THE UNIVERSITY OF MELBOURNE

Melbourne Institute Working Paper Series

Working Paper No. 18/07

A Metropolis-in-Gibbs Sampler for Estimating Equity Market Factors

Sarantis Tsiaplias

MELBOURNE INSTITUTE
of Applied Economic and Social Research

A Metropolis-in-Gibbs sampler for estimating equity market factors*

Sarantis Tsiaplias
Melbourne Institute of Applied Economic and Social Research
The University of Melbourne

Melbourne Institute Working Paper No. 18/07

ISSN 1328-4991 (Print)

ISSN 1447-5863 (Online)

ISBN 978-0-7340-3250-8

June 2007

* I thank Guay Lim, Peter Summers, Michael Chua, and John Creedy for their helpful comments and suggestions. This work was funded by a Melbourne Research Scholarship.

Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
Victoria 3010 Australia
***Telephone* (03) 8344 2100**
***Fax* (03) 8344 2111**
***Email* melb-inst@unimelb.edu.au**
***WWW Address* http://www.melbourneinstitute.com**

Abstract

A model incorporating common Markovian regimes and GARCH residuals in a persistent factor environment is considered. Given the intractable and approximate nature of the likelihood function, a Metropolis-in-Gibbs sampler with Bayesian features is constructed for estimation purposes. The common factor drawing procedure is effectively an exact derivation of the Kalman filter with a Markovian regime component and GARCH innovations. To accelerate the drawing procedure, approximations to the conditional density of the common component are considered. The model is applied to equity data for 18 developed markets to derive global, European, and country specific equity market factors.

JEL classification: C32, C51

Keywords: Common factors, Kalman filter, Markov switching, Monte Carlo, GARCH, Equities

1. Introduction

This paper seeks to construct a unified factor model incorporating various facets of return persistence, as well as co-movement possibilities in the time and discrete regime-dependent contexts. The theoretical capacity for persistence and feedback in asset prices is discussed extensively in the behavioural finance literature (see, for example, De Long et al., 1990a,b; Daniel, Hirshleifer, and Subrahmanyam, 1998, 2001; Hong and Stein, 1999). In turn, the specification of discrete regime-dependent volatility enables assessment of the effects of common shocks across the various volatility-regime structures, while asset-specific idiosyncratic volatility provides the capacity for deriving time and volatility-dependent measures of asset co-movement. To incorporate the desired persistence and co-movement properties, the factor model allows for persistent common factors and common regime switching, while sensitivity to the common factors is not restricted to the contemporaneous set of common information. Similarly, the asset-specific (or idiosyncratic) components are not restricted to the zero-persistence or homoscedastic states. In this respect, asset-specific volatilities are specified as GARCH processes thereby implicitly accounting for residual heteroscedasticity.

To facilitate the estimation and obtain the requisite factors, a dynamic common factor model is constructed and applied to developed national equity market returns. In theory, the model may be exactly estimated using Kalman filter with Markovian regimes and GARCH innovations. An exact likelihood function for such a filter cannot, however, be constructed (King, Sentana, and Wadhwani, 1994; Kim and Nelson, 1998). Given the intractable and approximate nature of the adopted model's likelihood function, a Metropolis-in-Gibbs sampler is constructed to obtain exact Bayesian inferences for the dynamic factor model in the presence of persistence in the common and idiosyncratic components and their respective volatility structures. The sampler effectively constructs an exact Kalman filter in the presence of Markovian regimes and GARCH innovations. As such, the model may be interpreted in terms of King, Sentana, and Wadhwani (1994), but extended to incorporate common Markovian regimes and exact estimation.

The paper is structured as follows. Section 2 describes the adopted model structure and contrasts its properties with those of a basic model. Section 3 defines the data used in the empirical application of the adopted model. The estimation procedure is detailed in Section 4, while Section 5 considers approximations for accelerating the common factor drawing process. Section 6 examines the convergence properties of the sampler and the estimation results. The paper concludes with Section 7.

2. The model structure

This section defines a common factor structure that treats equity market volatilities as functions of Markovian regime switching and GARCH processes while concurrently providing for persistent common and idiosyncratic asset factors. The approach is contrasted with a basic model to facilitate comparison of the new model with generally adopted approaches. Global and European common factors are specified and common volatility is determined using global and European discrete Markovian regime processes.

2.1 A basic model

In the single factor case, the *N*-vector r_t, representing excess returns for the *N* markets, is treated as a function of a single latent factor common to each market.

$$r_t = \begin{bmatrix} B & \mathrm{I}_N \end{bmatrix} \begin{bmatrix} W_t & u_t' \end{bmatrix}' = BW_t + u_t \,, \tag{1}$$

$$u_t = G^{1/2} z_t \,, \tag{2}$$

$$W_t = \mu_w + w_t \,, \tag{3}$$

$$w_t \sim N(0, \sigma^2_{w,t}) \,, \tag{4}$$

$$E\left(W_a u_b\right) = 0 \;\; \forall a, b \in \underset{\sim}{t} \,, \tag{5}$$

where B is a vector of factor sensitivities, I_N is an identity matrix of order N, W_t is a scalar representing information common to all markets at time t, $\underset{\sim}{t}$ is the set of observations to time T, G is a diagonal matrix of order N and z_t is a multivariate standard normal vector $z_t \sim iidMVN\left(0_N, \mathrm{I}_N\right)$. The identity loading on u_t identifies the idiosyncratic component.

In the basic model, returns are assumed to be uncorrelated over time in both the idiosyncratic and common components. Accordingly, the lag polynomials in u_t and W_t, $\psi_i\left(L\right)=1-\psi_{i,1}L-...-\psi_{i,k(i)}L^{k(i)}$ and $\phi_w\left(L\right)=1-\phi_{w,1}L-...-\phi_{w,k(w)}L^{k(w)}$, are subject to the restriction $\psi_i\left(L\right)=\phi_w\left(L\right)=1$. The idiosyncratic components are deemed to incorporate information pertinent only to the relevant markets and are thereby presumed to be uncorrelated with the common factor. Given the lack of persistence in the idiosyncratic components, it is also assumed that investors place no emphasis on the historical path of the idiosyncratic component in constructing its expected value. In this respect, the conditional expectation of the idiosyncratic component is always zero and the model accords with the rational pricing forms that reject the existence of idiosyncratic pricing. Common factor pricing is accommodated through the provision for non-zero μ_w.

The volatility of the common component $\sigma^2_{w,t}$ is modelled as a discrete switching process:

$$\sigma^2_{w,t} = \sum_{m=1}^{M(w)} \sigma^2_{w,m} S_{w,m,t}, \tag{6}$$

$$\sigma_{w,m+1} > \sigma_{w,m} \forall m. \tag{7}$$

The latent state variable $S_{w,m,t}$ takes the value unity if state $m = m\left(w\right)$, $m\left(w\right) \in M_w$, $M_w = \left\{m \in \mathcal{N} : m \le M\left(w\right) \in \mathcal{N}\right\}$, and zero otherwise. The probability of state m prevailing is determined in accordance with the Markovian transition matrix:

$$\mathit{Pr}_w = \begin{bmatrix} P_{1,1,w} & \cdots & P_{M(w),1,w} \\ \vdots & \ddots & \vdots \\ P_{1,M(w),w} & \cdots & P_{M(w),M(w),w} \end{bmatrix}, \tag{8}$$

where $Pr_w' 1_{M(w)} = 1_{M(w)}$ and $1_{M(w)}$ is an $M(w)$-dimensional column vector of ones. The $P_{a,b,w}$, $a,b \in M_w$, represent individual transition probabilities such that $P_{a,b,w} = \Pr\left(S_{w,m,t} = b \mid S_{w,m,t-1} = a\right)$ represents the probability of a transition from state *a* to state *b*.

Pursuant to the time variation in the volatility component for the latent global factor and the time-invariance of *G* the representation for r_t is equivalent to the variance decomposition:

$$V\left(r_t \left| S_{w,m,t}, I_{t-1}\right.\right) = BP_tB' + G = \Gamma_t\Gamma_t' + G, \tag{9}$$

where $\Gamma_t = BP_t^{1/2}$ represents a time-varying matrix of factor loadings.

In accordance with the unrealistic restriction on *G*, and along the lines of the regime-switching or GARCH factor models (Diebold and Nerlove, 1989; Engle and Susmel 1993; Kim and Nelson, 1998), variation in the scedastic function is provided for by the time-varying common volatility term $\sigma_{w,t}^2$. The conditional decomposition of the variance-covariance matrix as per $\Gamma_t\Gamma_t' + \Sigma$ implies that, conditional on the *t*th state-contingent volatility $\sigma_{w,t}^2$, the variance of the world factor is unity. Given the equivalence $BP_tB' = \Gamma_t\Gamma_t'$ and observation of $S_{w,t}$, the extent of each market's time-*t* sensitivity to the standardised common factor is determined by reference to the prevailing state at time *t*. In turn, given the unconditional treatment of *G*, all shocks to the *i*th idiosyncratic component belong to a single *i*th-market specific regime. It is, therefore, assumed that the time-varying degree of influence exerted on returns by the common factor depends solely on the common regime. The assumption implies that integration levels among the markets under consideration are determined by $\Gamma_t\left(S_{w,t}\right)$ and identified by reference to the states (or regimes) defined by the state variable $S_{w,t}$.

The adoption of a sole global factor presumes that market relationships are not subject to regional or other factors. This approach is, however, consistent with the global derivation of the Capital Asset Pricing Model. Given a sole global factor and purely idiosyncratic factors (viz. idiosyncratic factors whose shocks do not covary), the covariance of any two markets may be explained pursuant to their state-dependent sensitivity to global information. Consequently, in accordance with generally observed financial market behaviour (consider, for example, King and Wadhwani, 1990; Bollerslev, Chou, and Kroner, 1992; King, Sentana, and Wadhwani, 1994), markets are presumed to covary at a level proportionate to the state of general volatility.

2.2 Extensions to the basic model

In the extended variant, the *N*-vector r_t is treated as a function of two latent factors. The first latent factor is equivalent to the world factor in the basic model and is common to each market, while the second factor is common to the European markets only (i.e. the second factor's sensitivity to non-European markets is always zero).

$$r_t = \left[C(L) \quad \mathrm{I}_N \right] \left[\tilde{f}_t' \quad u_t' \right]' = C(L)\tilde{f}_t + u_t\,, \tag{10}$$

$$\Psi(L)u_t = G^{1/2} z_t\,, \tag{11}$$

$$G_{t|t-1} = diag\left(\sigma^2_{t|t-1}\right), \tag{12}$$

$$\sigma^2_{t|t-1} = \left[\sigma^2_{1,t|t-1} \quad \sigma^2_{2,t|t-1} \quad \dots \quad \sigma^2_{N,t|t-1} \right]', \tag{13}$$

where $C(L)$ is a polynomial in the lag operator *L*, $\tilde{f}_t$ is a *K* = 2 dimensional vector comprising the world and European factors W_t and E_t respectively, $\Psi(L) = \mathrm{I}_N - \Psi_1 L$, Ψ_1, G are diagonal matrices of order *N* and z_t is a multivariate standard normal vector $z_t \sim iidMVN\left(0_N, \mathrm{I}_N\right)$. The Euro-specific factor is partially identified through the imposition of zero-restrictions on the European component of $C(L)$ for the non-European markets in the dataset. In contrast to the basic approach, the idiosyncratic factors are not restricted to the zero-persistence property. In this respect, for $\Psi(L) \neq \mathrm{I}_N$, the N-dimensional idiosyncratic vector u_t will be autocorrelated. The presence of a significant and persistent idiosyncratic term indicates that a historic pricing error is considered relevant in the construction of future prices. Equivalently, the persistent idiosyncrasy may be interpreted as a form of idiosyncratic feedback (see, for example, Daniel, Hirshleifer, and Subrahmanyam 1998, 2001).

The model acknowledges the potential for time-varying idiosyncratic variances through the conditionally time-varying matrix G. The diagonal elements of $G_{t|t-1}$ are modelled as GARCH processes such that, given $G_{t|t-1} \neq G$, idiosyncratic shocks are conditionally heteroscedastic. In the basic scenario, idiosyncratic volatility follows the stylised constant form implying that time-varying volatility in asset returns is restricted to the common volatility component. The conditional extension to G, in addition to providing certain estimation advantages, enables the treatment of volatility across the common and idiosyncratic components in a time-varying context.

The diagonality of $\Psi(L)$ restricts the manner in which lagged effects may be transmitted; idiosyncratic persistence is permitted within but not between markets. Additionally, the diagonal form for G treats idiosyncratic shocks endogenously thereby restricting the transmission effects of an idiosyncratic shock to the market responsible for generating the shock. The joint diagonality of $\Psi(L)$, G has three broad implications: 1) the idiosyncratic factors do not contain information stemming from other markets, 2) lagged effects and shocks particular to market i are restricted to market i, and 3) the transmission of information across markets is restricted to lagged effects and shocks common to the market set.

The global (or world) factor is extended as per the following:

$$\phi_w(L) f_{w,t} = \phi_w(L) W_t = \mu_{w,t} + w_t, \tag{14}$$

$$\mu_{w,t} = \sum_{m=1}^{M(w)} \mu_{w,m} S_{w,m,t}. \tag{15}$$

The basic model restrictions, $\phi_w(L)=1$ and $\mu_{w,t}=\mu_w$, imply that the expected value of the global factor is captured entirely by the time-invariant intercept term μ_w. The relaxation of the persistence restriction introduces an additional source of information for the construction of expected returns. In this respect, $\phi_w(L)\neq 1$ recognises the potential for $k^{(w)}$th-order effects in W_t and provides a capacity for treating common persistence as a relevant pricing construct. Persistent common effects allow for the possibility of serial correlation in equity returns and suggest that the pricing effects of common shocks may be dissipative (assuming that the roots of $\phi_w(L)$ are outside the unit circle). This approach is contrasted with the restriction $\phi_w(L)=1$ and the associated implication that common effects are immediately and fully digested by financial markets. In the case of non-zero $\phi_{w,k|k\geq 1}$, the global factor is associated with a time-varying expectation and the information retrieved through $\phi_w(L)$ may be interpreted as capturing a dimension of autocorrelation in the risk-premium. Persistence effects are transmitted to all markets with applicable non-zero loading terms in $C(L)$. The extension to a time-varying intercept term $\mu_{w,t}$ also acknowledges the potential for time variation in the risk-premium. In the case, $\phi_w(L)=1$ and $C(L)=C$, the expected return attributable to the global factor is time varying pursuant to its proportionality to the global intercept $\mu_{w,t}$.

