1 (AXP)</th>
<th>(1)</th>
<th>(5)</th>
<th>(10)</th>
<th>(22)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margin 1 (BAC)</td>
<td>1.79</td>
<td>0.89</td>
<td>-0.23</td>
<td>0.43</td>
</tr>
<tr>
<td>Margin 1 (COF)</td>
<td>5.37</td>
<td>3.43</td>
<td>0.77</td>
<td>2.64</td>
</tr>
<tr>
<td>Margin 1 (GS)</td>
<td>2.40</td>
<td>1.72</td>
<td>2.49</td>
<td>2.30</td>
</tr>
<tr>
<td>Margin 1 (JPM)</td>
<td>2.35</td>
<td>1.80</td>
<td>1.83</td>
<td>3.23</td>
</tr>
<tr>
<td>Margin 1 (MS)</td>
<td>6.08</td>
<td>1.57</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>Margin 1 (USB)</td>
<td>2.88</td>
<td>0.35</td>
<td>-0.64</td>
<td>0.03</td>
</tr>
<tr>
<td>Correlation Contribution</td>
<td>3.08</td>
<td>2.72</td>
<td>3.60</td>
<td>2.54</td>
</tr>
<tr>
<td>Multivariate test</td>
<td>3.14</td>
<td>2.66</td>
<td>4.54</td>
<td>2.92</td>
</tr>
</tbody>
</table>

Test of Predictive ability: Loss(DCC) - Loss(RSDD). A positive value means the RSDD has better predictive ability. The critical values are: 2.5758 (1 percent), 1.96 (5 percent), 1.6449 (10 percent)
Figure 1: Data Retention Percentages
Figure 3: Annualized Realized Volatility - Asset 5 to 7
Figure 4: Annualized Realized Volatility Detail - Asset 1 to 4: September 1, 2008 to June 1, 2009
Figure 5: Annualized Realized Volatility Detail - Asset 5 to 7: September 1, 2008 to June 1, 2009

- AXP
- BAC
- COF
- GS
Figure 6: Smoothed Probability That the State is in Regime 1 or 2
Figure 7: Filtered Probability That the State is in Regime 1 or 2
Figure 8: Smoothed Probability That the State is in Regime 1 or 2: September 1, 2008 to June 1, 2009
Figure 10: Realized and Conditional co-volatilities and correlation JPM and BAC

Realized volatility BAC

Realized volatility JPM

Correlations: – Realized, – Conditional

Covariances: – Realized, – Conditional
Figure 11: Realized and Conditional co-volatilities and correlation JPM and BAC: September 1, 2008 to June 1, 2009
Appendix A  Proof of Expressions (18) and (19)

The log-likelihood functions in the conditional variance case is:

\[ \mathcal{L}_c = \frac{1}{2} \sum_{t=1}^{T} \left\{ -n \ln(2\pi) - \ln|H_t| - Z_t' H_t^{-1} Z_t \right\} \]

While the realized covariances implies the following:

\[ \mathcal{L}_r = \frac{1}{2} \sum_{t=1}^{T} \left\{ (\nu - n - 1) \ln|X_t| - (n \nu) \ln(2) - \nu \ln|\Sigma_t / \nu| 
- 2 \ln \Gamma_n(\nu/2) - \text{tr}\left(X_t \nu \Sigma_t^{-1}\right) \right\} \]

Now, replacing \( H_t \) by its decomposition in \( \mathcal{L}_c \) we have:

\[ \frac{1}{2} \sum_{t=1}^{T} \left\{ -n \ln(2\pi) - 2 \ln |D_t| - \tilde{O}_t' \tilde{O}_t \right\} + \frac{1}{2} \sum_{t=1}^{T} \left\{ \tilde{O}_t' \tilde{O}_t - \ln |\Gamma_t| - \tilde{O}_t' \Gamma_t^{-1} \tilde{O}_t \right\} \]

where \( \tilde{O}_t = D_t^{-1} Y_t \) is the diagonal matrix of standardized residual returns. \( \tilde{O}_t \) consists of the elements of \( Y_t \) in equation (2) divided by the corresponding square root of the conditional variance. Noticing that \( 2 \ln |D_t| = \sum_{i=1}^{n} \ln h_{i,t} \) and \( \tilde{O}_t' \tilde{O}_t = \sum_{i=1}^{n} \frac{y_{i,t}^2}{h_{i,t}} \) the final expressions is:

\[ \mathcal{L}_c = \mathcal{L}_c(\text{univariate}) + \mathcal{L}_c(\text{multivariate}) \]

\[ \mathcal{L}_c(\text{univariate}) = \frac{1}{2} \sum_{t=1}^{T} \left\{ \sum_{i=1}^{n} -\ln(2\pi) - \ln h_{i,t} - \frac{y_{i,t}^2}{h_{i,t}} \right\} \]

\[ \mathcal{L}_c(\text{multivariate}) = \frac{1}{2} \sum_{t=1}^{T} \left\{ \tilde{O}_t' \tilde{O}_t - \ln |\Gamma_t| - \tilde{O}_t' \Gamma_t^{-1} \tilde{O}_t \right\} \]

