Regime Switching for Dynamic Correlations*

Denis Pelletier †
North Carolina State University

First version: March 2002
This version: April 2004
Compiled: 29th April 2004

---

*I thank my co-advisor Nour Meddahi for his generous advice. I also thank Peter Christoffersen, William McCausland, Jean-Marie Dufour, Yongil Jeon, Doug Pearce, James McKinnon, Lynda Khalaf, Todd Smith, Sean Campbell, Neil Shephard and Michael McAleer for several helpful comments. I am grateful to the Editor Giampiero M. Gallo and two anonymous referees for suggestions and constructive comments that have led to improvement of the paper. I want to thank participants at the CIREQ-CIRANO Univariate and Multivariate Models for Asset Pricing conference (2002), the 36th CEA annual conference (University of Calgary, 2002), the 42nd SCSE annual conference (Aylmer,2002) and the New Frontiers in Financial Volatility Modelling conference (Florence, 2003). I also want to thank seminar participants at Carleton University, NCSU, Queen's University, University of Alberta, UBC, Bank of Canada, LSU and Texas A&M for valuable discussions. Financial support by the Social Sciences and Humanities Research Council of Canada, the Fonds FCAR (Government of Québec), CIRANO and CIREQ is greatly acknowledged.

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

ABSTRACT

We propose a new model for the variance between multiple time series, the Regime Switching Dynamic Correlation. We decompose the covariances into correlations and standard deviations and the correlation matrix follows a regime switching model; it is constant within a regime but different across regimes. The transitions between the regimes are governed by a Markov chain. This model does not suffer from a curse of dimensionality and it allows analytic computation of multi-step ahead conditional expectations of the variance matrix when combined with the ARMACH model [Taylor (1986) and Schwert (1989)] for the standard deviations. We also present an empirical application which illustrates that our model can have a better fit of the data than the Dynamic Conditional Correlation model proposed by Engle (2002).

Key words: dynamic correlation, regime switching, Markov chain, ARMACH.
Journal of Economic Literature Classification: C32, C53, G0, G1.

1. Introduction

It is a well known fact that the variance and covariance of most financial time series are time-varying. Modeling time-varying variance is not just a statistical exercise where someone tries to increase the value of the likelihood; it has important impacts in terms of asset allocation, asset pricing, computation of Value-at-Risk (VaR). A lot of work has been done to model univariate financial time series since the introduction of the ARCH model by Engle (1982). However, we face additional problems when we try to write a multivariate model of volatility. Not only must the variances be positive, the variance matrix must also be positive semi-definite (PSD) at every point in time. Another important problem is the curse of dimensionality. We want models that can be applied to more than a few time series. This rules out the direct generalizations of univariate GARCH models such as the BEKK model of Engle and Kroner (1995).

The most popular multivariate volatility model so far is certainly the Constant Conditional Correlation (CCC) model of Bollerslev (1990) [a survey of multivariate GARCH model is given by Bauwens, Laurent, and Rombouts (2003)]. In this model, the covariances of a vector of returns are decomposed into standard deviations and correlations. The major hypothesis in this model is that the conditional correlations are constant through time. With this hypothesis, it is easy to get PSD variance matrices because we only have to ensure that the correlation matrix is PSD and that the standard deviations are non-negative. It also breaks the curse of dimensionality because the likelihood can be seen as a set of SURE equations, i.e. a two-step estimation procedure where univariate volatility models are estimated in a first step that will yield consistent estimates. However, the hypothesis of constant correlations is not always supported by the data [e.g. Engle and Sheppard (2001)].

In this work, we present a new multivariate volatility model, the Regime Switching Dynamic Correlation (RSDC) model. We also decompose the covariances into standard deviations and correlations, but these correlations are dynamic. The correlation matrix follows a regime switching model; it is constant within a regime but different across regimes. The transitions between the regimes are governed by a Markov chain. The CCC model is a special case of ours where we take the number of regimes to be one.

The RSDC model has many interesting properties. First, it is easy to impose that the variance matrices are PSD. Second, it does not suffer from a curse of dimensionality because it can be estimated with a two-step procedure[1]. Third, when combined with the ARMACH model [see Taylor (1986) and Schwert (1989)] for the standard deviations, this correlation model allows analytic computation of multi-step ahead conditional expectations of the whole variance matrix. Fourth, it can produce smooth patterns for the correlations. We also present an empirical application to exchange rate time series which illustrates that it can have a better fit of the data than the Dynamic Conditional Correlation (DCC) model recently proposed in Engle (2002).

[1]We say that a model does not suffer from a curse of dimensionality if it is possible to obtain consistent estimates of the parameters even when the number of time series is large. These estimates may not be efficient.

The model of Engle (2002) and the model proposed in Tse and Tsui (2002) use the same decomposition for the variance matrix as in Bollerslev (1990), but instead of taking constant correlations they propose a GARCH-type dynamic. Because a correlation must lie between -1 and 1, these models must include a rescaling that introduces non-linearities. One side effect of this rescaling is that we can't analytically compute multi-step ahead conditional expectations of the correlation and variance matrices. We can also ask ourselves if a GARCH-type model is appropriate for the correlations because the dynamic of a correlation can be intrinsically different than the behavior of a covariance, e.g. a correlation is bounded from below and above while a covariance is not.

Another approach for breaking the curse of dimensionality of the multivariate GARCH is Ledoit, Santa Clara, and Wolf (2003)'s that proposes a flexible estimation procedure for the Diagonal-Vech model of Bollerslev, Engle, and Wooldridge (1988). The maximization of the likelihood of this model is not computationally feasible if the number of time series is greater than five [see Ding and Engle (2001)]. They propose a way to combine the estimates from univariate and bivariate model so as to get consistent estimates of the parameters of the full multivariate Diagonal-Vech and insure that the variance matrices are PSD. Although their results are very interesting, this estimation procedure is only valid for the Diagonal-Vech model.

The paper is organized as follows. The second section presents the RSDC model and its properties. Section three describes the estimation of this model and the theoretical properties of the estimates. Section four outlines the computation of one-step and multi-step ahead conditional expectations of the variance matrix. Section five presents an application of the model to multiple exchange rates series. Section six contains a few concluding remarks. Finally, proofs are in the appendix.

2. The RSDC model

In this section we present the Regime Switching Dynamic Correlation (RSDC) model. Assume that the K-variate process Y_t has the form:

$$Y_t \quad = \quad H_t^{1/2} U_t \tag{2.1}$$

where U_t is an i.i.d. $(0, I_K)$ process. The time varying covariance matrix H_t can be decomposed into:

$$H_t \quad \equiv \quad S_t \Gamma_t S_t \tag{2.2}$$

where S_t is a diagonal matrix composed of the standard deviations $s_{k,t}$, $k = 1, \ldots, K$ and the matrix Γ_t contains the correlations. Both S_t and Γ_t are time varying. This decomposition of the covariance matrix has previously been used by Bollerslev (1990), Tse and Tsui (2002), Engle (2002) and Barnard, McCulloch, and Meng (2000). The series Y_t could be a filtered process.

With this decomposition the log-likelihood can be written

$$
\begin{aligned}
L &= -\frac{1}{2}\sum_{t=1}^{T}\left(K\log(2\pi)+\log(|H_t|)+Y_t'H_t^{-1}Y_t\right)\\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(K\log(2\pi)+\log(|S_t\varGamma_t S_t|)+Y_t'S_t^{-1}\varGamma_t S_t^{-1}Y_t\right)\\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(K\log(2\pi)+2\log(|S_t|)+\log(|\varGamma_t|)+\tilde{U}_t'\varGamma_T^{-1}\tilde{U}_t\right) \qquad (2.3)
\end{aligned}
$$

where $\tilde{U}_t = [\tilde{u}_{1,t}, \ldots, \tilde{u}_{K,t}]'$ is a zero-mean process with covariance matrix $\varGamma_t$ and $|H_t| = \det(H_t)$. This is the first building block of our RSDC model: to model the full covariance matrix we model the variances and the correlations separately.

2.1. Regime switching for the correlations

In this work we will argue for a regime switching model for the correlations. This can be seen as a midpoint between the CCC model of Bollerslev (1990) and models such as the DCC of Engle (2002) where the correlations change every period. This model will have the appealing property of constant correlations within a regime but will still have dynamic correlations because of the regime switching. More specifically, the time-varying correlation matrix $\varGamma_t$ follows:

$$\varGamma_t = \sum_{n=1}^{N} \mathbb{1}_{\{\varDelta_t=n\}}\varGamma_n \qquad (2.4)$$

with $\varDelta_t$ an unobserved Markov chain process independent of U_t which can take N possible values ($\varDelta_t = 1, 2, \ldots, N$). The symbol $\mathbb{1}$ is the indicator function. The $K \times K$ matrices $\varGamma_n$ are correlation matrices (symmetric, PSD, ones on the diagonal, off-diagonal elements between -1 and 1) with $\varGamma_n \neq \varGamma_{n'}$ for $n \neq n'$. The probability law governing $\varDelta_t$ is defined by its transition probability matrix, denoted by $\varPi$. The probability of going from regime i in period t to regime j in period $t+1$ is denoted by $\pi_{i,j}$ and the limiting probability of being in regime n is π_n. The element on row j and column i of $\varPi$ is $\pi_{i,j}$. We make the standard assumptions on the Markov chain [aperiodic, irreducible and ergodic. See Ross (1993, Chapter 4)][2].

[2]The transition probabilities could be a function of weakly exogenous variables, as in by Diebold, Lee, and Weinbach (1994). This could be an interesting extension but we do not pursue it in this work. It could be one way of allowing the variances and correlations to move jointly. Feasibility of this extension will be briefly discussed in Section 3.2.

Beside its very intuitive interpretation, this model has many appealing properties. It is easy to impose that Γ_t is a correlation matrix because we only have to impose it for every Γ_n. Imposing that the diagonal elements are equal to one and that the off-diagonal elements are in $[-1, 1]$ does not guarantee that Γ_n is PSD. One way to impose that Γ_n will be a correlation matrix is to take its Choleski decomposition, i.e. $\Gamma_n = P_n P_n'$ where P_n is a lower triangular matrix, and to impose constraints on P_n so that we get ones on the diagonal. These constraints will automatically give off-diagonal elements between -1 and 1. Consider a trivariate example. In this case, the Choleski decomposition gives:

$$\Gamma = \begin{bmatrix} p_{1,1}^2 & p_{1,1}\,p_{2,1} & p_{1,1}\,p_{3,1} \\ p_{1,1}\,p_{2,1} & p_{2,1}^2 + p_{2,2}^2 & p_{2,1}\,p_{3,1} + p_{2,2}\,p_{3,2} \\ p_{1,1}\,p_{3,1} & p_{2,1}\,p_{3,1} + p_{2,2}\,p_{3,2} & p_{3,1}^2 + p_{3,2}^2 + p_{3,3}^2 \end{bmatrix}.$$

Imposing the additional constraint that the elements on the diagonal P_n are positive, the restrictions $\Gamma_{j,j} = 1$, for $j = 1, \ldots, K$, becomes

$$p_{j,j} = \sqrt{1 - \sum_{i=1}^{j-1} p_{j,i}^2} \quad (j = 1, \ldots, K) \tag{2.5}$$

where the sum is zero for $j = 1$.

