Internet Appendix

International Correlation Asymmetries: Frequent-but-Small and Infrequent-but-Large Equity Returns

BRUNO SOLNIK*and THAISIRI WATEWAI†

*Finance Department, LSK Business School, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. E-mail: solnik@ust.hk.

†Department of Banking and Finance, Chulalongkorn Business School, Chulalongkorn University, Bangkok, Thailand 10330. Tel.: +66 22185691 ext. 13. E-mail: thaisiri@cbs.chula.ac.th. (Corresponding Author)

Appendix A EM-based Estimation Method

This section derives the complete-data log-likelihood function and describes the expectation and maximization steps for the hidden Markov model developed in Section 3. We first consider the complete-data log-likelihood function:

$$\begin{aligned}\log L(\mathbb{C}_T \mid \Theta) = \log L(Y_1, Z_1, \Delta N_1, \delta_1 \mid \Theta) + \log L(Y_2, Z_2, \Delta N_2, \delta_2 \mid \mathbb{C}_1, \Theta) \\ + \cdots + \log L(Y_T, Z_T, \Delta N_T, \delta_T \mid \mathbb{C}_{T-1}, \Theta)\end{aligned} \tag{A.1}$$

Let $\mathbf{1}(A)$ denote an indicator function taking value 1 if A is true, and 0 otherwise. Using the Markov property, we have

$$\begin{aligned}\log L(Y_t, Z_t, \Delta N_t, \delta_t \mid \mathbb{C}_{t-1}, \Theta) &= \sum_{y=1}^{K}\sum_{k=1}^{K} \mathbf{1}(Y_{t-1} = y, Y_t = k) \log p_{yk} \\ &+ \sum_{k=1}^{K} \mathbf{1}(Y_t = k) \left\{ -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma(k)| - \frac{1}{2}(Z_t - b(k))'\Sigma(k)^{-1}(Z_t - b(k)) \right\} \\ &+ \sum_{k=1}^{K}\sum_{c=0}^{\infty} \mathbf{1}(Y_t = k, \Delta N_t = c) \left\{ -\lambda(k) + c\log\lambda(k) - \log(c!) \right\} \\ &+ \sum_{k=1}^{K}\sum_{c=0}^{\infty} \mathbf{1}(Y_t = k, \Delta N_t = c) \sum_{l=1}^{c} \left\{ -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Omega(k)| - \frac{1}{2}(\delta_t^l - \eta(k))'\Omega(k)^{-1}(\delta_t^l - \eta(k)) \right\}\end{aligned} \tag{A.2}$$

Substituting (A.2) into (A.1) and taking the expectation conditional on $\mathbb{X}_T$ and $\Theta^{(p)}$, we have

$$\begin{aligned}Q(\Theta, \Theta^{(p)}) &= \mathbb{E}\left[\log L(\mathbb{C}_T \mid \Theta) \;\middle|\; \mathbb{X}_T, \Theta^{(p)} \right] \\ &= \sum_{k=1}^{K} P(Y_1 = k|\mathbb{X}_T, \Theta^{(p)}) \log \varrho_k + \sum_{t=2}^{T}\sum_{y=1}^{K}\sum_{k=1}^{K} P(Y_{t-1} = y, Y_t = k|\mathbb{X}_T, \Theta^{(p)}) \log p_{yk} \\ &- \frac{nT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K} P(Y_t = k|\mathbb{X}_T, \Theta^{(p)}) \log|\Sigma(k)| \\ &- \frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K} P(Y_t = k|\mathbb{X}_T, \Theta^{(p)}) \mathbb{E}\left[(Z_t - b(k))'\Sigma(k)^{-1}(Z_t - b(k)) \;\middle|\; Y_t = k, \mathbb{X}_T, \Theta^{(p)} \right] \\ &+ \sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{c=0}^{\infty} P(Y_t = k, \Delta N_t = c|\mathbb{X}_T, \Theta^{(p)}) \left\{ -\lambda(k) + c\log\lambda(k) - \log(c!) \right\} \\ &- \frac{n}{2}\log(2\pi)\sum_{t=1}^{T}\sum_{c=0}^{\infty} cP(\Delta N_t = c|\mathbb{X}_T, \Theta^{(p)}) - \frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{c=0}^{\infty} cP(Y_t = k, \Delta N_t = c|\mathbb{X}_T, \Theta^{(p)}) \log|\Omega(k)| \\ &- \frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{c=0}^{\infty} cP(Y_t = k, \Delta N_t = c|\mathbb{X}_T, \Theta^{(p)}) \mathbb{E}\left[(\delta_t^1 - \eta(k))'\Omega(k)^{-1}(\delta_t^1 - \eta(k)) \;\middle|\; Y_t = k, \Delta N_t = c, \mathbb{X}_T, \Theta^{(p)} \right]\end{aligned} \tag{A.3}$$

