A Bayesian factor model with dynamic shrinkage for time-varying correlation matrices with an application to Financial Crises

Daniel Andrew Coulson[1] and David S. Matteson[2]Martin T. Wells[3]

[1,2,3]Department of Statistics and Data Science, Cornell University, Ithaca, NY

Abstract

In this paper we propose a novel approach to quantifying the risk from a co-movement of stocks in a portfolio through time, derived from time varying model-based correlation matrices. The correlation matrices are estimated in a Bayesian fashion utilizing a dynamic shrinkage prior process for the state variables to be estimated and a multivariate factor stochastic volatility process for the observation error covariance matrices. To summarize the information in correlation matrices we create an intuitive and simple scalar score. We prove a posterior concentration result to validate our modeling approach theoretically. Through a simulation study we demonstrate our estimation approach achieves superior performance in terms of several metrics and has an ability to rapidly adapt to changing market conditions compared to competing methods. Through real world examples we demonstrate the new insights provided by our proposed framework in identifying known periods of financial instability and the additional information it provides beyond existing measures such as the VIX index. We subsequently compare the static minimum variance portfolio to that of a dynamically changing minimum variance portfolio in times of financial crisis. Our model provides a new overall

measure of correlation in a system which can be utilized by both practitioners and researchers as a tool to quantify the correlation risks in a portfolio and the degree of toxicity in the financial system.
Key words: Bayesian methods, Financial risk management,Portfolio allocation, Time Series, Posterior Concentration

1 Introduction

Crises such as the U.S. subprime mortgage crisis or the COVID-19 pandemic can have a large detrimental impact on an investor's portfolio especially through increased market volatility, as discussed in Chen et al. [2022] and Foo and Witkowska [2024]. Thus, understanding the impact of these market instabilities is essential for investors to have a fuller insight into the risks of their portfolio.

During a financial crisis correlations both within, and across markets increase. Lin et al. [1994], Solnik et al. [1996], and Junior and Franca [2012] find that the correlations between international stock markets increase during periods of high volatility. Silvennoinen and Teräsvirta [2005] observe that correlations among S&P 500 stocks are greater during periods of more volatility.

Therefore, understanding the correlations within a portfolio is important in order to mitigate the impact of time periods of greater market volatility. The benefits of diversification have been exalted in the literature, such as Wagner and Lau [1971] and Lumby and Jones [1998]. In this paper we propose a methodology for quantifying the time varying correlation in a stock portfolio and demonstrate that diversification does not help in times of financial crisis.

The standard method for computing time varying correlation matrices of a portfolio rely on rolling correlation estimators. Rolling correlation estimators suffer from several problems. Lag effects induced by using a rolling window, which contains several data points unrelated to the current data point, can lead to previous data points having an undue impact on estimated correlation matrices. These lag effects further induce a pronounced inability of the estimator to respond to changes in market conditions in a sufficiently responsive manner. Furthermore, rolling estimators are inherently high rank estimators due to over parameterization, which naturally leads to estimators with a larger variance than is usually desired. Rolling estimators also have an inability to estimate instantaneous correlation matrices which is undesirable, especially when alternative methods exist that can do this. Another popular approach to modeling time varying correlation matrices is the Dynamic conditional correlation model (DCC). To fit a DCC model you first fit a univariate GARCH model to each of the time series to obtain the scaled residuals. From the residuals then estimate the correlation matrices. However, due to the two stage approach, if there is model misspecification in the first stage this will then impact the correlation estimates obtained in the second stage. Furthermore, none of these methods can provide us with uncertainty quantification which is undesirable, especially in financial applications where we are making sensitive decisions, and additional knowledge on the uncertainty of our correlation matrices could lead to more prudent decision making such as with respect to portfolio diversification.

Bayesian methods can provide us with the desired uncertainty quantification, such as the construction of credible intervals. Multivariate factor stochastic volatility models such as Chib et al. [2006] and Hosszejni and Kastner [2021] utilize a latent factor model approach. However, they do not utilize time dependent shrinkage which leads to very loose credible intervals.

To account for these difficulties, we propose a novel Bayesian approach to estimating correlation matrices from multivariate factor models. We utilize dynamic shrinkage processes (Kowal et al. [2019]) for Bayesian estimation of the idiosyncratic parameter variances and a multivariate factor stochastic volatility model for the observation error covariances. That is, we assume the state variables across assets are independent, with dependence across series through the unobserved observation error covariances. Through the DSP priors this allows the model to locally adapt to changing market conditions and gives tighter credible intervals.

A problem with estimating time series of correlation matrices is how do we summarize this information? The most popular solution is to plot the estimated pair wise correlations through time; however this can become cumbersome, and ultimately uninformative in even moderate dimension. Therefore, we propose a novel scalar score which ranges between -1 to +1, to summarize individual correlation matrices. We then derive posterior samples of this score and track this through time. This score is based upon the concept of scalar projection and possesses several desirable properties, most importantly simplicity and interpretability. The score along with the proposed estimation framework provides both industry practitioners and researchers with a novel tool for quantifying risk in a portfolio linked to a co-movement of its constituent stocks. This portfolio specific measure can provide portfolio managers with a clearer understanding of the structure of their portfolio and help them decide on future portfolio allocation. This also helps to reduce the amount of computer storage required to perform Bayesian analysis of the correlation matrix of a given stock portfolio over time by only tracking a scalar quantity through time instead of a matrix quantity over time.

This paper will proceed as follows: in Section 2 we will introduce the methodology including specification of priors and observation equations. Section 3 will discuss the estimation of correlation matrices and state how we summarize these correlation matrices through our novel score and some properties of this score. In section 4 we show that our model satisfies a posterior concentration result. Section 5 will discuss the computational details involved with estimation such as the details of our Gibbs sampling algorithm. Section 6 will discuss results of our proposed method, both in a simulation study and real-world examples where we demonstrate that the impacts of financial crises on an investor's portfolio cannot be diversified away in a portfolio of stocks. Section 7 will conclude and suggest future research directions.

2 Methodology

2.1 Notation and definitions

In this Section we present our modeling methodology based on multivariate time varying parameter linear factor models with an application focus on asset return.

Definition 1. *The net return of an asset adjusted for dividends at time t is given by $R_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}}$, where P_t is the price of the asset at time t and D_t is the dividend paid before time t.*

We will also be utilizing the concept of risk-free rate. This refers to the return someone can earn on an asset, where the variance of this return is zero, for example some fixed income securities such as U.S. treasury bills, although such assets are not truly risk free.

Definition 2. *The excess market return is the return you can make by investing in the market portfolio (theoretical collection of all investable assets) minus the risk free rate. Similarly, the excess return of an asset is the return of the asset minus the risk-free rate.*

Definition 3. *A multivariate linear factor model for N assets is given by:*

$$\boldsymbol{r}_t = \boldsymbol{\alpha}_t + \boldsymbol{\beta}_{1,t} F_{1,t} + \ldots + \boldsymbol{\beta}_{z,t} F_{z,t} + \boldsymbol{\epsilon}_t. \tag{1}$$

Where $t \in \mathbb{N}$, $\boldsymbol{r}_t = (r_{1,t}, \ldots, r_{N,t})'$ is the vector of N excess asset returns at time t. The vector of intercept terms is given by $\boldsymbol{\alpha}_t = (\alpha_{1,t}, \ldots, \alpha_{N,t})'$. The factors in the model are $F_{1,t}, \ldots, F_{z,t}$ which in this paper are observed, and $\boldsymbol{\beta}_{1,t}, \ldots, \boldsymbol{\beta}_{z,t}$ are the factor loading vectors. The vector of idiosyncratic observation errors is given by $\boldsymbol{\epsilon}_t$. In Section 2.2 we place further constraints on this model such as prior distributional assumptions.