A regional European factor is also introduced as a means of evaluating the impact of region-specific information. The regional component provides the capacity for treatment of European markets in a homogeneous manner such that European market prices are determined by reference to global, idiosyncratic or regional sources. Clearly, the consideration of regional factors may be extended to the non-European case. The orthogonal European factor is modelled using the same approach adopted for the global factor.

Given the adoption of two factors, the representation for r_t is equivalent to the variance decomposition:

$$V\left(r_t \left| S_{w,\{t\}}, S_{e,\{t\}}, I_{t-1}\right.\right) = \left[C(L) \quad \mathrm{I}_N\right] \tilde{P}_t^* \left[C(L) \quad \mathrm{I}_N\right]' = C^* P_t^* C^{*\prime} = \Gamma_t^* \Gamma_t^{*\prime}, \qquad (16)$$

where $\tilde{P}_t^* = V\left(\left[\tilde{f}_t' \quad u_t'\right]' \left| S_{\{t\}}, I_{t-1}\right.\right)$ is a block-diagonal matrix containing the variance-covariances of the common and idiosyncratic components, $\Gamma_t^* = C^* P_t^{*1/2}$, and $\{S_t\}$ is the set of states constituting the state path to time *t*. $\tilde{P}_t^*$ depends on the entire volatility path to time *t*, $\left\{\sigma_{w,i}^2, \sigma_{e,i}^2, \sigma_{1,i}^2, ..., \sigma_{N,i}^2 : i = 0,1,2,...t\right\}$, and the common and idiosyncratic autocovariances captured through $\Phi(L)$, $\Psi(L)$. Note that P_t^* is augmented with the relevant lagged variance terms such that the lag operator in $C(L)$ is unnecessary.

The transmission of lagged effects and shocks between the European and world factors is determined as:

$$\begin{bmatrix} \phi_w(L) & 0 \\ 0 & \phi_e(L) \end{bmatrix} \begin{bmatrix} W_t \\ E_t \end{bmatrix} = \begin{bmatrix} \mu_{w,t} \\ \mu_{e,t} \end{bmatrix} + \begin{bmatrix} w_t \\ e_t \end{bmatrix}, \qquad (17)$$

$$\tilde{H}_t = E\left[\begin{pmatrix} w_t \\ e_t \end{pmatrix}\begin{pmatrix} w_t \\ e_t \end{pmatrix}' \left| S_{w,t}, S_{e,t}\right.\right] = \begin{bmatrix} \sigma_{w,t}^2 & 0 \\ 0 & \sigma_{e,t}^2 \end{bmatrix}, \qquad (18)$$

$$Pr = \begin{bmatrix} P_{w=1,w=1} & \cdots & P_{e=M(e),w=1} \\ \vdots & \ddots & \vdots \\ P_{w=1,e=M(e)} & \cdots & P_{e=M(e),e=M(e)} \end{bmatrix} = \begin{bmatrix} Pr_w & \emptyset_{M(w),M(e)} \\ \emptyset_{M(e),M(w)} & Pr_e \end{bmatrix}, \qquad (19)$$

where $\emptyset_{a,b}$ is an *a* by *b* matrix of zeros.

The transmission restrictions imply: 1) W_t, E_t are purely autoregressive processes in that lagged effects may not be transmitted across factors, and 2) The first diagonal element of $\tilde{H}_t$ and the first diagonal partition of Pr are solely responsible for shocks to the world factor (with an analogous interpretation for European shocks). Given the restrictions, the European factor is uncorrelated with the world factor. Clearly, the two independent state vectors with transition matrices Pr_w and Pr_e are equivalent to a single state vector with transition matrix Pr. In this sense, the common factor shocks can be interpreted as arising from an appropriately restricted system of $M(w)+M(e)$ states.

Conditioning on the factors, the expected return structure differs in the extended formulation in terms of the addition of a regional factor, time variation in the common factor intercepts, the consideration of historic pervasive information, and the modelling of the idiosyncratic component as a persistent process. In the situation, $c_{i,e}=0$, $c_i(L)=c_i$, and $\psi_i(L)=1$, expected returns are determined by contemporaneous association with the global factor as per the basic case. It should be noted, however, that the expected value for the global factor is influenced by $\mu_{w,t}$ and $\phi_w(L)$ such that the restrictions $\mu_{w,t}=\mu_w$, $\phi_w(L)=1$ are also relevant in collapsing to the basic expectations structure.

Finally, the *i*th market's volatility under the basic model is determined as a function of sensitivity to the sole common factor and the constant volatility of the market specific component. In the extended scenario, volatility is approximated by reference to the *i*th market's sensitivity to both contemporaneous and historic common information, the covariance structure of the common and idiosyncratic factors, the common volatility regimes and the conditional Markovian volatilities of the idiosyncratic components.

3. Data

Returns data are obtained from the set of U.S.-dollar denominated MSCI (Morgan Stanley Capital International) developed country indices.[1] The indices are used to construct weekly returns during the period commencing the first week of Jan. 1980 and ending the second week of Sept. 2004 (T = 1289) for 18 of the developed markets. The 18 markets are Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hong Kong, Italy, Japan, the Netherlands, Norway, Singapore, Spain, Sweden, Switzerland, the

[1] Refer to www.mscibarra.com.

U.K. and the U.S. Five of the 23 markets designated as developed are omitted due to lack of data.[2]

Excess returns are constructed by reference to the appropriately adjusted 13-week U.S. Treasury Bill such that the continuously compounded excess return (multiplied by 100) for country *i* at time *t* is:

$$r_{i,t} = 100 \ln\left(1 + \tilde{r}_{i,t} - \tilde{r}_{f,t}\right), \tag{20}$$

where $\tilde{r}_{i,t} = \left(z_{i,t} / z_{i,t-1}\right) - 1$, $\tilde{r}_{f,t} = \left(1 + r_{f,t}\right)^{1/52} - 1$, $z_{i,t}$ is the index value for country *i* at time *t*, and $r_{f,t}$ is the annualized decimal yield on the 13-week (U.S.) Treasury Bill at time *t*.

4. Estimation framework

Given the Markovian properties of the asset factor levels, the common regimes and the idiosyncratic volatilities, an exact likelihood equation is not available. In any case, approximate maximum likelihood estimation over the set of possible outcomes poses an intractable problem. A Metropolis-in-Gibbs sampler is, therefore, constructed to overcome the intractable nature of the maximum likelihood estimator and provide draws from the exact joint posterior density of the parameter set. The MCMC sampler spends the greatest portion of its time drawing from the full conditional density of the common factors. In seeking to accelerate the estimation time for each pass of the sampler, approximations to the conditional density of the common factors are constructed and their empirical accuracy is assessed in the next section.

The sampler is used to obtain draws from the posterior density of the parameter set by iterating through the following steps:

Step 1 Draw the common factor set W_t, E_t using the procedure detailed in Section 4.1.

[2] Finland, Greece, Ireland, New Zealand, and Portugal are omitted.

Step 2 Draw the factor loading matrix C and the N idiosyncratic persistence terms ψ_i, $i = 1, 2, ..., N$, (Appendix A).

Step 3 Draw the factor persistence vectors ϕ and the factor intercepts μ for the global and European factors (Appendix A).

Step 4 Obtain the idiosyncratic variance matrix G_t by drawing the GARCH terms $\varpi_i, \alpha_i, \beta_i$ for each market i, $i = 1, 2, ..., N$, (Section 4.2).

Step 5 Draw the regime-dependent global and European factor variances (i.e. the vector σ_w^2 for the global factor and the vector σ_e^2 for the European factor) (Appendix A).

Step 6 Draw the regime vectors $S_{w,t}, S_{q,t}$ for the global and European factors and use regimes to obtain the global and European probability transition matrices (Appendix A).

4.1 Generating the common factors

(a) Derivation of the data generating process

The chosen model assumes a single universal factor and a regional factor pertaining to European markets as the $K = 2$ common factors in addition to N country factors for a total of $N + K$ factors. Adopting $c_{i,1}(L) = c_{i,1,1} + c_{i,1,2}L,\ c_{i,2}(L) = c_{i,2,1} + c_{i,2,2}L$, the model is given by:

$$\begin{aligned} r_{i,t} &= c_{i,1}(L) f_{universal,t} + c_{i,2}(L) f_{Europe,t} + u_{i,t} \\ &= c_{i,1,1} f_{universal,t} + c_{i,1,2} f_{universal,t\text{-}1} + c_{i,2,1} f_{Europe,t} + c_{i,2,2} f_{Europe,t\text{-}1} + u_{i,t}. \end{aligned} \tag{21}$$

To capture common and idiosyncratic persistence, the common factors are modelled as third-order stochastic difference equations whereas the country factors are modelled as first-order stochastic difference equations. This model may be represented in state-space form as:

$$r_t = \left[C_1(L) \quad C_2(L) \quad \mathrm{I}_N\right]\left[W_t \quad E_t \quad u_t'\right]', \tag{22}$$

where $C_1(L) = C_{1,w} + C_{2,w}L + C_{3,w}L^2$, $C_2(L) = C_{1,e} + C_{2,e}L + C_{3,e}L^2$, and The restriction $C_{3,(\cdot)} = 0$ is imposed thereby restricting the observed vector r_t to the contemporaneous and immediately preceding common factor values. For convenience, the state-space form is re-arranged as per:

$$r_t = \left[C_1 \quad C_2 \quad \emptyset_{N,K} \quad \mathrm{I}_N\right]\left[\tilde{f}_t' \quad \tilde{f}_{t-1}' \quad \tilde{f}_{t-2}' \quad u_t'\right]' = \left[C \quad \emptyset_{N,K} \quad \mathrm{I}_N\right]f_t^*, \tag{23}$$

$$f_t^* = \mu_t^* + \Phi^* f_{t-1}^* + v_t^*, \tag{24}$$

$$f_t^* = \left[W_t \quad E_t \quad W_{t-1} \quad E_{t-1} \quad W_{t-2} \quad E_{t-2} \quad u_{1,t} \;\ldots\; u_{17,t} \quad u_{18,t}\right]', \tag{25}$$

$$\mu_t^* = \left[\mu_{w,t} \quad \mu_{e,t} \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \ldots \quad 0 \quad 0\right]', \tag{26}$$

$$\Phi^* = \begin{bmatrix} \Phi & \emptyset \\ \emptyset & \Phi_2^* \end{bmatrix}, \tag{27}$$

$$\Phi = \begin{bmatrix} \phi_{w,1} & 0 & \phi_{w,2} & 0 & \phi_{w,3} & 0 \\ 0 & \phi_{e,1} & 0 & \phi_{e,2} & 0 & \phi_{e,3} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \tag{28}$$

$$\Phi_2^* = diag\left(\left[\psi_1 \quad \psi_2 \quad \psi_3 \quad \cdots \quad \psi_{16} \quad \psi_{17} \quad \psi_{18}\right]\right), \tag{29}$$

$$E\left(v_t^* v_t^{*\prime}\right) = \begin{bmatrix} Q_{1,t} \\ Q_{2,t} \end{bmatrix}, \tag{30}$$

$$Q_{1,t} = \begin{pmatrix} \sigma_{w,t}^2 & 0 & \\ 0 & \sigma_{e,t}^2 & \\ 0 & 0 & \\ 0 & 0 & \emptyset_{3K,N+2K} \\ 0 & 0 & \\ 0 & 0 & \end{pmatrix}, \tag{31}$$

$$Q_{2,t} = \left(\emptyset_{N,3K} \quad diag\left(\left[\sigma_1^2 \quad \sigma_2^2 \quad \sigma_3^2 \ldots \sigma_{16}^2 \quad \sigma_{17}^2 \quad \sigma_{18}^2 \right] \right) \right), \tag{32}$$

where $diag(x)$ is a diagonal matrix with the elements of *x* on its diagonal, $E\left(v_t^* v_t^{*\prime}\right)$ is an *N*+3*K* by *N*+3*K* matrix, and $\emptyset_{L_1,L_2}$ is an L_1 by L_2 matrix of zeros.

