Stochastic Volatility Models using Seemingly Unrelated Regressions

Selina Chen: 20515953, Patrick Li: 20557430

April 21, 2018

Abstract

Continuous time diffusion processes such as the stochastic volatility model are proven to be highly effective tools for forecasting asset prices in large systems. Unfortunately, these models are incredibly computationally intensive to fit for long time horizons. Modelling asset dependence is critical for portfolio managers. Mild increases in asset correlation can greatly increase the potential losses at the tails of a distribution, implying that modelling asset covariance is imperative for creating stock price forecasts. Unfortunately, many popular models such as GARCH are only adequate in stable market conditions. Excessive reliance on independence in these models create overly optimistic forecasts of tail events. To remedy the situation, we fit the stochastic volatility model using seemingly unrelated regression (SUR) models, which can model the correlation of multiple equations in a system without excessive amounts of computational time. In addition, we will use Direct Monte Carlo (DMC) and Markov Chain Monte Carlo (MCMC) sampling schemes to complement our inferences from the SUR model.

1 Introduction

Modelling asset returns, one of the most popular applications of statistics, is the foundation of most financial institutions for portfolio selection and risk assessment. Due to its popularity, a plethora of forecasting models are available. Although no model is provably correct or optimal in the modelling of the dynamic stock market, many models such as GARCH fall short in inferring the dependence structure of assets [2]. As exhibited in historical bull and bear markets, various groups of financial assets are highly correlated [3] ; excessive reliance on heteroskedacity in financial models greatly underestimates the potential losses that can occur at the tails of a distribution. It is thus imperative that investors try to understand the dependence relationships between key asset groups when selecting a portfolio.

The use of Brownian motion to model movement of financial asset prices was first presented by Black, Merton, and Scholes [4][5][6]. Using stochastic models, unprecedented success has been achieved in financial forecasting in the past few decades [7]. Using the stochastic

volatility model, we hope to accomplish two goals: 1) Create a computationally efficient forecasting model to estimate price trajectories under different market conditions over a 14 year time horizon on a portfolio of 46 stocks from the S&P 500. 2) Infer the dependencies between asset returns of the same portfolio under varying market conditions. Unfortunately, calculating the covariance of a system of SDE is difficult unless we create independence assumptions (e.g. independent Gaussian white noise between systems)[8]. Furthermore, the fitting of a stochastic differential equation for long time horizons is computationally demanding if we try to integrate out the volatility terms $\sigma(X_t)$, so we look to modify the standard SDE model to better fit our needs.

Leveraging the capabilities of Seemingly Unrelated Regression (SUR) models, we can concurrently analyze multiple systems of equations to reveal interactions between different assets over time. Furthermore, SUR allows us to create non-diagonal covariance matrices to analyze a dynamic system [9], and is far more computationally tractable than fitting an SDE. Although SUR models are inherently powerful tools for estimating covariance structure, they do not capture model uncertainty as well as Bayesian approaches. Building upon Zellner's ideas for Bayesian analysis on SUR [10] [11], we outline Markov Chain Monte Carlo (MCMC) Gibbs sampling, and Direct Monte Carlo (DMC) approaches to analyze SUR equations.

Outline of Paper: In section 2, we introduce the stochastic volatility model and SUR in more depth, and explicitly outline how we combined these two models. Section 3 continues to describe the mathematics behind models we used by discussing Bayesian approaches (i.e. DMC and MCMC models) for system of SUR equations. Section 4 then addresses the differences in forecasting capabilities of the aforementioned models, and also analyzes asset dependencies under varying time horizons to see if dependencies change under differing market conditions. We conclude the paper in section 5 by discussing potential improvements and areas to explore with our SUR and stochastic volatility models. See the appendix for computational challenges, implementation tests, and auxiliary mathematical formulas.

2 Methodology

2.1 Stochastic Volatility

Let $X_t^{(i)} > 0$ represent the value of asset i on day t, and $Y_t^{(i)} \in$ denote its return. The return metric we will be using is the log returns, i.e. $y_t^{(i)} = log(X_t^{(i)}) - log(X_{t-1}^{(i)})$.