More information on financial time series, and fundamental quantitative finance methods can be found in several texts including Tsay [2005] and Ruppert and Matteson [2011] .

2.2 Model

We are proposing a Bayesian time series model utilizing dynamic shrinkage prior (DSP) processes put forward in Kowal et al. [2019]. DSP processes build on the Horseshoe prior of Carvalho et al. [2009] which is a global-local shrinkage prior using normal scale mixtures. Global-local shrinkage priors are continuous priors which impose a global level of shrinkage on all the state variables in a model, but also allow for parameter specific levels of shrinkage and are an alternative to exact sparsity inducing priors such as the spike and slab prior.

DSP processes extend this idea to the four parameter Z-distribution which provide a natural extension as they can be written as Normal mean-scale mixtures (Barndorff-Nielsen et al. [1982]) and therefore can provide additional flexibility in the shape of the shrinkage, which includes horseshoe shaped shrinkage as a special case. However, when applied to time series analysis prior distributions such as the horseshoe prior suffer from a lack of temporal adaptability, that is the shrinkage is constant with respect to time.

Kowal et al. [2019] instead proposes a prior process for the amount of shrinkage which has the advantage of having the shrinkage be locally adaptive with respect to time. This is very helpful from the perspective of time series analysis; for example suppose we are moving from one time point to the next time point in a random walk fashion, if there is little change in the signal we would desire the innovation of the process to be shrunk strongly towards zero, alternatively if there is suddenly a large innovation then we would prefer very little shrinkage, see theorem 2 and 3 in Kowal et al. [2019]. By modeling the shrinkage through a prior stochastic process, DSP processes utilize the previous observations to determine a good amount of shrinkage but can also adapt the shrinkage to sudden large innovations.

The second component of our model relies on multivariate factor stochastic volatility (MFSV) processes. As highlighted in Section 1, estimation of covariance matrices can suffer from the curse of dimensionality, since we have several free parameters but only one data point for a single moment in discrete time. Therefore, to make estimation feasible we need to make some low rank inducing assumptions. To do this the model assumes that the time series of covariances is driven by a small set of common latent factors. This results in a computationally tractable model which we utilize for the time varying observation error covariances between the assets in a portfolio.

Throughout the rest of this article we will refer to the model which combines DSP with MFSV as DSP-MFSV with observation equation and prior distribution specification discussed in Sections 2.2.1 and 2.2.2, respectively. We discuss the computation of drawing samples from the posterior distribution of our model in Section 5.

2.2.1 Observation equation

For ease of explanation we will focus on rank one factor models, but our methodology works for higher rank factor models too. Consider the capital asset pricing model of Sharpe [1964], Lintner [1965a], Lintner [1965b], and Mossin [1966].

$$r_{a,t} = \alpha_{a,t} + \beta_{a,t} r_{M,t} + \epsilon_{a,t} \tag{2}$$

$$r_{M,t} = exp\{\frac{h_{M,t}}{2}\}\epsilon_{M,t}, \epsilon_{M,t} \sim N(0,1) \tag{3}$$

The excess return of asset $a \in 1, ..., N$, in our portfolio at time $t \in \mathbb{N}$ is given by $r_{a,t}$ and $r_{M,t}$ is the excess market return at time t, $\epsilon_{a,t}$ is the idiosyncratic observation error, with $\alpha_{a,t}$ and $\beta_{a,t}$ being unobserved state variables. We assume the state variables are independent across assets, but allow the observation errors to exhibit dependence. That is, we assume that the observation errors follow a MFSV process as discussed in Hosszejni and Kastner [2021]. In this model we assume the process of time-varying covariance matrices is driven by a small number of latent factors, which is appropriate for modeling the observation errors due to their unobserved nature and the fact that there are attributes in the wider economy which cannot be accounted for in our model but are leading to co-movement of stock returns. Particularly,

$$\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, ..., \epsilon_{N,t})^{'}|(\boldsymbol{\Lambda}, \boldsymbol{f}_t, \bar{\boldsymbol{\Sigma}}_t) \sim N_N(\boldsymbol{\Lambda}\boldsymbol{f}_t, \bar{\boldsymbol{\Sigma}}_t) \text{ with } \boldsymbol{f}_t|\tilde{\boldsymbol{\Sigma}}_t \sim N_m(\mathbf{0}_m, \tilde{\boldsymbol{\Sigma}}_t) \tag{4}$$

The m latent factors in the model at time t are given by $\boldsymbol{f}_t$. The priors for the diagonal matrices $\bar{\boldsymbol{\Sigma}}_t$ and $\tilde{\boldsymbol{\Sigma}}_t$, and the static matrix of factor loadings $\boldsymbol{\Lambda}$ will be discussed in Section 2.2.2. Note, we may write that at a given time point the covariance matrix of $\boldsymbol{\epsilon}_t$ is given by $\boldsymbol{\Lambda}\tilde{\boldsymbol{\Sigma}}_{\mathbf{t}}\boldsymbol{\Lambda}^{'} + \bar{\boldsymbol{\Sigma}}_{\mathbf{t}}$.

2.2.2 Priors

In this Section we discuss our prior distribution specification for the observation equation in Section 2.2.1. We assume the state variables between the assets in our model evolve independently of each other, and for each state variable in a given assets observation equation we further assume that they too are independent and evolve according to their own specific dynamic shrinkage prior process. These assumptions seem valid as the α and β of one stock should not influence the α and β of another stock as they could be considered a defining property of a given stock. Then for a given asset a in our portfolio we have:

$$\beta_{a,t+1} = \beta_{a,t} + \omega_{\beta_a,t}, \tag{5}$$

$$\omega_{\beta_a,t}|\tau_{a,0}\tau_{\beta_a}, \{\lambda_{\beta_a,s}\} \sim N(0, \tau_{a,0}^2\tau_{\beta_a}^2\lambda_{\beta_a,t}^2), \tag{6}$$

$$h_{\beta_a,t} = log(\tau_{a,0}^2\tau_{\beta_a}^2\lambda_{\beta_a,t}^2), \tag{7}$$

$$h_{\beta_a,t} = \mu_{\beta_a} + \phi_{\beta_a}(h_{\beta_a,t-1} - \mu_{\beta_a}) + \eta_{\beta_a,t}, \tag{8}$$

$$\tau_{a,0} \sim C^+(0, \frac{1}{\sqrt{T}}), \tau_{\beta_a} \sim C^+(0,1), \eta_{\beta_a,t} \sim Z(\frac{1}{2}, \frac{1}{2}, 0, 1), \tag{9}$$

$$\frac{\phi_{\beta_a} + 1}{2} \sim Beta(10, 2). \tag{10}$$

$$h_{M,t+1} = \mu_M + \phi_M(h_{t,M} - \mu_M) + \sigma_M\eta_t, \eta_t \sim N(0,1) \tag{11}$$

$$\mu \sim N(0,100), \frac{\phi_M+1}{2} \sim Beta(10,3) \tag{12}$$

The prior specification for the other state variables in a given observation equation are the same as above. By stating our state variables evolve according to a normal random walk in equation (5) of our prior specification the problem becomes tractable and still allows us to capture a rich collection of dynamics while maintaining algebraic simplicity.

We then allow the innovation of the process to follow a global local shrinkage prior, where $\tau^2_{a,0}$ determines the global level of shrinkage across all the state variables, and across all of time. Similarly, $\tau^2_{\beta_a}$ determines the amount of shrinkage over all time of the parameter β of asset a in our portfolio, with shrinkage being implied by the half-Cauchy distribution. Finally $\lambda^2_{\beta_a,t}$, the local shrinkage parameter, determines the amount of shrinkage of the β parameter of asset a, at a particular point in time, that is it determines the amount of temporally local shrinkage.