In the expectation step (E-step), we need to compute the conditional probabilities and expectations in $Q(\Theta, \Theta^{(p)})$ for a given parameter set $\Theta^{(p)}$. In the maximization step (M-step), we find Θ that maximizes $Q(\Theta, \Theta^{(p)})$. Then we set $\Theta^{(p+1)} = \arg\max_{\Theta} Q(\Theta, \Theta^{(p)})$. The algorithm starts with an initial set of parameters $\Theta^{(0)}$ and the E-step and M-step are run alternately until a termination (convergence) condition is met.

We use a modified version of the forward-backward algorithm of Baum et al. (1970) to compute the conditional probabilities and expectations in the E-step. In particular, we use the modified forward probabilities of Lystig and Hughes (2002)

$$\omega(t, k) = P(Y_t = k, r_t \mid \mathbb{X}_{t-1}, \Theta) \tag{A.4}$$

and the backward (or smoothed) probabilities

$$\gamma(t,k) = P(Y_t = k \mid \mathbb{X}_T, \Theta). \tag{A.5}$$

With this choice of the forward and backward probabilities we can avoid the underflow problem and easily compute the log-likelihood value of the incomplete data:

$$L(\mathbb{X}_T \mid \Theta) = \sum_{t=1}^{T} \log \left(\sum_{k=1}^{K} \omega(t,k) \right). \tag{A.6}$$

Due to the choices of our complete data, we are able to solve for $\Theta^{(p+1)}$ in the M-Step explicitly, avoiding the use of a computationally intensive search algorithm.

Finally, the asymptotic standard errors of the parameters can be obtained from the Fisher information matrix. Specifically, the asymptotic distribution of the estimates of parameters in Θ is normal with mean Θ_0 and variance $\mathcal{I}_\Theta^{-1}$ where Θ_0 is the set of the true parameters, and $\mathcal{I}_\Theta$ is the information matrix[1]. The re-parameterization technique can be used to obtain the asymptotic distribution of the estimates of a new set of parameters. Lystig and Hughes (2002) provide a recursive method to compute the first and second derivatives of the log-likelihood function for hidden Markov models, which can be used to compute the Fisher information matrix.

[1] In Appendix B we show that our model is identifiable so the Fisher information is nonsingular (see Rothenberg, 1971).

Appendix B Model Identification

For a given parameter set for the diffusion and jump components $\theta = (\mu, \Sigma, \lambda, \eta, \Omega)$, let $F_\theta : \mathbb{R}^n \to [0,1]$ denote the cumulative distribution function of the $n-$dimensional vector of returns

$$R_t = Z_t + \sum_{m=1}^{\Delta N_t} \delta_t^m. \tag{B.1}$$

That is, F_θ represents the distribution of the returns for a given regime characterized by the parameter set θ. Let $\mathcal{F}$ denote the class of all distribution functions F_θ, and $\phi(s) = \mathbb{E}[e^{s'R_t}], s \in \mathbb{R}^n$ the moment generating function of R_t. It is easy to show that

$$\phi(s) = \exp\left(\frac{1}{2}s'\Sigma s + s'\mu + \lambda\left(e^{\frac{1}{2}s'\Omega s + s'\eta} - 1\right)\right). \tag{B.2}$$

With the moment generating function above, one can show that any finite mixture of $F_{\theta_1}, \ldots, F_{\theta_k}$ is identifiable for the univariate case ($n = 1$) using a similar approach as in the proof of Proposition 1 of Teicher (1963), which derives the result for the Gaussian distribution. The identifiability of the finite mixture of the multivariate case can be proved from the univariate case similarly to the proof of the Gaussian distribution in Proposition 2 of Yakowitz and Spragins (1968). More specifically, by assuming the finite mixture of the multivariate case is not identifiable, we can show along the line of the Gaussian distribution case that it contradicts with the fact that the univariate case is identifiable. Finally, the identifiability of the finite mixture of distribution F_θ can be extended to the identifiability of our hidden Markov model using Theorem 2 of Teicher (1967). See also Section 12.4 of Cappe, Moulines, and Ryden (2005) for detailed explanations of how to apply Theorem 2 of Teicher (1967) to prove the identifiability of general hidden Markov models from finite mixtures.