Note that $X_t^{(i)}$ is a continuous-time stochastic process such that:

1. Almost every sample path is continuous
2. $X_t^{(i)}$ has the homogeneous Markov property

Then, $X_t^{(i)}$ is a diffusion process which can be represented by a stochastic differential equation

(SDE):

$$dX_t^{(i)} = \mu(X_t^{(i))}dt + \sigma(X_t)dZ_t$$

Applying Euler discretization, we can approximate SDE on $[0, T]$ at intervals of $\Delta t = T/N$:

$$Y_t^{(i)} = (\alpha_i - \frac{1}{2}\tau_i V_t)\Delta t + (\tau_i V_t)^{1/2} Z_t^{(i)} \tag{1}$$

$$Y_t^{(i)} \overset{iid}{\sim} N((\alpha_i - \frac{1}{2}\tau_i V_t)\Delta t, \tau_i V_t \Delta t)$$

Where V_t is the volatility index (VIX) at time t, and $Z_t^{(i)}$ is the Brownian motion.

Since $\Delta t = 1$, we can rewrite equation (1) as:

$$\frac{Y_t^{(i)}}{V_t^{1/2}} = \frac{\alpha_i}{(\tau_i V_t)^{1/2}} + (\tau_i)^{1/2} V_t^{1/2} + E_t^{(i)} \tag{2}$$

We also want to understand daily changes in VIX. Let A_t represent $log(V_t)$, then:

$$\Delta A_t = \eta + \gamma V_t + E_t^{(0)} \tag{3}$$

$$E_t \overset{iid}{\sim} (0, \Sigma)$$

2.2 Seemingly Unrelated Regression

We are interested in the following parameters: $\alpha, \tau, \eta, \gamma, \Sigma$. To obtain these parameters, we use seemingly unrelated regression (SUR) by defining the following:

$$\begin{aligned} y_t^{(0)} &= \Delta A_t, \quad X_t^{(0)} = (1, V_t), \quad \beta^{(0)} = (\eta, \gamma) \\ y_t^{(i)} &= \Delta X_t, \quad X_t^{(i)} = (V_t^{(-1/2)}, V_t^{(1/2)}), \quad \beta^{(i)} = (\frac{\alpha_i}{\tau_i^{1/2}}, \tau_i^{1/2}) \end{aligned}$$

Under the SUR framework, we have m

$$\begin{aligned} \mathbf{y}_1 &= X_1 \boldsymbol{\beta_1} + \mathbf{u}_1 \\ \mathbf{y}_2 &= X_2 \boldsymbol{\beta_2} + \mathbf{u}_2 \\ &... \\ \mathbf{y}_m &= X_m \boldsymbol{\beta_m} + \mathbf{u}_m \end{aligned}$$

where $\mathbf{u}_t \sim (0, \sigma)$, and $E[\mathbf{u}_t \mathbf{u}_t' | X] = \Sigma$.

2.3 SUR Estimation using Feasible Generalized Least Squares

Since the error terms are correlated, a generalized least squares (GLS) estimation would be an efficient estimator for the β's. This GLS estimator is also known as the SUR estimator [17]. However, running GLS requires the true covariance matrix of the residuals, which we do not know. A typical solution is to use feasible generalized least squares (FGLS) estimation. The FGLS estimator takes in an estimated covariance matrix based on ordinary least squares (OLS). Notice that each equation is a linear regression on its own. Thus, the coefficients of each equation can be estimated consistently using OLS. After obtaining the OLS estimations, we use the error terms to construct the estimated covariance matrix.

The FGLS procedure is as follows:

1. Estimate each equation using OLS individually, and obtain error terms, $u_i = [u_{i1}, ..., u_{iT}]$, for $i = 1, ..., N$.
2. Calculate the estimated covariance matrix for the error terms, $\hat{\Sigma}$, where

$$\hat{\Sigma}_{ij} = \frac{1}{T} \sum_{t=1}^{T} u_{it} u_{jt}$$

3. Run GLS regression using the estimated covariance matrix $\hat{\Omega} = \hat{\Sigma} \otimes I_T$:

$$\hat{\beta} = \left(X' \hat{\Omega}^{-1} X\right)^{-1} X' \hat{\Omega}^{-1} Y$$

This estimator links all equations together using the error terms. There is an existing rpackage *systemfit* that performs SUR fitting. However, we found the execution time to be long due to the large number of observations and stocks (i.e. 3407 observations and 46 stocks including VIX). Since SUR estimation involves many inefficient matrix operations - e.g. $(X' \hat{\Omega}^{-1} X)^{-1}$, which takes the inverse of a 3407 by 3407 matrix - we decided to re-implement the code using more efficient operations (see Appendix for details).