As is common practice, the model specifies the prior distribution in terms of the log conditional variance of the innovation from equation (7). We assume the log-variance process evolves according to an autoregressive one model (8), but with Z-distributed errors (9), rather than the typical normally distributed errors, due to their ability to induce shrinkage. Particularly, we utilize horeshoe like shrinkage by specifying $Z(\frac{1}{2}, \frac{1}{2}, 0, 1)$ due to its symmetric level of shrinkage, which mostly applies either a lot of shrinkage or a little shrinkage. See Kowal et al. [2019] for a discussion on some of the other shrinkage types available. For the observation error covariance we use

$$\bar{\mathbf{\Sigma}}_t = diag(exp(\bar{h}_{t,1}), ..., exp(\bar{h}_{t,m})), \tilde{\mathbf{\Sigma}}_t = diag(exp(\tilde{h}_{t,1}), ..., exp(\tilde{h}_{t,r})), \tag{13}$$

$$\bar{h}_{t,i} \sim N(\bar{\mu}_i + \bar{\psi}_i(\bar{h}_{t-1,i} - \bar{\mu}_i), \bar{\sigma}^2_i), i = 1, ..., m, \tag{14}$$

$$\tilde{h}_{t,j} \sim N(\tilde{\mu}_j + \tilde{\psi}_j(\tilde{h}_{t-1,j} - \tilde{\mu}_j), \tilde{\sigma}^2_j), j = 1, ..., r, \tag{15}$$

$$\mathbf{\Lambda}_{i,j}|\tau^2_{i,j} \sim N(0, \tau^2_{i,j}), \tau^2_{i,j}|\lambda^2_i \sim Ga(0.1, \frac{0.1\lambda^2_i}{2}), \lambda^2_i \sim Ga(1,1), \tag{16}$$

$$\bar{\sigma}_i \sim Ga(\frac{1}{2}, \frac{1}{2}), \tilde{\sigma}_i \sim Ga(\frac{1}{2}, \frac{1}{2}), \tag{17}$$

$$\bar{\mu}_i \sim N(0, 10), \tilde{\mu}_j \sim N(0, 10), i = 1, ..., m, j = 1, ..., r, \tag{18}$$

$$\frac{\bar{\psi}_i + 1}{2} \sim Beta(10, 3), \frac{\tilde{\psi}_i + 1}{2} \sim Beta(10, 3), i = 1, .., m, j = 1, ..., r. \tag{19}$$

As can been seen in equations (13)-(15) the entries of the conditional covariance matrices from Section 2.2.1 follow independent stochastic volatility processes of order one which allows the model to capture time-varying covariances of the observation errors of our observation equations. The prior utilizes normal gamma shrinkage priors (Griffin and Brown [2010]) for the entries of the matrix of factor loadings $\mathbf{\Lambda}$. Finally, we place a beta prior on a function of the persistence parameters to ensure they remain between 0 and 1 for stationarity, see (10), (12), and (19). In addition to the observation equation and prior distribution specification above, we further assume the excess market return $r_{M,t}$ follows a stochastic volatility process of order one (Taylor [1982]), independently of the other terms in (2).

3 Estimation and summary of the correlation matrices

We utilize a Gibbs sampling algorithm (see Section 5) to perform posterior inference. For each MCMC sample we compute an estimated model based covariance matrix, where there is one covariance matrix for each observed time point. We then standardize these covariance matrices to obtain correlation matrices, and finally we use our proposed scalar summary to summarize the correlation matrices in the series. After this procedure we will then have posterior draws consisting of times series of the estimated scalar summaries of correlation matrices.

3.1 Construction of the Covariance matrices

To obtain good estimators of covariance matrices we propose to use those derived from the estimated multivariate linear factor model which allows us to utilize the dynamic shrinkage prior processes and take advantage of the time dependent shrinkage. By parameterizing our covariance estimators using low rank linear factor models (Section 2.2.1) this induces approximate low rank structure. This low rankness combined with dynamic shrinkage provide low variance covariance estimators. Furthermore, by reserving the use of latent factor models for the unobserved observation error, our covariance estimators are more explainable and easier to interpret.

Proceeding from the CAPM where $\boldsymbol{r}_t = \boldsymbol{\alpha}_t + r_{M,t}\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t$ we obtain the formula for the model-based covariance matrix of the returns as

$$var(\boldsymbol{r}_t) = var(\boldsymbol{\alpha}_t) + var(r_{M,t})E[\boldsymbol{\beta}_t\boldsymbol{\beta}_t^{'}] + var(\boldsymbol{\epsilon}_t). \tag{20}$$

The derivation of the above formula is provided below where we assume mutual independence and the excess market returns follow a stochastic volatility process of order one.

3.1.1 Derivation

$\boldsymbol{r}_t = \boldsymbol{\alpha}_t + r_{M,t}\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t$

$\implies var(\boldsymbol{r}_t) = var(\boldsymbol{\alpha}_t) + var(r_{M,t}\boldsymbol{\beta}_t) + var(\boldsymbol{\epsilon}_t)$, by assuming mutual independence

$var(r_{M,t}\boldsymbol{\beta}_t) = E[r_{M,t}^2]E[\boldsymbol{\beta}_t\boldsymbol{\beta}_t^{'}] - E[r_{M,t}]^2E[\boldsymbol{\beta}_t]E[\boldsymbol{\beta}_t]^{'}$, by mutual independence

$\implies var(r_{M,t}\boldsymbol{\beta}_t) = (var(r_{M,t}) + E[r_{M,t}]^2)E[\boldsymbol{\beta}_t\boldsymbol{\beta}_t^{'}] - E[r_{M,t}]^2E[\boldsymbol{\beta}_t]E[\boldsymbol{\beta}_t]^{'}$

$\implies var(r_{M,t}\boldsymbol{\beta}_t) = var(r_{M,t})E[\boldsymbol{\beta}_t\boldsymbol{\beta}_t^{'}]$, since $r_{M,t}$ follows a mean zero stochastic process

$\implies var(\boldsymbol{r}_t) = var(\boldsymbol{\alpha}_t) + var(r_{M,t})E[\boldsymbol{\beta}_t\boldsymbol{\beta}_t^{'}] + var(\boldsymbol{\epsilon}_t)$.

3.2 Summary of Correlation matrices

In a given MCMC sample, once we have an estimated covariance matrix we can derive the associated correlation matrix by standardization. However, for correlation matrices of even moderate dimension, say 5 or 10 assets, how do we gain valuable insights from our correlation matrices to aid portfolio allocation? The common practice is to plot the