Appendix C Test Statistics for Model Selection

We identify the number of regimes and the existence of jumps based on Davies (1987) who derives the upper bound of the p-value for the hypothesis testing when nuisance parameters (those associated with jumps and additional regimes) are present only under the alternative hypothesis. We test the hypothesis based on the assumption that the likelihood ratio defined by

$$LR = 2(\text{log-likelihood under alternative} - \text{log-likelihood under null}) \tag{C.1}$$

has a single peak over the space of the additional parameters under the alternative hypothesis (see, for example, Garcia and Perron, 1996) when the model under the null hypothesis is nested in the model under the alternative hypothesis. Our model has a larger set of estimated parameters, but the test adjusts for the number of parameters. For the non-nested case (one regime with jumps vs. two regimes without jumps), we obtain an upper bound of the $p-$value from the test statistic of Rivers and Vuong (2002):

$$z_T = \frac{\sqrt{T}}{\hat{\sigma}_T}\left[\frac{1}{T}\sum_{t=1}^{T}\ln\left(\frac{l_A(r_t\ |\mathbb{X}_{t-1}, \hat{\Theta}_A)}{l_N(r_t\ |\mathbb{X}_{t-1}, \hat{\Theta}_N)}\right)\right] \tag{C.2}$$

where l_N and $\hat{\Theta}_N$ (l_A and $\hat{\Theta}_A$) are the one-period likelihood function and estimated parameters under null (alternative) hypothesis, respectively, and $\hat{\sigma}_T$ is the estimate of standard deviation of the difference of the one-period log-likelihoods. We compute $\hat{\sigma}_T$ based on Newey and West (1987) using the Bartlett weights with various lag values, and choose the highest $p-$value as the upper bound.

Appendix D Benckmark Models

This section describes the benchmark models used in Section 5. Throughout this appendix, let $r_{i,t}$ denote the return of country i at time t for $i = 1, \ldots, n, t = 1, \ldots, T$.

D.1 Multivariate GARCH

We assume that each univariate process follows the GJR-GARCH(p, o, q) model of Glosten, Jagannathan, and Runkle (1993):

$$r_{i,t} = \mu_i + \sigma_{i,t}\epsilon_{i,t} \tag{D.1}$$

$$\sigma^2_{i,t} = \omega_i + \sum_{j=1}^{p} \alpha_{i,j} z^2_{i,t-j} + \sum_{j=1}^{o} \gamma_{i,j} \eta^2_{i,t-j} + \sum_{j=1}^{q} \beta_{i,j} \sigma^2_{i,t-j} \tag{D.2}$$

where $z_{i,t} = \sigma_{i,t}\epsilon_{i,t}$, $\eta_{i,t} = z_{i,t}\mathbf{1}(z_{i,t} < 0)$ and $\epsilon_{i,t}$ is normally distributed with mean 0 and variance 1. In the CCC model, the correlation matrix of $\epsilon_t = [\epsilon_{1,t}, \ldots, \epsilon_{n,t}]'$ is a constant matrix S. In the asymmetric DCC(m, l, k) model, the correlation matrix at time t of ϵ_t is S_t, which the implied correlation matrix of Q_t, where Q_t satisfies

$$Q_t = \left(1 - \sum_{j=1}^{m} a_j - \sum_{j=1}^{k} b_j\right) \bar{Q} - n \sum_{j=1}^{l} g_j + \sum_{j=1}^{m} a_j z_{t-j} z'_{t-j} + \sum_{j=1}^{l} g_j \eta_{t-j} \eta'_{t-j} + \sum_{j=1}^{k} b_j Q_{t-j} \tag{D.3}$$

and $z_t = [z_{1,t}, \ldots, z_{n,t}]'$, $\eta_t = [\eta_{1,t}, \ldots, \eta_{n,t}]'$. Note that a_j, b_j, g_j are constants and $\bar{Q}$ is a constant matrix. We fit the models with various choices of $p, o, q, m, l, k \in \{0, 1, 2\}$. Based on the AIC, we select the CCC model with GJR-GARCH(1,1,1), and the asymmetric DCC(2,0,2) model with GJR-GARCH(1,1,1).