Similarly to \( H_t \), the realized covariance matrix can be decomposed into realized standard deviations and correlation matrices:

\[ X_t = V_t S_t V_t \quad (A.1) \]

where \( V_t = \text{diag}(\sqrt{x_{1,t}}, \sqrt{x_{2,t}}, ..., \sqrt{x_{n,t}}) \) and \( S_t \) is the realized correlation matrix. Therefore, \( \{x_{i,t}\}_{t=1}^{T} \) are the daily realized variances for the \( i^{th} \) asset. if \( X_t \) and \( \Sigma_t \) are replaced by their
decompositions, \( \mathcal{L}_r \) becomes:

\[
\mathcal{L}_r = \sum_{i=1}^{r} \left\{ \frac{y-n-1}{2} \ln|V_i| + \frac{n-y-1}{2} \ln|\Sigma_i| + \frac{n}{2} \ln(y/2) - \frac{y}{2} \ln|\tilde{D}_i| \right.

\]

\[
\left. - \frac{y}{2} \ln|\Gamma_i| - \ln|\Sigma_n| - \frac{y}{2} \text{tr}(\tilde{S}_i \Gamma_i^{-1}) \right\}
\]

Upon noticing also that \( 2\ln|\tilde{D}_i| = \sum_{i=1}^{n} \ln \tilde{h}_{i,t} \) and that \( \text{tr}(\nu \tilde{S}_t) = \nu \sum_{i=1}^{n} \frac{\tilde{h}_{i,t}}{\tilde{h}_i} \), the final expression for the log-likelihood function is obtained:

\[
\mathcal{L}_r = \mathcal{L}_r(\text{univariate}) + \mathcal{L}_r(\text{multivariate})
\]

\[
\mathcal{L}_r(\text{univariate}) = \sum_{i=1}^{r} \left\{ \sum_{i=1}^{n} \frac{y}{2} \ln(y/2) - \ln|\Gamma_i| + \frac{y}{2} \ln(\tilde{h}_{i,t}) + \frac{y-2}{2} \ln x_{i,t} - \frac{y}{2} \frac{x_{i,t}}{\mu_{i,t}} \right\}
\]

\[
\mathcal{L}_r(\text{multivariate}) = \sum_{i=1}^{r} \left\{ \frac{1-n}{2} \ln|V_i| - \ln|\Gamma_i| + n \ln|\Gamma_n(y/2)| - \ln|\Sigma_n(y/2)| + \frac{y}{2} \text{tr}(\tilde{S}_i \Gamma_i^{-1}) \right\}
\]

Therefore combining \( \mathcal{L}_r(\text{univariate}) \) and \( \mathcal{L}_r(\text{univariate}) \) yields expression (18). And combining \( \mathcal{L}_r(\text{multivariate}) \) and \( \mathcal{L}_r(\text{multivariate}) \) yields expression (19).

**Appendix B  Proof of equation (20) and (21)**

### 2.1 Notations

Let \( X = \{X_t\}_{t=1}^{T} \), \( Y = \{Y_t\}_{t=1}^{T} \) be the sets of observed realized covariance matrices and daily log returns. Let \( \mathcal{S} = (s_T, s_{T-1}, ..., s_1) \) be the set of unobserved states. We assume that \( s_t \) can take a particular value in \( \{1, 2, ..., M\} \). Additionally, we assume that the transition between states follows a first order Markov chain such that \( \pi_{ji} = P(s_t = j|s_{t-1} = i) \).

The objective is to find the values of the parameters \( \theta_2 = \{\Gamma_1, \Gamma_2, ..., \Gamma_M; \pi_{11}, \pi_{12}, ..., \pi_{MM}\} \) such that the likelihood, \( f(Y, X; \theta_2) \), is maximized. \( \mathcal{S} \) is not observed and we use a hidden markov process framework to estimate the likelihood by the multiple integral \( \int_{x \in \mathcal{S}} f(Y, X; \mathcal{S}; \theta_2) dS \).