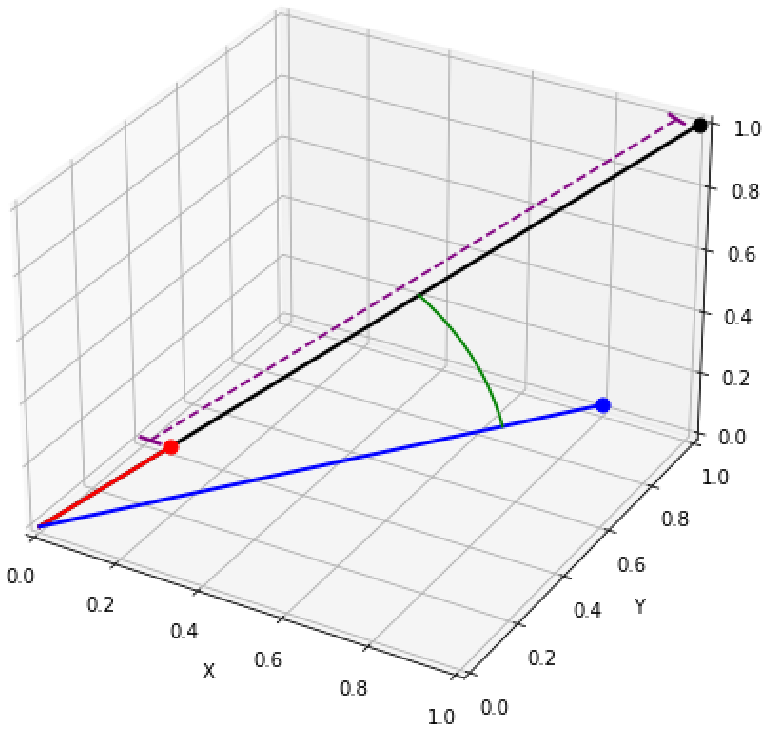

Figure 1: Plot of three vectors. The vector $(1, 1, 1)^T$ is in black, the vector $(0.2, 0.2, 0.2)^T$ is in red, and the vector $(0.9, 0.8, 0.2)^T$ is in blue. The difference in the length between $(0.2, 0.2, 0.2)^T$ and $(1, 1, 1)^T$ is represented by the purple dashed line, and the angle between the vectors $(0.9, 0.8, 0.2)^T$ and $(1, 1, 1)^T$ is given by the green arc.

time series of the estimated pairwise correlations. However, having plots of multiple estimated pairwise correlation series can become cumbersome and ultimately uninformative, particularly in understanding the overall amount of correlation in a portfolio. Therefore, we propose a simple scalar summary of correlation matrices based on scalar projection.

The scalar projection of the vector $\boldsymbol{a}$ onto the vector $\boldsymbol{b}$ says how much of the vector $\boldsymbol{a}$ is in the direction of vector $\boldsymbol{b}$ and is given by the dot product of the two vectors divided by the vector norm of $\boldsymbol{b}$. Basic algebra shows that scalar projection takes into account both the length of the vector $\boldsymbol{a}$ but also the cosine of the angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

In our application we are interested in how close a correlation matrix is to being perfectly correlated in accordance with the literature which demonstrates that during times of financial crisis market correlations converge towards one. A correlation matrix is perfectly correlated when all the pairwise correlations are one and therefore all the columns are one vectors. Therefore, to measure how close a matrix is to the all one matrix we propose to see how close each column vector of the matrix is to the all one vector. In Figure 1 we can see that the $(0.2, 0.2, 0.2)^T$ vector is perfectly aligned with the $(1, 1, 1)^T$ vector but it is noticeably shorter. Consequently our proposed summary should take into account the length of the column vector. We also have the $(0.9, 0.8, 0.2)^T$ vector. This vector is longer than the $(0.2, 0.2, 0.2)^T$ vector but it has a non-zero angle with the all one vector. Thus to see how close a given column of a correlation matrix is to the all one vector we also need to consider the angle between the column vector and $(1, 1, 1)^T$. A quantity that takes both of these aspects into account is the scalar projection of the respective column of the correlation matrix onto the all one vector. Based upon this we propose the following score to summarize correlation matrices.

Definition 4. *For an $N \times N$ correlation matrix $\boldsymbol{A}$ with column vectors $\boldsymbol{x}_1, ..., \boldsymbol{x}_N$ our proposed scalar score is given by*

$$score(\mathbf{A}) = \frac{\sum_{j=1}^{N} \boldsymbol{x}_j^{'} \mathbf{1}_N - N}{N(N-1)}, \mathbf{1}_N = (1, 1, ..., 1)^{'}. \tag{21}$$

This score is the sum of the scalar projections of the columns of the correlation matrix onto the all one vector scaled to the range $[-1, 1]$ to allow the measure to not depend on the dimension of the matrix. This score builds upon the notion that correlation matrices can be seen as a matrix of cosines of angles. The $score(\cdot)$ also has some simple but desirable properties.

Property 1. *For the correlation matrix $\boldsymbol{A} = c\mathbf{1}_N\mathbf{1}_N^{'}$, where $-1 \leq c \leq 1$, $score(\boldsymbol{A}) = c$. That is, $score(\boldsymbol{A}) = \frac{1}{N(N-1)} \sum_{j=1}^{N} [c(N-1)+1] - N = \frac{1}{N(N-1)}(cN(N-1)+N-N) = c$.*

Property 2. *The score$(\cdot)$ is invariant to the dimension of the matrix. For example if all the pairwise correlations in an $N \times N$ matrix $\boldsymbol{B}$ are equal to c, then from Property 1, the score$(\cdot)$ is c. That is, $\frac{1}{N(N-1)}(\sum_{j=1}^{N}[c(N-1)+1]-N) = c$. Then if we add an additional row and column where all the pairwise correlations are c, then*
$score(\boldsymbol{B}) = \frac{1}{N(N+1)}(\sum_{j=1}^{N+1}[cN+1]-(N+1)) = c$.

Beyond these properties, this score summarizes different pairwise correlations in an intuitive manner. For example if there is a 3×3 correlation matrix with pairwise correlations equal to 0.9,0.9,0.7 then the score is 0.833 to three decimal places (3.d.p); for the 4×4 correlation matrix with pairwise correlations 0.9,0.2,-0.1,0.87,0.5,0.52 the score is 0.481 to 3.d.p. Hence, once we have the estimated correlation matrix time series for a given MCMC sample we can then summarize this matrix time series as a scalar time series using the above score function.

4 Posterior Concentration

To prove our posterior concentration result (Theorem 9) we will be utilizing Theorem 1 of Ghosal and van der Vaart [2007]. To utilize this we need to prove that our model satisfies three properties which we show in Theorem 4, Theorem 5, and Theorem 7. The first property, Theorem 4, requires that there exists tests which have exponentially decaying probability of type 1 and type 2 error. In order to verify this we first need to deduce the mixing properties of our prior stochastic process of the correlation matrices. Firstly we define two types of mixing of stochastic processes.

Definition 5. *Let $(Y(t))_{t\in\mathbb{Z}}$ be a stationary process in $\mathbb{R}^l$. Denote the σ-algebra generated by $\{Y(t) : t \leq 0\}$ by $\mathcal{A}_0$ and $\mathcal{A}^k$, the σ-algebra generated by $\{Y(t) : t \geq k\}$. Then the complete regularity coefficient,for $k > 0$ is given by $\beta(k) = \mathbb{E}[\sup_{B\in\mathcal{A}^k} |\mathbb{P}(B/\mathcal{A}_0) - \mathbb{P}(B)|]$. Similarly, the strong mixing coefficient is defined by $\alpha(k) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{A}_0, B \in \mathcal{A}^k\}$. We say $(Y(t))_{t\in\mathbb{Z}}$ is geometrically completely regular if $\exists 0 < \rho < 1$ such that $\beta(k) = O(\rho^k)$. Mokkadem [1988]*

The stochastic process of the correlation matrices involves exponential functions of stochastic processes. Now we will show that a measurable function of a stochastic process is geometrically beta mixing if the input stochastic process is also geometrically beta mixing.