D.2 Factor copula

Similar to Oh and Patton (2012), we assume that each univariate process follows the AR(1)-GJR-GARCH(1,1,1):

$$r_{i,t} = a_i + b_i x_{i,t-1} + \sigma_{i,t}\epsilon_{i,t} \tag{D.4}$$

$$\sigma^2_{i,t} = \omega_i + \alpha_i z^2_{i,t-1} + \gamma_i \eta^2_{i,t-1} + \beta_i \sigma^2_{i,t-1} \tag{D.5}$$

where $z_{i,t} = \sigma_{i,t}\epsilon_{i,t}$ and $\eta_{i,t} = z_{i,t}\mathbf{1}(z_{i,t} < 0)$. Note that $a_i, b_i, \omega_i, \alpha_i, \gamma_i$ and β_i are constants. $\epsilon_{i,t}$ is assumed to be independent across t and has marginal distribution $\hat{F}_i$ which is the empirical distribution of estimated $\epsilon_{i,t}$ from the AR(1)-GJR-GARCH(1,1,1). Based on Oh and Patton (2012), the joint CDF of $\epsilon_t = [\epsilon_{1,t}, \ldots, \epsilon_{n,t}]'$ is modeled by a copula $C(\hat{F}_1, \ldots, \hat{F}_n)$ implied from the copula of $Y = [Y_1, \ldots, Y_n]'$ in the following factor model:

$$Y_i = \sum_{j=1}^{k} c_{i,j} f_j + u_i$$

where f_j and u_i are all independent, and $c_{i,j}$ are the parameters representing the factor loadings. We assume that f_1 follows the skewed-t distribution of Hansen (1994) with parameters (ν, λ), and f_2, f_3, f_4 (in the four-factor model) and u_i follows the same Student-t distribution with ν degrees of freedom. For the one-factor model with the same factor loading, we impose the condition $c_{i,1} = c$ for all $i = 1, \ldots, n$. For the four-factor model, we impose no restriction on $c_{i,1}, i = 1, \ldots, n$, but impose $c_{i,2} = c_2$ if country i is in the Asia-Pacific region, and 0 otherwise; $c_{i,3} = c_3$ if country i is in Europe, and 0 otherwise; and $c_{i,4} = c_4$ if country i is in North America, and 0 otherwise. We use the simulated method of moments of Oh and

Patton (2013) to fit the copula models based on the Spearman's rank correlation and quantile dependence at quantiles 0.15, 0.20, 0.80 and 0.85 fitting criteria. Due to high sensitivity to the starting values, we randomly choose the starting values and fit each model for at least 30 times, and select the best model.

D.3 Multivariate factor stochastic volatility

The returns are assumed to follow the multivariate factor stochastic volatility model of Omori and Ishihara (2012) with q factors:

$$r_{i,t} = \sum_{j=1}^{q} b_{i,j} f_{j,t} + \lambda_t^{-1} e^{\alpha_{i,t}/2} \epsilon_{1,i,t} \qquad i = 1, \ldots, n \tag{D.6}$$

$$f_{j,t} = e^{\alpha_{n+j,t}/2} \epsilon_{2,j,t} \qquad j = 1, \ldots, q \tag{D.7}$$

$$\alpha_{k,t+1} = \phi_k \alpha_{k,t} + \eta_{k,t} \qquad k = 1, \ldots, n+q,\ t \geq 1 \tag{D.8}$$

where λ_t is i.i.d. with gamma$(\nu/2, \nu/2)$ distribution, $\alpha_{k,1}$ is normally distributed with mean 0 and variance $\sigma_{k,\epsilon}^2/(1-\phi_k^2)$. Let $\epsilon_{j,t} = [\epsilon_{j,1,t}, \ldots, \epsilon_{j,n,t}]'$ for $j = 1, 2$, $\epsilon_t = [\epsilon_{1,t}', \epsilon_{2,t}']'$ and $\eta_t = [\eta_{1,t}, \ldots, \eta_{n+q,t}]'$. Assume that the vector $[\epsilon_t', \eta_t']'$ is jointly normally distributed with mean 0 and variance-covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{\epsilon\epsilon} & \Sigma_{\epsilon\eta} \\ \Sigma_{\epsilon\eta} & \Sigma_{\eta\eta} \end{bmatrix} \tag{D.9}$$

where

$$\Sigma_{\epsilon\epsilon} = diag(\sigma_{1,\epsilon}^2, \ldots, \sigma_{n+q,\epsilon}^2) \tag{D.10}$$

$$\Sigma_{\eta,\eta} = diag(\sigma_{1,\eta}^2, \ldots, \sigma_{n+q,\eta}^2) \tag{D.11}$$