3 Bayesian inference for SUR

We introduce two distinct methods for drawing a finite set of β and Σ samples.

3.1 Gibbs Sampler on SUR

The Bayesian approach for inference on SUR was first introduced in Zellner's 1971 book [10]. In matrix form, our parameters can be denoted as:

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}, \mathbf{u} \sim \mathcal{N}(0, \Omega \otimes I)$$

where Ω is a $m \times m$ covariance matrix, and $\mathbf{y}, X, \boldsymbol{\beta}, \mathbf{u}$ denote the concatenation of all m $\mathbf{y_i}, X_i, \boldsymbol{\beta_i}, \mathbf{u_i}$ matrices, $i \in 1, .., m$.

The conditional likelihood function under this model is:

$$\begin{aligned}\mathcal{L}(\mathbf{y_i}, X_i\boldsymbol{\beta}, \Omega) &= \frac{1}{(2\pi)^{nm/2}}|\Omega|^{n/2}exp\Big[-\frac{1}{2}(\mathbf{y} - X\boldsymbol{\beta})'(\Omega^{-1})(\mathbf{y} - X\boldsymbol{\beta})\Big] \\ &= \frac{1}{(2\pi)^{nm/2}}|\Omega|^{n/2}exp\Big[-\frac{1}{2}tr[R\Omega^{-1}]\Big]\end{aligned}$$

where $|\Omega|$ denotes the determinant of Ω, and $r_{ij} = (\mathbf{y_i} - X_i\boldsymbol{\beta_i})'(\mathbf{y_i} - X_i\boldsymbol{\beta_i})$.For this paper, we assume no prior knowledge, and use non-informative priors. As commonly used in practice as a non-informative prior, we use Jeffrey's invariant prior [16] on $\boldsymbol{\beta}, \Omega$.

$$\pi_j(\boldsymbol{\beta}, \Omega) = \pi_j(\boldsymbol{\beta})\pi_j(\Omega) \propto |\Omega|^{-\frac{m+1}{2}}$$

From this, the joint posterior density becomes

$$p(\boldsymbol{\beta}, \Omega|D) \propto |\Omega|^{-(n+m+1)/2}exp\Big[-\frac{1}{2}tr[R\Omega^{-1}]\Big]$$

, and the conditional posteriors are:

$$\begin{aligned}p(\boldsymbol{\beta}|\Omega, D) &= \mathcal{N}(\hat{\boldsymbol{\beta}}, \hat{\Omega}) \\ p(\Omega|\boldsymbol{\beta}, D) &= \mathbf{InvWishart}(R, n)\end{aligned}$$

where $\hat{\boldsymbol{\beta}} = X'(\Omega \otimes I)X^{-1}X'(\Omega^{-1} \otimes I)\mathbf{y}$ and $\hat{\Omega} = (X'(\Omega^{-1} \otimes I)X)^{-1}$

Although the Gibbs sampling approach is most popular tool for obtaining a sample of SUR parameters [12], [13], it has many disadvantages. Most notably:

1. β and Σ draws are dependent on one another, resulting in highly autocorrelated outputs.
2. Burn-in period length, and there is no universal rule for the number of MCMC samples to draw largely due to the difficulty in checking for convergence.
3. Inability to check for convergence does not provide adequate testing methods to ensure that draws are from the desired posterior distribution.

3.2 Direct Monte Carlo Simulation on SUR

To remedy the issue, Zellner and Ando (2008) proposed a new algorithm for Bayesian inference. A reformulated SUR equation is formed in order to create an alternative DMC procedure as follows:

$$
\begin{aligned}
\mathbf{y}_1 &= X_1\boldsymbol{\beta_1} + \mathbf{e}_1 \\
\mathbf{y}_2 &= X_2\boldsymbol{\beta_2} + \rho_{21}\mathbf{u}_1 + \mathbf{e}_2 \\
&... \\
\mathbf{y}_m &= X_m\boldsymbol{\beta_m} + \sum_{j=1}^{m-1}\rho_{mj}\mathbf{u}_j + \mathbf{e}_m
\end{aligned}
$$

with $\Sigma = diag(\sigma_1^2, ..., \sigma_m^2)$ and

$$
E[\mathbf{e}_i\mathbf{e}_\mathbf{j}'] = \begin{cases} 0 & (\text{i} \neq \text{j}) \\ \\ \sigma_i\mathbf{I} & (\text{i} = \text{j}) \end{cases}
$$