Conditional on *Ψ*, Φ^* is no longer dependent on *N* and the system transition equations may be represented by the 3*K* < *N* dimensions of *Φ*. This alternative state-space form will be used to estimate the latent common factors and is represented by (33)-(43).

$$\Psi(L) r_t = \begin{bmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{17,t} \\ r_{18,t} \end{bmatrix} - diag\begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{17} \\ \psi_{18} \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \\ \vdots \\ r_{17,t-1} \\ r_{18,t-1} \end{bmatrix} = B \begin{bmatrix} f_{w,t} \\ f_{e,t} \\ f_{w,t-1} \\ f_{e,t-1} \\ f_{w,t-2} \\ f_{e,t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{17,t} \\ \varepsilon_{18,t} \end{bmatrix}, \tag{33}$$

$$B = \left[c_{1,1} \quad c_{2,1} \quad c_{1,2} - \Phi_2^* c_{1,1} \quad c_{2,2} - \Phi_2^* c_{2,1} \quad -\Phi_2^* c_{1,2} \quad -\Phi_2^* c_{2,2} \right], \tag{34}$$

$$E\left(\varepsilon_t \varepsilon_t' \middle| I_{t-1} \right) = diag\left(E\left(\sigma^2 \middle| I_{t-1} \right) \right), \tag{35}$$

$$\sigma_{i,t|t-1}^2 = E\left(\sigma_i^2 \middle| I_{t-1} \right) = \varpi_i + \sum_{p=1}^{P} \alpha_{i,p} \varepsilon_{i,t-p}^2 + \sum_{q=1}^{Q} \beta_{i,q} \sigma_{i,t-q|t-q-1}^2 , \tag{36}$$

$$\begin{bmatrix} f_{w,t} \\ f_{e,t} \\ f_{w,t-1} \\ f_{e,t-1} \\ f_{w,t-2} \\ f_{e,t-2} \end{bmatrix} = \begin{bmatrix} \mu_{w,t} \\ \mu_{e,t} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \phi_{w,1} & 0 & \phi_{w,2} & 0 & \phi_{w,3} & 0 \\ 0 & \phi_{e,1} & 0 & \phi_{e,2} & 0 & \phi_{e,3} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} f_{w,t-1} \\ f_{e,t-1} \\ f_{w,t-2} \\ f_{e,t-2} \\ f_{w,t-3} \\ f_{e,t-3} \end{bmatrix} + \begin{bmatrix} w_t \\ e_t \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{37}$$

$$\mu_{w,t} = \sum_{m=1}^{M(w)} \mu_{w,m} S_{w,m,t} , \tag{38}$$

$$\mu_{e,t} = \sum_{m=1}^{M(e)} \mu_{e,m} S_{e,m,t} , \tag{39}$$

$$E\left(v_t v_t' \middle| I_{t-1}, S_t\right) = H_t = \begin{pmatrix} \sigma^2_{w,t} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{e,t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{40}$$

$$v_t = \begin{bmatrix} w_t & e_t & 0 & 0 & 0 & 0 \end{bmatrix}', \tag{41}$$

$$\sigma^2_{w,t} = \sum_{m=1}^{M(w)} q_{w,m} S_{w,m,t} \, , \tag{42}$$

$$\sigma^2_{e,t} = \sum_{m=1}^{M(e)} q_{e,m} S_{e,m,t} \, , \tag{43}$$

where $c_{1,1}, c_{2,1}$ are the contemporaneous loadings on the global and European factors respectively (such that $c_{1,2}, c_{2,2}$ are the lagged loadings). It is important to note that the loadings for the non European markets are, by definition, set to zero for $c_{2,1}$, $c_{2,2}$. To ensure identification of the common factors, the contemporaneous universal loading for the U.S. market and the contemporaneous European loading for the German market are normalised to unity. The values $M_1 = 3$ and $M_2 = 2$, pertaining to equations (38)-(39) and (42)-(43), are chosen as the minimal set of regimes that account for conditional heteroscedasticity in the global and European factors respectively.

(b) Derivation of the full conditional density

The full conditional density for the common component is:

$$\begin{aligned} f\left(F \middle| R, S, \sigma_\alpha, \sigma_\beta, \sigma_\varpi, \theta^*\right) \propto\ & f\left(R \middle| F, C, a, \psi, \sigma_\alpha, \sigma_\beta, \sigma_\varpi\right) \\ & \times f\left(F \middle| S, \phi, \gamma, \varpi\right) \\ & \times f\left(S \middle| P\right) \\ & \times f\left(C, a, \psi, \sigma_\alpha, \sigma_\beta, \sigma_\varpi, \phi, \gamma, \varpi, P\right), \end{aligned} \tag{44}$$

where $\theta^* = \{C, a, \psi, \phi, \gamma, \varpi, P\}$, $F = \{f_1, f_2, ..., f_T\}$, $R = \{r_1, r_2, ..., r_T\}$, S and P are the common regimes and transition probabilities respectively, C is the set of loadings on F, $\{\sigma_\alpha, \sigma_\beta, \sigma_\varpi\}$ is the set of GARCH parameters and γ, ϖ define the intercept and volatility parameters for F.[3]

Given the T-dimensionality of F, the dependence on the regime process S and the GARCH set $\{\sigma_\alpha, \sigma_\beta, \sigma_\varpi\}$, the provision of an exact draw from the entire block F is an intractable problem. Consequently, the entire block F is drawn in T elements pursuant to the full conditional density of f_t.

$$\begin{aligned} f\left(f_t \middle| F_{\neq t}, R, S, \sigma_\alpha, \sigma_\beta, \sigma_\varpi, \theta^*\right) \propto\ & f\left(R \middle| F, C, a, \psi, \sigma_\alpha, \sigma_\beta, \sigma_\varpi\right) \\ & \times f\left(F \middle| S, \phi, \gamma, \varpi\right) \\ & \times f\left(S \middle| P\right) \\ & \times f\left(h, a, \psi, \sigma_\alpha, \sigma_\beta, \sigma_\varpi, \phi, \gamma, \varpi, P\right). \end{aligned} \tag{45}$$

Given the Markovian nature of F the conditional density $f\left(F \middle| S, \phi, \gamma, \varpi\right)$ may be written as $\prod_i f\left(f_i \middle| f_{i-1}, S, \phi, \gamma, \varpi\right)$. Accordingly, the density $f\left(F \middle| S, \phi, \gamma, \varpi\right)$ depends on f_t only via the product of the terms $f\left(f_{t+1} \middle| f_t, S, \phi, \gamma, \varpi\right)$ and $f\left(f_t \middle| f_{t-1}, S, \phi, \gamma, \varpi\right)$. In turn, neither $f\left(S \middle| P\right)$ nor $f\left(C, a, \psi, \sigma_\alpha, \sigma_\beta, \sigma_\varpi, \phi, \gamma, \varpi, P\right)$ depend on the value taken by f_t. The full conditional density of f_t is, therefore, proportionate to:

$$\begin{aligned} f\left(f_t \middle| F_{\neq t}, R, S, \sigma_\alpha, \sigma_\beta, \sigma_\varpi, \theta^*\right) \propto\ & f\left(R \middle| F, C, a, \psi, \sigma_\alpha, \sigma_\beta, \sigma_\varpi\right) \\ & \times f\left(f_{t+1} \middle| f_t, S, \phi, \gamma, \varpi\right) \\ & \times f\left(f_t \middle| f_{t-1}, S, \phi, \gamma, \varpi\right). \end{aligned} \tag{46}$$

[3] The simulation is undertaken pursuant to redefinition of the common intercept as $\mu_{k,t} = \gamma_{k,t}' x_{t-1}$. Accordingly, a return intercept a may therefore be identified.

(c) Drawing from the full conditional density

Given the difficulty of drawing from the full conditional density of the *t*th element of *F*, draws are made from a proposal density. At every *t*th element of the block *F*, a proposed value $\tilde{f}_t^{\,p}$ is accepted according to the Metropolis-Hastings algorithm (see Tierney, 1994; Chib and Greenberg, 1995). The algorithm accepts draws with probability:

$$\begin{aligned}\Pr\left(\tilde{f}_t,\tilde{f}_t^{\,p}\right) &= \min\left[\frac{f\left(\tilde{f}_t^{\,p}\middle|F_{\neq t},R,S,C,a,\psi,\sigma_\alpha,\sigma_\beta,\sigma_\varpi,\phi,\gamma,\varpi,P\right)\Big/p\left(\tilde{f}_t^{\,p}\right)}{f\left(\tilde{f}_t\middle|F_{\neq t},R,S,C,a,\psi,\sigma_\alpha,\sigma_\beta,\sigma_\varpi,\phi,\gamma,\varpi,P\right)\Big/p\left(\tilde{f}_t\right)},1\right] \\ &= \min\left[\frac{f^*\left(\tilde{f}_t^{\,p}\right)\Big/p\left(\tilde{f}_t^{\,p}\right)}{f^*\left(\tilde{f}_t\right)\Big/p\left(\tilde{f}_t\right)},1\right],\end{aligned} \tag{47}$$

where $f^*\left(\tilde{f}_t\right)=f\left(R\middle|F,C,a,\psi,\sigma_\alpha,\sigma_\beta,\sigma_\varpi\right)f\left(\tilde{f}_{t+1}\middle|\tilde{f}_t,S,\phi,\gamma,\varpi\right)f\left(\tilde{f}_t\middle|\tilde{f}_{t-1},S,\phi,\gamma,\varpi\right)$, $p\left(\tilde{f}_t\right)$ is a proposal density in the support of the full conditional density $f\left(\tilde{f}_t\middle|\cdot\right)$ and $\tilde{f}_t$ is the first *K*-vector in f_t.

The terms in $f^*\left(\cdot\right)$ are evaluated as:

$$\begin{aligned}f\left(R\middle|F,C,a,\psi,\sigma_\alpha,\sigma_\beta,\sigma_\varpi\right) &= \prod_{\forall i\geq t} f\left(r_i\middle|f_i,R_{i-1},F_{i-1},C,a,\psi,\sigma_\alpha,\sigma_\beta,\sigma_\varpi\right) \\ &= \prod_{\forall i\geq t} MVN\left(y_i-Bf_i\middle|0_N,G_i\right),\end{aligned} \tag{48}$$

$$f\left(\tilde{f}_t\middle|\tilde{f}_{t-1},S,\phi,\gamma,\varpi\right)=MVN\left(\begin{bmatrix}w_t\\ e_t\end{bmatrix}\middle|0_2,\begin{bmatrix}\sigma_{w,t}^2 & 0\\ 0 & \sigma_{e,t}^2\end{bmatrix}\right), \tag{49}$$

where $R_{i-1}=\left\{r_1,r_2,\ldots,r_{i-1}\right\}$ and $F_{i-1}=\left\{f_1,f_2,\ldots,f_{i-1}\right\}$. $f\left(\tilde{f}_{t+1}\middle|\tilde{f}_t,S,\phi,\gamma,\varpi\right)$ is evaluated in an analogous fashion to (49). Draws are proposed in natural order commencing at t = 1 and ending at $t = T$.

(d) Derivation of the proposal density $p(\cdot)$

Assume, foremost, that the set of idiosyncratic GARCH variances R_v is observed. Pursuant to this assumption the GARCH set $\{\sigma_\alpha, \sigma_\beta, \sigma_\varpi\}$ is irrelevant. The full conditional density of the set of state vectors $\{f_1, f_2, ..., f_T\}$ may be written as:

$$\begin{aligned} p\left(F_T \left| R_T, R_v, \cdot\right.\right) &= p\left(f_1, f_2, ..., f_T \left| R_T, R_v, \cdot\right.\right) \\ &= \prod_{t=1}^{T} p\left(f_t \left| f_{t+1}, f_{t+2}, ..., f_T, R_T, R_v, \cdot\right.\right) \\ &= \prod_{t=1}^{T} p\left(f_t \left| F^{t+1}, R_T, R_v, \cdot\right.\right). \end{aligned} \tag{50}$$

Given R_v, and suppressing further reference to the conditional set R_v for notational convenience, the *t*th conditional density $p\left(f_t \left| F^{t+1}, R_T, R_v, \cdot\right.\right)$ may be rewritten according to (see Chib and Greenberg, 1996):

$$p\left(f_t \left| F^{t+1}, R_T, \cdot\right.\right) \propto p\left(f_t \left| R_t, \cdot\right.\right) p\left(f_{t+1} \left| f_t, R_t, \cdot\right.\right) p\left(R^{t+1}, F^{t+1} \left| f_t, f_{t+1}, R_t, \cdot\right.\right). \tag{51}$$

It is instructive to rewrite $p\left(R^{t+1}, F^{t+1} \left| f_t, f_{t+1}, R_t, \cdot\right.\right)$ as:

$$p\left(R^{t+1} \left| F^{t+1}, f_t, f_{t+1}, R_t, \cdot\right.\right) p\left(F^{t+1} \left| f_t, f_{t+1}, R_t, \cdot\right.\right). \tag{52}$$

The right hand side density of (52) can be written as a product of conditional densities:

$$\begin{aligned} & p\left(F^{t+1} \left| f_t, f_{t+1}, R_t, \cdot\right.\right) \\ &= p\left(f_{t+1}, f_{t+2}, ..., f_T \left| f_t, f_{t+1}, R_t, \cdot\right.\right) \\ &= p\left(f_T \left| f_t, f_{t+1}, ..., f_{T-1}, R_t, \cdot\right.\right) p\left(f_{T-1} \left| f_t, f_{t+1}, ..., f_{T-2}, R_t, \cdot\right.\right) ... \\ & ... p\left(f_{t+2} \left| f_t, f_{t+1}, R_t, \cdot\right.\right) p\left(f_{t+1} \left| f_t, f_{t+1}, R_t, \cdot\right.\right) \\ &= p\left(f_T \left| f_{T-1}, \cdot\right.\right) p\left(f_{T-1} \left| f_{T-2}, \cdot\right.\right) ... p\left(f_{t+2} \left| f_{t+1}, \cdot\right.\right) p\left(f_{t+1} \left| f_{t+1}, \cdot\right.\right). \end{aligned} \tag{53}$$

In accordance with the Markovian nature of the state vector, $p\left(f_{t+k}\left|f_t,f_{t+1},\ldots,f_{t+k-1},R_t,\cdot\right.\right)$ collapses to $p\left(f_{t+k}\left|f_{t+k-1},\cdot\right.\right)$ such that $f_{t+k}\left|\{f_{t+k-1},F_{t+k-2},\cdot\}\sim MVN\left(\mu_{t+k}+\Phi f_{t+k-1},H_{t+k}\right)\right.$. This ensures the validity of moving from the third to the fourth line of (53) such that only the immediately preceding state is relevant and the distribution of the current state f_t is unaffected by $p\left(F^{t+1}\left|f_t,f_{t+1},R_t,\cdot\right.\right)$.

The left hand side density of (52), $p(R^{t+1}\,|\,F^{t+1},f_t,f_{t+1},R_t,\cdot)$, may also be derived as a product of conditional densities:

$$p\left(R^{t+1}\left|F^{t+1},f_t,f_{t+1},R_t,\cdot\right.\right)=p\left(r_T\left|f_T,\cdot\right.\right)p\left(r_{T-1}\left|f_{T-1},\cdot\right.\right)\ldots p\left(r_{t+2}\left|f_{t+2},\cdot\right.\right)p\left(r_{t+1}\left|f_{t+1},\cdot\right.\right). \quad (54)$$

Given the form for r_{t+k}, the density of the (t+k)th returns vector conditional on the state f_{t+k} is independent of any other state or the returns history up to time t+k. Therefore a draw from $p\left(f_t\left|F^{t+1},R_T,\cdot\right.\right)$ is based on $p\left(f_t\left|R_t,\cdot\right.\right)p\left(f_{t+1}\left|f_t,R_t,\cdot\right.\right)=p\left(f_t\left|R_t,\cdot\right.\right)p\left(f_{t+1}\left|f_t,\cdot\right.\right)$.