$$\Sigma_{\epsilon,\eta} = diag(\rho_1 \sigma_{1,\epsilon}\sigma_{1,\eta}, \ldots, \rho_{n+q}\sigma_{n+q,\epsilon}\sigma_{n+q,\eta}) \tag{D.12}$$

and $diag(a_1, \ldots, a_n)$ is the diagonal matrix whose (i,i) entry is a_i. For identification of the factor loadings, assume that

$$b_{i,j} = 0, \quad i < j, \quad i = 1, \ldots, q \tag{D.13}$$

$$b_{i,i} = 1, \quad i = 1, \ldots, q \tag{D.14}$$

Note that $b_{i,j}, \phi_k, \sigma_{i,\epsilon}, \sigma_{i,\eta}$ and ρ_i are constants. We estimate the model parameters for one, three and five factors using the MCMC method described in their paper.

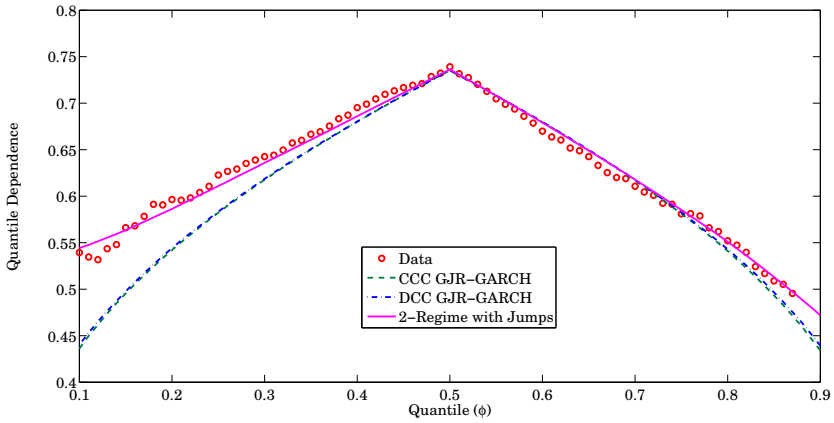
Appendix E Quantile Dependence

Oh and Patton (2013), among others, use quantile dependence to illustrate the dependence between stock returns. We now investigate how well our two-regime model with jumps fits the quantile dependence implied from the data compared to the benchmark models. This provides a robustness check for a different measure of extreme dependence. In particular, let r_1 and r_2 denote the returns of countries 1 and 2 whose marginal cumulative distribution functions are F_1 and F_2. The quantile dependence at quantile $\phi \leq 0.5$ is $P(F_1(r_1) < \phi \mid F_2(r_2) < \phi)$ and at quantile $\phi > 0.5$ is $P(F_1(r_1) > \phi \mid F_2(r_2) > \phi)$. Note that it does not depend on the order of r_1 and r_2. For each model, we simulate 500,000 observations and compute its empirical quantile dependence[2]. Figure E.1 shows the implied quantile dependences of multivariate GARCH models (Panel (a)), factor copula models (Panel (b)), and multivariate factor stochastic volatility models (Panel (c)) compared to those from the actual data, and the two-regime model with jumps. Observe that the implied quantile dependence of our two-regime model with jumps matches that of the data very well for all quantile levels. The quantile dependence from the multivariate GARCH models fits well to the data for the upper tail ($\phi > 0.5$) but underestimates those for the lower tail ($\phi < 0.5$). It is also symmetric due to conditionally normal assumption. The factor copula models, on the other hand, generate asymmetric quantile dependence with higher values for the lower tail. However, they underestimate the quantile dependences for both lower and upper tails, especially for the extreme values of ϕ. The multivariate factor stochastic volatility models provide similar results as the factor copula models except that they underestimate the quantile dependences for all values of ϕ. This emphasizes the limitation of these models on capturing dependency of extreme returns.

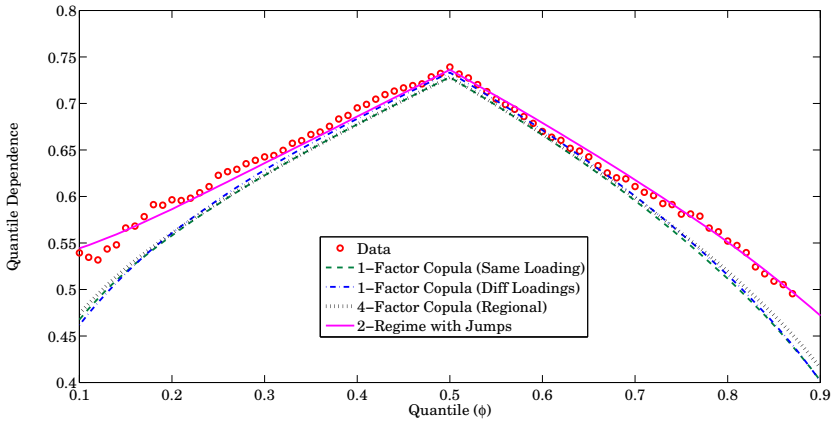
[2]The empirical quantile dependence of observations $(r_{1,t}, r_{2,t}), t = 1, \ldots T$ at quantile ϕ is