The likelihood of our model is equivalent to

$$
L(D|\mathbf{b}, \Sigma) = \prod_{j=1}^{m} \frac{1}{(2\pi\sigma_j^2)^{\frac{n}{2}}} exp\Big[- \frac{(\mathbf{y}_j - Z_j\mathbf{b}_j)'(\mathbf{y}_j - Z_j\mathbf{b}_j)}{2\sigma_j^2}\Big]
$$

Priors and posterior densities of parameters can calculated using Jeffrey's (1946) diffuse prior. We obtain:

$$
\pi(\mathbf{b}, \Sigma) \propto \prod_{j=1}^{m}(\sigma_j)^{-1}
$$

Jeffrey's invariant prior implies that the joint posterior parameter densities of $\{\mathbf{b}, \Sigma\}$ is:

$$
g(\mathbf{b}, \Sigma|D) \propto \prod_{j=1}^{m}(\sigma_j)^{-(n+1)} exp\Big[- \frac{(\mathbf{y}_j - Z_j\mathbf{b}_j)'(\mathbf{y}_j - Z_j\mathbf{b}_j)}{2\sigma_j^2}\Big]
$$

which equivalent to the following conditional normal and inverse-gamma posteriors:

$$
\begin{aligned}
g(\mathbf{b}_j|\mathbf{b}_{j-1}, ..., \mathbf{b}_1, \sigma_j^2, D) &= N\big(\hat{\mathbf{b}}, \sigma_j^2(Z_j'Z_j)^{-1}\big) \\
g(\sigma_j^2|\mathbf{b}_{j-1}, ..., \mathbf{b}_1, D) &= IG\big(\frac{1}{2}(\mathbf{y}_j - Z_j\mathbf{b}_j)'(\mathbf{y}_j - Z_j\mathbf{b}_j), \frac{1}{2}(n - p_j - j + 1)\big)
\end{aligned}
$$

Although the DMC approach resolves the issues of the Gibbs sampler, the Cholesky style setup presents potential problems. By parameterizing the equations in a non-exchangeable order, we force the sampler to draw each sample in a distinct order. Without knowing the nature of asset dependence, we cannot ascertain how to choose which stock to be $\mathbf{y}_1$, and which stock to be $\mathbf{y}_m$. By creating the assumption, we add an unjustifiable bias to the model. We cannot quantify the impact of the Cholesky style equation setup in this paper, but we do test the assertions of improved auto-correlation in the following section.

Algorithm 1 One Step of Direct Monte Carlo Sampling

1. Set j = 1
2. Generate $\sigma_1^{2(k)}$, k =1,..N and calculate $\pi(\mathbf{b}_1|\sigma_1^{2(k)}, D)$ and draw $\mathbf{b}$ terms from the distribution.
3. Increment j = j + 1. Draw $\sigma_j^{(k)}$ from $g(\sigma_j^2|\mathbf{b}_{j-1}^{(k)}, .., \mathbf{b}_1^{(k)}, D)$, and then generate $\mathbf{b}_j^{(k)}$ from $g(\mathbf{b}_j|\mathbf{b}_{j-1}^{(k)}, ..., \mathbf{b}_1^{(k)}, \sigma_j^{(k)}, D)$
4. Repeat step 2 sequentially until j = m

4 Results

4.1 Forecasting

Using the estimated coefficients from SUR, we obtain our fitted SDE model:

$$\Delta logV_t = \hat{\beta_1^{(0)}} + \hat{\beta_2^{(0)}} V_t + u_t^{(0)} \tag{4}$$

$$\Delta logX_t^{(i)} = V_t^{1/2} \times \left[\hat{\beta_1^{(i)}} V_t^{-1/2} + \hat{\beta_2^{(i)}} V_t^{1/2} + u_t^{(i)}\right] \tag{5}$$

To perform one simulation of the forecast, we start by sampling the error terms, u_t, from $N(X\hat{\beta}, \Sigma)$. Next, we calculate $V_{t+1} = V_t \times exp(\Delta logV_t)$, then plug V_{t+1} into (5) and back out S_{t+1}. Note for multi-period forecasting, we repeatedly do this operation until we arrive at the specified time.

We are interested in comparing our models' ability to forecast. We run 100 simulations using fitted coefficients and covariance matrix under the FGLS and the MCMC algorithm.