(e) Procedure for drawing from the proposal density $p\left(f_t\left|F^{t+1},R_T,\cdot\right.\right)$

Run the Kalman filter to generate $f_{t|t}$, $P_{t|t}$. B,Φ,G_t,μ_t and H_t form the system matrices used for the Kalman filter extraction of the state vector.

The relevant equations are (suppressing the implied conditioning on R_v and the regime set S):

$$\eta_{1,t|s}=f_t-E_s\left(f_t\right)=f_t-f_{t|s}, \quad (55)$$

$$\eta_{2,t}=y_t-E_{t-1}\left(y_t\right)=y_t-y_{t|t-1}=y_t-Bf_{t|t-1}, \quad (56)$$

$$f_{t|t-1}=E_{t-1}\left(f_t\right)=\mu_t+\Phi f_{t-1|t-1}, \quad (57)$$

$$P_{t|t-1}=E_{t-1}\left(\eta_{1,t|t-1}\eta_{1,t|t-1}{}'\right)=\Phi P_{t-1|t-1}\Phi'+H_t, \quad (58)$$

$$f_{t|t}=E_t\left(f_t\right)=f_{t|t-1}+K_t\eta_{2,t}, \quad (59)$$

$$P_{t|t} = E_t\left(\eta_{1,t|t}\eta_{1,t|t}{}'\right) = P_{t|t-1} - P_{t|t-1}B'Q_{t|t-1}^{-1}BP_{t|t-1} = P_{t|t-1} + K_t BP_{t|t-1}, \tag{60}$$

$$Q_{t|t-1} = E_{t-1}\left(\eta_{2,t}\eta_{2,t}{}'\right) = BP_{t|t-1}B' + G_t, \tag{61}$$

$$K_t = E_{t-1}\left(\eta_{1,t|t-1}\eta_{2,t}{}'\right)E_{t-1}\left(\eta_{2,t}\eta_{2,t}{}'\right)^{-1} = P_{t|t-1}B'Q_{t|t-1}^{-1}. \tag{62}$$

The filter is initialised as per:

$$f_{0|0} = \left(\mathrm{I}_{3K} - \Phi\right)^{-1}\mu^*, \tag{63}$$

$$\mu^* = \begin{bmatrix} \mu_w{}'P_w^* & \mu_e{}'P_e^* & 0 & 0 & 0 & 0 \end{bmatrix}', \tag{64}$$

$$vec\left(P_{0|0}\right) = \left(\mathrm{I}_{(3K)^2} - \Phi \otimes \Phi\right)^{-1} vec\left(H^*\right), \tag{65}$$

$$H^* = diag\left(\begin{bmatrix} \sigma_w^{2\prime}P_w^* & \sigma_e^{2\prime}P_e^* & 0 & 0 & 0 & 0 \end{bmatrix}\right), \tag{66}$$

where P_q^* is the steady state transition probability for the qth common factor (refer to Appendix A).

The term $G_t = E\left(\varepsilon_t\varepsilon_t{}' \middle| I_{t-1}\right)$ is derived using the incumbent values of all parameters for the relevant iteration of the Markov chain. As such $E = Y - F^{inc}B'$, where F^{inc} is the incumbent draw of F and the ith row of E contains ε_i'. Given the Gaussian assumption on the error terms, the filter gives us the optimal time-t estimates of the factors and their variances. The estimates are also the optimal time-t linear estimates of the factors irrespective of the validity of the Gaussian assumption.

To simulate the factors, the time-t estimates $f_{t|t}$ are updated using the entire set of observed information I_T to obtain the estimates $f_{t|T} = E_T\left(f_t\right)$, $P_{t|T} = E_T\left(\eta_{1t|T}\eta_{1t|T}{}'\right)$. A procedure for obtaining the 'smoothed' estimate $f_{t|T}$ is given in Kim and Nelson (1998).

4.2 *Generating the idiosyncratic variance G_t*

The idiosyncratic variance G_t, is modelled as a set of GARCH (*P*, *Q*) processes:

$$\sigma_{i,t}^2 = \sigma_{i,t|t-1}^2 = E\left(\sigma_i^2 \mid I_{t-1}\right) = \varpi_i + \sum_{p=1}^{P} \alpha_{i,p} \varepsilon_{i,t-p}^2 + \sum_{q=1}^{Q} \beta_{i,q} \sigma_{i,t-q|t-q-1}^2 , \tag{67}$$

where the parameters $\{\omega, \alpha, \beta\}_i$ are estimated using a straightforward amendment of the approach in Nakatsuma (1998, 2000).

Conditional on $u_{i,t}$ and ψ_i, $\varepsilon_{i,t}$ is observed as per $\varepsilon_{i,t} = u_{i,t} - \psi_i u_{i,t-1}$. Given the diagonality of G_t, and $\varepsilon_{i,t}$, market *i*'s idiosyncratic component is unaffected by shocks originating in other markets such that:

$$E\left(\sigma_i^2 \middle| I_{t-1}\right) = E\left(\sigma_i^2 \middle| \varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots, \varepsilon_{i,0}, \varepsilon_{\neq i,t-1}, \varepsilon_{\neq i,t-2}, \ldots, \varepsilon_{\neq i,0}\right) = E\left(\sigma_i^2 \middle| \varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots, \varepsilon_{i,0}\right).$$

The posterior density for the set of GARCH parameters pertaining to market $i, \{\omega, \alpha, \beta\}_i$, is given by:

$$p\left(\{\varpi, \alpha, \beta\}_i \middle| r_i, F, c_i, \psi_i\right) \propto f\left(r_i \middle| I_t\right) p\left(\{\varpi, \alpha, \beta\}_i\right) \tag{68}$$

$$f\left(r_i \middle| I_T\right) = f\left(r_i \middle| F, c_i, a_i, \psi_i, \{\varpi, \alpha, \beta\}_i\right) = 2\pi^{-T/2} \prod_t \sigma_t^{-1} \exp\left(-\frac{\varepsilon_{i,t}^2}{2\sigma_t^2}\right) \tag{69}$$

where $p\left(\{\varpi, \alpha, \beta\}_i\right)$ is the prior density.[4]

In contrast to Nakatsuma's adoption of a *κ*-variate normal proposal density, a *κ*-variate *t* proposal density is used (see, for example, Bauwens and Lubrano, 1998):

$$g\left(\{\varpi, \alpha, \beta\}_i \middle| r_i, F, c_i, a_i, \psi_i, \lambda\right) = \frac{\Gamma\left(\frac{1}{2}(\lambda + \kappa)\right)}{\Gamma(\lambda/2)\pi^{\kappa/2}\lambda^{\kappa/2}} \left|\tilde{\Delta}\right|^{-1/2} \left(1 + \frac{\tilde{\delta}}{\lambda}\right)^{-\frac{\lambda+\kappa}{2}}, \tag{70}$$

[4] A normal prior density is used for the GARCH terms. The independent prior mean (variance) for each term is zero (ten).

where $\tilde{\delta} = (\delta - \mu)' \tilde{\Delta}^{-1} (\delta - \mu)$, $\mu = E(\delta | r_i, F, c_i, \psi_i)$, $\Delta = \mathrm{cov}(\delta | r_i, F, c_i, \psi_i) = \frac{\lambda}{\lambda - 2} \tilde{\Delta}$, $\delta = \begin{pmatrix} \varpi_i & \alpha_i' & \beta_i' \end{pmatrix}'$, λ is the degree of freedom parameter for the $t(\delta | \mu, \tilde{\Delta}, \lambda)$ density and κ is the dimension of δ.[5]

To ensure that the unconditional idiosyncratic volatility is always defined the draw δ is accepted subject to its satisfaction of the restriction:

$$\frac{\varpi_i}{1 - \sum_{\forall p} \alpha_i - \sum_{\forall q} \beta_i} \gg 0 .$$

(a) Sampling $\delta = \{\varpi, \alpha\}_i$

The hyper-parameters $\mu_{\omega,\alpha}$, $\Delta_{\omega,\alpha}$ are obtained as per Nakatsuma (1988, 2000). Given estimates of $\mu_{\omega,\alpha} = E(\{\omega, \alpha\}_i | \beta_i, r_i, F, \cdot)$ and $\Delta_{\omega,\alpha} = \mathrm{cov}(\{\omega, \alpha\}_i \mid \beta_i, r_i, F, \cdot)$, a draw $\hat{\delta} = \{\hat{\omega}, \hat{\alpha}\}_i$, $\hat{\delta} \sim t(\mu_{\varpi,\alpha}, \tilde{\Delta}_{\varpi,\alpha}, \lambda)$ is obtained by:

$$\hat{\delta} = \mu_{\varpi,\alpha} + W^{1/2} Z , \tag{71}$$

$$W = diag(\chi_\lambda^2(\kappa))^{-1} (\lambda - 2) \Delta_{\varpi,\alpha} = diag(\chi_\lambda^2(\kappa))^{-1} \lambda \tilde{\Delta}_{\varpi,\alpha}, \tag{72}$$

$$Z \sim N(0_\kappa, \mathrm{I}_\kappa), \tag{73}$$

where $\chi_\lambda^2(\kappa)$ is a κ-vector draw from the chi-square distribution with λ degrees of freedom and κ is the dimension of $\delta = \{\omega, \alpha\}_i$.

The draw $\hat{\delta}$ is accepted with Metropolis-Hastings based probability:

[5] The estimation was undertaken using $\lambda = 5$.

$$\Pr\left(\delta,\hat{\delta}\right)=\min\left(\frac{p\left(\hat{\delta}_i\,|\beta_i,r_i,F,c_i,a_i,\psi_i\right)}{p\left(\delta_i\,|\beta_i,r_i,F,c_i,a_i,\psi_i\right)}\frac{g\left(\delta_i\,|\beta_i,r_i,F,c_i,a_i,\psi_i,\lambda\right)}{g\left(\hat{\delta}_i\,|\beta_i,r_i,F,c_i,a_i,\psi_i,\lambda\right)},1\right), \tag{74}$$

where δ_i is the incumbent $\{\omega,\alpha\}_i$, and $\Pr\left(\delta,\hat{\delta}\right)$ is the probability of replacing the incumbent with the draw $\delta^p=\hat{\delta}$.

(b) Sampling $\delta=\beta_i$

Given the estimates $\mu_\beta=E\left(\beta_i\left|\{\omega,\alpha\}_i,r_i,F,\cdot\right.\right)$, $\Delta_\beta=\operatorname{cov}\left(\beta_i\left|\{\omega,\alpha\}_i,r_i,F,\cdot\right.\right)$ a draw $\hat{\delta}=\hat{\beta}_i$ is obtained by:

$$\hat{\delta}=\mu_\beta+W^{1/2}Z, \tag{75}$$

$$W=diag\left(\chi^2_\lambda(\kappa)\right)^{-1}\lambda\tilde{\Delta}_\beta, \tag{76}$$

$$Z\sim N\left(0_\kappa,\mathrm{I}_\kappa\right). \tag{77}$$

where κ is the dimension of β_i. The estimates μ_β, Δ_β are obtained pursuant to Nakatsuma (1988, 2000). As per the case for $\{\omega,\alpha\}_i$, the draw $\hat{\delta}$ is accepted with Metropolis-Hastings based probability:

$$\Pr\left(\delta,\hat{\delta}\right)=\min\left(\frac{p\left(\hat{\delta}_i\left|\{\varpi,\alpha\}_i,\cdot\right.\right)}{p\left(\delta_i\left|\{\varpi,\alpha\}_i,\cdot\right.\right)}\frac{g\left(\delta_i\left|\{\varpi,\alpha\}_i,\lambda,\cdot\right.\right)}{g\left(\hat{\delta}_i\left|\{\varpi,\alpha\}_i,\lambda,\cdot\right.\right)},1\right), \tag{78}$$

where δ_i is the incumbent β_i.

5. Approximating the conditional density of the common component

A draw from the exact full conditional density $h\left(f_t\,|\,F_{\neq t},R,\cdot\right)$ requires computation of the set of conditional volatilities $\left\{\sigma^2_{i,k}\forall i\left|k\geq t\right.\right\}$. To obtain the entire common set F the set of conditional volatilities must, therefore, be computed T times. As a means of speeding up the sampling process, two approximations to the full condition density $h\left(F\middle|R,\cdot\right)$ are derived. In the first case $h_1\left(F\middle|R,S,R_v,\cdot\right)$, R_v is the set of conditional variances determined using the existing draw of the common component F and the relevant parameter set. In the second case $h_2\left(F\middle|R,S,R_v,\cdot\right)$, R_v is determined by replacing e_t^2 with its conditional expectation $E\left(e_t^2\middle|I_{t-1}\right)$ (see King, Sentana, and Wadhwani, 1994). In contrast to the *T*-block approach adopted in the exact case, the proposal densities draw the common factor as a single block.

Three models of sample size T = 415 and number of variables N = 15 are simulated to assess the accuracy of the common factor conditional density approximations (note T = 415, N = 15 is less than the T = 1289, N = 18 pertaining to the weekly equity market returns). For the simulated models, the μ term is generalised such that the factor intercept depends on exogenous data as per $\mu_t = \gamma_{S_t}{}' x_t$.[6] The a,ψ parameters are restricted to zero for the first two models but set to non-zero values for the third model. The first two models are distinguished on the basis of the persistence parameters for the common component and conditional volatility terms. Model 1 restricts the common factor persistence ϕ to zero but sets higher values for $\sigma_\alpha,\sigma_\beta$. Model 2 induces non-zero common factor persistence coupled with low values of $\sigma_\alpha,\sigma_\beta$ such that $\sigma_\alpha+\sigma_\beta$ is clearly less than unity. Model 3 induces persistence in both the common factor and conditional volatility components and sets $\sigma_\alpha+\sigma_\beta$ to values closer to unity than either of the previous models. The third model is also estimated using draws from the full conditional density $h\left(F\middle|Y,\cdot\right)$ as a benchmark for assessing the performance of the two approximate densities for a full-scale model.

[6] The exogenous variables are the difference of the log of: Monthly New One Family Houses Sold (United States), Monthly Industrial Production Index (United States). The resulting log-differences span the period December, 1969 to June, 2004.