Theorem 1. *For any measurable function $g : \mathbb{R} \to \mathbb{R}$ and geometrically beta mixing process $(X_t)_{t\in\mathbb{Z}} \in \mathbb{R}$, the process $g(X_t)$ is also geometrically beta mixing.*

Proof. Firstly we define the σ-algebras we will be working with in view of Definition 5. Let $A_0^X = \sigma(X(t) : t \leq 0)$ and $A_k^X = \sigma(X(t) : t \geq k)$. Then $B_X^k = \mathbb{E}[\sup_{B\in A_k^X} |\mathbb{P}(B|A_0) - \mathbb{P}(B)|]$. Now, consider $Y(t) = g(X(t))$. Then, $A_0^Y = \sigma(Y(t) : t \leq 0) = \sigma(g(X(t)) : t \leq 0) \subseteq A_0^X$ and $A_k^Y = \sigma(Y(t) : t \geq k) = \sigma(g(X(t)) : t \geq k) \subseteq A_k^X$.

Note that, $\mathbb{P}(B|A_0^Y) = \mathbb{E}[\mathbf{1}_B|A_0^Y] = \mathbb{E}[\mathbb{E}[\mathbf{1}_B|A_0^X]|A_0^Y]$, by the law of total expectation, which equals $\mathbb{E}[\mathbb{P}(B|A_0^X)|A_0^Y]$. This implies $\mathbb{P}(B|A_0^Y) - \mathbb{P}(B) = \mathbb{E}[\mathbb{P}(B|A_0^X)|A_0^Y] - \mathbb{P}(B) = \mathbb{E}[\mathbb{P}(B|A_0^X) - \mathbb{P}(B)|A_0^Y]$. By Jensen's inequality we then obtain $|\mathbb{E}[\mathbb{P}(B|A_0^X) - \mathbb{P}(B)|A_0^Y]| \leq \mathbb{E}[|\mathbb{P}(B|A_0^X) - \mathbb{P}(B)||A_0^Y]$. This implies $\sup_{B\in A_k^Y} |\mathbb{P}(A_0^Y - \mathbb{P}(B)| \leq \sup_{B\in A_k^Y} \mathbb{E}[|\mathbb{P}(B|A_0^X) - \mathbb{P}(B)||A_0^Y] \leq \sup_{B\in A_k^X} \mathbb{E}[|\mathbb{P}(B|A_0^X) - \mathbb{P}(B)||A_0^Y] \leq \mathbb{E}[\sup_{B\in A_k^X} |\mathbb{P}(B|A_0^X) - \mathbb{P}(B)||A_0^Y]$. This implies $\mathbb{E}[\sup_{B\in A_k^Y} |\mathbb{P}(B|A_0^Y) - \mathbb{P}(B)|] \leq \mathbb{E}[\mathbb{E}[\sup_{B\in A_k^X} |\mathbb{P}(B|A_0^X) - \mathbb{P}(B)||A_0^Y]] = \mathbb{E}[\mathbb{E}[\sup_{B\in A_k^X} |\mathbb{P}(B|A_0^X - \mathbb{P})||A_0^Y]|\{\phi, \Omega\}] = \mathbb{E}[\sup_{B\in A_k^X} |\mathbb{P}(B|A_0^X) - \mathbb{P}(B)||\{\phi, \Omega\}]$, by the law of total expectation. This equals $\mathbb{E}[\sup_{B\in A_k^X} |\mathbb{P}(B|A_0^X) - \mathbb{P}(B)|] = \beta_X(k)$, Therefore, $B_Y(k) \leq \beta_X(k) = O(\rho^k)$. Therefore, Y is geometrically beta mixing. □

Using Definition 5 and Theorem 1 we will show that the process is geometric beta mixing. By establishing the prior correlation matrix process is beta mixing this will allow us to derive a Bernstein-type inequality which allow us to prove Theorem 4.

Theorem 2. *The process $cov(\mathbf{r}_t)_{ij}|\mathcal{F}_{t-1} = \delta_{ij}exp\{h_{\alpha_{i,t-1}}\} + exp\{\mu_M + \phi_M(h_{M,t-1} - \mu_M) + \frac{1}{2}\sigma_M^2\}\beta_{i,t-1}\beta_{j,t-1} + \sum_{k=1}^{r} \Lambda_{ik}\Lambda_{jk}exp\{\tilde{\mu}_k + \tilde{\phi}_k(\tilde{h}_{k,t-1} - \tilde{\mu}_k) + \frac{1}{2}\tilde{\sigma}_k^2\} + \delta_{ij}exp\{\bar{\mu}_i + \bar{\phi}_i(\bar{h}_{i,t-1} - \bar{\mu}_i) + \frac{1}{2}\bar{\sigma}_i^2\}$ and $\mathcal{F}_{t-1}$ is the information set at time t-1, is geometrically beta mixing.*

Proof. The time evolution of the stochastic process is governed by the latent log variance processes. Each of the log variance processes are auto regressive processes of order 1 with innovations which have probability distributions which are absolutely continuous with respect to the Lebesgue measure. In general the processes have the form $Y_t = \mu + \phi(Y_{t-1} - \mu) + \eta_t$ which may be written in the mean centered form $Y_t - \mu_t = \phi(Y_{t-1} - \mu) + \eta_t$. Let $X_t := Y_t - \mu$. Then, $X_t - \phi X_{t-1} = \eta_t$. In view of theorem 1 of Mokkadem [1988] $B(0) = 1, B(1) = -\phi, B(i) = 0$ for $i \geq 2$. Similarly, $A(0) = 1, A(k) = 0$ for $k \geq 1$. This give us $P(Z) = 1 - \phi z$. The roots of this equation are given by $1 - \phi z = 0 \iff \phi z = 1 \iff z = \frac{1}{\phi}$ which is greater than 1, since all the specified AR processes in our model have $|\phi| < 1$. Then by theorem 1 of Mokkadem [1988] each of the log variance processes are geometrically beta mixing. Then by theorem 1 above we have that the covariance process is entry wise geometrically beta mixing. $\square$

Now that we have established the mixing properties of the stochastic process we can now deduce a Bernstein type inequality which we subsequently use to show the existence of the hypothesis tests with exponentially decaying type 1 and type 2 error.

Theorem 3. *Let $X_t^{(i,j)} = R_{t,\theta}^{(i,j)} - R_{t,\theta_0}^{(i,j)}$, where $R_{t,\theta}^{(i,j)}$ is the (i,j)th entry of the observed correlation matrix at time t, and $R_{t,\theta_0}^{(i,j)}$ is the (i,j)th entry of the true correlation matrix at time t, in view of the subsequent posterior concentration result. Let $\mu = E_\theta[X_t]$. Then, $\mathbb{P}_\theta(|\frac{1}{n}\sum_{t=1}^{n} X_t^{(i,j)}| > \epsilon_n + |\mu^{(i,j)}|) \leq exp\{-C_4\sqrt{n}\epsilon_n^2\}$, for fixed n, constant $C_4 > 0$, and $\epsilon_n = \frac{1}{log(n)log(log(n))}$.*

Proof. We first need to verify some conditions in order for us to apply Theorem 1 of Merlevède et al. [2009]. Firstly, $|X_t^{(i,j)}| = |R_{t,\theta}^{(i,j)} - R_{t,\theta_0}^{(i,j)}| \leq |R_{t,\theta}^{(i,j)}| + |R_{t,\theta_0}^{(i,j)}| \leq 2 < \infty$, by the triangle inequality and the fact that the largest value of any entry in a correlation matrix is at most 1. This then gives $|X_t^{(i,j)} - \mu^{(i,j)}| \leq |X_t^{(i,j)}| + |\mu^{(i,j)}| \leq 4 \leq \infty$ by the triangle inequality.