We could think that estimation of the RSDC model would be complicated by the possibly high number of parameters coming from each Γ_n. Fortunately we will see later on that we can use the EM algorithm [Dempster, Laird, and Rubin (1977)] as presented in Hamilton (1994, chapter 22) so that increasing the number of time series, to which the model is applied will not complicate the estimation.

This specification has three additional interesting properties. The first is that because this model for the correlations is basically linear due to the Markov chain we are able to compute multi-step ahead conditional expectations of the correlation matrix. Also, if we use an appropriate model for the standard deviations, we will also be able to perform these computations for the whole variance matrix. We present such a model in Section 2.3. This is in contrast to the models of Engle (2002) and Tse and Tsui (2002) where the rescaling that is used to keep the correlations between -1 and 1 introduces non-linearities that forbid the computation of multi-step ahead conditional expectations. It is unfortunate because one reason why we study volatility is to be able to forecast it. The second property comes from the Markov chain. If there is some general form of persistence in the chain (high probability of staying in a given regime for more than one period), then this will lead to smooth time-varying correlations. This could have important impacts namely for the computation of VaR and dynamic portfolio allocation because the benefits of portfolio diversification would be less volatile. The third is that by having a regime switching for the correlations, the variances and covariances are not bounded which is the case when they are the ones following a regime switching [e.g. see Geweke and Amisano (2001)]

An alternative way of modelling dynamic correlations we mentioned above is the use of multivariate GARCH models as in Engle (2002) and Tse and Tsui (2002). For example, in Engle (2002), the conditional correlation matrix Γ_t follows

$$\tilde{\Gamma}_t = (1 - \sum_{i=1}^{q} a_i - \sum_{j=1}^{p} b_j)\Gamma + \sum_{i=1}^{q} a_i(\tilde{U}_{t-i}\tilde{U}'_{t-i}) + \sum_{j=1}^{p} b_j\tilde{\Gamma}_{t-j}, \quad (2.6)$$

$$\Gamma_t = D_t^{-1}\tilde{\Gamma}_t D_t^{-1} \quad (2.7)$$

where D_t is a diagonal matrix with $\sqrt{\tilde{\Gamma}_{i,i,t}}$ on row i and column i, and a_i and b_j are scalars. Since a correlation matrix must have ones on the diagonal and off-diagonal elements between -1 and 1, we must rescale the correlation matrix [equation (2.7)] because $\tilde{U}_{t-i}\tilde{U}'_{t-i}$ is not constrained to have elements between -1 and 1. The matrix Γ gives the unconditional value of the correlation matrix (the correlation matrix of the CCC model). The theoretical and empirical properties of this model are developed in Engle and Sheppard (2001).

Regime switching has been often employed to model univariate heteroskedastic time series. The most straightforward approach is to have constant variance (and mean) in each regime, e.g. see Garcia and Perron (1996). This model can be generalized by combining regime switching and GARCH models:

$$\sigma_t^2 = \omega(\Delta_t) + \alpha(\Delta_t)y_{t-1}^2 + \beta(\Delta_t)\sigma_{t-1}^2.$$

Recursive substitution shows that σ_t^2 depends on the entire history of Δ_t. For a sample of length T, the evaluation of the likelihood would require the integration over all the N^T possible paths, which cannot be done for meaningful sample sizes. One solution to this problem is to use an ARCH(p) model instead of a GARCH model [see Cai (1994), Hamilton and Sumsel (1994)]. A second solution is to modify the GARCH dynamic by assuming that σ_t^2 is a function of a weighted sum of the past variances over the possible regimes:

$$\sigma_t^2 = \omega(\Delta_t) + \alpha_1(\Delta_t)y_{t-1}^2 + \beta_1(\Delta_t)\tilde{\sigma}_{t-1}^2$$

where $\tilde{\sigma}_{t-1}^2 = \sum_{n=1}^{N} \mu_{t-1}(n)\sigma_{t-1}^2(\Delta_{t-1} = n)$. Different weights are given in Gray (1996), Dueker (1997) and Klaassen (2002). A third solution proposed by Haas, Mittnik, and Paolella (2003)) is to have N GARCH equations evolving in parallel according to different sets of parameters. The regime switching is then over the possible N values of σ_t^2. A final solution is to use linear representations for powers of y_t^2 [see Francq and Zakoïan (2004)].

These models have been generalized to bivariate time series. In Edwards and Susmel (2001, 2003), the authors uses the decomposition in (2.2) and assume that the variances are following a regime switching ARCH model as in Hamilton and Sumsel (1994). The correlation is constant within each regime associated to one of the variances. Two additional multivariate generalization are presented in Ang and Chen (2002). The first is a mixture of multivariate Normal distribution, similar to the model employed in Garcia and Perron (1996). The second is a direct generalization of Gray (1996) where the GARCH equation has a BEKK representation.

We see that the model presented in this work is a combination of a mixture model for the correlation matrix (instead of the variance) and the usual GARCH-type models to allow interesting dynamics for the variances and covariances. An alternative way of introducing a regime switching for the correlations which would be a generalization of Gray (1996) would be to let the a, b and Γ parameters in DCC equation (2.6) be a function of the regimes.

An interesting question is whether models with regime switching for the variances and models with regime switching for the correlations can be distinguished? We should be able to differentiate them. Since the correlation between $y_{i,t}$ and $y_{i,t}$ is always 1, a regime switching model for the correlations will not introduce a regime switching for the variances. On the other hand, if the regime switching is only for the variances and we have the decomposition in (2.2), then dividing the series by their standard deviations should wash out the regime switching (we would not observe any regime switching dynamic across pairs of series divided by their standard deviations). We could also envision a model where both the standard deviations and the correlations are driven by a regime switching.

2.2. A parsimonious model

We next present a restricted version of the general regime switching model which will have a reduced number of parameters and will remain easy to estimate. For the matrix Γ_t we propose the following form:

$$\Gamma_t = \Gamma\lambda(\Delta_t) + I_K(1 - \lambda(\Delta_t)) \tag{2.8}$$

where Γ is a fixed correlation matrix, I_K is a $K \times K$ identity matrix, $\lambda(\Delta_t) \in [0, 1]$ is a univariate random process governed by an unobserved Markov chain process Δ_t that can take N possible values ($\Delta_t = 1, 2, \ldots, N$) and is independent of U_t. The probability law governing Δ_t is defined by its transition probability matrix, denoted by Π.

The correlation matrix at time t is a weighted average of two extreme states of the world. In one state, the returns are uncorrelated [$\lambda(\Delta_t) = 0$] and in the other the returns are (highly) correlated [$\lambda(\Delta_t) = 1$]. We then have regimes of generally higher or lower correlations and the changes across correlations in a given regime are proportional. The variable $\lambda(\Delta_t)$ can be related to the notion of common features and factor models [Engle and Susmel (1993), Bollerslev and Engle (1993), King, Sentana, and Wadhwani (1994), Diebold and Nerlove (1989), Engle, Ng, and Rothschild (1990), Ng, Engle, and Rothschild (1992)] where the factor affects the variance matrix instead of the correlation matrix.

Note that for the off-diagonal elements only the product of Γ and λ can be identified (by construction the diagonal elements of Γ_t are equal to 1). To solve this identification problem we can consider two natural sets of constraints. The first is:

$$\lambda(1) = 1,\ \lambda(1) > \lambda(2),\ \ldots,\ \lambda(N-1) > \lambda(N), \tag{2.9}$$

In this case, fixing one of the $\lambda(n)$ to be one identifies the product of Γ and λ. We also restrict the $\lambda(n)$s to be a decreasing sequence to remove the possibility of relabelling regime i as regime j and vice versa. An alternative identification assumption is:

$$\max_{i \neq j} |\Gamma_{i,j}| = 1 \quad \text{with} \quad 1 > \lambda(1), \lambda(1) > \lambda(2), \ldots, \lambda(N-1) > \lambda(N). \tag{2.10}$$

In this case, instead of fixing the highest value of $\lambda(n)$ to be one, we impose this restriction on an off-diagonal element of Γ. The second identification scheme does not impose that one correlation is equal to 1 or -1 because we multiply Γ by $\lambda(\Delta_t)$. Depending on the estimation scheme that we use, one of the two sets of constraints will be more appropriate.

The assumption $\lambda(\Delta_t) \in [0, 1]$ can be very restrictive. It is tempting to allow $\lambda(\Delta_t)$ to take negative values to allow the correlations to change sign, which might happen in periods of market distress for example. However we don't have a result for a lower bound on $\lambda(\Delta_t)$ that would guarantee that Γ_t is PSD. To understand the problem, consider the correlation matrix of a trivariate time series. If all the correlations are 0.99 then the correlation matrix is PSD; if all the correlations are -0.99, then it will not be PSD. We see that a lower bound on $\lambda(\Delta_t)$ would depend on Γ. In empirical applications, negative values for $\lambda(\Delta_t)$ could be allowed if we instead impose during the maximization that Γ_t is PSD.

2.3. Univariate volatility models

To complete the RSDC model we have to specify the dynamic for the standard deviations. The most common one for the volatility of univariate processes is certainly the GARCH model of Bollerslev (1986) where the conditional variance at time t, s_t^2, is a linear function of past squared innovations and past conditional variances:

$$s_t^2 = \omega + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j s_{t-j}^2. \tag{2.11}$$

We should notice that our RSDC model is not written in terms of variances but in terms of standard deviations; a covariance is a correlation times the standard deviations. By using a model such as the GARCH for the variance, the covariance becomes the product of a correlation and the square-root of the product of two variances. The square-root introduces non-linearities that will prohibit analytic computation of conditional expectations.

One model for the volatility of univariate time series that would not have this problem is the GARCH in absolute innovations of Taylor (1986) and Schwert (1989). This class of model is also referred to as ARMACH process in Taylor (1986). In these models the conditional standard deviations follows:

$$s_t = \omega + \sum_{i=1}^{q} \tilde{\alpha}_i |y_{t-i}| + \sum_{j=1}^{p} \beta_j s_{t-j} \tag{2.12}$$

with $\tilde{\alpha}_i = \alpha_i / E|\tilde{u}_t|$. The conditional standard deviations (instead of the conditional variances) are a recursive function of absolute value of past innovations (instead of squared innovations).