$$\hat{q}(\phi) = \begin{cases} \frac{1}{\phi T} \sum_{t=1}^{T} \mathbf{1}(\hat{F}_1(r_{1,t}) \leq \phi \text{ and } \hat{F}_2(r_{2,t}) \leq \phi) & \phi \in (0, 0.5] \\ \frac{1}{(1-\phi)T} \sum_{t=1}^{T} \mathbf{1}(\hat{F}_1(r_{1,t}) > \phi \text{ and } \hat{F}_2(r_{2,t}) > \phi) & \phi \in (0.5, 1) \end{cases}$$

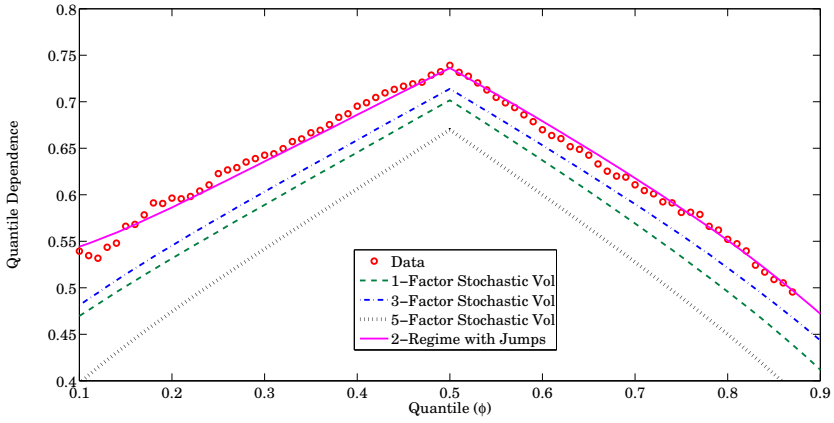
where $\mathbf{1}(A)$ is an indicator function equal to 1 if A is true, and 0 otherwise, and $\hat{F}_i$ is the empirical CDF of return r_i.



(a) Multivariate GARCH models



(b) Factor Copula Models



(c) Multivariate Factor Stochastic Volatility Models

Figure E.1

Quantile dependence of GARCH, factor copula and factor stochastic volatility models.

This figure shows the average quantile dependences from forty-five country pairs implied from data, our two-regime model with jumps and the benchmark models: GARCH (Panel (a)), factor copula (Panel (b)) and factor stochastic volatility (Panel (c)).

Appendix F Optimal Portfolio Weight Derivation and Calculation

This appendix provides the derivation of Theorem 6.1 and how to compute the optimal portfolio weights numerically. First, let $\mathcal{F}_t = \{R_1, \ldots, R_t, W_0, \ldots, W_t\}$ denote the information set at time t. Based on $\mathcal{F}_t$, the investor forms her belief about the likelihood of the market regime $q_t = [q_{1,t}, \ldots, q_{K,t}]'$, and optimally chooses her portfolio weight x_t. Then the returns R_{t+1}, whose distribution depends on the next-period *unobservable* regime Y_{t+1}, are realized, and the investor updates her belief using the Bayes' rule

$$q_{y,t+1}(q_t, R_{t+1}) = \frac{\sum_{j=1}^{K} q_{j,t} p_{j,y} f_y(R_{t+1})}{\sum_{z=1}^{K}\sum_{j=1}^{K} q_{j,t} p_{j,z} f_z(R_{t+1})} \tag{F.1}$$

where $f_y(r)$ is the likelihood function of the return R_{t+1} at r given that the next-period regime Y_{t+1} is y. Observe that $q_{y,t+1}$ depends on probability vector q_t and return vector R_{t+1}.