Secondly, by theorem 2, the correlation matrix process is entry wise geometrically completely regular this implies that the entry wise process is α-mixing. Particularly, there exists constants κ and ρ where $0 < \rho < 1$ such that $\alpha(n) \leq \kappa\rho^n = exp\{ln(\kappa\rho^n)\} =$

$exp\{ln(\kappa) - n(-ln(\rho))\}$. Upon choosing any C_1 such that $C_1 < -log(\rho)$ we see $\alpha(n) \leq exp\{ln(\kappa) - C_1 n\}$. If $\kappa \leq 1$ then $ln(\kappa) \leq 0 \implies \alpha(n) \leq exp\{-C_1 n\}$. Let $C_1 = 2C_2$, for $C_2 > 0$. Then we obtain $\alpha(n) \leq exp\{-2C_2 n\}$ which satisfies condition 1.3 of Merlevède et al. [2009].

Then by theorem 1 of Merlevède et al. [2009] we obtain $\mathbb{P}(|\sum_{t=1}^{n} X_t^{(i,j)} - \mu^{(i,j)}| \geq x) \leq exp\{-\frac{C_3 x^2}{4^2 n + 4x(ln(n))(ln(ln(n)))}\}$ for $x \geq 0, n \geq 4$. Let $x = n\epsilon_n$, where $\epsilon_n = \frac{1}{log(n)log(log(n))}$. Then, $\mathbb{P}(|\sum_{t=1}^{n} X_t^{(i,j)} - \mu^{(i,j)}| \geq n\epsilon_n) \leq exp\{-\frac{C_3 n^2 \epsilon_n^2}{16n + \frac{4nlog(n)log(log(n))}{log(n)log(log(n))}}\}$ which implies $\mathbb{P}(|\frac{1}{n}\sum_{t=1}^{n} X_t^{(i,j)} - \mu^{(i,j)}|) \leq exp\{-C_4 n\epsilon_n^2\}$, where $C_4 := \frac{C_2}{20}$. Then, by the triangle inequality we obtain $\mathbb{P}(|\frac{1}{n}\sum_{t=1}^{n} X_t^{(i,j)}| \geq \epsilon_n + |\mu^{(i,j)}|) \leq exp\{-C_4 n\epsilon_n^2\}$. □

Using Theorem 3 we can now verify the existence of tests with exponentially decaying type 1 and type 2 error probabilities which is required so we can apply Theorem 1 of Ghosal and van der Vaart [2007]. This allows us to separate the true parameters from alternative parameter values.

Theorem 4. *There exists a test ϕ_n such that $\mathbb{P}_{\theta_0}^{(n)}(\phi_n = 1) \leq exp\{-Kn\epsilon_n^2\}$ and $\sup_{d_n(\theta,\theta_1)<\xi\epsilon_n} \mathbb{P}_{\theta}^{(n)}(1-\phi_n) \leq exp\{-Kn\epsilon_n^2\}$, where $K, \xi > 0$ are constants, and $d_n(\theta, \theta') = ||R_\theta - R_{\theta'}||_F$, and ϵ_n is given in Theorem 3.*

Proof. Consider the test $\phi_n = \mathbf{1}\{\max_{1\leq i,j\leq N} |\bar{X}_n^{(i,j)}| > \tau\}$, and $\tau = \frac{\epsilon_n}{N}$. Then $\mathbb{P}_{\theta_0}^{(n)}(\phi_n = 1) = \mathbb{P}_{\theta_0}^{(n)}(\max_{1\leq i,j\leq N} |\bar{X}_n^{(i,j)}| > \tau) = \mathbb{P}_{\theta_0}^{(n)}(\bigcup_{i,j=1}^{N} |\bar{X}_n^{(i,j)}| > \tau) \leq \sum_{i,j=1}^{N} \mathbb{P}_{\theta_0}^{(n)}(|\bar{X}_n^{(i,j)}| > \tau)$, by the union bound. Then by Theorem 2, and the fact that the probability is taken with respect to the null hypothesis we obtain $\mathbb{P}_{\theta_0}(\phi_n = 1) \leq N^2 exp\{-C_4 n\tau^2\} = N^2 exp\{-C_4 \frac{n\epsilon_n^2}{N^2}\} = exp\{log(N^2) - C_4\frac{n\epsilon_n^2}{N^2}\} \leq exp\{-K_1 n\epsilon_n^2\}$, where $\frac{C_4 n\epsilon_n^2}{N^2} - log(N^2) \geq K_1 n\epsilon_n^2 \iff \frac{C_4}{N^2} - \frac{log(N^2)}{n\epsilon_n^2} \geq K_1$.

Now suppose we have an alternative hypothesis satisfying $||R_\theta - R_{\theta_0}||_F > \epsilon_n$. That is, $\sqrt{\sum_{i,j=1}^{N} |R_\theta^{(i,j)} - R_{\theta_0}^{(i,j)}|^2} > \epsilon_n \implies \exists(i^*, j^*)$ such that $|R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)}| \geq \frac{\epsilon_n}{N}$. Suppose to the contrary, that is $|R_\theta^{(i,j)} - R_{\theta_0}^{(i,j)}| < \frac{\epsilon_n}{N}, \forall 1 \leq i,j \leq N$. Then, $||R_\theta - R_{\theta_0}||_F = \sqrt{\sum_{i,j=1}^{N} |R_\theta^{(i,j)} - R_{\theta_0}^{(i,j)}|^2} < \sqrt{\sum_{i,j=1}^{N} |\frac{\epsilon_n}{N}|^2} = \epsilon_n$. That is, $||R_\theta - R_{\theta_0}|| < \epsilon_n$

which is a contradiction. Furthermore, if $\phi_n = 0$ then $\max_{1\leq i,j\leq N} |\bar{X}_n^{(i,j)}| \leq \tau$ which implies $|\bar{X}_n^{(i^*,j^*)}| \leq \frac{\epsilon_n}{N}$. Then, $|\bar{X}_n^{(i^*,j^*)} - (R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)})| = |(R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)}) - \bar{X}_n^{(i^*,j^*)}| \geq ||R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)}| - |\bar{X}_n^{(i^*,j^*)}||$, by the reverse triangle inequality. This leads to $|\bar{X}_n^{(i^*,j^*)} - (R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)})| \geq ||\frac{\epsilon_n}{N}| - |0|| = \frac{\epsilon_n}{N}$. Then, $\mathbb{P}_\theta^{(n)}(1-\phi_n) = \mathbb{P}_\theta(\phi_n = 0) = \mathbb{P}_\theta^{(n)}(\max_{1\leq i,j\leq N} |\bar{X}_n^{(i,j)}| \leq \tau) \leq \mathbb{P}_\theta^{(n)}(|\bar{X}_n^{(i^*,j^*)}| \leq \tau)$. Let $A^*_{(i,j)}$ be the event that $\{|\bar{X}_n^{(i^*,j^*)}| \leq \tau\}$. On $A^*_{(i,j)}$ we have $|R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)}| \geq \frac{\epsilon_n}{N}$ and $|\bar{X}_n^{(i^*,j^*)}| \leq \tau$. Therefore, on $A^*_{(i,j)}$ we have $|\bar{X}_n^{(i^*,j^*)} - (R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)})| \geq \frac{\epsilon_n}{N}$. Thus, $A^*_{(i,j)} \subseteq \{|\bar{X}_n^{(i^*,j^*)} - (R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)})| \geq \frac{\epsilon_n}{N}\}$. This implies that, $\mathbb{P}_\theta^{(n)}(|\bar{X}_n^{(i^*,j^*)}| \leq \tau) \leq \mathbb{P}_\theta^{(n)}(|\bar{X}_n^{(i^*,j^*)} - (R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)})| \geq \frac{\epsilon_n}{N}) = \mathbb{P}_\theta^{(n)}(|\bar{X}_n^{(i^*,j^*)} - \frac{1}{n}\sum_{t=1}^{n} E[X_t^{(i^*,j^*)}]| \geq \frac{\epsilon_n}{N}) \leq exp\{-C_4 n(\frac{\epsilon_n}{N})^2\} = exp\{-C_4 n\frac{\epsilon_n^2}{N^2}\} = exp\{-K_2 n\epsilon_n^2\}$, $K_2 = \frac{C_4}{N^2}$.