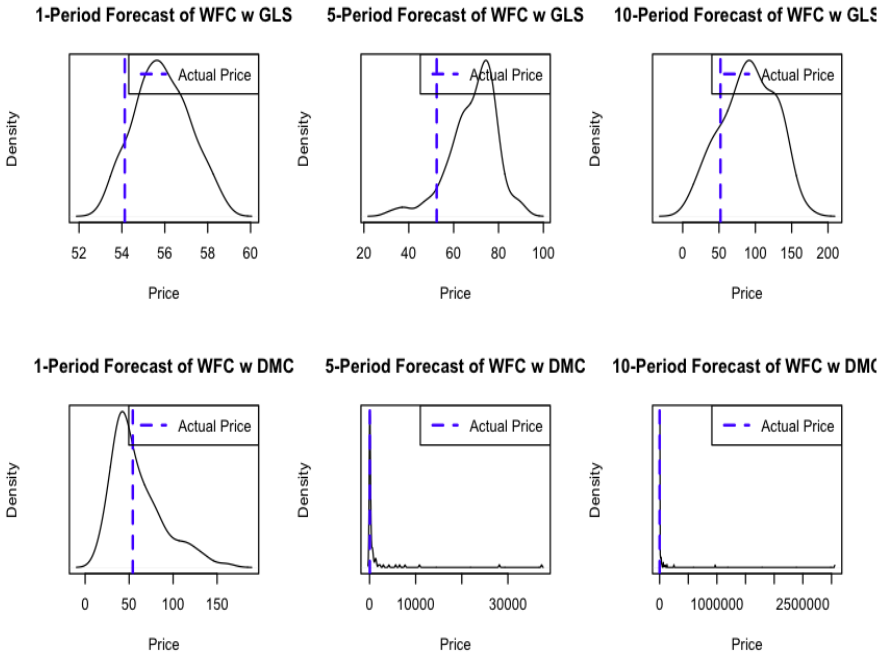


Figure 1: Forecasting results using 100 simulations.

We can see that forecast simulations drawn under the DMC sampler are significantly wider than intervals using the GLS estimates, accounting for the higher amount of uncertainty from the Bayesian parameterization of the problem. Forecasting intervals created from MCMC samples using *bayesm* were highly similar to intervals created by the DMC. All models were able perform 1-day forecast on stock prices reasonably well. That is, the true price falls within reasonable range of the simulated prices. However, the models failed for longer time horizons (e.g. 10-period forecast), where the prediction intervals become significantly wider, and predictions become less meaningful.

4.2 Clustering of Financial Sector Asset Correlations

Due to the change in consumer confidence in financial institutions between the bull markets of 2007 to the 2008 recession, we thought that financial institution correlations would change drastically over different periods. Using the covariance matrices fit from our GLS procedures and the *pvclust* package, we performed bootstrapped agglomerative clustering of financial assets based on their correlations. The following graphs are generated using 1000 bootstrap samples using the distance metric on the j^{th} and kth assets:

$$1 - \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x_j})(x_{ik} - \bar{x_k})}{\sqrt{\sum_{i=1}^{n}(x_{ij} - \bar{x_j})^2}\sqrt{\sum_{i=1}^{n}(x_{ik} - \bar{x_k})^2}}$$

where $\bar{x_k} = \frac{1}{n}\sum_{i=1}^{n} x_{ik}$, $\bar{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$.