Using the Markov property, it can be seen that the investor requires only her current wealth W_t and the regime probability vector q_t to make her allocation. Let $V(t, q, w)$ denote the value function at time t when $q_t = q$ and $W_t = w$

$$V(t, q, w) = \max_x \mathbb{E}\left[\frac{W_T^{1-\gamma}}{1-\gamma} \;\middle|\; q_t = q, W_t = w \right]. \tag{F.2}$$

The associated Bellman equation for the optimality condition is given by

$$V(t, q, w) = \max_x \mathbb{E}[V(t+1, q_{t+1}, W_{t+1}) \mid \mathcal{F}_t]. \tag{F.3}$$

It can be shown that the value function is of the form

$$V(t, q, w) = h(t, q)\frac{w^{1-\gamma}}{1-\gamma} \tag{F.4}$$

where $h(T, q) = 1$ for all probability vector q. Substituting (F.4) into (F.3), and using (7) and (F.1), we obtain

$$h(t, q) = (1-\gamma)\max_x \sum_{z=1}^{K}\sum_{y=1}^{K} q_z p_{z,y} \mathbb{E}\left[\frac{\left(e^{r_f} + x'\left(e^{R_{t+1}} - e^{r_f}\mathbf{1}\right)\right)^{1-\gamma}}{1-\gamma} h(t+1, q_{t+1}(q, R_{t+1})) \;\middle|\; Y_{t+1} = y \right] \tag{F.5}$$

where $q = [q_1, \ldots, q_K]'$. From the optimality condition for the Bellman equation (F.3), the maximizer in (F.5) is the optimal portfolio weight. This proves Theorem 6.1. When a one-regime model is assumed, the resulting optimal portfolio weight reduces to a constant vector

$$x^* = \arg\max_x \mathbb{E}\left[\frac{\left(e^{r_f} + x'\left(e^{R_{t+1}} - e^{r_f}\mathbf{1}\right)\right)^{1-\gamma}}{1-\gamma} \right] \tag{F.6}$$

as the hedging demand for stochastic regime disappears.

The optimal portfolio weights can be obtained from solving the optimization problem (8). In order to make computation possible, the return distribution is discretized based on an approximate integral formula provided in Stroud (1971). Specifically, we use an accurate approximation to an integral of the form

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-(r_1^2 + \ldots + r_n^2)} g(r_1, \ldots, r_n) dr_1 \cdots dr_n \approx \sum_{h=1}^{H} \bar{w}_h g(\bar{r}_{1,h}, \ldots, \bar{r}_{n,h}) \tag{F.7}$$

for some function g where the points $\bar{r}_h = [\bar{r}_{1,h}, \ldots, \bar{r}_{n,h}]'$ and their associated weights $\bar{w}_h$, $h = 1, \ldots, H$ are chosen so that the approximation becomes exact for any multinomial g of degree 5 or less[3]. For 10 countries, we need $H = 1,044$ points. This choice of approximation guarantees that all weights $\bar{w}_h$ are positive. Without this property, the expectation in (8) after the discretization may assign negative probabilities to some discrete return realizations, and consequently the numerical algorithm for maximization problem (8) will try to minimize the utility at those realizations causing undesired errors in the resulting x. See also Haber (1970) for the reasons supporting positive weights in integral approximations. Other approximations of the form as in (F.7) fail to ensure this positive-weight property for degrees higher than 5, while approximations of other forms such as a product-rule quadrature will generally require a much larger number of points for the same level of accuracy[4].

To approximate the expectation in (8), we first condition on the number of jumps ΔN_t so that the conditional distribution of R_t given $Y_t = y$ and $\Delta N_t = m$ is normal with mean $b(y) + m\eta(y)$ and covariance matrix $\Sigma(y) + m\Omega(y)$

$$\begin{aligned}
&\mathbb{E}\left[\frac{\left(e^{r_f} + x'\left(e^{R_{t+1}} - e^{r_f}\mathbf{1}\right)\right)^{1-\gamma}}{1-\gamma} h(t+1, q_{t+1}(q, R_{t+1})) \;\middle|\; Y_{t+1} = y\right] \\
&\quad = \sum_{m=0}^{\infty} \frac{e^{-\lambda(y)}\lambda(y)^m}{m!} \mathbb{E}\left[\frac{\left(e^{r_f} + x'\left(e^{R_{t+1}} - e^{r_f}\mathbf{1}\right)\right)^{1-\gamma}}{1-\gamma} h(t+1, q_{t+1}(q, R_{t+1})) \;\middle|\; Y_{t+1} = y, \Delta N_{t+1} = m\right].
\end{aligned} \tag{F.8}$$