Let $K = min(K_1, K_2)$, then $\mathbb{P}_{\theta_0}^{(n)}(\phi_n = 1) \leq exp\{-Kn\epsilon_n^2\}$ and $\mathbb{P}_\theta^{(n)}(1-\phi_n) \leq exp\{-Kn\epsilon_n^2\}$ which implies $\sup_{d_n(\theta_1,\theta_0)} \mathbb{P}_\theta^{(n)}(1-\phi_n) \leq exp\{-Kn\epsilon_n^2\}$, as required. □

Next, we need to prove a metric entropy condition for our model in order to apply Theorem 1 of Ghosal and van der Vaart [2007]. Particularly, it establishes that the parameter space near the true value of the parameters is not too complex such that the data would not be able get us close to the true value.

Theorem 5. $\sup_{\epsilon>\epsilon_n} logN(\frac{1}{2}\epsilon\xi, \{\theta \in \Theta_n : ||R_n(\theta) - R_n(\theta_0)||_F < \epsilon\}, ||\cdot||_F) \leq n\epsilon_n^2$, *where* $\epsilon_n = \sqrt{\frac{\xi^2(a+b+\frac{1}{2}+\frac{\alpha}{2})\#\theta log(n)}{n}}$ *for* $\xi > 1$ *and* $\Theta_n = \{\theta \in \Theta_n : \max_{1\leq i\leq N, 1\leq t\leq n-1} |h_{\alpha_{i,t-1}}| \leq |\mu_{\alpha_i}| + \frac{clog(n)+log(2k(s))}{s}, \max_{1\leq t\leq n} |h_{M,t-1}| \leq |\mu_M| + plog(n), |\mu_M| \leq log(n), |\phi_M| \leq \sqrt{1 - \frac{1}{log(log(n))}}, \sigma_m^2 \leq log(n)^{0.5}, \max_{1\leq i\leq N, 1\leq t\leq n} |\beta_{i,t-1}| \leq \sqrt{(\kappa+1)C_0 n^\alpha log(n)}, ||\Lambda||_F \leq log(n), \max_{1\leq k\leq r} |\tilde{\mu}_k| \leq log(n), \max_{1\leq k\leq r} \tilde{\phi}_k \leq \sqrt{1 - \frac{1}{log(log(n))}}, \max_{1\leq k\leq r} \tilde{\sigma}_k^2 \leq log(n)^{0.5}, \max_{1\leq k\leq r, 1\leq t\leq n} |\tilde{h}_{k,t-1}| \leq |\tilde{\mu}_k| + plog(n), \max_{1\leq i\leq N} |\bar{\mu}_i| \leq log(n), \max_{1\leq i\leq N} \bar{\phi}_i \leq \sqrt{1 - \frac{1}{log(log(n))}},$
$\max_{1\leq i\leq N, 1\leq t\leq n} |\bar{h}_{i,t-1}| \leq |\bar{\mu}_i| + plog(n), \max_{1\leq i\leq N} \bar{\sigma}_i^2 \leq log(n)^{0.5}, \max_{1\leq i\leq N} |\mu_{\alpha_i}| \leq log(n^4)\}$ *for* $p > 0, \alpha = \frac{c}{s} + 5$, *with* $\mathbb{P}(\Theta_n^c) \to 0$.

Proof. To start this proof we aim to bound the probability of the compliment of the sieve. To does this we bound the probability of each parameter in the sieve.

For the auto regressive process of order 1 with normally distributed innovations

($\tilde{h}_t$,$\bar{h}_t$, and $h_{M,t}$) we have $h_t = \sum_{j=0}^{\infty} \phi^j \eta_{t-j}$ which implies $M_{h_t}(s) = \prod_{j=0}^{\infty} M_\eta(s\phi^j) = \prod_{j=0}^{\infty} exp\{\frac{\sigma^2 s^2 \phi^{2j}}{2}\} = exp\{\frac{1}{2}\sigma^2 s^2 \sum_{j=0}^{\infty} \phi^{2j}\} = exp\{\frac{\sigma^2}{2(1-\phi^2)}\}$. Then, the Chernoff bound gives us $\mathbb{P}(h_t > u) \leq exp\{\frac{\sigma^2}{2(1-\phi^2)} - su\}$ and $\mathbb{P}(h_t < -u) = \mathbb{P}(-h_t > u) \leq exp\{\frac{\sigma^2}{2(1-\phi^2)} - su\}$. Then, $\mathbb{P}(|h_t| > u) = \mathbb{P}(h_t > u) + \mathbb{P}(h_t < -u) \leq 2exp\{\frac{\sigma^2 s^2}{2(1-\phi^2)} - su\}$. Let $s = \frac{(1-\phi^2)u}{\sigma^2}$, then $\mathbb{P}(|h_t| > u) \leq 2exp\{\frac{\sigma^2(1-\phi^2)u^2}{2\sigma^4(1-\phi^2)} - \frac{(1-\phi^2)u^2}{\sigma^2}\} = 2exp\{\frac{(1-\phi^2)u^2}{2\sigma^2} - \frac{(1-\phi^2)u^2}{\sigma^2}\} = 2exp\{-\frac{(1-\phi^2)u^2}{2\sigma^2}\}$. Now, let $u = |\mu| + plog(n)$ for $p > 0$. Then, $\mathbb{P}(|h_t| > u|\sigma^2, \phi) \leq 2exp\{-\frac{(1-\phi^2)(|\mu|+plog(n)-|\mu|)}{2\sigma^2}\} = 2exp\{-\frac{(1-\phi^2)(plog(n))^2}{2\sigma^2}\}$. Then, union bounding in order to bound the maximum, we see that $\mathbb{P}(\max_{1\leq k\leq r, 1\leq t\leq n-1} |\tilde{h}_{k,t-1}| \leq |\tilde{\mu}_k| + plog(n)|\tilde{\phi}, \tilde{\sigma}^2) \leq 2rexp\{-\frac{(1-\phi^2)p^2 log(n)^2}{2\sigma^2} + ln(n)\}$. On the sieve, $|\phi| \leq \sqrt{1 - \frac{1}{loglog(n)}}$ which implies $1-\phi^2 \geq \frac{1}{loglog(n)}$. This implies, $\mathbb{P}(\max_{1\leq k\leq r, 1\leq t\leq n-1} |\tilde{h}_{k,t-1}| \leq |\tilde{\mu}_k| + plog(n)|\tilde{\phi}, \tilde{\sigma}^2) \leq 2rexp\{ln(n) - \frac{p^2}{2}\frac{log(n)^{1.5}}{log(log(n))}\} \to 0$. Similarly, $\mathbb{P}(\max_{1\leq i\leq N, 1\leq t\leq n-1} |\bar{h}_{i,t-1}| \leq |\bar{\mu}_i| + plog(n)|\bar{\phi}, \bar{\sigma}^2) \leq 2Nexp\{ln(n) - \frac{p^2}{2}\frac{log(n)^{1.5}}{log(log(n))}\} \to 0$.