There are numerous reasons why a volatility model based on absolute values instead of squared innovations could be a good thing. One reason can be linked to the least absolute deviations versus least squares approach. As argued by Davidian and Carroll (1987), the model could be more robust if we use the absolute value instead of the squared innovation. However, we must reckon that the interpretation of an outlier in a volatility model is not as straightforward as in a regression context. It could also be that the absolute return is a better measure of risk than the squared return. This question is studied by Granger and Ding (1993).

Using the ARMACH model for the volatility of univariate time series is not a prerequisite of our model. We consider this model because it allows the computation of multi-step ahead conditional expectations of the variance matrix. If conditional expectations are not a point of interest or if the ARMACH gives a clearly inferior fit of the data then another model could be used.

3. Estimation

The estimation of the RSDC model can in theory be done in one step but if we have more than a few time series the high number of parameters will prohibit us from doing so. Fortunately, we can use a two-step estimation procedure as in Engle (2002). In a first step, we can estimate the univariate volatility models and in a second step, we can estimate the parameters in the correlation matrix conditional on the first step estimates.

In the first subsection we review the one-step estimation and explain how the likelihood can be evaluated. In the following subsection we present estimation methods which can greatly ease the estimation problem due to the high number of parameters.

3.1. One-step estimation

To maximize the likelihood we need to evaluate

$$QL(\boldsymbol{\theta}; \boldsymbol{Y}) = \sum_{t=1}^{T} \log f(Y_t | \underline{Y_{t-1}}), \tag{3.1}$$

where $\underline{Y_{t-1}} = \{Y_{t-1}, Y_{t-2}, \ldots\}$ and $\boldsymbol{\theta}$ is the vector of parameter values. To do this we use Hamilton's filter [Hamilton (1989), Hamilton (1994, chapter 22)] because the Markov chain Δ_t is unobserved. Inference on the state of the Markov chain is given by the following equations:

$$\hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t|t-1} \odot \eta_t)}{\mathbf{1}'(\hat{\xi}_{t|t-1} \odot \eta_t)}, \tag{3.2}$$

$$\hat{\xi}_{t+1|t} = \Pi \hat{\xi}_{t|t}, \tag{3.3}$$

$$\eta_t = \begin{bmatrix} f(Y_t|\underline{Y_{t-1}}, \Delta_t = 1; \boldsymbol{\theta}) \\ \vdots \\ f(Y_t|\underline{Y_{t-1}}, \Delta_t = N; \boldsymbol{\theta}) \end{bmatrix}, \tag{3.4}$$

where $\hat{\xi}_{t|t}$ is an $(N \times 1)$ vector which contains the probability of being in each regime at time t conditional on the observations up to time t. The $(N \times 1)$ vector $\hat{\xi}_{t+1|t}$ gives these probabilities at time $t+1$ conditional on observations up to time t. The n-th element of the $(N \times 1)$ vector η_t is the density of Y_t conditional on past observations and being in regime n at time t, $\mathbf{1}$ is an $(N \times 1)$ vector of 1s, and $\odot$ denotes elements-by-elements multiplication. Given a starting value $\hat{\xi}_{1|0}$ and parameter values $\boldsymbol{\theta}$, one can iterate over (3.2) and (3.3) for $t = 1, \ldots, T$. The likelihood is obtained as a by-product of this algorithm:

$$QL(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log \left(\mathbf{1}'(\hat{\xi}_{t|t-1} \odot \eta_t) \right). \tag{3.5}$$

Smoothing inference on the state of the Markov chain can also be computed using an algorithm developed by Kim (1994). The probability of being in each regime at time t conditional on observations up to time T is given by the following equation:

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot \left\{ \Pi' \left[\hat{\xi}_{t+1|T} \, (\div) \, \hat{\xi}_{t+1|t} \right] \right\} \tag{3.6}$$

where $(\div)$ denotes element-by-element division. One would start iterating over (3.6) with $t = T$, where $\hat{\xi}_{T|T}$ is given by (3.2).

What remains is deciding how to start up the algorithm, i.e. specifying $\hat{\xi}_{1|0}$. One approach would be to add this vector to the parameter space and estimate these initial probabilities. This would add N parameters, $p_1, \ldots, p_N \geq 0$ with $p_1 + \cdots + p_N = 1$. Another approach would be to use the limiting probabilities $(\pi_1, \pi_2, \ldots, \pi_N)$ of the Markov process [Ross (1993, Chapter 4)]. These probabilities are the solution of the following system of equations:

$$\begin{bmatrix} \pi_1 \\ \vdots \\ \pi_N \end{bmatrix} = \Pi \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_N \end{bmatrix},$$
$$\sum_{n=1}^{N} \pi_n = 1.$$

In the two-regime case the solution is

$$\pi_1 = \frac{1-\pi_{2,2}}{(1-\pi_{1,1})+(1-\pi_{2,2})}\,; \quad \pi_2 = \frac{1-\pi_{1,1}}{(1-\pi_{1,1})+(1-\pi_{2,2})}.$$

In this work both approaches will be used, depending on the estimation method. As we will see below, when using the EM algorithm there is an advantage in treating $\xi_{1|0}$ as unknown parameters. If we are not using the EM algorithm then we will use the limiting probabilities of the Markov chain because in this case these extra parameters would complicate the estimation.

In the evaluation of the likelihood, notice that the correlation matrix can only take N possible values in our model so we only have to invert N times a $K \times K$ matrix. When the number of time series is large this can be a computational advantage over models such as Engle (2002) and Tse and Tsui (2002) where a different correlation matrix has to be inverted for every observation.

The (quasi) maximum likelihood estimators are consistent and have the usual asymptotic normal distribution. One-step estimation is not really practicable if the number of time series is more than a few because of the high number of parameters. In this case, we need an estimation method which does not require non-linear optimization of a number of parameters which depends on the number of time series. This is what we present in the next subsection.

3.2. Two-step estimation

We first begin by introducing elements of notation. The complete parameter space $\boldsymbol{\theta}$ is split into $\boldsymbol{\theta_1}$ for the parameters in the univariate volatility model and $\boldsymbol{\theta_2}$ for the parameters in the correlation model. We denote by QL_1 the likelihood where the correlation matrix is taken to be an identity matrix:

$$QL_1(\boldsymbol{\theta_1}; \boldsymbol{Y}) = -\frac{1}{2}\sum_{t=1}^{T}\left(K\log(2\pi) + 2\log(|S_t|) + U_t'U_t\right). \tag{3.7}$$

We denote by QL_2 the likelihood given $\boldsymbol{\theta_1}$ where we have concentrate out S_t:

$$QL_2(\boldsymbol{\theta_2}; \boldsymbol{Y}, \boldsymbol{\theta_1}) = -\frac{1}{2}\sum_{t=1}^{T}\left(K\log(2\pi) + \log(|\Gamma_t|) + U_t'\Gamma_T^{-1}U_t\right). \quad (3.8)$$

Notice two important features of QL_1. Firstly, it is the sum of K univariate log-likelihood so maximizing it is equivalent to maximizing each univariate log-likelihood separately. Secondly, the evaluation of these log-likelihood is straightforward since it does not involve the use of Hamilton's filter. To maximize QL_2 we again have to use Hamilton's filter since Δ_t is unobserved.

Because the number of parameters in the correlation model grows at a quadratic rate with the number of time series, direct maximization of QL_2 is not practicable if we analyze more than a few series. To bypass this problem, we present two estimation methods, one for the non-restricted model (2.4) and one for the restricted model (2.8), which do not rely on the simultaneous non-linear maximization of all the parameters.

For the non-restricted model, it turns out that maximization of the likelihood QL_2 for the correlation model can be done with the EM algorithm. Using the results of Hamilton (1994, chapter 22) we know that the MLE estimates of the transition probabilities and the correlation matrices satisfy the following equations if the initial probabilities $\hat{\xi}_{1|0}$ are not a function of Π and Γ_i:

$$\hat{\pi}_{i,j} = \frac{\sum_{t=2}^{T} P[\Delta_t = j, \Delta_{t-1} = i|\underline{\hat{U}_T}; \hat{\boldsymbol{\theta}}_2]}{\sum_{t=2}^{T} P\left[\Delta_{t-1} = i|\underline{\hat{U}_T}; \hat{\boldsymbol{\theta}}_2\right]}, \quad (3.9)$$

$$\hat{\Gamma}_n = \frac{\sum_{t=1}^{T}(\hat{U}_t\hat{U}_t')P[\Delta_t = n|\underline{\hat{U}_T}; \hat{\boldsymbol{\theta}}_2]}{\sum_{t=1}^{T} P[\Delta_t = n|\underline{\hat{U}_T}; \hat{\boldsymbol{\theta}}_2]}. \quad (3.10)$$

Starting with an initial value $\hat{\boldsymbol{\theta}}_2^{(0)}$ for the vector $\boldsymbol{\theta_2}$, we can compute a new vector $\hat{\boldsymbol{\theta}}_2^{(1)}$ using equations (3.9) and (3.10). We then continue the iteration until the difference between successive vectors $\hat{\boldsymbol{\theta}}_2^{(m)}$ and $\hat{\boldsymbol{\theta}}_2^{(m+1)}$ is small. For the vector of initial probabilities $\hat{\xi}_{1|0}$, it is also shown that their MLE estimates are given by the smoothed probabilities of the first observation [see Hamilton (1994, p. 695)].

Notice that the dimension of Γ_n (i.e. the number of time series) does not affect the complexity of the estimation because we only have to take sums in (3.9) and (3.10). We should also mention that equation (3.10) cannot be used directly because typically it does not provide correlation matrices, i.e. the elements on the diagonal of $\hat{\Gamma}_n$ are not imposed to be one. One should rescale these matrices as in equation (2.7) after each iteration so they are correlation matrices. By doing this transformation, the estimates obtained with these equations will not exactly be the numerical maximum of the likelihood, but very close to it. From our experience, a limited number of Newton-type iterations are necessary to obtain the exact numerical maximum.

For the restricted model (2.8) we can estimate the matrix Γ, up to a scale factor, by doing correlation targeting. This leaves a number of parameters to be non-linearly estimated that does not depend on the number of time series. To see how it can be done, we first compute the unconditional expectation of the correlation matrix:

$$E[\Gamma_t] = \Gamma \sum_{n=1}^{N} \lambda(n)\pi_n + I_K \sum_{n=1}^{N} (1-\lambda(n))\pi_n.$$

Notice that the off-diagonal elements of $E[\Gamma_t]$ are the product of Γ and the scalar $\sum_{n=1}^{N} \lambda(n)\pi_n$.