Then we write conditional R_{t+1} in terms of n i.i.d. standard normal random variables $z_1, \ldots, z_n$ using the Cholesky decomposition of the covariance matrix $\Sigma(y) + m\Omega(y)$:

$$R_{t+1}^{(y,m,z)} = b(y) + m\eta(y) + Lz \tag{F.9}$$

where $z = [z_1, \ldots, z_n]'$, and L is the lower triangular matrix obtained from the Cholesky decomposition of $\Sigma(y) + m\Omega(y) = LL'$. Applying approximation (F.7) to expectations on the right-hand-side of (F.8), we have

$$\begin{aligned}
&\mathbb{E}\left[\frac{\left(e^{r_f} + x'\left(e^{R_{t+1}} - e^{r_f}\mathbf{1}\right)\right)^{1-\gamma}}{1-\gamma} h(t+1, q_{t+1}(q, R_{t+1})) \;\middle|\; Y_{t+1} = y, \Delta N_{t+1} = m\right] \\
&\quad \approx \frac{1}{\pi^{n/2}} \sum_{h=1}^{H} \bar{w}_h g_{y,m}(\sqrt{2}\bar{r}_{1,h}, \ldots, \sqrt{2}\bar{r}_{n,h}; x)
\end{aligned} \tag{F.10}$$

where

$$g_{y,m}(z_1, \ldots, z_n; x) = \frac{\left(e^{r_f} + x'\left(e^{R_{t+1}^{(y,m,z)}} - e^{r_f}\mathbf{1}\right)\right)^{1-\gamma}}{1-\gamma} h(t+1, q_{t+1}(q, R_{t+1}^{(y,m,z)})). \tag{F.11}$$

[3] A multinomial function of degree d only contains terms of the form $r_1^{d_1} r_2^{d_2} \cdots r_n^{d_n}$ such that all $d_1, \ldots, d_n$ are nonnegative integers, and $d_1 + \ldots + d_n \leq d$ with at least one term having the sum equal to d.

[4] See, for example, Cools (1999) and the update of the list of all available approximations of the same type as in (F.7) on the author's website. A product-rule such as the Gaussian-Hermite quadrature requires $H = \left(\frac{d+1}{2}\right)^n$ points for exact approximation of multinomial of degree d, or 59,049 points for $n = 10$ countries with $d = 5$.

References

Baum, L. E., T. P. Petrie, G. Soules, and N. Weiss. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics* 41:164–171.

Cappe, O., E. Moulines, and T. Ryden. 2005. *Inference in Hidden Markov Models*. New York: Springer.

Cools, R. 1999. Monomial cubature rules since "Stroud": a compilation-part 2. *Journal of Computational and Applied Mathematics* 112:21–27.

Davies, R. 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74:33–43.

Garcia, R., and P. Perron. 1996. An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics* 78:111–125.

Glosten, L. R., R. Jagannathan, and D. E. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48:1779–1801.

Haber, S. 1970. Numerical evaluation of multiple integrals. *SIAM Review* 21:481–526.

Hansen, B. E. 1994. Autoregressive conditional density estimation. *International Economic Review* 35:705–730.

Lystig, T. C., and J. P. Hughes. 2002. Exact computation of the observed information matrix for hidden Markov models. *Journal of Computational and Graphical Statistics* 11:678–689.

Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708.

Oh, D. H., and A. J. Patton. 2012. Modelling dependence in high dimensions with factor copulas. Working paper, Duke University.

Oh, D. H., and A. J. Patton. 2013. Simulated method of moments estimation for copula-based multivariate models. *Journal of the American Statistical Association* 108:689–700.

Omori, Y., and T. Ishihara. 2012. Multivariate stochastic volatility models. In L. Bauwens, C. M. Hafner, and S. Laurent (eds.), *Handbook of Volatility Models and Their Applications*, pp. 221–248. John Wiley and Sons.

Rivers, D., and Q. Vuong. 2002. Model selection tests for nonlinear dynamic models. *Econometrics Journal* 5:1–39.

Rothenberg, T. J. 1971. Identification in parametric models. *Econometrica* 39:577–591.

Stroud, A. H. 1971. *Approximate Calculation of Multiple Integrals*. Englewood Cliffs, N.J.: Prentice-Hall.

Teicher, H. 1963. Identifiability of finite mixtures. *Annals of Mathematical Statistics* 34:1265–1269.

Teicher, H. 1967. Identifiability of mixtures of product measures. *Annals of Mathematical Statistics* 38:13001302.

Yakowitz, J. S., and J. D. Spragins. 1968. On the identification of finite mixtures. *Annals of Mathematical Statistics* 39:209–214.