The estimates for the parameters of greatest interest (the common component parameters) across the three models are presented in Tables 1-4.[7] Pursuant to Table 1, it is readily observed that the latent regime structures are accurately captured by both approximations to the full conditional density of the common component for all three simulated models. The regime selection accuracy is above 90% for both approximations, and is equal to the accuracy of the full conditional density for the third and most comprehensive simulated model. The 95% highest posterior density (HPD) bands for the first regime count (i.e.: the number of times the first regime is observed in the simulated regime vector) also encompass the true values across all three simulated models irrespective of the approximation adopted.

Table 1 *Descriptive statistics regarding common regimes across the three simulated models*

		$h_1(F\|R,S,R_v,\cdot)$			$h_2(F\|R,S,R_v,\cdot)$			$h(F\|R,\cdot)$		
	Actual	Median	95% HPD		Median	95% HPD		Median	95% HPD	
Simulated model 1										
obs(regime 1)[a]	345	352	340	363	352	338	362	-	-	-
state accuracy[b]		0.939			0.944			-		
Simulated model 2										
obs(regime 1)	291	289	269	304	287	268	303	-	-	-
state accuracy		0.918			0.915			-		
Simulated model 3										
obs(regime 1)	290	289	272	305	288	270	304	288	270	303
state accuracy		0.918			0.920			0.920		

a. Returns the number of times the first regime is observed in the simulated regime vector.
b. Returns the accuracy of the regime estimates as a percentage. The regime estimate is taken as the mode of the average regime probabilities constructed using the MCMC output.

In terms of the first model, the true values are confined to the space determined by the 95% HPD levels for 96 of the 99 model parameters in the case of $h_1\left(F\middle|R,S,R_v,\cdot\right)$, and 95 of the 99 parameters given $h_2\left(F\middle|R,S,R_v,\cdot\right)$ (refer to Table 2 for the first model's common component convergence statistics). The estimates fail to identify any substantive differences between the approximating conditional densities and it is clear that both methods provide accurate results. The model is also estimated a second time using a

[7] Estimates for the remaining parameters are available upon request.

dispersed set of initial values.[8] Both sets of starting conditions yield almost identical results.

Table 2 *Descriptive statistics for the common component parameters using approximate estimators (simulated model 1)*

		$h_1(F\|R,S,R_v,\cdot)$			$h_2(F\|R,S,R_v,\cdot)$		
	Actual	Median	95% HPD		Median	95% HPD	
$\gamma_{1,1}$	5.000	5.038	4.733	5.353	4.897	4.540	5.249
$\gamma_{2,1}$	2.000	2.116	1.911	2.335	2.043	1.822	2.268
$\gamma_{1,2}$	1.000	0.716	-0.219	1.668	0.699	-0.207	1.623
$\gamma_{2,2}$	7.000	7.903	6.626	9.019	7.629	6.342	8.758
ϕ	0.000	-0.028	-0.060	0.002	-0.026	-0.057	0.007
$\sigma^2_{w,1}$	2.000	2.416	1.883	2.976	1.922	1.404	2.476
$\sigma^2_{w,1}$	6.000	6.135	3.322	10.020	5.251	2.698	8.701
$P_{1,1}$	0.900	***0.942***	***0.908***	***0.971***	***0.940***	***0.904***	***0.970***
$P_{2,2}$	0.600	0.665	0.503	0.808	0.659	0.497	0.805

a. $\gamma_{i,j}$ represents pervasive factor sensitivity to explanatory variable *i* in state *j*.

b. $P_{b,a}$ is the probability of moving to state *a* from state *b* (equiv. to the *a*th row, *b*th column of the transition matrix, $Pr_{a,b}$).

The second model incorporates non-zero persistence in the common component. In contrast to the estimates produced for the first model, the estimates given by $h_1\left(F\middle|R,S,R_v,\cdot\right)$ clearly dominate those obtained using $h_2\left(F\middle|R,S,R_v,\cdot\right)$. In the former case, the 95% HPD levels encompass the true parameter values 98 of 99 times, as opposed to 78 of 99 times where $h_2\left(F\middle|R,S,R_v,\cdot\right)$ is adopted (Table 3). Both methods accurately estimate the persistence imposed on the common factor. The greatest disparity across the methods is observed in the estimates of the factor sensitivities (or loadings) c_i. In this respect, the use of the conditional expectation $\hat{e}_t^2$ in place of e_t^2, pursuant to $h_2\left(F\middle|R,S,R_v,\cdot\right)$, appears to be associated with a negative bias in the resulting estimates.

---

[8] The distance between the initial conditional log-likelihoods using the first estimation method h_1, $\|L_2 - L_1\|$, is approximately 27.

Accordingly, the derivation of e_t^2 based on the existing common factor seems preferable to the method advocated by King, Sentana, and Wadhwani (1994) for the type of data process considered here. A second estimation, using dispersed initial values, produces results almost identical to the first set of estimates.[9]

Table 3 *Descriptive statistics for the common component parameters using approximate estimators (simulated model 2)*

		$h_1\left(F\left\|R,S,R_v,\cdot\right.\right)$			$h_2\left(F\left\|R,S,R_v,\cdot\right.\right)$		
	Actual	Median	95% HPD		Median	95% HPD	
$\gamma_{1,1}$	5.000	5.189	4.905	5.464	***5.614***	***5.249***	***5.994***
$\gamma_{2,1}$	2.000	2.051	1.820	2.281	2.198	1.939	2.466
$\gamma_{1,2}$	1.000	0.768	0.132	1.427	0.845	0.189	1.565
$\gamma_{2,2}$	7.000	6.868	6.357	7.374	7.439	6.842	8.062
ϕ	0.200	0.201	0.170	0.230	0.201	0.171	0.232
$\sigma_{w,1}^2$	2.000	2.331	1.835	2.919	2.473	1.839	3.154
$\sigma_{w,1}^2$	6.000	6.598	4.573	9.030	7.062	4.749	9.745
$P_{1,1}$	0.800	0.809	0.735	0.872	0.811	0.737	0.875
$P_{2,2}$	0.600	0.623	0.500	0.737	0.623	0.501	0.735

The third model is estimated using both approximate forms and the exact conditional density (Table 4). As per the second model, the posterior densities for four of the fifteen factor loadings are negatively skewed in the case $h_2\left(F\left|R,S,R_v,\cdot\right.\right)$. In contrast, the 95% HPD intervals encompass the true values for all factor loadings when using $h_1\left(F\left|R,S,R_v,\cdot\right.\right)$ and for fourteen of the fifteen loadings in the exact case $h\left(F\left|R,\cdot\right.\right)$. Although the 95% HPD estimates for the intercept terms *a* capture the true values in all cases for $h_2\left(F\left|R,S,R_v,\cdot\right.\right)$ and in all but a single case for $h_1\left(F\left|R,S,R_v,\cdot\right.\right)$, three of the fifteen parameters are negatively skewed when drawing from the exact conditional density for the common component. The median estimates for a_2 are significantly larger than the true value for all three common component conditional density forms. This

[9] The distance between the initial conditional log-likelihoods using the first estimation method h_1, $\left\|L_2 - L_1\right\|$, is approximately 47.

appears to be associated with the near violation of the stationarity assumption in the second variable's conditional idiosyncratic volatility pursuant to which $\sigma_\alpha + \sigma_\beta = 0.993$.

Table 4 *Descriptive statistics for the common component parameters using approximate and exact estimators (simulated model 3)*

		$h_1(F\|R,S,R_v,\cdot)$			$h_2(F\|R,S,R_v,\cdot)$			$h(F\|R,\cdot)$		
	Actual	Median	95% HPD		Median	95% HPD		Median	95% HPD	
$\gamma_{1,1}$	5.000	5.078	4.822	5.348	5.239	4.971	5.522	5.158	4.905	5.393
$\gamma_{2,1}$	2.000	2.055	1.822	2.288	2.096	1.850	2.340	2.078	1.853	2.317
$\gamma_{1,2}$	1.000	0.692	0.089	1.319	0.727	0.100	1.393	0.734	0.128	1.375
$\gamma_{2,2}$	7.000	6.813	6.344	7.296	7.009	6.512	7.504	6.933	6.487	7.412
ϕ	0.200	0.197	0.167	0.228	0.198	0.166	0.228	0.195	0.164	0.226
$\sigma^2_{w,1}$	2.000	2.341	1.847	2.918	2.238	1.689	2.845	2.437	1.947	3.016
$\sigma^2_{w,1}$	6.000	5.914	4.079	8.194	6.189	4.156	8.702	6.070	4.192	8.328
$P_{1,1}$	0.800	0.821	0.753	0.882	0.818	0.748	0.879	0.821	0.751	0.880
$P_{2,2}$	0.600	0.585	0.447	0.713	0.582	0.442	0.711	0.592	0.459	0.725

The posterior estimates for the remaining parameters are similar across the conditional density forms. As is the case for the previous two models, the estimates for the conditional volatility terms appear to be the least accurate. Notwithstanding, the 95% HPD levels for $\{\sigma_\varpi, \sigma_\alpha, \sigma_\beta\}$ fail to encompass their true values for only one of 45 parameters in the cases of $h_1(F|R,S,R_v,\cdot)$ and $h(F|R,\cdot)$ and three of 45 parameters for $h_2(F|R,S,R_v,\cdot)$. In total, the 95% highest posterior density estimates capture the true values in 97 of 99 cases for the form $h_1(F|R,S,R_v,\cdot)$, 90 of 99 cases for $h_2(F|R,S,R_v,\cdot)$ and 94 of the 99 parameters for the exact case $h(F|R,\cdot)$. Interestingly, the approximate form $h_1(F|R,S,R_v,\cdot)$ appears to provide more accurate estimates of the posterior distribution of the parameter set than either the second approximation $h_2(F|R,S,R_v,\cdot)$ or the exact scenario $h\left(F\left|R,\cdot\right.\right)$. The third model is also estimated using dispersed initial conditions for all three conditional densities with almost identical outcomes.[10]

The exact sampler appears more sensitive to initial conditions and slower to converge than either of the samplers based on the common factor conditional density approximates. The slower convergence observed for the third model when using $h\left(F\left|R,\cdot\right.\right)$ seems to stem from the necessity to draw F as T separate blocks rather than a single block as in the approximate cases. It would, therefore, appear prudent to commence the sampler using the approximate $h_1\left(F\left|R,S,R_v,\cdot\right.\right)$ even where the exact full conditional $h\left(F\left|R,\cdot\right.\right)$ is adopted. Pursuant to the simulation output, the real-data model is estimated using the approximate common factor conditional density $h_1\left(F\left|R,S,R_v,\cdot\right.\right)$ and verified using a smaller run based on the exact full conditional density $h\left(F\left|R,\cdot\right.\right)$.

6. Convergence of the sampler and assessment of the model

Chain convergence

To assess convergence, the Metropolis-in-Gibbs sampler is run twice, using dispersed initial conditions, to obtain two sets of draws of length 100,000 each. Convergence statistics for the parameters of interest are provided in Appendix B.[11] The autocorrelation levels and standard errors presented in Appendix B are estimated using the second run, while the Gelman-Rubin R statistics (R statistics) are estimated using the output from both runs. The standard errors are constructed using overlapping and non-overlapping batch means (Song and Schmeiser, 1993, 1995). The non-overlapping batches are derived subject to the condition that first-order autocorrelation fails to differ from zero at the 5% significance level. Since both methods provided similar estimates, only the standard errors determined using the non-overlapping batch method are provided.

---

[10] The distance between the initial conditional log-likelihoods using the first estimation method h_1, $\left\|L_2 - L_1\right\|$, is approximately 163.

[11] To preserve space, the convergence diagnostics in Appendix B are limited to the common component parameters. Convergence diagnostics for the remaining parameters are available upon request.

The autocorrelation statistics suggest that the idiosyncratic persistence, common persistence, regime-dependent intercept and global transition terms mix reasonably well with most autocorrelation tapering off by the tenth lag. The European transition parameters, regime-dependent volatilities and factor loadings exhibit greater autocorrelation suggesting that the retrieval of approximately uncorrelated draws requires intervals of greater than ten lags. The mixing properties are poorest for the idiosyncratic factor GARCH terms with autocorrelation tapering off extremely slowly. To investigate whether the convergence properties of the GARCH terms are a remnant of the drawing method, the parameters $\{\varpi, \alpha, \beta\}_i$ are also drawn as a single block using a multivariate normal proposal density. The hyper-parameters for the proposal density are determined using an initial run of the sampler where draws for each element in the block $\{\varpi, \alpha, \beta\}_i$ are undertaken pursuant to both normal and *t* distributed random walk proposal densities (see Vrontos, Dellaportas, and Politis, 2000). The resulting convergence properties (and posterior densities) for the GARCH parameters remain similar to those observed using the drawing method proposed in Section 4.

The R statistics approach their limiting value (unity) for all parameters estimated. The largest R statistic is 1.06 and corresponds to the contemporaneous French loading on the global factor. The R statistics suggest that both runs produce draws from the same posterior distribution. The standard errors suggest that a run of 100,000 draws produces estimates of the expected parameter values with typical order of precision 1/1000. The largest standard error is 0.59 for the third volatility regime associated with global innovations and represents approximately 0.6% of the expected value of the relevant parameter.

Residual ARCH effects

A preliminary model coupling common regime-dependent volatility (under the single-global factor or joint global and European factor hypotheses) with constant idiosyncratic volatility fails to sufficiently explain the heteroscedasticity observed in the weekly returns for any of the markets considered. The presence of heteroscedasticity in the weekly returns data is, therefore, inadequately explained by the chosen common volatility structure. It is clear that, at least in the case of returns observed on a weekly frequency, the idiosyncratic components exhibit heteroscedasticity that should be explicitly accounted for in a comprehensive model of the data.