Therefore, a sample correlation matrix computed with the standardized residuals from the first step estimation will provide an estimate $\hat{\Gamma}$ of Γ up to the scale factor $\sum_{n=1}^{N} \lambda(n)\pi_n$ for the off-diagonal elements. The scale indetermination can be solved by using the constraints on Γ and $\lambda(n)$ described in equation (2.10): We would divide the off-diagonal elements of $\hat{\Gamma}$ by the highest in absolute value[3], so as to get a 1 or -1 off the diagonal, and we would take $\lambda(1) > 1$. This leaves a number of parameters to be non-linearly estimated which increase with the number of regimes, not with the number of time series. The properties of the two-step estimation are described in the following theorem.

Theorem 3.1 TWO-STEP MAXIMUM LIKELIHOOD ESTIMATION. *If the usual assumptions for the validity of QMLE are satisfied, then the two-step estimates are consistent and their asymptotic distribution is:*

$$\sqrt{T}\left(\begin{bmatrix} \hat{\boldsymbol{\theta}}_1 \\ \hat{\boldsymbol{\theta}}_2 \end{bmatrix} - \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix}\right) \longrightarrow N(0; V)$$

with

$$V = \begin{bmatrix} G_{\boldsymbol{\theta}_1}^{-1} & -G_{\boldsymbol{\theta}_1}^{-1} G_{\boldsymbol{\theta}_2} M^{-1} \\ 0 & M^{-1} \end{bmatrix} E\left[\frac{\partial \ln f}{\partial \boldsymbol{\theta}} \frac{\partial \ln f}{\partial \boldsymbol{\theta}'}\right] \begin{bmatrix} G_{\boldsymbol{\theta}_1}^{-1} & -G_{\boldsymbol{\theta}_1}^{-1} G_{\boldsymbol{\theta}_2} M^{-1} \\ 0 & M^{-1} \end{bmatrix}'$$

where

$$G_{\boldsymbol{\theta}_1} = E\left[\frac{\partial g(Y, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1{}'}\right]; \quad G_{\boldsymbol{\theta}_2} = E\left[\frac{\partial g(Y, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2{}'}\right]; \quad M = E\left[\frac{\partial m(Y, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2{}'}\right]$$

$$g(Y, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{\partial \ln f(Y_t | \underline{Y_{t-1}})}{\partial \boldsymbol{\theta}_1}; \quad m(Y, \boldsymbol{\theta}_2) = \frac{\partial \ln f(Y_t | \underline{Y_{t-1}})}{\partial \boldsymbol{\theta}_2}$$

[3] Consider the following trivariate example. Suppose the correlations computed from the standardized residuals are $\Gamma_{1,2} = 0.2$, $\Gamma_{1,3} = 0.4$ and $\Gamma_{2,3} = 0.5$. The highest correlation being 0.5, we would take $\hat{\Gamma}_{1,2} = 0.2/0.5 = 0.4$, $\hat{\Gamma}_{1,3} = 0.4/0.5 = 0.8$ and $\hat{\Gamma}_{2,3} = 0.5/0.5 = 1$, then we would maximize the likelihood over $\lambda(1)$, $\lambda(2)$, $\lambda(3)$ and Π.

The matrix V can be consistently estimated by its plug-in estimate.

The proof is in the appendix.

Using the general results summarized in Pagan (1986) on two-step estimation we can compute efficient estimates from the two-step estimates by doing one step of a Newton-Raphson estimation of the full likelihood using our two-step estimates as the starting point. Notice that the computation of these estimates could be costly in computing time when dealing with very large systems because of the need to compute the matrix of second derivatives. We nonetheless say that this model does not have a curse of dimensionality by opposition to models for which consistent estimates cannot be obtained in practice when the number of series is large.

For the restricted model, estimation would not be substantially more complicated if the transition probabilities are a function of an exogenous variable [see Diebold, Lee, and Weinbach (1994)] because we could still do correlation targeting. This would not be the case for the unrestricted model. The two-step estimation presented in equations (3.9) and (3.10) could not applied if Π is not constant.

The remaining problem in this work is to specify the number of regimes in the Markov chain. It is well known that testing for the number of regimes in a Markov chain is a hard problem to tackle. We leave this problem for further work. The asymptotic theory of an LR test of $N+1$ versus N regimes is complicated by the fact that some parameters are not identified under the null hypothesis and we are testing parameter values that are on the boundary of the maintained hypothesis [see Andrews (1999, 2001)]. The asymptotic properties of this test are unknown for the moment, although bounds are given by Davies (1987). A solution could be the use of Monte Carlo test procedures [see Dufour (2002)]. An alternative procedure could be the specification tests presented in Hamilton (1996).

4. Multi-step ahead conditional expectations

In this section we study one-step and multi-step ahead conditional expectations of the variance matrix. To compute these we must take the conditional expectations of the product of a correlation and two standard deviations. We begin by introducing a notation for the matrix Γ_t that covers both the restricted and unrestricted model. We will denote by $\Gamma(\Delta_t = n)$ the value taken by Γ_t when the chain is in regime n at time t. All the calculus will be presented for the case where the univariate volatility model is an ARMACH(1,1). Extension to a more general ARMACH(p,q) would not introduce new difficulties.

One-step ahead conditional expectations are straightforward. Using the fact that tomorrow's conditional standard deviations are known, we get

$$E_t[H_{t+1}] = S_{t+1}\Gamma_{t+1|t}S_{t+1}$$

where

$$\Gamma_{t+1|t} = \sum_{n=1}^{N} \Gamma(\Delta_{t+1} = n)\xi_{n,t+1|t} \, .$$

We next sketch how we can compute d-step ahead conditional expectations. To compute $E_t[H_{t+d}]$ we have to compute elements of the following form, for $i, j = 1, 2, \ldots, K$,

$$E_t[s_{i,t+d}s_{j,t+d}\Gamma_{i,j}(\Delta_{t+d})].$$

The ARMACH model described in equation (2.12) can be rewritten in an ARMA-type representation and for an ARMACH(1,1) we get:

$$s_{k,t} = \omega_k + (\alpha_k + \beta_k)s_{k,t-1} + \alpha_k s_{k,t-1}\tilde{v}_{k,t-1} \tag{4.1}$$

where

$$\tilde{v}_{k,t-1} = \left(\frac{|\tilde{u}_{k,t-1}|}{E|\tilde{u}_{k,t-1}|} - 1 \right) \tag{4.2}$$

is a martingale difference sequence. Using the fact that the Markov chain is independent of the process U_t, we can first compute the expectation conditional on the Markov chain and then integrate it out:

$$E_t[s_{i,t+d}s_{j,t+d}\Gamma_{i,j}(\Delta_{t+d})] = E_t^{\Delta}\left[\Gamma_{i,j}(\Delta_{t+d})E_t^{U}[s_{i,t+d}s_{j,t+d}|\Delta]\right]$$

where $E_t^U[\cdots|\Delta]$ is the expectation with respect to the innovations U_t conditional on the present and future values of Δ_t, and $E_t^{\Delta}[\cdots]$ is the expectation with respect to the process Δ_t. We can now treat the correlations as known for the computation of $E_t^U[\cdots|\Delta]$. Doing so, we get a recursive expression that we can solve:

$$\begin{aligned} E_t^U\left[s_{i,t+d}s_{j,t+d}|\Delta\right] &= \sum_{l=1}^{d-1} a_{i,j,t+d-l}\left(\prod_{m=1}^{l-1} b_{i,j,t+d-m}(n_{d-m})\right) \\ &\quad + \prod_{m=1}^{d-1} b_{i,j,t+d-m}(n_{d-m})s_{i,t+1}s_{j,t+1} \end{aligned} \tag{4.3}$$

where $\prod_{m=1}^{l-1} b_{i,j,t+d-m}(n_{d-m})$ is equal to one when $l = 1$ and where

$$a_{i,j,t+d-1} = \omega_i\omega_j + \omega_i(\alpha_j + \beta_j)a_{j,t+d-1} + \omega_j(\alpha_i + \beta_i)a_{i,t+d-1} \, , \tag{4.4}$$

$$a_{k,t+d} = \omega_k \frac{1 - (\alpha_k + \beta_k)^{d-1}}{1 - (\alpha_k + \beta_k)} + (\alpha_k + \beta_k)^{d-1}s_{k,t+1} \, , \tag{4.5}$$

$$b_{i,j,t+d-1}(n_{d-1}) = (\alpha_i + \beta_i)(\alpha_j + \beta_j) + \alpha_i \alpha_j f_{i,j,t+d-1}(n_{d-1}), \quad (4.6)$$

$$f_{i,j,t+d}(n_d) = \frac{1}{(E|\tilde{u}_{i,t}|E|\tilde{u}_{j,t}|)} E\left[|\tilde{u}_{i,t+d}||\tilde{u}_{j,t+d}|\middle|\Delta_{t+d} = n_d\right] - 1. \quad (4.7)$$

For the expectation in (4.7), if we assume that the U_t's are jointly Gaussian then, it has a closed-form solution which involves a hyper-geometric function with the correlation between $\tilde{u}_{i,t+d}$ and $\tilde{u}_{j,t+d}$, which is known, as an argument[4]:

$$f_{i,j,t+d}(n_d) = \frac{2\left((1-\Gamma_{i,j}(n_d)^2)^2 + 2\Gamma_{i,j}(n_d)^2 HG\left(\frac{1}{2}, 2, \frac{3}{2}, \frac{-\Gamma_{i,j}(n_d)^2}{1-\Gamma_{i,j}(n_d)^2}\right)\right)}{2\sqrt{1-\Gamma_{i,j}(n_d)^2}} - 1,$$

$$HG(a,b,c,z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k z^k}{(c)_k \, k!}$$

where $(x)_k = x(x+1)\cdots(x+k)$.

In the case where U_t is not Gaussian and a closed-form solution cannot be found, $f_{i,j,t+d}(n_d)$ could be evaluated by numerical integration. However this would have to be done only N times because n_d can take only N possible values. In any case, for the form of the distribution of U_t, a stronger stand must be taken than only saying that it has mean zero and an identity matrix for the variance.

Keeping in mind that $b_{i,j,t+d-m}(n_{d-m})$ in equation (4.3) depends on the state of the Markov chain at time $t+d-m$ we next integrate out the Markov chain. Doing so we get

$$\begin{aligned}
& E_t[s_{i,t+d}s_{j,t+d}\Gamma_{i,j}(\Delta_{t+d})] \\
= & \sum_{l=1}^{d-1} a_{i,j,t+d-l} \sum_{n_d=1}^{N} \Gamma_{i,j}(n_d) \sum_{n_{d-1}=1}^{N} b_{i,j,t+d-1}(n_{d-1})^{\mathbb{1}_{\{l>1\}}} \pi_{n_{d-1},n_d} \cdots \times \\
& \qquad \sum_{n_1=1}^{N} b_{i,j,t+1}(n_1)^{\mathbb{1}_{\{l>d-1\}}} \pi_{n_1,n_2}\, \xi_{n_0,t+1|t} \\
+ & \; s_{i,t+1}s_{j,t+1} \sum_{n_d=1}^{N} \Gamma_{i,j}(n_d) \sum_{n_{d-1}=1}^{N} b_{i,j,t+d-1}(n_{d-1}) \pi_{n_{d-1},n_d} \cdots \times \\
& \qquad \sum_{n_1=1}^{N} b_{i,j,t+1}(n_1) \pi_{n_1,n_2}\, \xi_{n_0,t+1|t}. \quad (4.8)
\end{aligned}$$

We are able to compute multi-step ahead conditional expectations of the whole variance matrix for two reasons. The first is that the conditional expectations of the correlation matrix are given by the summation of a constant times a probability which is linearly updated.