To implement the EM - algorithm, first define \( Q(\theta_2^{k+1}|\theta_2^k) \), the intermediate quantity of EM, such that:

\[
Q(\theta_2^{k+1}|\theta_2^k) = \int_{x \in \mathcal{S}} \ln f(X, Y; \mathcal{S}; \theta_2^{k+1}) f(\mathcal{S}|X, Y; \theta_2^k) dS \quad \text{(B.1)}
\]
For the following derivation it is important to notice that:

\[
f(Y, X, \mathcal{Y}; \theta_2) = f(Y_T, X_T | Z_t; \theta_2)p(s_T | s_{t-1}; \theta_2)f(Y_{T-1}, X_{T-1} | Z_{T-1}; \theta_2)p(s_{T-1} | s_{T-2}; \theta_2) \ldots p(Y_2, X_2 | Z_1; \theta_2)p(s_2 | s_1; \theta_2)f(Y_1, X_1 | Z_0; \theta_2) \rho_0
\] (B.2)

The function \( f(Y_t, X_t | Z_{t-1}; \theta_2) \) is the conditional distribution of \( X_t \) at time \( t \) given the information \( Z_{t-1} = (s_{t-1}, s_{t-2}, \ldots, s_1) \) and \( \rho_0 \) is the initial distribution of states.

### 2.2 Proof of equation (20)

To implement the Maximization step of the EM algorithm we maximize \( Q(\theta_2^{k+1} | \theta_2^k) \) with respect to \( \pi_{ij} \) and subject to the condition that \( \sum_{j=1}^{M} \pi_{ij} = 1 \) by introducing a Lagrange multiplier \( \lambda_i \). We have then the maximization problem and the first order conditions:

\[
\mathcal{L} = Q(\theta_2^{k+1} | \theta_2^k) - \lambda_i \left( \sum_{j=1}^{M} \pi_{ij}^{k+1} - 1 \right)
\] (B.3)

From the equation of the expansion of \( f(Y, X, \mathcal{Y}; \theta_2) \) above,

\[
\frac{\partial \ln f(Y, X, \mathcal{Y}; \theta_2^{k+1})}{\partial \pi_{ij}^{k+1}} = \frac{1}{\pi_{ij}^{k+1}} \sum_{t=1}^{T} \mathbb{I}_{(s_t = j, s_{t-1} = i)}
\]

\[
\frac{\partial Q(\theta_2^{k+1} | \theta_2^k)}{\partial \pi_{ij}^{k+1}} = \int_{s \in S} \frac{1}{\pi_{ij}^{k+1}} \sum_{t=1}^{T} \mathbb{I}_{(s_t = j, s_{t-1} = i)} f(\mathcal{Y} | X, Y; \theta_2^k) dS
\]

\[
\frac{\partial Q(\theta_2^{k+1} | \theta_2^k)}{\partial \pi_{ij}^{k+1}} = \frac{1}{\pi_{ij}^{k+1}} \sum_{t=1}^{T} f(s_t = j, s_{t-1} = i | X, Y; \theta_2^k)
\] (B.4)

To obtain (A.4), we use the Fubini theorem and the fact that

\[
\int_{s \in S} \mathbb{I}_{(s_t = j, s_{t-1} = i)} f(\mathcal{Y} | \theta_2^k) dS = f(s_t = j, s_{t-1} = i | X, Y; \theta_2^k). \]

Using (A.4) in (A.3) yields:

\[
\sum_{t=1}^{T} f(s_t = j, s_{t-1} = i | X, Y; \theta_2^k) = \pi_{ij}^{k+1} \lambda_i
\] (B.5)
Summing over all \( j \)'s:

\[
\sum_{j=1}^{M} \sum_{t=1}^{T} f(s_t = j, s_{t-1} = i | Y, X; \theta_2) = \sum_{j=1}^{M} \pi_{ij}^{k+1} \lambda_i
\]

\[\implies \sum_{t=1}^{T} f(s_{t-1} = i | X, Y; \theta_2) = \lambda_i \quad \text{ (B.6)}\]

Using (A.6) in (A.5) to substitute for \( \lambda_i \) and solving for \( \pi_{ij}^{k+1} \) gives equation (20).