In the majority of cases, given the incorporation of common volatility shifts, a first-order GARCH process suffices in explaining the residual heteroscedasticity present in the returns data. Extensions to the first-order GARCH process, however, are necessary to satisfactorily accommodate the remaining heteroscedasticity in the Australian and Belgian markets. In this respect, the Engle Lagrange Multiplier test is applied to the standardised idiosyncratic residuals to examine for any remaining ARCH effects. The idiosyncratic residuals are standardised using the mean of the draws of GARCH conditional variances obtained pursuant to the Metropolis-in-Gibbs sampler used to estimate the model. The probability levels associated with the test statistics are presented in Table 5.

Table 5 *LM test for idiosyncratic residual ARCH effects: p-levels*

Market / Lag	*1*	*2*	*3*	*4*	*5*	*6*
Austria	0.601	0.872	0.865	0.850	0.927	0.871
Belgium	0.626	0.232	0.348	0.506	0.503	0.603
Denmark	0.963	0.988	0.993	0.742	0.611	0.704
France	0.873	0.431	0.615	0.609	0.672	0.785
Germany	0.109	0.146	0.123	0.172	0.264	0.372
Italy	0.333	0.592	0.249	0.098	0.167	0.252
Netherlands	0.951	0.823	0.640	0.781	0.731	0.820
Norway	0.880	0.826	0.932	0.972	0.992	0.978
Spain	0.377	0.675	0.835	0.926	0.833	0.905
Sweden	0.853	0.983	0.994	0.907	0.774	0.798
Switzerland	0.252	0.433	0.606	0.758	0.860	0.908
U.K.	0.830	0.889	0.802	0.708	0.758	0.843
Canada	0.076	0.165	0.228	0.215	0.279	0.474
U.S.	0.653	0.690	0.089	0.142	0.165	0.248
Australia	0.501	0.755	0.905	0.417	0.525	0.639
Hong Kong	0.095	0.243	0.414	0.467	0.615	0.582
Japan	0.603	0.848	0.955	0.748	0.827	0.805
Singapore	0.548	0.823	0.902	0.939	0.931	0.969

The null hypothesis of no ARCH effects is not rejected at the 5% level for any market at any of the lags under consideration (the test statistic was obtained for the first twenty lags; only the results for the first six lags are presented). Accordingly, one may conclude that no remaining ARCH effects are present in the idiosyncratic residuals. The absence of residual ARCH effects suggests that the volatility specifications are reasonable and suffice for the conditional homoscedasticity assumption underlying the regression estimates used by the sampler. The idiosyncratic residuals for all markets (save Australia and Belgium) are modelled as GARCH (P = 1, Q = 1) processes. The idiosyncratic residuals for Belgium are modelled using $P = 2$, $Q = 1$, while the Australian idiosyncratic residuals are modelled according to $P = Q = 2$.

The properties of the extracted factors

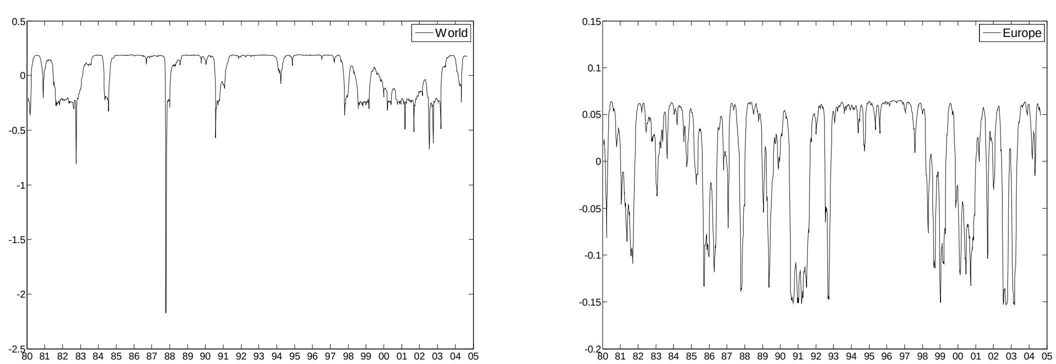
The ARCH test is also undertaken for the world and European factor residuals. The results are displayed in Table 6. In terms of the world factor, the null hypothesis is not rejected at the 5% level for any of the first twenty lags. It is, therefore, reasonable to conclude that, in the case of the world factor, volatility persistence and the associated heteroscedasticity is adequately accounted for by the three-regime Markov-switching volatility specification. The null hypothesis of no ARCH effects is rejected at the 5% level for one of the first twenty lags (lag eight) of the European factor. As such, it also appears reasonable to conclude that heteroscedasticity for the European composite is adequately accounted for by the two-regime Markov switching volatility specification.

Table 6 *LM test for common factor residual ARCH effects: p-levels*

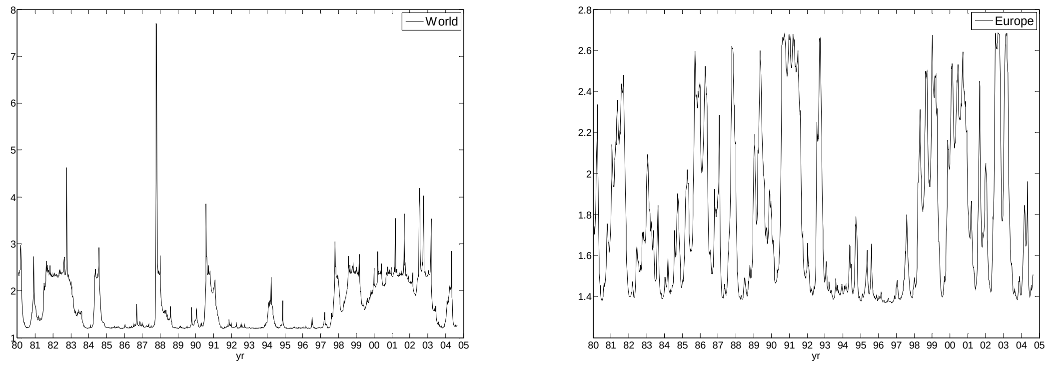
Lag/Factor	*World*	*Europe*
1	0.167	0.675
2	0.294	0.437
3	0.452	0.581
4	0.581	0.445
5	0.677	0.498
6	0.770	0.412
7	0.529	0.521
8	0.602	0.047
9	0.555	0.068
10	0.625	0.075
11	0.571	0.109
12	0.645	0.140

The global factor for developed equity markets exhibits significant, positive first order dependence while the European factor observations do not appear to exhibit any persistence (refer to Appendix C for descriptive statistics on the common factor coefficients). Both the first and second global and European regimes are highly persistent suggesting that present day volatility is likely to continue into the future for low to medium volatility levels. As expected, the high global volatility regime is the least persistence of all six volatility regimes. The positive global and European intercept terms for the low volatility regime suggest that returns tend to be positive during periods of low volatility, steering towards the negative ranges as volatility levels rise.

Figures 1 and 2 identify a number of important financial market events. The effects of rising U.S. unemployment (the U.S. unemployment rate reached 10.1% in 1982; the highest since 1940), and the subsequent surge in global markets following successive easing of U.S. interest rates is evident in 1982. The October, 1987 crash and the spike in oil prices following the 1990 U.S. invasion of Iraq are clearly observed. Higher levels of volatility associated with the 1997 Asian crisis and the continuing fear of a U.S. recession (leading to a 45-year U.S. interest rate of 1% in June, 2003) are also evident in the latter period of the study. The equivalent results for the European factor are less clear suggesting that much of the volatility in developed equity markets is embedded in the global factor. An obvious fall in volatility, coupled with sustained European specific gains, is however evident in the period 1993-1998; probably emanating from lower European interest rates following early 1990 recessionary fears for the major European economies.



Figures 1a,b. Global and European time-varying intercept levels



Figures 2a,b. Time-varying volatility for the global and European factors

7. Conclusion

This paper presents a unified model allowing for return persistence and regime and time dependent co-movement levels. The model incorporates persistent common and idiosyncratic factors, coupled with Markovian common volatility regimes and GARCH idiosyncratic volatilities and may be used to jointly assess both persistence in asset returns and the time-varying impact of common regimes on asset co-movement levels. The incorporation of GARCH idiosyncratic volatilities in a persistent, Markovian regime-dependent common factor framework renders maximum likelihood estimation an intractable problem. Instead, estimation is facilitated by construction of a Metropolis-in-Gibbs sampler. The most time consuming element of the sampler pertains to the provision of draws from the conditional density of the common factors. In this respect, the single-block approach (see, Chib and Greenberg, 1996; Kim and Nelson, 1998) is not available and draws are obtained according to their time-dependent individual conditional densities. The common factor drawing procedure is effectively an exact derivation of the Kalman filter model with GARCH innovations (cf. the approximation in King, Sentana, and Wadhwani, 1994).

To accelerate the procedure for drawing from the conditional density of the common component, two approximations to the conditional density of the common component are derived and examined. The approaches take the GARCH idiosyncratic volatilities as given by their conditional expectations or as a function of idiosyncratic volatilities determined by using the sampler history. The latter of the two approaches provides draws from the posterior density of the parameter set that are not distinguishable from their counterparts determined using the exact conditional density of the common factors.

The procedure is applied to developed equity markets to derive global, European and country-specific factors for 18 markets. There is little evidence of any residual heteroscedasticity in the country-specific or common factors suggesting that that volatility structure adequately accounts for the time-varying volatilities observed in equity prices.

Appendix A: Parameter estimation

(1) Generating the loading matrix C

As it stands, the effective error term in the equation $r_t = C(L)\mathbf{f}_t + u_t$ is both heteroscedastic and serially correlated as per:

$$\psi_i(L)u_{i,t} = \varepsilon_{i,t}, \tag{A.1}$$

$$\varepsilon_{i,t} \sim N\left(0, \sigma_{i,t}^2\right), \tag{A.2}$$

where the i subscript is a reference to the ith (i = 1 to N) market.

Multiplying both sides of the equation $r_t = C(L)\mathbf{f}_t + u_t$ by $\psi(L)$ removes the autocorrelation in the error term. For $\psi_i(L) = 1 - \psi_i L$, this implies:

$$y_{i,t} = r_{i,t} - \psi_i L r_{i,t} = c_{i,1,1}\left(\mathbf{f}_{w,t} - \psi_i L \mathbf{f}_{w,t}\right) + \ldots + c_{i,2,2}\left(\mathbf{f}_{e,t} - \psi_i L \mathbf{f}_{e,t}\right) + \varepsilon_{i,t}. \tag{A.3}$$

Conditional on $\sigma_{i,t}^2$, $\varepsilon_{i,t}$ is independent such that $\varepsilon_{i,t}\sigma_{i,t}^{-1} = z_{1,t} \sim iidN(0,1)$. Given the diagonality of G_t and $\sigma_{i,t}^2$, the vectors $c_i \subset C, \forall i \in \{1, 2, \ldots, N\}$, may be estimated equation by equation using a generalised least squares (GLS) approach. Bayesian GLS estimates are given by:

$$c_i^p \left| a, \psi_i, \sigma_{i,t}^2, F, R_i \sim N\left(\overline{a}_i, \overline{A}_i\right)\right., \tag{A.4}$$

$$\overline{A}_i = \left(\underline{A}_i^{-1} + X'X\right)^{-1}, \tag{A.5}$$

$$\overline{a}_i = \overline{A}_i\left(\underline{A}_i^{-1}\underline{a}_i + X'y\right), \tag{A.6}$$

$$X_t = \left[\psi_i(L)W_t \quad \psi_i(L)E_t \quad \psi_i(L)W_{t-1} \quad \psi_i(L)E_{t-1}\right] / \sigma_{i,t} \tag{A.7}$$

$$y_t = y_{i,t} / \sigma_{i,t} \tag{A.8}$$

where $F = \{W_1,...,W_T, E_1,...,E_T\}$ and $R_i = \{r_{i,1},...,r_{i,T}\}$. The conditional volatilities $\sigma_{i,t}$ are derived using the parameters pertaining to the current pass of the MCMC chain (see Section 5). The hyper parameters $\underline{a}_i, \underline{A}_i$ are priors for the mean and covariance of c_i such that $p(c_i) \sim N(\underline{\alpha}_i, \underline{A}_i)$.[12]

The draw c_i is accepted in accordance with the Metropolis-Hastings probability:

$$\Pr\left(c_i, c_i^p\right) = \min\left[\frac{f^*\left(r_i \middle| c_i^p, \cdot\right) p(c_i^p) \big/ p^*\left(c_i^p\right)}{f^*\left(r_i \middle| c_i, \cdot\right) p(c_i) \big/ p^*\left(c_i\right)}, 1\right], \tag{A.9}$$

where:

$$\begin{aligned} f^*\left(r_i \middle| c_i, \cdot\right) &= \prod_{\forall t} f\left(r_{i,t} \middle| F, c_i, a_i, \psi_i, \sigma_{\alpha,i}, \sigma_{\beta,i}, \sigma_{\varpi,i}\right) \\ &= \prod_{\forall t} N\left(y_{i,t} - Bf_{i,t} \middle| 0, \sigma_{i,t}^2\right). \end{aligned} \tag{A.10}$$

The prior density $p(c_i)$ is evaluated as $N\left(c_i \middle| \underline{\alpha}_i, \underline{A}\right)$ while the proposal density $p^*(c_i)$ is given by $N\left(c_i \middle| \overline{a}_i, \overline{A}_i\right)$. If the draw is accepted, c_i' forms the ith row of the N by $2K$ loading matrix C.

(2) Generating the idiosyncratic persistence ψ

The idiosyncratic persistence ψ is drawn in a similar fashion to C. In this respect, conditional on R, F, C, G and the form for $\psi(L)$, u_t is observed such that the proposed draw is straightforward.

Conditional on F and c_i, the idiosyncratic component for market i is:

$$u_{i,t} = r_{i,t} - c_{i,1,1} f_{w,t} - \ldots - c_{i,2,2} f_{e,t-1}, \tag{A.11}$$

[12] The independent prior mean (variance) for each element of c_i is zero (ten).