[4]Computed with Mathematica.

The second is the use of a model for the conditional standard deviation (ARMACH) instead of the variance. Note that the use of the ARMACH model is not required. If another univariate model for the conditional volatility is obviously better and if analytic computation of multi-step ahead conditional expectations are not of interest then this model should be used.

5. Application to exchange rate data

In this section we apply the RSDC model to the exchange rate dataset used by Harvey, Ruiz, and Shephard (1994) and Kim, Shephard, and Chib (1998). This dataset contains four weekdays close exchange rates (Pound, Deutschmark, Yen, Swiss-Franc all against the U.S. dollar) over the period 1/10/81 to 28/6/85. The number of observation is 946. We first take 100 times the first difference of the logarithm of each series, minus the sample mean, before applying directly our variance model (these are our filtered series). We employ this dataset because Harvey, Ruiz, and Shephard (1994) use it to present a multivariate stochastic volatility model where they assume that correlations are constant through time. Using our model we can check if their assumption was reasonable.

To perform the out-of-sample analysis, we used the next 255 observations of the four series (from 01/07/85 to 02/07/86). These data are taken from Weigend, Huberman, and Rumelhart (1992). The augmented series are plotted in Figure 1.

The results are generated using Ox version 3.30 on Linux [see Doornik (1999)]. The estimation results that we present in the various tables are for full one-step maximum likelihood estimation.

5.1. RSDC model with two regimes

We first present results for the models with two regimes. Models with three regimes are studied in the following subsection. The results for the unrestricted and restricted models are presented in Table 1. We only present the detailed results for the RSDC with ARMACH since the results for the correlation models do not depend on the univariate model for the standard deviations. It is an indication that we can replace the traditional GARCH by the ARMACH or that the correlation models are robust to the specification of the standard deviations. For the restricted RSDC model we present the correlation matrix in each regime and their standard deviations computed with the Delta method instead of the matrix Γ and the value of $\lambda(2)$ [we use the identification scheme of equation (2.9) when doing the one-step estimation] so that the results are directly comparable to those of the unrestricted model.

The results for the univariate volatility models are not reported since they are similar to the usual findings with this type of financial series. The impact on the log-likelihood of replacing the ARMACH model by the GARCH model is an increase of about 15 points. An information criterion such as the BIC would indicate that the GARCH model is much better than the ARMACH model. The value of the log-likelihood for all the models estimated in this work are presented in Table 4.

Looking at the Table 1 and Figure 2 where we have plotted for the unrestricted model the smoothed probabilities of being in regime one at each point in time, we see that the correlations appear to be dynamic. Figure 2 shows that we frequently move between both regimes and there is little uncertainty about the regime we are in at each point in time. The process is spending more time in regime one and spells in regime two are shorter on average than in regime one. This is explained by the estimate of the transition probability matrix, which is very similar across the various models with two regimes. The probability of being in regime one at time $t+1$ conditional on being in regime one at time t, $\pi_{1,1}$, is around 0.93. That means a high level of persistence in the Markov chain because the probability of spending the next five days in regime one is $0.93^5 = 0.70$. In comparison, for regime two this probability is $0.67^5 = 0.14$. This illustrates that 0.93 and 0.67, although both high probabilities, are very different.

As for the value of the correlations in each regime, the results for the restricted model are similar to those of the unrestricted model. Under the unrestricted model, the magnitude of all the correlations in regime two is smaller than in regime one. So the hypothesis of the restricted version of the model that there is an ordering in the magnitude of the correlations across the different regimes seems plausible. The hypothesis that they all decrease in the same proportion is less supported by the data. The LR test for this hypothesis is 27.2 and should follow a Chi-square distribution with five degrees of freedom. The 1 % critical value being 15.09, we reject the restricted version of the model.

We mentioned at the end of Section 3 that a LR test of one regime versus two does not asymptotically follow a Chi-square distribution with degrees of freedom equal to the number of extra parameters. The increase in the likelihood by going from one regime [which is the CCC model of Bollerslev (1990)] to two regimes is so high, more than 250 points, that we don't need a formal test to reject the model with one regime. Using the results of Davies (1987), as applied in Garcia and Perron (1996, Appendix A), we know that the unknown p-value is less than 0.1%.

We also ran the following Monte Carlo simulation to gauge the loss of efficiency of the two-step estimation procedure relative to the one-step estimation. We simulated samples of size 1000 from the unrestricted RSDC model with ARMACH using the estimates in Table 1 as parameter values and estimated the model using the two-step and one-step estimation methods. We repeated the experiment until we obtained 250 samples for which all the likelihood maximizations converged.

The loss of efficiency differs across the "types" of parameters. The biggest loss of efficiency occurs for the ARMACH parameters where the mean absolute deviation (MAD) for the two-step estimates is 83 % higher than for the one-step estimates (the constant has a positive bias and the parameter for the lagged standard deviation has a negative bias). The MAD of the correlation matrix parameters are 45 % higher while the factor λ has a MAD 22 % higher. The MADs of the transition probabilities are almost equal (1.6 % higher).

5.2. RSDC model with three regimes

We next allow a third regime in the Markov chain. The estimation results for the unrestricted and restricted models are presented in Table 2. The increase of the log-likelihood is about 40 points for the unrestricted model and 50 points for the restricted model, while the third regime adds respectively eleven and five parameters. Using Davies' bound, we know that the p-value of a LR test of two regimes versus three is less than 1% for both models.

The addition of a third regime now allows the data to identify two regimes with high correlations and one regime of very low correlations. Again, we have in general the same ranking of the correlations across the regimes with the unrestricted model. The correlations in regime one are larger than in regime two, which are larger than in regime three. We can again test the restricted model versus the unrestricted. In this case, we compare twice the difference of the likelihood, i.e. 8 for the ARMACH, to a Chi-square with ten degrees of freedom and doing so we don't reject the restricted model.

Looking at Figure 3, we see that the Markov chain is spending most of its time in regimes of high correlations (regime one and two). Very rarely does the chain goes in the regime of low correlation. Again, we see that most of the time we have a strong idea about which regime we are in at every point in time as the smoothed probabilities are close to either zero or one most of the time. Examining more closely the correlation matrix for each regime, the smoothed probabilities and the smoothed correlations in Figure 4, we see that with a third regime, the Markov chain is beginning to identify what could be outliers[5]. The chain is going rarely in a regime which is very different from the others. This could be seen as an indicator that three regimes is enough.

5.3. DCC model

To evaluate the relative performance of our model to fit the data we estimate the DCC model of Engle (2002), as presented in Section 2.1, taking $p = 1$ and $q = 1$. The results for the model with ARMACH models for the standard deviations are in Table 3 (the results are again robust to the univariate volatility model). We report the results for the full one-step maximum likelihood estimation.

What is interesting is to compare the log-likelihood of the different models (all the results are in Table 4). There is a big difference in the level of the log-likelihood between the RSDC model and the DCC model. For our restricted model with two regimes (and GARCH model) the log-likelihood is 100 points higher than the DCC-GARCH while the regime switching model has only one more parameter. The difference in the log-likelihood is 114.5 points between the unrestricted RSDC model with two regimes and the DCC-GARCH at the cost of seven additional parameters.

Because our regime switching model and the DCC model are not nested we cannot

[5]An event study cannot explain most of the periods when we are in the third regime of low correlations.

perform a likelihood ratio test to verify if the increase in the likelihood is significant. One valid test for testing non-nested models is proposed by Rivers and Vuong (2002, Section 4). With this test, we reject at the 10% level[6] the hypothesis that the DCC model is as close to the true model as the RSDC model. Another approach for choosing one model over the other could be the use of information criteria. Ultimately, we are not interested in rejecting a model. A better solution would be to combine the forecasts from these different models.

Another interesting comparison is the correlations extracted from both models. If we compare the smoothed correlations from the unrestricted RSDC model (Figure 4) with the correlations from the DCC (Figure 5), we see that the correlations are generally smoother with the switching regime model. One interesting implication of smoother patterns for the correlations is for the computation of VaR and portfolio allocation. If the time-varying correlations are smoother, then the gain from portfolio diversification will also be smoother which might imply a smoother pattern for the VaR and portfolio weights.

It might be intriguing that the regime switching gives a higher value for the likelihood than the DCC because both models imply a VARMA dynamic for the outer-product of the standardized innovations. The DCC equation (2.6) can be rewritten as

$$vech(\tilde{U}_t\tilde{U}_t') = \bar{\Gamma}_1 + \sum_{i=1}^{\max(p,q)} (a_i + b_i) vech(\tilde{U}_{t-i}\tilde{U}_{t-i}') + V_t - \sum_{j=1}^{p} b_j V_{t-j}$$

where $V_t = vech(\tilde{U}_t\tilde{U}_t') - vech(\tilde{\Gamma}_t)$. The operator $vech$ stacks in a vector the elements on and below the diagonal of each column of a matrix. From this equation we see that both the AR and MA operators are scalar.

The VARMA representation of the regime switching model for the correlations presented in this work is derived in Dufour and Pelletier (2003):

$$\prod_{n=1}^{N-1} (1 - e_n\ L) vech(\tilde{U}_t\tilde{U}_t') = \bar{\Gamma}_2 + V_t + \sum_{n=1}^{N-1} B_n V_{t-n}$$

with V_t a white noise process and the e_ns are the eigenvalues of the transition matrix different than 1. The matrices of parameters B_n are function of the correlation matrices and the transition matrix. From this, we see that one reason why the regime switching model can be doing better is because the MA operator is not restricted to be scalar.

5.4. Out-of-sample

Out-of-sample evaluation of volatility models are complicated by the fact that the conditional variance matrix is unobserved. We can still construct a proxy. A common and

[6] No parameter is treated as a nuisance parameter. We use the suggested Newey and West (1987) estimator for the variance. We tried a wide range of values for the truncation lag in the computation of the variance.

successful approach is to use cumulative cross-product of intraday returns over the forecast horizon; e.g. see Andersen, Bollerslev, Diebold, and Labys (2001) and Andersen, Bollerslev, and Lange (1999). Another approach presented in Brandt and Diebold (2004) combines the daily range for the exchange rate between every pair of currencies and no-arbitrage conditions. Unfortunately, we don't have intra-day returns or information on the range for the ten possible pairs of currencies for the period following our in-sample analysis. The out-of-sample analysis will then be performed with daily returns.