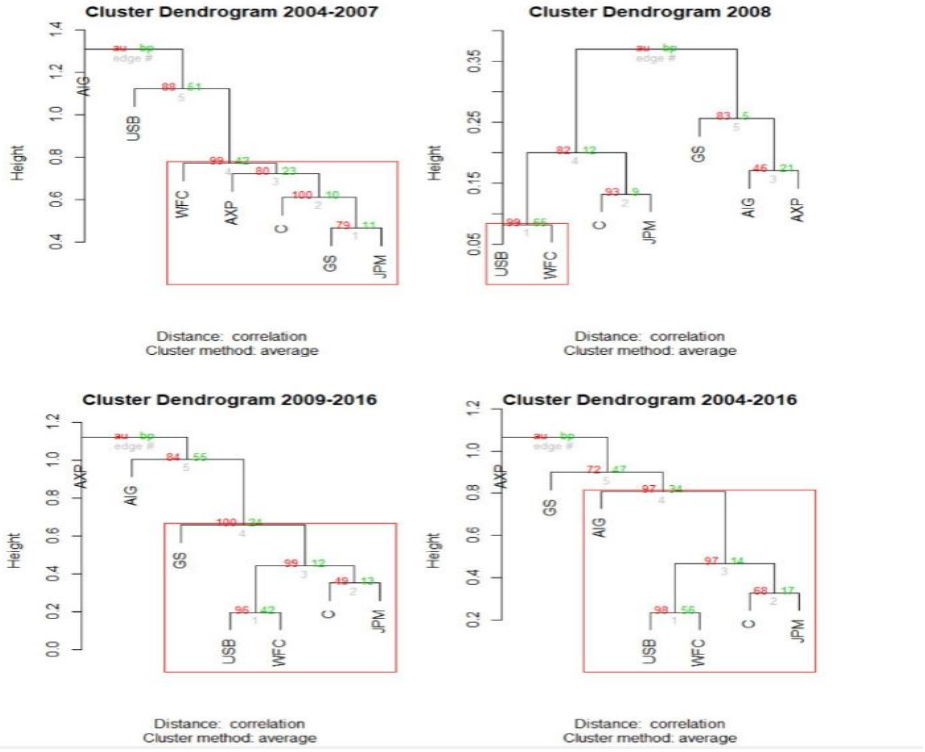
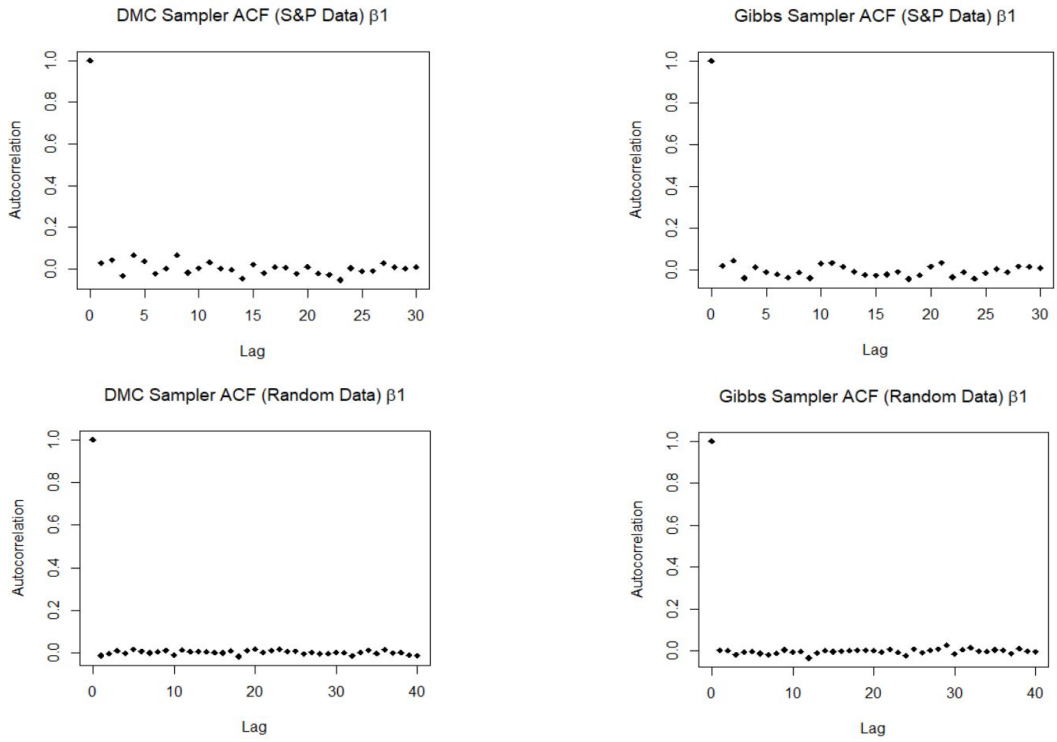


Figure 2: Cluster dendrograms. Boxes are drawn around clusters with p-values less than 0.05.

Red and green numbers in figure 1 denote the level of significance between edges between two clusters. These numbers are equal to $1 - pvalue$. Red numbers are calculated by AU (Approximately Unbiased) multiscale bootstrapping, and green numbers are computed using regular bootstrapping. Edges in the dendrogram contained in red boxes imply a statistically significant relationship. Heights above 1 indicate a negative correlation, while heights below 1 indicate a positive correlation, with distance away from 1.0 indicating stronger correlations. As we can see from these plots, differing market conditions cause the correlations between companies to evolve. Thus, investors must be wary of how they interpret fitted covariance matrices from forecasting models, as asset dependencies continually change based on market conditions. In particular, these plots reveal a high amount of log return correlation in the financial collapse of 2008, but slightly varying clusters during other fitted periods which include greater amounts of market optimism. Cluster dendrograms on other groups of assets exhibited similar behaviour; under varying market conditions, the estimates of covariance for almost all sectors, and assets between different sectors changed.