From $\psi_i(L) = 1 - \psi_i L$, and given the diagonality of G_t, ψ_i is dependent only on $\{u_{i,t}, \sigma_{i,t}^2\}$ such that a proposal density based on GLS estimation is naturally employed. Bayesian GLS estimates are given by:

$$\psi_i^p \left| u_i, \sigma_{i,t}^2 \sim N\left(\overline{a}_i, \overline{A}_i\right)\right., \tag{A.12}$$

$$\overline{A}_i = \left(\underline{A}_i^{-1} + X'X\right)^{-1}, \tag{A.13}$$

$$\overline{a}_i = \overline{A}_i\left(\underline{A}_i^{-1}\underline{a}_i + X'y\right), \tag{A.14}$$

$$X_t = u_{i,t-1} / \sigma_{i,t}, \tag{A.15}$$

$$y_t = u_{i,t} / \sigma_{i,t}, \tag{A.16}$$

where $\underline{a}_i, \underline{A}_i$ are priors for the mean and covariance of ψ_i such that $p(\psi_i) \sim N\left(\underline{\alpha}_i, \underline{A}_i\right)$.[13] Rejection sampling is used to ensure the stationarity of the *i*th idiosyncratic component u_i (i.e.: $|\psi_i| < 1$). The draw ψ_i^p is accepted on a Metropolis-Hastings basis in the same manner as c_i (refer to equations (A.9)-(A.10) replacing c_i, c_i^p with ψ_i, ψ_i^p).

(3) Generating the factor persistence Φ

The common factors are modelled as:

$$\Phi(L)\tilde{f}_t - \tilde{\mu}_t = \begin{bmatrix} v_{w,t} \\ v_{e,t} \end{bmatrix}, \tag{A.17}$$

$$\Phi(L) = \begin{bmatrix} \phi_w(L) & 0 \\ 0 & \phi_e(L) \end{bmatrix}, \tag{A.18}$$

$$\begin{bmatrix} v_{w,t} \\ v_{e,t} \end{bmatrix} = \begin{bmatrix} w_t \\ e_t \end{bmatrix} \sim N\left(0, \tilde{H}_t\right). \tag{A.19}$$

$$E\left(v_{q,t}\varepsilon_{i,s}\right) = 0, \forall q \in \{w, e\}, i \in \{1, 2, ..., N\}, s \in \{1, 2, ..., T\}. \tag{A.20}$$

[13] The independent prior mean (variance) for ψ_i is zero (ten).

Given (A.20) and $\tilde{H}_t = diag\left(\left[\sigma^2_{w,t} \quad \sigma^2_{e,t}\right]\right)$, the distribution of ϕ_q depends only on the qth factor:

$$v_{q,t} = \mathbf{f}_{q,t} - \mu_{q,t} - \phi_{q,1}\mathbf{f}_{q,t-1} - \phi_{q,2}\mathbf{f}_{q,t-2} - \phi_{q,3}\mathbf{f}_{q,t-3}\,, \quad \text{(A.21)}$$

$$v_{q,t} \sim N\left(0, \sigma^2_{q,t}\right), \quad \text{(A.22)}$$

$$q \in \{w, e\}\,.$$

Conditional on the state system $S_t = \{S_{w,t}, S_{e,t}\}$ and the state-specific variances $\sigma^2_{q,m}, \forall m \in \{1, 2, ..., M(q)\}$, GLS estimates may be used to obtain draws of ϕ_q as per:

$$\phi_q \left| \mathbf{f}_q, \sigma^2_q, \mu_q \sim N\left(\overline{a}_q, \overline{A}_q\right)\right., \quad \text{(A.23)}$$

$$\overline{A}_q = \left(\underline{A}_q^{-1} + X'X\right)^{-1}, \quad \text{(A.24)}$$

$$\overline{a}_q = \overline{A}_q\left(\underline{A}_q^{-1}\underline{a}_q + X'y\right), \quad \text{(A.25)}$$

$$X_t = \left[\mathbf{f}_{q,t-1} \quad \mathbf{f}_{q,t-2} \quad \mathbf{f}_{q,t-3}\right] / \sigma_{q,t}\,, \quad \text{(A.26)}$$

$$y_t = \mathbf{f}_{q,t} / \sigma_{q,t}\,, \quad \text{(A.27)}$$

where $\underline{a}_q, \underline{A}_q$ are priors for the mean and covariance of ϕ_q such that $p\left(\phi_q\right) \sim N\left(\underline{\alpha}_q, \underline{A}_q\right)$.[14] To ensure identification of the regime dependent variances, rejection sampling is employed such that the roots of $\phi_q(L)$ lie outside the unit circle.

(4) Generating the factor intercept μ

The intercept terms are defined as:

$$\Phi(L)\tilde{f}_t = \begin{bmatrix} \mu_{w,t} \\ \mu_{e,t} \end{bmatrix} + \begin{bmatrix} v_{w,t} \\ v_{e,t} \end{bmatrix}, \quad \text{(A.28)}$$

[14] The independent prior mean (variance) for each element in ϕ_q is zero (ten).

$$\mu_{q,t} = \mu_q{}' S_{q,t} \ ,\forall q \in \{w, e\} \ , \tag{A.29}$$

The absence of any dependence between the world and European factors implies that the distribution of μ_q depends only on the qth factor:

$$v_{q,m,t} = \mathbf{f}_{q,t} - \mu_{q,m} - \phi_{q,1}\mathbf{f}_{q,t-1} - \phi_{q,2}\mathbf{f}_{q,t-2} - \phi_{q,3}\mathbf{f}_{q,t-3} \ , \tag{A.30}$$

$$v_{q,m,t} \sim N\left(0, \sigma^2_{q,m}\right), \tag{A.31}$$

$$m \in \{1, 2, ..., M(q)\},\ q \in \{w, e\} \ . \tag{A.32}$$

Conditional on the qth set of states $S_{q,t}$, the draw μ_q may, therefore, be obtained from its full conditional density using:

$$\mu_q^* \left| \mathbf{f}_q, \sigma_q^2, \phi_q \sim N\left(\overline{a}_q, \overline{A}_q\right)\right., \tag{A.33}$$

$$\overline{A}_q = \left(\underline{A}_q^{-1} + X'X\right)^{-1}, \tag{A.34}$$

$$\overline{a}_q = \overline{A}_q \left(\underline{A}_q^{-1} \underline{a}_q + X'y\right), \tag{A.35}$$

$$X_t = \begin{bmatrix} 1 & S_{q,2,t} & S_{q,3,t} \end{bmatrix} \Big/ \sigma_{q,t} \ , \tag{A.36}$$

$$y_t = \mathbf{f}_{q,t} \big/ \sigma_{q,t} \ , \tag{A.37}$$

where $\mu_q^* = \begin{bmatrix} \mu_{q,1} & \mu_{q,2}^* & \mu_{q,3}^* \end{bmatrix}'$ and $\underline{a}_q, \underline{A}_q$ are priors for the mean and covariance of μ_q^* such that $p\left(\mu_q^*\right) \sim N\left(\underline{\alpha}_q, \underline{A}_q\right)$.[15] The indicator variable $S_{q,m,t}$ is equal to unity in regime $m(q)$ and zero otherwise. Pursuant to the construction (A.36), $\mu_{q,2} = \mu_{q,1} + \mu_{q,2}^*$ and $\mu_{q,3} = \mu_{q,1} + \mu_{q,3}^*$. Note that in the case of μ_e, $M(e) = 2$ such that the regression matrix X_t is replaced by $X_t = \begin{bmatrix} 1 & S_{q,2,t} \end{bmatrix} \Big/ \sigma_{q,t}$. Given the identification restriction

[15] The independent prior mean (variance) for each element of μ_q^* is zero (ten).

$\sigma_{q,m+1}^2 \gg \sigma_{q,m}^2 \ \forall m$, the regime-dependent intercepts may be estimated without further restriction.

(5) Generating the common variance H_t

Given $E\left(v_{q,t}\varepsilon_{i,s}\right)=0,\ \forall q,i,s$, $\tilde{H}_t = diag\left(\left[\sigma_{w,t}^2 \quad \sigma_{e,t}^2\right]\right)$ and the restrictions on the state generating process, the distribution of $\sigma_{q,t}^2$ depends on $F_{q,}\ S_q$, ϕ_q and μ_q:

$$v_{q,t} = f_{q,t} - \mu_{q,t} - \phi_{q,1}f_{q,t-1} - \phi_{q,2}f_{q,t-2} - \phi_{q,3}f_{q,t-3}\,, \tag{A.38}$$

$$v_{q,t} \sim N\left(0,\sigma_{q,t}^2\right), \tag{A.39}$$

$$q \in \{w,e\}\,. \tag{A.40}$$

The variance $\sigma_{q,t}^2 = \sigma_{q,1}^2 S_{q,1,t} + \sigma_{q,2}^2 S_{q,2,t} + \sigma_{q,3}^2 S_{q,3,t}$ is modelled as (suppressing the q subscript in $\sigma_{q,m}^2, \tilde{\sigma}_{q,m}^2$) (see Kim and Nelson, 1998):

$$\sigma_1^2\left(1+\tilde{\sigma}_2^2 S_{q,2,t}\right)\left(1+\tilde{\sigma}_2^2 S_{q,3,t}\right)\left(1+\tilde{\sigma}_3^2 S_{q,3,t}\right), \tag{A.41}$$

where $S_{q,m,t}$ is an indicator variable taking on the value unity if $S_{q,t}=m$ and zero otherwise. Given the form adopted for $\sigma_{q,t}^2$, σ_m^2 is sampled conditional on $\sigma_{\neq m}^2$ as follows:

$$\sigma_1^{-2} \sim gamma\left(\tfrac{1}{2}N^{(1)}, \tfrac{1}{2}S^{(1)}\right), \tag{A.42}$$

$$N^{(1)} = S_{q,1} + S_{q,2} + S_{q,3}, \tag{A.43}$$

$$S^{(1)} = v^{(1)\prime}v^{(1)}, \tag{A.44}$$

$$v_t^{(1)} = v_{q,t} \,/ \sqrt{\left(1+\tilde{\sigma}_2^2 S_{q,2,t}\right)\left(1+\tilde{\sigma}_2^2 S_{q,3,t}\right)\left(1+\tilde{\sigma}_3^2 S_{q,3,t}\right)}\,, \tag{A.45}$$

$$\left(1+\tilde{\sigma}_2^2\right)^{-1} \sim gamma\left(\tfrac{1}{2}N^{(2)}, \tfrac{1}{2}S^{(2)}\right), \tag{A.46}$$

$$N^{(2)} = S_{q,2} + S_{q,3}, \tag{A.47}$$

$$S^{(2)} = v^{(2)\prime} v^{(2)}, \tag{A.48}$$

$$v_t^{(2)} = \left(S_{q,2,t} + S_{q,3,t}\right) v_{q,t} \Big/ \sqrt{\sigma_1^2 \left(1 + \tilde{\sigma}_3^2 S_{q,3,t}\right)}, \tag{A.49}$$

$$\left(1 + \tilde{\sigma}_3^2\right)^{-1} \sim gamma\left(\tfrac{1}{2} N^{(3)}, \tfrac{1}{2} S^{(3)}\right), \tag{A.50}$$

$$N^{(3)} = S_{q,3}, \tag{A.51}$$

$$S^{(3)} = v^{(3)\prime} v^{(3)}, \tag{A.52}$$

$$v_t^{(3)} = S_{q,3,t}\, v_{q,t} \Big/ \sqrt{\sigma_1^2 \left(1 + \tilde{\sigma}_2^2 S_{q,2,t}\right)\left(1 + \tilde{\sigma}_2^2 S_{q,3,t}\right)}, \tag{A.53}$$

where $S_{q,m} = \sum_{\forall t} S_{q,m,t}$.

The qth state vector S_q is identified through the restrictions $\sigma_{q,3}^2 > \sigma_{q,2}^2 > \sigma_{q,1}^2 \Leftrightarrow \tilde{\sigma}_{q,3}^2 > 0, \tilde{\sigma}_{q,2}^2 > 0$. Note that in the case $q = e$ only the first two states are relevant. The process for obtaining the conditional draws, however, remains the same given the appropriate imposition $S_{e,3,t} = 0 \forall t$.

(6) Generating the state (or regime) vectors $S_{w,t}, S_{e,t}$

Define $J_{q,m,t}$ as the probability of the latent factor $q \in \{w, e\}$ being in state (or regime) m at time t. State m is an element of the discrete set $m \in \{1, 2, ..., M(q)\}$. $J_{q,t}$ is the $M(q)$-vector of probabilities $\{J_{q,m,t} \forall m\}$. Pr_q is the time-invariant state transition matrix for common factor q. The process for drawing S_q is as follows:

1. Obtain $J_{q,t|t-1}$.
2. Update $J_{q,t|t-1}$ with the time-t information set to obtain $J_{q,t|t}$.
3. Given $\{J_{q,t|t} \forall t, t \in \{1, 2, ..., T\}\}$, update $J_{q,t|t}$ with information I_T to obtain $J_{q,t|T}$:

   3.0 Initialise the updating process at time $t = T$ by using the random value $c \sim U(0, 1)$ to draw $S_{q,T}$ from its discrete probability distribution $J_{q,T|T}$.

3.1 Set $t = t$-1 while t>1.

3.2 Use $S_{q,t+1}$ to obtain the update $J_{q,t|T}$.

3.3 Draw $S_{q,t}$ using $J_{q,t|T}$.

3.4 Goto 3.1.

Obtaining $J_{q,t|t-1}$, $J_{q,t|t}$

Conditional on $Pr_{q.t} = Pr_q$ all time-t information pertinent to the time t+1 state estimate is already incorporated in $J_{q,t|t}$ and the one-step ahead estimate is given by,

$$J_{q,t+1|t} = P\left(S_{q,t+1} \left| I_t \right.\right) = Pr_q J_{q,t|t}. \tag{A.54}$$

Upon realization of the time t+1 information set, the probabilities $J_{q,t+1|t}$ may be updated using the state-contingent density $h_{t+1}(q,m) = h\left(f_{q,t+1} \left| S_{q,m}, I_t \right.\right)$ (see Hamilton, 1989, 1990, for further information).

The filtering process is initiated by setting the unconditional state probabilities determined through Pr_q as $J_{q,0|0}$.[16] Given a third-order persistence process for f_q, it is clear that $u_{q,m,t+1}$ is observed only for $t \geq 3$. This problem may be overcome by treating $h_{t+1}(q,m)$ as a diffuse density ($h_{t+1}(q,m)$ = c for $t = 0,1,2$ pursuant to a professed ignorance about $u_{q,m,t+1}$ for $t < 3$) such that $J_{q,t+1|t+1} = Pr_q J_{q,t|t}$ for $t = 0,1,2$ and proceeding as normal for $t \geq 3$.