We compute one-day and five-day ahead forecasts of the variance matrix. For the models which don't have an analytic expression for the five-day ahead forecasts (DCC model and models using GARCH for the variances), we assume that the innovations U_t have a joint Gaussian distribution and compute the expectations using 5000 simulated sample paths. We also assume a Gaussian distribution for the expectation in (4.7). The parameter estimates of the different models are updated every 5 days to reduce computation time. We performed the out-of-sample analysis over 250 days, starting five days after the end of the in-sample so the first five-day ahead forecast does not overlap with the in-sample.

Denote by $H_{t+d|t}$ the d-day ahead forecast of the variance matrix. As in Andersen, Bollerslev, and Lange (1999), we use the following two criteria to compare the quality of the volatility forecasts:

$$RMSE_d = \left(\frac{1}{K^2} \sum_{i,j} E\left[(H_{i,j,t+d|t} - y_{i,t+d} y_{j,t+d})^2 \right] \right)^{1/2} \tag{5.1}$$

$$MAD_d = \frac{1}{K^2} \sum_{i,j} E \left| H_{i,j,t+d|t} - y_{i,t+d} y_{j,t+d} \right| \tag{5.2}$$

Criteria based on absolute deviations are sometimes preferred because they are more robust to outliers. To isolate the impact of the correlation model we only present results for models with ARMACH for the standard deviations, with the exception of the CCC-GARCH so we can compare the GARCH and ARMACH models.

Table 5 reports estimates of the two criteria for the two horizons. Comparing the results for the CCC-ARMACH and CCC-GARCH model, we see that for out-of-sample forecasts the ARMACH is performing better than the GARCH model. As for the correlation models, according to MAD_1, MAD_5 and $RMSE_1$ the four regime switching models are better than the DCC model. For the $RMSE_5$ criterion, all the dynamic correlation models are about equal. Over this sample period, the model with constant correlations is performing the best. One puzzling result is that the criteria for five-day ahead forecasts are lower than for one-day ahead forecasts. This could be an indication that there are some uncaptured dynamics in the first moment of the series.

6. Conclusion

In this work we propose a new model for the variance between multiple time series, the Regime Switching Dynamic Correlation (RSDC) model. We decompose the covariances into correlations and standard deviations and both the correlations and the standard deviations are dynamic. For the correlation matrix, we propose a regime switching model. It is constant within a regime but different across regimes. The transitions between the regimes are governed by a Markov chain. This property of constant correlation could have important impacts, namely for the computation of Value-at-Risk and for dynamic portfolio allocation. We also present a restricted version of our model where the changes across correlations in a given regime are proportional. This regime switching model can be seen as a mid-point between the CCC model of Bollerslev (1990) where the correlations are constant and models such as the DCC model of Engle (2002) where the correlation matrix change at every point in time.

One appealing feature of this model for the correlations is that when combined with the ARMACH model [Taylor (1986) and Schwert (1989)] for the conditional standard deviations, it allows analytic computation of multi-step ahead conditional expectations of the whole variance matrix. The ARMACH model is a GARCH-type model for the conditional standard deviations instead of the conditional variance.

The evaluation of the likelihood is done with Hamilton's filter because of the unobserved Markov chain. By decomposing the variance matrix into a diagonal matrix of standard deviations and a correlation matrix, we can use a two-step estimation procedure as in Engle (2002). Combining this two-step estimation procedure with either correlation targeting (for the restricted model) or the EM algorithm (for the unrestricted model) breaks the curse of dimensionality, i.e. the number of parameters in every non-linear estimation is not a function of the number of time series.

An application of this model to four major exchange rate series illustrates its good behavior. A comparison of our regime switching model with the DCC model of Engle (2002) shows that our model has a better performance in and out-of-sample. An interesting aspect of our regime switching model is that we find strong persistence in the Markov chain, which produces smoother time-varying correlations than the DCC model.

Possible extensions in future work includes the addition of relations between correlations and standard deviations as the work of Andersen, Bollerslev, Diebold, and Labys (2001) seems to indicate. Identification of the number of regimes in the Markov chain is also an ongoing research project.

References

ANDERSEN, T., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

ANDERSEN, T. G., T. BOLLERSLEV, AND S. LANGE (1999): “Forecasting financial market volatility: sample frequency vis-à-vis forecast horizon,” *Journal of Empirical Finance*, 6, 457–477.

ANDREWS, D. W. K. (1999): “Estimation when a parameter is on a boundary,” *Econometrica*, 67(6), 1341–1383.

——— (2001): “Testing when a parameter is on the boundary of the maintained hypothesis,” *Econometrica*, 69(3), 683–734.

ANG, A., AND J. CHEN (2002): “Asymmetric Correlations of Equity Portfolios,” *Journal of Financial Economics*, 63(3), 443–494.

BARNARD, J., R. MCCULLOCH, AND X.-L. MENG (2000): “Modelling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage,” *Statistica Sinica*, 10(4).

BAUWENS, L., S. LAURENT, AND J. ROMBOUTS (2003): “Multivariate GARCH models: a survey,” Discussion Paper 2003/31, CORE discussion paper.

BOLLERSLEV, T. (1986): “Generalized Autoregressive Conditional Heteroskedasticity,” *journal of Econometrics*, 31, 307–327.

——— (1990): “Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model,” *Review of Economics and Statistics*, 72, 498–505.

BOLLERSLEV, T., AND R. F. ENGLE (1993): “Common Persistence in Conditional Variances,” *Econometrica*, 61(1), 167–186.

BOLLERSLEV, T., R. F. ENGLE, AND J. M. WOOLDRIDGE (1988): “A capital asset pricing model with time-varying covariances,” *The Journal of Political Economy*, 96(1), 116–131.

BRANDT, M., AND F. DIEBOLD (2004): “A no-arbitrage approach to range-based estimation of return covariances and correlations,” *Journal of Business*, forthcoming.

CAI, J. (1994): “A Markov model of switching-regime ARCH,” *JBES*, 12, 309–316.

DAVIDIAN, M., AND R. J. CARROLL (1987): “Variance Function Estimation,” *Journal of the American Statistical Association*, 82, 1079–1091.

DAVIES, R. B. (1987): “Hypothesis testing when a nuisance parameter is present only under the alternative,” *Biometrika*, 74(1), 33–43.

DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN (1977): “Maximum likelihood from incomplete data via the EM algorithm,” *J. Roy. Statist. Soc. Ser. B*, 39(1), 1–38, With discussion.

DIEBOLD, F., J.-H. LEE, AND G. WEINBACH (1994): “Regime switching with time-varying transition probabilities,” in *Nonstationary time series analysis and cointegration*, ed. by C. P. Hargreaves, Advanced Texts in Econometrics, pp. 283–302, New York. The Clarendon Press Oxford University Press.

DIEBOLD, F. X., AND M. NERLOVE (1989): "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model," *Journal of Applied Econometrics*, 4(1), 1–21.

DING, Z., AND R. F. ENGLE (2001): "Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing," Discussion Paper FIN-01-029, Stern School of Business, New York University.

DOORNIK, J. A. (1999): "Object-Oriented Matrix Programming Using Ox, 3rd ed.," Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik.

DUEKER, M. (1997): "Markov-switching in GARCH processes and mean-reverting stock market volatility," *JBES*, 15, 26–34.

DUFOUR, J.-M. (2002): "Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics," *Journal of Econometrics*, forthcoming.

DUFOUR, J.-M., AND D. PELLETIER (2003): "Linear estimation of univariate and multivariate volatility models," Unpublished manuscript.

EDWARDS, S., AND R. SUSMEL (2001): "Volatility dependence and contagion in emerging equity markets," *Journal of Development Economics*, 66, 505–532.

——— (2003): "Interest-rate volatility in emerging markets," *Review of Economics and Statistics*, 85, 328–348.

ENGLE, R. (2002): "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business and Economic Statistics*, 20(3), 339–350.

ENGLE, R., AND K. SHEPPARD (2001): "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH," UCSD Discussion Paper 2001-15.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

ENGLE, R. F., AND K. F. KRONER (1995): "Multivariate simultaneous generalized ARCH," *Econometric Theory*, 11, 122–150.

ENGLE, R. F., V. K. NG, AND M. ROTHSCHILD (1990): "Asset pricing with a factor-ARCH covariance structure, empirical estimates for treasury bills," *Jopurnal of Econometrics*, 45, 213–237.

ENGLE, R. F., AND R. SUSMEL (1993): "Common Volatility in International Equity Markets," *Journal of Business and Economic Statistics*, 11(2), 167–176.

FRANCQ, C., AND J.-M. ZAKOÏAN (2004): "The L^2-structures of standard and switching-regime GARCH models," Unpublished, http://www.crest.fr/pageperso/ls/zakoian/zakoian.htm.

GARCIA, R., AND P. PERRON (1996): "An analysis of the real interest rate under regime shifts," *Review of Economics and Statistics*, 78(1), 111–125.

GEWEKE, J., AND G. AMISANO (2001): "Compound Markov Mixture Models with Applications in Finance," Unpublished manuscript.

GRANGER, C. W. J., AND Z. DING (1993): "Some Properties of Absolute Return - An Alternative Measure of Risk," Discussion Paper 93-38, University of California, San Diego.

GRAY, S. (1996): "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42, 27–62.

HAAS, M., S. MITTNIK, AND M. S. PAOLELLA (2003): "Volatility Dynamics in Exchange Rates: Markov Switching GARCH-Mixtures," Institute of Statistics, University of Munich.

HAMILTON, J., AND R. SUMSEL (1994): "Autoregressive conditional heteroskedasticity and change in regime," *Journal of Econometrics*, 64, 307–333.

HAMILTON, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

——— (1994): *Time Series Analysis*. Princeton University Press.

——— (1996): "Specification Testing in Markov-Switching Time-Series Models," *Journal of Econometrics*, 70, 127–157.

HARVEY, C., E. RUIZ, AND N. SHEPHARD (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

KIM, C.-J. (1994): "Dynamic linear models with Markov-switching," *Journal of Econometrics*, 60(1-2), 1–22.

KIM, S., N. SHEPHARD, AND S. CHIB (1998): "Stochastic Volatility: Likelihood Inference and Comparison With ARCH Models," *Review of Economic Studies*, 65, 361–393.

KING, M., E. SENTANA, AND S. WADHWANI (1994): "Volatility and Links Between National Stock Markets," *Econometrica*, 62(4), 901–933.

KLAASSEN, F. (2002): "Improving GARCH volatility forecasts with Regime-Switching GARCH," *Empirical Economics*, 27, 363–394.