### 2.3 Proof of equation (21)

We maximize (A.1) with respect to the parameter \( \theta \in \theta_2 \). We then have the first order condition for the \( i^{th} \) state:

\[
\frac{\partial Q(\theta_2^{k+1}|\theta_2)}{\partial \Gamma_{s_i = i}^{k+1}} = 0 \quad \text{(B.7)}
\]

From (A.2) we have:

\[
\frac{\partial \ln f(Y, X, \gamma; \theta_2^{k+1})}{\partial \Gamma_{s_i = i}^{k+1}} = \sum_{t=2}^{T} \frac{\partial \ln f(Y_t, X_t | Z_t; \theta_2^{k+1})}{\partial \Gamma_{s_i = i}^{k+1}} \tag{B.8}
\]

Using (A.1), (A.7) and (A.8) and the Fubini theorem one more time we get:

\[
\sum_{t=2}^{T} \sum_{i=1}^{M} \frac{\partial \ln f(Y_t, X_t | Z_t; \theta_2^{k+1})}{\partial \Gamma_{s_i = i}^{k+1}} f(s_t = i | X, Y; \theta_2) = 0 \tag{B.9}
\]

\( f(X, Y, \gamma; \theta_2^{k+1}) \) here is the joint probability distribution of the returns and the realized covariance defined in expression (15). We can express the log-likelihood to maximize (omitting terms which are fixed in the maximization problem) as \( \mathcal{L}_c(\text{multivariate}) + \mathcal{L}_r(\text{multivariate}) \) because \( \Gamma_i \) does not appear in the univariate part. Differentiating w.r.t. \( \Gamma_i \):

\[
\frac{\partial \mathcal{L}_c(\text{multivariate})}{\partial \Gamma_i} = -\frac{1}{2} \Gamma^{-1}_i + \frac{1}{2} \Gamma^{-1}_i \hat{U}_i \hat{U}_i' \tag{B.10}
\]

\[
\frac{\partial \mathcal{L}_r(\text{multivariate})}{\partial \Gamma_i} = -\frac{v}{2} \Gamma^{-1}_i + \frac{v}{2} \Gamma^{-1}_i S_i \Gamma^{-1}_i \tag{B.11}
\]

Omitting the superscript \( k + 1 \) for notation clarity, and combining (A.10) and (A.11) we
get:
\[
\frac{\partial \ln f(Y_t, X_t|Z_t; \theta_2^{k+1})}{\partial s_{i,l}} = -\frac{1}{2} \Gamma_{s,=i}^{-1} + \frac{1}{2} \Gamma_{s,=i}^{-1} \hat{U}_i \hat{U}_i' \Gamma_{s,=i}^{-1} - \frac{1}{2} \Gamma_{s,=i}^{-1} + \frac{1}{2} \Gamma_{s,=i}^{-1} \hat{S}_i \Gamma_{s,=i}^{-1}
\]
\[
= -\frac{1}{2} \Gamma_{s,=i}^{-1} + \frac{1}{2} \Gamma_{s,=i}^{-1} (\hat{U}_i \hat{U}_i' + v \hat{S}_i) \Gamma_{s,=i}^{-1}
\]

(B.12)

Applying (A.12) to (A.9) we have:
\[
\sum_{t=2}^{T} \left\{ -\frac{v+1}{2} \Gamma_{s,=i}^{-1} + \frac{1}{2} \Gamma_{s,=i}^{-1} (\hat{U}_i \hat{U}_i' + v \hat{S}_i) \Gamma_{s,=i}^{-1} \right\} f(s_t = i|X,Y; \theta^k) = 0
\]

(B.13)

Solving for \(\Gamma_{s,=i}\) gives equation (21) with observed \(C_t = \frac{1}{\|\hat{U}_i \hat{U}_i' + v \hat{S}_i\|} \left\{ \hat{U}_i \hat{U}_i' + v \hat{S}_i \right\} \).

### Appendix C  Details of the Probabilities Computation

The Probabilities are computed by the algorithm of Hamilton (1989) and Kim (1994). Inference on the unobserved state of the Markov Chain is given by the following equations:

\[
\hat{\xi}_{t+1|t} = \Pi \hat{\xi}_{t|t}, \quad \text{(C.2)}
\]

\[
\eta_t = \left[ f(Y_t, X_t|\mathcal{F}_{t-1}, s_t = 1; \theta) \right] \quad \text{etc.}
\]

\[
\hat{\xi}_{t|t} = (\hat{\xi}_{t|t-1} \odot \eta_t), \quad \text{(C.1)}
\]

where \(\hat{\xi}_{t|t}\) is an \((M \times 1)\) vector which contains the probability of occurrence of each regime at time \(t\) conditional on the observations up to time \(t\). The \((M \times 1)\) vector \(\hat{\xi}_{t+1|t}\) gives the forecasted probabilities of time \(t+1\) conditional on observations up to time \(t\). The \(m\)-th element of the \((M \times 1)\) vector \(\eta_t\) is the density of \((Y_t, X_t)\) conditional on past observations and the regime \(m\) at time \(t\), \(s_t = 1\) is an \((M \times 1)\) vector of 1s, \(\odot\) denotes elements-by-elements multiplication.