4.3 Autocorrelation Comparison between DMC and MCMC Samplers

To autocorrelation assertions presented in Zellner's DMC paper, we plot the ACF of the DMC and MCMC samples on both randomly generated data, and the S&P data.



We notice that under both cases of random data, and S&P data, there is a negligible dif-

ference in the amount of autocorrelation between samplers. Thus, we are unable to confirm the assertions from Zellner's DMC paper.

5 Discussion

Leveraging the predictive power of the stochastic volatility model, and combining it with the ability to analyze dependent systems using SUR, we have implemented an R package with Frequentist and Bayesian approaches to forecast asset returns. The methods include the use of GLS to fit SUR, and a DMC sampler to draw a set of samples on SUR parameters. Under stable financial conditions and short time horizons it appears that our models are able to accurately forecast the returns of financial portfolios, but forecast intervals begin to fall apart under volatile markets (e.g. 2008), and when we increase the time horizon. Naturally, DMC incorporates greater uncertainty into its forecasts, giving a more conservative estimates of potential losses. Although these models are satisfactory under mild market conditions, investors should be wary of over-reliance on the covariance matrices outputted by forecasting models. As seen by figure 1, asset dependence frequently change, especially when market conditions move.

There are many potential improvements to different components of our model. Firstly, financial markets are highly complex systems, and linear stochastic differential equations may not be sufficiently flexible to model the behaviours. The use of non-linear SUR could improve the fit of parameters [14]. Secondly, the independent increments for Brownian motion terms in the stochatic volatility model indicates that the present price of a stock does not affect future prices. This is often not the case, as bull market behaviour and bear market behaviour, as shown by the change in cluster relationships in figure 1, meaning that these models need to be adjusted. One remedy is fractional Brownian motion model with adaptive parameters [15]. Lastly, there are still improvements to be made onto the DMC and MCMC models; the DMC constrains the order of asset draws, leading to additional model bias, while convergence of the MCMC model is not guaranteed. Further reparameterization and improvements to these approaches is necessary to produce better, more representative samples of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.

References

[1] Greene, W. (2003): Econometric Analysis, 5th ed. New Jersey, Prentice Hall.

[2] Francq C, Horvath L, Zakoian JM. 2009. Merits and drawbacks of variance targeting in GARCH models. MPRA Paper no.15143, Munich, Germany.

[3] Kenett, D.Y., Huang, X., Vodenska, I., Havlin, S. and Stanley, H.E., Partial correlation analysis: Applications for financial markets. Quant. Finance, 2015

[4] F. Black and M. Scholes, "The pricing of options and corporate liabilities," Journal of Political Economy, vol. 81, no. 3, pp. 637–654, 1973.

[5] F. Black and M. Scholes, "Taxes and the pricing of options," Journal of Finance, vol. 31, no. 2, pp. 319–332, 1976.

[6] R. C. Merton, "Theory of rational option pricing," The Rand Journal of Economics, vol. 4, pp. 141–183, 1973.

[7] GONZÁLES A., NOVALES, A. and RUBIO, G., (2011). "Estimation of Stochastic Volatility Models with Jumps in Returns for Stock Market Indices", August 2011, Working Paper, p.1-28.

[8] Fan, H., Shan, X., Yuan, J., and Ren, Y. (2011). Covariances of linear stochastic differential equations for analyzing computer networks. Tsinghua Science and Technology, 16(3), 264–271.

[9] Zellner, A., 1962. An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. Journal of the American Statistical Association 57, 348–368.

[10] Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics.Wiley, New York.

[11] Zellner A, Ando T. A direct Monte Carlo approach for Bayesian analysis of the seemingly unrelated regression model. J Econom. 2010;159(1):33–45.

[12] Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernado, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 4. Clarendon Press, Oxford, pp. 169–193.

[13] Raftery, A.E., Lewis, S.M., 1992. One long run with diagnostics: implementation strategies for Markov chain–Monte Carlo. Statistical Science 7, 493–497

[14] Moon , H. R. , Perron , B. (2006). Seemingly Unrelated Regressions. Mimeo

[15] Areerak, T. (2014). Mathematical model of stock prices via a fractional brownian motion model with adaptive parameters. ISRN Applied Mathematics, 2014, 791418.