LEDOIT, O., P. SANTA CLARA, AND M. WOLF (2003): "Flexible Multivariate GARCH Modeling With an Application to International Stock Markets," *Review of Economics and Statistics*, 85, forthcoming.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large sample estimation and hypothesis testing," in *Handbook of econometrics, Vol. IV*, pp. 2111–2245. North-Holland, Amsterdam.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

NG, V., R. F. ENGLE, AND M. ROTHSCHILD (1992): "A multi-dynamic-factor model for stock returns," *Journal of Econometrics*, 52, 245–266.

PAGAN, A. (1986): "Two Stage and Related Estimators and Their Applications," *The Review of Economic Studies*, 53(4), 517–538.

RIVERS, D., AND Q. VUONG (2002): "Model selection tests for nonlinear dynamic models," *Econometrics Journal*, 5, 1–39.

ROSS, S. R. (1993): *Introduction to Probability Models*. Academic Press, fifth edn.

SCHWERT, G. W. (1989): "Why Does Stock Market Volatility Change Over Time," *Journal of Finance*, 44(5), 1115–1153.

TAYLOR, S. (1986): *Modelling Financial time Series*. John Wiley and Sons.

TSE, Y. K., AND K. C. TSUI (2002): "A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations," *Journal of Business and Economic Statistics*, 20(3), 351–362.

WEIGEND, A. S., B. A. HUBERMAN, AND D. E. RUMELHART (1992): "Predicting sunspots and exchange rates with connectionist networks," in *Nonlinear Modeling and Forecasting*, ed. by M. Casdagli, and S. Eubank, pp. 395–432. Addison-Wesley.

Appendix

A. Proofs

PROOF OF THEOREM **3.1**

Scaling (3.7) by $1/T$, the uniform strong law of large numbers implies that a.s. we get

$$\mathcal{L}_1 = -\frac{1}{2}E_{\boldsymbol{\theta_0}}\left[\sum_{k=1}^{K}\left(\log 2\pi + 2\log s_{k,t} + \frac{y_{k,t}^2}{s_{k,t}^2}\right)\right] \tag{A.1}$$

where $E_{\boldsymbol{\theta_0}}$ is the expectation with respect to the true density. Similarly, scaling (2.3) by $1/T$, a.s. we get

$$\mathcal{L} = -\frac{1}{2}E_{\boldsymbol{\theta_0}}\left[K\log 2\pi + \log|\mathit{\Gamma}_t| + 2\sum_{k=1}^{K}\log s_{k,t} + \tilde{U}_t'\mathit{\Gamma}_t^{-1}\tilde{U}_t\right] \tag{A.2}$$

If we can show that both sets of first order conditions with respect to $\boldsymbol{\theta_1}$ are satisfied for the same vector of parameters then we can conclude that the estimates from (3.7) will converge to their true value.

Denoting by $\theta_{k,j}$ one of the parameters in $\boldsymbol{\theta_1}$ that appears in the expression of $s_{k,t}$, we can write the first order conditions for $\mathcal{L}_1$ as

$$\frac{\partial\mathcal{L}_1}{\partial\theta_{k,j}} = E_{\boldsymbol{\theta_0}}\left[-\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial\theta_{k,j}} + \tilde{u}_{k,t}^2\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial\theta_{k,j}}\right] = 0. \tag{A.3}$$

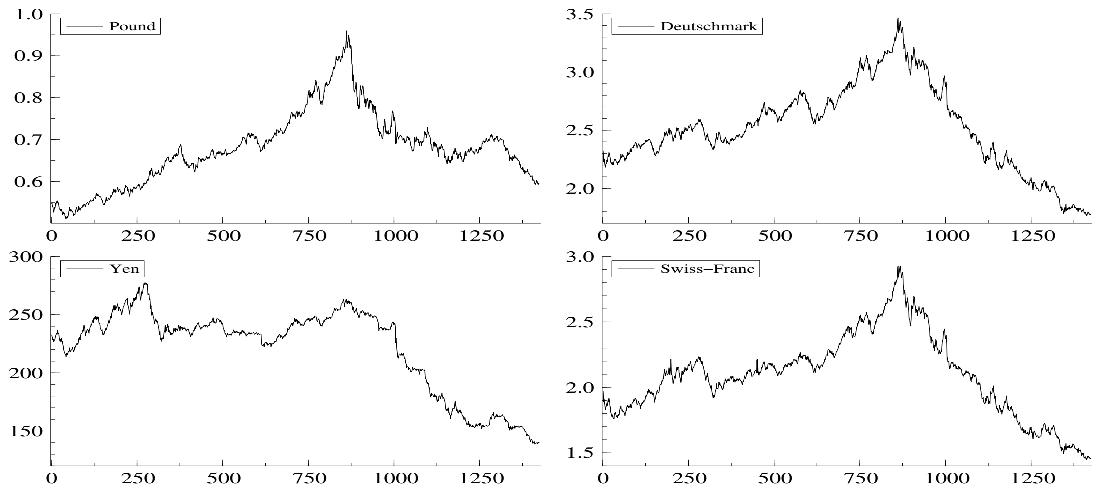
While the first order conditions for $\mathcal{L}$ are

$$\frac{\partial\mathcal{L}}{\partial\theta_{k,j}} = E_{\boldsymbol{\theta_0}}\left[-\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial\theta_{k,j}} + \tilde{U}_t'\mathit{\Gamma}_t^{-1}\begin{bmatrix}0\\ \vdots\\ \tilde{u}_{k,t}\\ \vdots\\ 0\end{bmatrix}\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial\theta_{k,j}}\right] = 0 \tag{A.4}$$

Using the trace operator we can easily see that $\tilde{U}_t'\mathit{\Gamma}_t^{-1}[0,\ldots,\tilde{u}_{k,t},\ldots,0]'$ is a random variable with unit mean, just like $\tilde{u}_{k,t}^2$. From this we see that the value of $\theta_{k,j}$ that will solve equation (A.4) will also solve equation (A.3). For the rest of the proof see Newey and McFadden (1994).

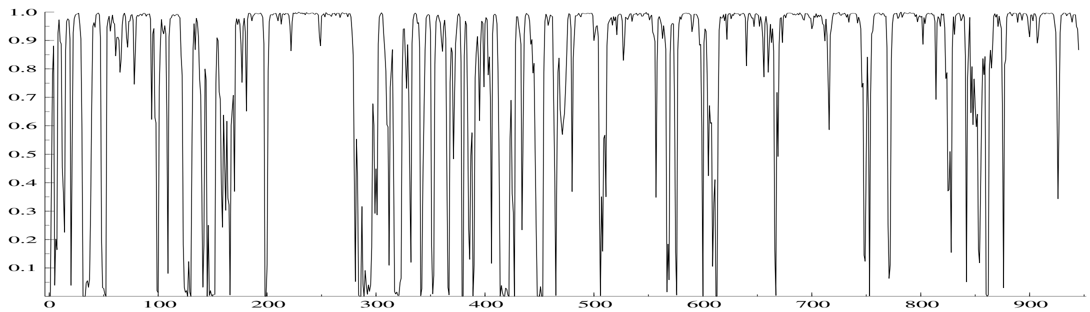
□

Figure 1: Exchange rate series



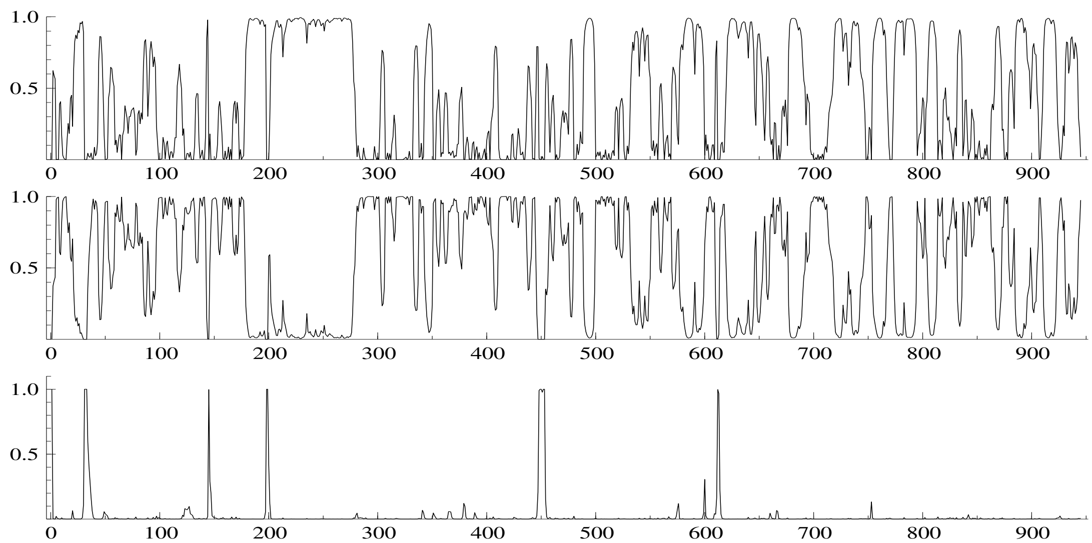
This figure presents the four exchange rate series (Pound, Deutschmark, Yen, Swiss-Franc all against the U.S. dollar) over the period October 1, 1981 to July 2, 1986. The source of the data is described at the beginning of Section 5.

Figure 2: Smoothed probabilities of the two-regime RSDC model



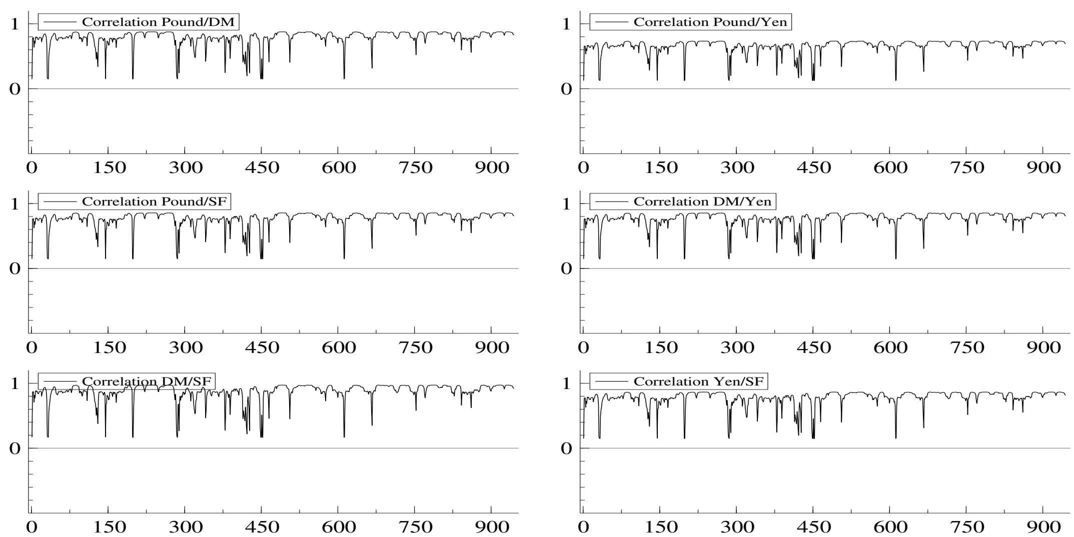
This figure presents the smoothed probabilities of being in regime 1 (high correlations) for the unrestricted RSDC model with two regimes and ARMACH models for the standard deviations. This model is described in Section 2.1. The results are based on the in-sample data, i.e. from October 1, 1981 to June 28, 1985.