Obtaining $J_{q,t|T}$ *and drawing* $S_{q,t}$

Conditional on $S_{q,t+1} = j$, the incremental information set $I_{\neq\{S_q^{t+1}\},\neq\{f_q^t\}}$ provides no further information regarding $S_{q,t}$ such that:

[16] The unconditional state probabilities may be obtained using the characteristic decomposition of the transition matrix. In this respect, $J_{q,0|0}$ is given by the appropriately scaled (to ensure the probabilities sum to unity) characteristic vector associated with the largest characteristic root (being equal to unity).

$$P\left(S_{q,t}\left|I_T\right.\right)=P\left(S_{q,t}\left|S_{q,t+1},I_t\right.\right), \tag{A.55}$$

where $\neq\{S_q^{t+1}\},\neq\{f^t\}$ refers to the states and qth factor values observed after times t+1 and t respectively.

Given $S_{q,t+1}=j$ and $J_{q,t|t}$, the conditional on I_T draw of $S_{q,t}$ can be obtained using:

$$J_{q,t|T}=f\left(Pr_{q,j},J_{q,t|t}\right)=diag\left(J_{q,t|t}\right)Pr_{q,j}{}'\Big/1_{M(q)}{}'\,diag\left(J_{q,t|t}\right)Pr_{q,j}{}', \tag{A.56}$$

where $Pr_{q,j}$ is the jth row of Pr_q, $1_{M(q)}$ is an $M(q)$-dimensional column vector of ones and $diag(x)$ is a square matrix with the vector x on its diagonal.

$S_{q,t}$ is sampled by selecting from the 'smoothed' probability distribution $J_{q,t|T}$. The selection is made by reference to a draw from the continuous U(0,1) distribution. Given $u_{q,m}\sim N\left(0,\sigma_{q,m}^2\right)$, the identification of the regime-switching volatility $\sigma_{q,m}^2$, $\forall m\in\left\{1,2,...,M\left(q\right)\right\}$, requires that each of the M states is observed for the qth common factor. Therefore an entire draw S_q is rejected if $\nexists S_{q,t}=m,\forall m\in\left\{1,2,...,M\left(q\right)\right\}$ for at least $\dim\left(x\right)$ time periods t. A corollary of the identification restriction on S_q is that the transition matrix Pr_q, conditional on $S_{q,t}=m,\forall m\in\left\{1,2,...,M\left(q\right)\right\}$, must not be closed with respect to a transition to any state m', $m'\neq m$. The aforementioned identification restriction on Pr_q may be imposed by the truncation $Pr_{q,i,j}\in\left(0,1\right)$.

(7) Generating the transition probability matrix Pr_q

Conditional on the qth regime vector S_q, the remaining conditional variables provide no further information such that:

$$
\begin{aligned}
g\left(Pr_{q,t}\middle|I_T\right) &= g\left(Pr_{q,t}\middle|S_{q,t},S_{q,t-1}\right) \\
&= Pr_{q,1,1}{}^{d_t^{1,1}} Pr_{q,1,2}{}^{d_t^{1,2}} Pr_{q,1,3}{}^{d_t^{1,3}} \\
&\quad x\; Pr_{q,2,1}{}^{d_t^{2,1}} Pr_{q,2,2}{}^{d_t^{2,2}} Pr_{q,2,3}{}^{d_t^{2,3}} \\
&\quad x\; Pr_{q,3,1}{}^{d_t^{3,1}} Pr_{q,3,2}{}^{d_t^{3,2}} Pr_{q,3,3}{}^{d_t^{3,3}},
\end{aligned}
\tag{A.57}
$$

where $d_t^{i,j}$ is equal to one if $S_{q,t} = i, S_{q,t-1} = j$ and zero otherwise (see, for example, Kim and Nelson, 1998).

Given time-invariant transition probabilities $Pr_{q,t} = Pr_q$, the full conditional density $g\left(Pr_q\middle|S_q\right) = \prod_{\forall t} g\left(Pr_{q,t}\middle|S_{q,t},S_{q,t-1}\right)$ simplifies to:

$$
\begin{aligned}
g\left(Pr_q\middle|I_T\right) &= g\left(Pr_q\middle|S_q\right) \\
&= \prod_{\forall t}\prod_{\forall j} Pr_{q,1,j}{}^{d_t^{1,j}} Pr_{q,2,j}{}^{d_t^{2,j}} Pr_{q,3,j}{}^{d_t^{3,j}} \\
&= \prod_{\forall j} Pr_{q,1,j}{}^{\sum_{\forall t} d^{1,j}} Pr_{q,2,j}{}^{\sum_{\forall t} d^{2,j}} Pr_{q,3,j}{}^{\sum_{\forall t} d^{3,j}} \\
&= \prod_{\forall j} Pr_{q,1,j}{}^{n^{1,j}} Pr_{q,2,j}{}^{n^{2,j}} Pr_{q,3,j}{}^{n^{3,j}},
\end{aligned}
\tag{A.58}
$$

where $n^{i,j}$, $i, j \in \{1, 2, \ldots M(q)\}$, is the total number of transitions to state i from state j for state-vector q. Clearly, the conditional value attributed to $Pr_{q,i,j} \in (0,1)$, being the (i,j)th element of Pr_q, is determined by the weight of the ith element in the transition bin $\{n^{i'j} : i' \in 1, 2, ..., M(q)\}$. Given $n^{i,j} > 0\ \forall i$, we obtain $Pr_{q,i,j} \in (0,1)$, $\sum_{\forall i} Pr_{q,i,j} = 1$.

Given a uniform prior, a draw from the posterior density $g\left(Pr_q\middle|S_q\right)$ may be obtained using the Dirichlet distribution:

$$
Pr_{q,j} \sim dirichlet\left(\begin{bmatrix} n^{1,j} & \cdots & n^{M(q),j}\end{bmatrix}'\right),
\tag{A.59}
$$

$$Pr_{q,j} = \left[Pr_{q,1,j} \quad \cdots \quad Pr_{q,M(q),j} \right]', \tag{A.60}$$

for $j = 1, 2, ..., M(q)$. $Pr_{q,j}$ is, therefore, the *j*th column of the transition probability matrix Pr_q. The Dirichlet density also addresses the a priori identification restriction $Pr_{q,i,j} \in (0,1)$.

Appendix B: Common convergence statistics

Variable	*Gelman-Rubin R*	*ACF(1)*	*ACF(10)*	*ACF(50)*	*Std. Error*	*Mean*
$\phi_{w,1}$	1.000	0.664	0.288	0.071	0.001	0.151
$\phi_{w,2}$	1.000	0.246	0.004	-0.004	0.000	0.020
$\phi_{w,3}$	1.000	0.229	0.008	0.003	0.000	0.014
$\phi_{e,1}$	1.000	0.733	0.298	0.070	0.001	0.034
$\phi_{e,2}$	1.000	0.282	0.001	0.004	0.000	0.051
$\phi_{e,3}$	1.000	0.295	0.005	0.001	0.000	0.017
$\mu_{w,1}$	1.000	0.397	0.008	0.001	0.000	0.189
$\mu_{w,2}$	1.000	0.357	0.044	0.003	0.001	-0.237
$\mu_{w,3}$	1.000	0.246	0.003	0.000	0.012	-2.232
$\mu_{e,1}$	1.000	0.649	0.131	0.022	0.001	0.070
$\mu_{e,2}$	1.000	0.545	0.137	0.009	0.003	-0.171
$\sigma^2_{w,1}$	1.000	0.845	0.281	0.044	0.002	1.423
$\sigma^2_{w,2}$	1.000	0.693	0.165	0.017	0.008	5.450
$\sigma^2_{w,3}$	1.000	0.060	0.004	0.006	1.014	96.176
$\sigma^2_{e,1}$	1.000	0.929	0.571	0.084	0.006	1.821
$\sigma^2_{e,2}$	1.001	0.824	0.411	0.056	0.025	7.440

Transition parameters	*Gelman-Rubin R*	*ACF(1)*	*ACF(10)*	*ACF(50)*	*Std. Error*	*Mean*
World						
1,1	1.000	0.540	0.076	0.006	0.000	0.982
1,2	1.000	0.645	0.106	0.007	0.000	0.014
1,3	1.000	0.564	0.050	0.003	0.000	0.004
2,1	1.000	0.631	0.105	0.004	0.000	0.034
2,2	1.001	0.689	0.156	0.005	0.000	0.955
2,3	1.001	0.835	0.301	0.007	0.000	0.011
3,1	1.000	0.484	0.018	-0.007	0.001	0.196
3,2	1.000	0.467	0.024	-0.008	0.001	0.437
3,3	1.000	0.409	0.013	-0.008	0.001	0.368
Europe						
1,1	1.001	0.895	0.573	0.092	0.000	0.964
2,2	1.000	0.892	0.501	0.055	0.001	0.901

Appendix C: The common factor coefficients

Variable	*Median*	*Mean*	*Std. deviation*	*95% BCI*		*Pr>0*
$\phi_{w,1}$	0.151	0.151	0.047	0.060	0.243	1.000
$\phi_{w,2}$	0.020	0.020	0.033	-0.044	0.084	0.732
$\phi_{w,3}$	0.014	0.014	0.031	-0.047	0.075	0.671
$\phi_{e,1}$	0.033	0.034	0.052	-0.068	0.137	0.739
$\phi_{e,2}$	0.051	0.051	0.034	-0.016	0.119	0.932
$\phi_{e,3}$	0.016	0.017	0.034	-0.051	0.084	0.685
$\mu_{w,1}$	0.189	0.189	0.052	0.088	0.292	1.000
$\mu_{w,2}$	-0.232	-0.237	0.146	-0.541	0.034	0.044
$\mu_{w,3}$	-2.212	-2.232	2.503	-7.231	2.802	0.173
$\mu_{e,1}$	0.068	0.070	0.074	-0.072	0.219	0.830
$\mu_{e,2}$	-0.154	-0.171	0.232	-0.682	0.238	0.224
$\sigma^2_{w,1}$ [17]	1.419	1.423	-	1.139	1.699	-
$\sigma^2_{w,2}$	5.401	5.450	-	4.150	6.860	-
$\sigma^2_{w,3}$	59.391	96.176	-	10.864	265.182	-
$\sigma^2_{e,1}$	1.816	1.821	-	1.285	2.364	-
$\sigma^2_{e,2}$	7.192	7.440	-	4.921	10.522	-

Transition Parameters	*Mean*	*Std. deviation*	*95% BCI*	*Pr>0*
World				
1,1	0.982	0.007	0.967	0.993
1,2	0.014	0.007	0.003	0.029
1,3	0.004	0.003	0.000	0.012
2,1	0.034	0.015	0.011	0.069
2,2	0.955	0.019	0.912	0.982
2,3	0.011	0.013	0.000	0.044
3,1	0.195	0.163	0.006	0.600
3,2	0.437	0.198	0.079	0.821
3,3	0.368	0.185	0.065	0.755
Europe				
1,1	0.964	0.019	0.916	0.989
2,2	0.900	0.052	0.767	0.967

[17] The 95% intervals for the common variance coefficients are 95% HPDs (highest posterior densities) rather than 95% BCI levels.

References

Aguilar, O., West, M., 2000. Bayesian dynamic factor models and portfolio allocation. Journal of Business and Economic Statistics 18, 338-357.

Bauwens, L., Lubrano, M., 1998. Bayesian inference on GARCH models using the Gibbs sampler. Econometrics Journal 1, 23-46.

Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modelling in finance: a review of the theory and empirical evidence. Journal of Econometrics 52, 5-59.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure and mean-variance analysis on large asset markets. Econometrica 51, 1281-1304

Chib, S., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in econometrics. Econometric Theory 12, 409-431.

Daniel, K.D., Hirshleifer, D., Subrahmanyam, A., 2001. Overconfidence, arbitrage, and equilibrium asset pricing. Journal of Finance 56, 921-965.

De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990a. Positive feedback investment strategies and destabilizing rational speculation. Journal of Finance 45, 379-395.

De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990b. Noise trader risk in financial markets. Journal of Political Economy 98, 703-738.

Diebold, F.X., Nerlove, M., 1989. The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. Journal of Applied Econometrics 4, 1-21.

Engle, R.F., Susmel, R., 1993. Common volatility in international equity markets. Journal of Business and Economic Statistics 11, 167-176.

Hamilton, J.D., Lin, G., 1996. Stockmarket volatility and the business cycle. Journal of Applied Econometrics 11, 573-593.

Hong, H., Stein, J.C., 1999. A unified theory of underreaction, momentum trading and overreaction in asset markets. Journal of Finance 55, 265-295.

Kim, C.J., Nelson, C.R., 1998. Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. Review of Economics and Statistics 80, 188-201.

King, M., Sentana, E., Wadhwani, S., 1994. Volatility and links between national stock markets. Econometrica 62, 901-933.

King, M., Wadhwani, S., 1990. Transmission of volatility between stock markets. Review of Financial Studies 3, 5-33.

Nakatsuma, T., 1998. A Markov-chain sampling algorithm for GARCH models. Studies in Nonlinear Dynamics and Econometrics 3, 107-117.

Nakatsuma, T., 2000. Bayesian analysis of ARMA-GARCH models: a Markov chain sampling approach. Journal of Econometrics 95, 57-69.

Roll, R., Ross, S.A., 1980. An empirical investigation of the Arbitrage Pricing Theory, Journal of Finance 5, 1073-1103.

Song, W.T., Schmeiser, B.W., 1993. Variance of the sample mean: properties and graphs of quadratic-form estimators. Operations Research 41, 501–517.

Song, W.T., Schmeiser, B.W., 1995. Optimal mean-squared-error batch sizes. Management Science 41, 110–123.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22, 1701-1728.

Vrontos, I.D., Dellaportas, P., Politis, D.N., 2000. Full Bayesian inference for GARCH and EGARCH models. Journal of Business and Economic Statistics 18, 187-198.