Given a starting value \(\hat{\xi}_{1|0}\) and parameter values \(\theta\), one can iterate over (C.1) and (C.2) for \(t = 1, \ldots, T\). The log-likelihood is obtained as a by-product of this algorithm:

\[
\mathcal{L}(\theta) = \sum_{t=1}^{T} \log \left( 1'(\hat{\xi}_{t|t-1} \odot \eta_t) \right).
\]

(C.4)

Smooth inference on the state of the Markov chain can also be computed by the backward algorithm developed by Kim (1994). The probability of being in each regime at time \(t\) con-
ditional on observations up to time \( T > t \) is given by the following equation:

\[
\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot \left\{ \Pi \left[ \hat{\xi}_{t+1|T} \left( \hat{\xi}_{t+1|t} \right) \right] \right\}
\]  

(C.5)

where (\( \odot \)) denotes element-by-element division. One would start iterating over (C.2) with \( t = T \), where \( \hat{\xi}_{T|T} \) is given by (C.1) until \( t = 1 \) is reached. \( \hat{\xi}_{t|T} \) is therefore the vector of smoothed probability, that is the probability of a particular state at time \( t \) given all the available information, \( \mathcal{F}_T \).

### Appendix D Proof of Equation (29)

Solving forward equation (5), we get for a \( L \)-ahead period:

\[
h_{i,T+L} = \sum_{j=0}^{L-1} \omega_i \beta_{i}^{j} + \beta_{i}^{L} h_{T} + \sum_{k=0}^{L-1} \alpha_i \beta_{i}^{L-1} x_{T+L-1-k}
\]  

(D.1)

Taking the expected value at time \( T \) of (D.1) we get

\[
E_T h_{i,T+L|T} = h_{i,T+L|T} = \sum_{j=0}^{L-1} \omega_i \beta_{i}^{j} + \beta_{i}^{L} h_{T} + \sum_{k=0}^{L-1} \alpha_i \beta_{i}^{L} x_{T+L-1-k|T}
\]  

(D.2)

The \( T + L - 1 - k \) step ahead expected value of the realized variance given the time \( T \) information is simply:

\[
x_{T+L-1-k|T} = E_T x_{T+L-1-k} = \gamma_0 + \gamma_1 E_T h_{T+L-1-k}
\]  

(D.3)
References


URL: http://restud.oxfordjournals.org/content/75/2/339.short


URL: http://linkinghub.elsevier.com/retrieve/pii/S0304407611000029
http://linkinghub.elsevier.com/retrieve/pii/S0304407610000734

URL: http://www.jstor.org/stable/2239727

URL: http://www.sciencedirect.com/science/article/pii/0304407686900631

URL: http://www.jstor.org/stable/2109358


URL: http://www.jstor.org/stable/2984875
REFERENCES

URL: http://amstat.tandfonline.com/doi/full/10.1198/073500102753410444

URL: http://www.jstor.org/stable/1912773

URL: http://www.tandfonline.com/doi/abs/10.1198/073500102288618487

URL: http://doi.wiley.com/10.1002/jae.683

URL: http://amstat.tandfonline.com/doi/full/10.1080/01621459.1979.10482508

URL: http://linkinghub.elsevier.com/retrieve/pii/S0304407608002169


URL: http://www.jstor.org/stable/1912559


URL: http://www.tandfonline.com/doi/abs/10.1198/073500106000000071

REFERENCES


URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1980.10477520


URL: http://jfec.oxfordjournals.org/cgi/doi/10.1093/jjjfinec/nbs022


URL: http://www.sciencedirect.com/science/article/pii/0304407694900361

Klebaner, F. C. (2005), *Introduction to stochastic calculus with applications*, Imperial College Press.

URL: http://books.google.com/books?hl=en&lr=&id=JYzW0uqQxBOC&oi=fnd&pg=PR5&dq=Introduction+to+Stochastic+Calculus+with+Applications&ots=SrqK8iWVY7&sig=T0A54GMsrprQlVuQ6kS0mJDi0Js


URL: http://linkinghub.elsevier.com/retrieve/pii/S0304407612001777


URL: http://www.fednewyork.org/research/staff_reports/research_papers/9524.pdf


URL: http://amstat.tandfonline.com/doi/full/10.1198/073500102753410417


URL: http://link.springer.com/chapter/10.1007/978-3-540-71297-8_36


URL: http://linkinghub.elsevier.com/retrieve/pii/S0304407605000187


URL: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/3540003134


**URL:** [http://www.tandfonline.com/doi/abs/10.1198/073500102288618496](http://www.tandfonline.com/doi/abs/10.1198/073500102288618496)