Figure 3: Smoothed probabilities of the three-regime RSDC model



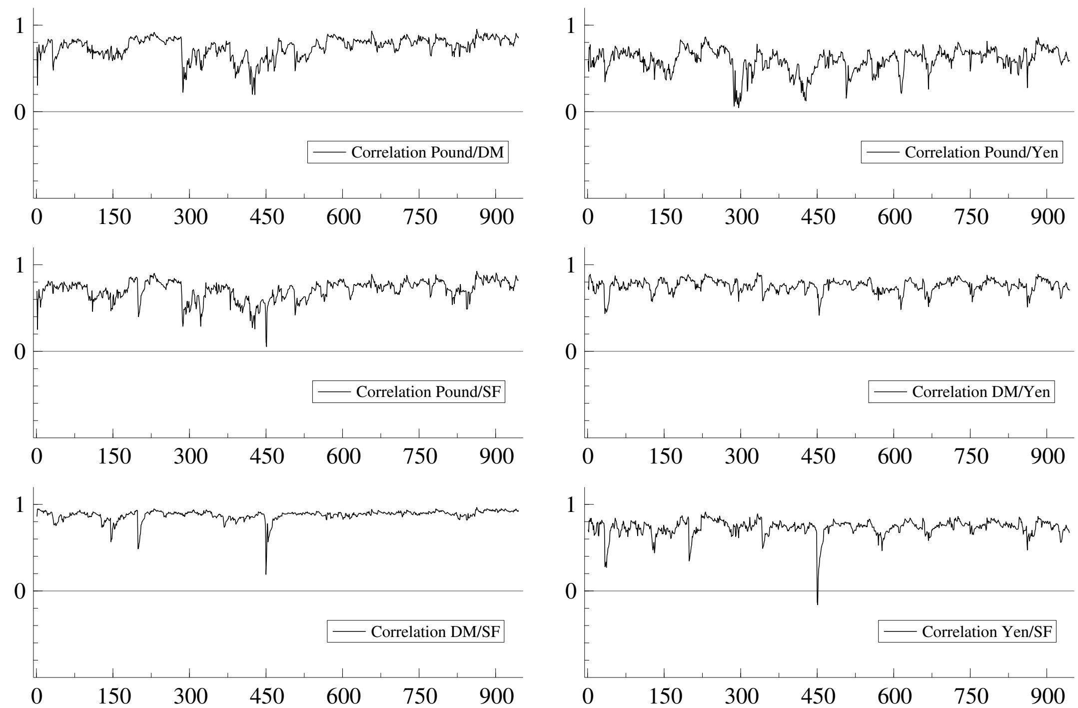
This figure presents the smoothed probabilities of being in each regime for the unrestricted RSDC model with three regimes and ARMACH models for the standard deviations. This model is described in Section 2.1. The results are based on the in-sample data, i.e. from October 1, 1981 to June 28, 1985. The probabilities of being in regime 1 (high correlations), regime 2 (medium correlations) and regime 3 (low correlation) are respectively in the top, middle and bottom panel.

Figure 4: Smoothed correlations of the three-regime RSDC model



This figure presents the smoothed correlations for the restricted RSDC model with three regimes and ARMACH models for the standard deviations. This model is described in Section 2.1. The results are based on the in-sample data, i.e. from October 1, 1981 to June 28, 1985.

Figure 5: Correlations of the DCC model



This figure presents the correlations for the DCC model with ARMACH models for the standard deviations. The DCC model is presented in Section 2.1. The results are based on the in-sample data, i.e. from October 1, 1981 to June 28, 1985.

Table 1: Parameters estimates of the two-regime RSDC models

	$\Gamma_{1,2}$	$\Gamma_{1,3}$	$\Gamma_{1,4}$	$\Gamma_{2,3}$	$\Gamma_{2,4}$	$\Gamma_{3,4}$
Unrestricted						
Regime 1	0.8754	0.7656	0.8569	0.8471	0.9510	0.8617
	(0.0292)	(0.0363)	(0.0283)	(0.0181)	(0.0061)	(0.0184)
Regime 2	0.4011	0.1859	0.3255	0.4739	0.5626	0.3250
	(0.0958)	(0.0996)	(0.1275)	(0.0843)	(0.1871)	(0.1666)
Restricted						
Regime 1	0.8549	0.7274	0.8347	0.8334	0.9479	0.8477
	(0.0233)	(0.0400)	(0.0241)	(0.0227)	(0.0069)	(0.0221)
Regime 2	0.3362	0.2861	0.3283	0.3278	0.3728	0.3334
	(0.1327)	(0.1138)	(0.1296)	(0.1294)	(0.1469)	(0.1316)

	$\pi_{1,1}$	$\pi_{2,2}$
Unrestricted	0.9291	0.6666
	(0.0356)	(0.0605)
Restricted	0.9473	0.6682
	(0.0254)	(0.0635)

This table presents the parameter estimates for the unrestricted and restricted RSDC models with two regimes and ARMACH models for the standard deviations. The unrestricted model is described in Section 2.1 while the restricted model is described in Section 2.2. The standard errors are in parenthesis. The results are based on the in-sample data, i.e. from October 1, 1981 to June 28, 1985.

Table 2: Parameters estimates of the three-regime RSDC models

	$\Gamma_{1,2}$	$\Gamma_{1,3}$	$\Gamma_{1,4}$	$\Gamma_{2,3}$	$\Gamma_{2,4}$	$\Gamma_{3,4}$
Unrestricted						
Regime 1	0.9491	0.8497	0.9298	0.8568	0.9251	0.8705
	(0.0101)	(0.0347)	(0.0667)	(0.0672)	(0.2257)	(0.0894)
Regime 2	0.6039	0.4189	0.5598	0.7222	0.8853	0.7238
	(0.0697)	(0.0831)	(0.1307)	(0.0381)	(0.1341)	(0.0590)
Regime 3	0.1850	0.0855	0.0730	0.2048	0.2199	0.0620
	(0.2592)	(0.0819)	(0.1263)	(0.0410)	(0.0989)	(0.0830)
Restricted						
Regime 1	0.8775	0.7348	0.8567	0.8550	0.9723	0.8649
	(0.0160)	(0.0275)	(0.0183)	(0.0143)	(0.0038)	(0.0141)
Regime 2	0.7835	0.6561	0.7649	0.7634	0.8682	0.7723
	(0.0225)	(0.0285)	(0.0236)	(0.0212)	(0.0195)	(0.0172)
Regime 3	0.1508	0.1262	0.1472	0.1469	0.1670	0.1486
	(0.0692)	(0.0581)	(0.0676)	(0.0674)	(0.0766)	(0.0682)
	$\pi_{1,2}$	$\pi_{1,3}$	$\pi_{2,1}$	$\pi_{2,3}$	$\pi_{3,1}$	$\pi_{3,2}$
Unrestricted	0.1248	0.0177	0.1153	0.0000	0.1271	0.2479
	(0.1168)	(0.0502)	(0.0746)	(0.0326)	(0.2189)	(0.2045)
Restricted	0.0686	0.0054	0.0797	0.0416	0.0305	0.4365
	(0.0200)	(0.0097)	(0.0288)	(0.0203)	(0.1151)	(0.1849)

This table presents the parameter estimates for the unrestricted and restricted RSDC models with three regimes and ARMACH models for the standard deviations. The unrestricted model is described in Section 2.1 while the restricted model is described in Section 2.2. The standard errors are in parenthesis. The results are based on the in-sample data, i.e. from October 1, 1981 to June 28, 1985.

Table 3: Parameter estimates of the DCC model

$\Gamma_{1,2}$	$\Gamma_{1,3}$	$\Gamma_{1,4}$	$\Gamma_{2,3}$	$\Gamma_{2,4}$	$\Gamma_{3,4}$	a	b
0.7554	0.6036	0.6849	0.7255	0.8682	0.6843	0.1088	0.8083
(0.0486)	(0.0524)	(0.0596)	(0.0383)	(0.0291)	(0.0515)	(0.0344)	(0.0571)

This table presents the parameter estimates for the DCC model with ARMACH models for the standard deviations. This model is described in Section 2.1. The standard errors are in parenthesis. The results are based on the in-sample data, i.e. from October 1, 1981 to June 28, 1985.

Table 4: Log-likelihood comparison of the various models

	Log-likelihood	Nb. par.
Unrestricted 3-regime GARCH	-1955.3	38
Restricted 3-regime GARCH	-1961.3	26
Unrestricted 3-regime ARMACH	-1971.7	38
Restricted 3-regime ARMACH	-1975.7	26
Unrestricted 2-regime GARCH	-1994.7	27
Restricted 2-regime GARCH	-2009.0	21
Unrestricted 2-regime ARMACH	-2011.6	27
Restricted 2-regime ARMACH	-2025.2	21
DCC-GARCH	-2109.2	20
DCC-ARMACH	-2137.8	20
CCC-GARCH	-2272.1	18
CCC-ARMACH	-2301.8	18

This table presents the value of the log-likelihood and the number of parameters for all the models estimated in this work. The rows are sorted in descending order of the log-likelihood value.

Table 5: Forecast criteria for variance matrices

	MAD_1	MAD_5	$RMSE_1$	$RMSE_5$
CCC-ARMACH	0.733	0.719	2.126	2.134
CCC-GARCH	0.756	0.730	2.140	2.136
DCC-ARMACH	0.787	0.785	2.128	2.123
2-regime restricted ARMACH	0.778	0.756	2.118	2.122
3-regime restricted ARMACH	0.776	0.754	2.116	2.122
2-regime unrestricted ARMACH	0.783	0.762	2.118	2.123
3-regime unrestricted ARMACH	0.758	0.741	2.119	2.129

This table presents the forecasted conditional variance matrices with the cross-product of the daily returns that serve as a proxy for the unobservable true conditional variance matrix. The criteria RMSE and MAD are described in equation (5.1) and (5.2). The results are based on the out-of-sample data, i.e. July 8, 1985 to July 2, 1986.