[16] Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems.Proceedings of the Royal Society of London, Series A 196, 453–461.

[17] Henningsen A, Hamann JD (2007a). "systemfit: A Package for Estimating Systems of Simultaneous Equations in R." Journal of Statistical Software, 23(4), 1–40.

6 Appendix

6.1 Computational Challenges

Each of our approaches to analyzing SUR models had their own share of computational challenges.

6.1.1 GLS

The GLS estimation of $\hat{\beta} = \left(X'\hat{\Omega}^{-1}X\right)^{-1}X'\hat{\Omega}^{-1}Y$, where $\hat{\Omega} = \hat{\Sigma} \otimes I_T$ requires large matrix multiplications. To resolve this, we create two helper functions.

1. **Calculating** $(X'\hat{\Omega}^{-1}X)$**.** Instead of performing a direct matrix multiplication, we transform it into $\sum_{t=1}^{T} X_t \Sigma^{-1} X_t'$. This reduces the size of the X matrix to $npxn$, where n is the number of equations, and p is the number of regressors per equation. In addition, we noticed that this operation returns a symmetric matrix. Take $p = 2$ as an example:

$$\begin{bmatrix} X_{11} & 0 \\ 0 & X_{21} \end{bmatrix} \begin{bmatrix} \Sigma_{11}^{-1} & \Sigma_{12}^{-1} \\ \Sigma_{21}^{-1} & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} X_{11}^T & 0 \\ 0 & X_{21}^T \end{bmatrix} = \begin{bmatrix} \Sigma_{11}^{-1} X_{11} X_{11}^T & \Sigma_{21}^{-1} X_{11} X_{21}^T \\ \Sigma_{12}^{-1} X_{21} X_{11}^T & \Sigma_{22}^{-1} X_{22} X_{22}^T \end{bmatrix}$$

Thus, we are able to calculate only the diagonal and the upper triangular entries, saving up to half the computational time. After looping through all periods, we get a upper triangular matrix. We add its transpose to itself and delete the diagonals to get the final matrix.

2. **Calculating** $X'\hat{\Omega}^{-1}Y$**.** Similar to the previous operation, we break the matrix multiplication into time periods, i.e. $\sum_{t=1}^{T} X_t \Sigma^{-1} Y_t$. Working out the math again, we get:

$$\begin{bmatrix} X_{11} & 0 \\ 0 & X_{21} \end{bmatrix} \begin{bmatrix} \Sigma_{11}^{-1} & \Sigma_{12}^{-1} \\ \Sigma_{21}^{-1} & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} Y_{11} \\ Y_{21} \end{bmatrix} = \begin{bmatrix} X_{11}(\Sigma_{11}^{-1} Y_{11} + \Sigma_{12}^{-1} Y_{21}) \\ X_{21}(\Sigma_{21}^{-1} Y_{11} + \Sigma_{22}^{-1} Y_{21}) \end{bmatrix}$$

6.1.2 Gibbs Sampler

The SUR Gibbs sampler is quite computationally expensive if we use the naive implementation due to the large dimension of the $\Omega \otimes I$ term, which grows to a matrix of size $156,000 \times 156,000$ with the full 14 year time series on 46 assets. Attempting to reduce the memory burden results in formulas that require multiple nested for loops. R is inherently slow with for loops, but the *RsurGibbs* function from the *bayesm* package uses *Rcpp* to speed up computation. Due to the time constraints of the project, we were unable to implement the Rcpp version of the sampler, so we left our naive implementation of the SUR Gibbs sampler out of our package, and used it as a tool to test the samples of *bayesm*'s much more efficient *RsurGibbs* function instead.

The naive implementation of the Gibbs sampler was necessary for testing certain components of the *bayesm* sampler. In particular, when looking at the conditional update formulas for Ω and β, each argument is dependent on the other. Furthermore, as outlined in Zellner's DMC paper, these dependencies make testing the posterior distribution difficult. We added in 'update_omega' and 'update_beta' flags to test the distribution of draws instead.

6.1.3 DMC Sampler

The DMC is also for loop intensive, but not as much as the SUR Gibbs sampler. Thus, although we were not able to leverage libraries such as RcppArmadillo or Rcpp to speed up matrix algebra, the penalty is less severe than the compuational constraints on the R implementation of the SUR Gibbs sampler.