For auto regressive process of order 1 with Z-distributed innovations we see $X_t = \sum_{j=0}^{\infty} \epsilon_{t-j}$ which implies $M_X(s) = \prod_{j=0}^{\infty} sec(\pi\phi^j s)$ (see the appendix). Note that for fixed $s \in (0, 0.5)$ the infinite produce converges. For such an s we have $|\pi\phi^j s| = \pi|\phi|^j s \leq \pi s < \frac{\pi}{2}$. So $\pi\phi^j s \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Consider $\sum_{j=1}^{\infty} log(sec(\pi\phi^j s))$ where by Taylor expanding we see $log(sec(x)) = \frac{x^2}{2} + O(x^4)$ which implies $log(sec(\pi\phi^j s)) = \frac{\pi^2\phi^{2j}s^2}{2} + O(\phi^{4j})$. Observe that, $\lim_{j\to\infty} \frac{\frac{1}{2}\pi^2 s^2 \phi^{2(j+1)} + O(\phi^{4j})}{\frac{1}{2}\pi^2 s^2 \phi^{2j} + O(\phi^{4j})} = \phi^2$, with $|\phi^2| < 1$. Therefore, by the ratio test the sum is absolutely convergent. Therefore the product converges, $\prod_{j=0}^{\infty} sec(\pi\phi^j s) = exp\{\sum_{j=0}^{\infty} log(sec(\pi\phi^j s))\} < \infty$. Now, $\mathbb{P}(|X_t| > u) = \mathbb{P}(X_t > u) + \mathbb{P}(X_t \leq u) \leq M_X(s)e^{-su} + M_X(-s)e^{-su} \leq 2K(s)e^{-su}, K(s) = max\{M_X(s), M_X(-s)\}$. This implies $\mathbb{P}(|h_t| > u) \leq \mathbb{P}(|h_t| > u - |\mu|) \leq 2K(s)exp\{-s(u - |\mu|)\}$. Let $u_n = |\mu| + \frac{clog(n)+log(2k(s))}{s}$, for some constant $c > 0$. This implies $\mathbb{P}(|h_t| > |\mu| + \frac{clog(n)+log(2k(s))}{s}) \leq 2K(s)exp\{-s(|\mu| + \frac{clog(n)+log(2K(s))}{s}) - |\mu|)\} = 2K(s)exp\{log(\frac{n^{-c}}{2K(s)})\} = 2K(s)\frac{n^{-c}}{2K(s)} = n^{-c}$. This implies $\mathbb{P}(\max_{1\leq a\leq N, 1\leq t\leq n} |h_{a,t}| > |\mu| + \frac{clog(n)+log(2k(s))}{s}) \leq Nn^{-C_1+1} = Nn^{-C_2} \to 0$, for some constants $C_1, C_2 > 0$.

Now consider the mean parameters in the autoregessive process in (8). We firstly deduce their probability density function. Note that if $\tau \sim C^+(0,b), then f(t) = \frac{2}{\pi b}\frac{1}{1+\frac{t^2}{b^2}}$ which implies $f_X = \frac{2}{\pi b}\frac{1}{1+\frac{e^{x^2}}{b^2}}e^x = \frac{2e^x}{\pi b(1+\frac{e^{2x}}{b^2})} = \frac{2be^x}{\pi(b^2+e^{2x})}$. Then, $f_s(s) = \int\limits_{-\infty}^{\infty} \frac{2b_0e^u}{\pi(b_0^2+e^{2u})}\frac{2b_1e^u}{\pi(b_1^2+e^{2u})}du.$ In our case $b_1 = 1$ which implies $f_s(s) = \frac{4b_0e^s}{\pi^2}\int\limits_{-\infty}^{\infty} \frac{1}{(b_0^2+e^{2u})(1+e^{2(s-u)})}du$. Let $v = e^{2u}$ which implies $\frac{dv}{du} = 2e^{2u}$ and therefore $du = \frac{1}{2v}dv$. This gives $f(s) = \frac{4b_0e^s}{\pi^2}\int\limits_0^{\infty}\frac{1}{(b_0^2+v)(1+\frac{e^{2s}}{v})}\frac{dv}{2v} = \frac{2b_0e^s}{\pi^2}\int\limits_0^{\infty}\frac{1}{v(b_0^2+v)(1+\frac{e^{2s}}{v})}dv = \frac{2b_0e^s}{\pi^2}\int\limits_0^{\infty}\frac{1}{(b_0^2+v)(v+e^{2s})}dv = \frac{2b_0e^s}{\pi^2}\int\limits_0^{\infty}\frac{1}{e^{2s}-b_0^2}(\frac{1}{b_0^2+v}-\frac{1}{v+e^{2s}})dv = \frac{2b_0e^s}{\pi^2(e^{2s}-b_0^2)}[\int\limits_0^{\infty}\frac{1}{b_0^2+v}dv - \int\limits_0^{\infty}\frac{1}{v+e^{2s}}dv] = \frac{2b_0e^{2s}}{\pi^2(e^{2s}-b_0^2)}[log(b_0^2+v)-log(e^{2s}+v)]_0^{\infty} = \frac{2b_0e^s}{\pi^2(e^{2s}-b_0^2)}\lim_{A\to\infty}[log(b_0^2+A)-ln(e^{2s}+A)-ln(b_0^2)+ln(e^{2s})] = \frac{2b_0e^s}{\pi^2(e^{2s}-b_0^2)}[log(\frac{b_0^2+A}{e^{2s}+A})+log(e^{2s})-log(b_0^2)] = \frac{2b_0e^s}{\pi^2(e^{2s}-b_0^2)}[log(e^{2s})-log(b_0^2)]$. Let $b_0 = \frac{1}{\sqrt{n}}$ which means $log(b_0^2) = log(\frac{1}{n}) = -log(n)$. This implies $f(s) = \frac{2e^s}{\pi^2\sqrt{n}(e^{2s}-\frac{1}{n})}[2s+log(n)] = \frac{2\sqrt{n}e^s}{\pi^2(Te^{2s}-1)}[2s+log(n)]$.Recall, $\mu = 2log(\tau_0\tau_1) = 2s$ which implies $s = \mu_2$. Therefore, $f(\mu) = \frac{\sqrt{n}e^{\frac{\mu}{2}}}{\pi^2(Te^{\mu}-1)}[\mu+log(n)]$

Now, we can bound the tail probability of the mean parameters.$\mathbb{P}(|\mu| > log(n^4)) = \mathbb{P}(\mu > log(n^4)) + \mathbb{P}(\mu < -log(n^4)) = \mathbb{P}(log(\tau_0^2\tau_1^2) > log(n^4)) + \mathbb{P}(log(\tau_0^2\tau_1^4) < log(\frac{1}{n^4})) = \mathbb{P}(\tau_0^2\tau_1^2 > n^4) + \mathbb{P}(\tau_0^2\tau_1^4 < \frac{1}{n^4})$. Observe, if $\tau_0^2\tau_1^2 > n^4$ then $\tau_0^2 > n^2$ or $\tau_1^2 > n^2$. Similarly, $\tau_0^2\tau_1^2 < \frac{1}{n^4}$ means either $\tau_0^2 < n^{-4}$ or $\tau_1^2 < n^{-4}$. Therefore,$\mathbb{P}(\tau_0^2\tau_1^2 > n^4) = \mathbb{P}(\tau_0^2 > n^2) + \mathbb{P}(\tau_1^2 > n^2) = \mathbb{P}(\tau_0 > n) + \mathbb{P}(\tau_1 > n)$. For $\tau \sim C^+(0,1), \mathbb{P}(\tau > t) = \int\limits_t^{\infty}\frac{2}{\pi b}\frac{1}{1+\frac{u^2}{b^2}}du = \frac{2}{\pi b}\int\limits_t^{\infty}\frac{1}{1+\frac{u^2}{b^2}}du = \frac{2}{\pi b}\int\limits_t^{\infty}\frac{b^2}{b^2+u^2}du = \frac{2b}{\pi}\int\limits_t^{\infty}\frac{1}{b^2+u^2}du \le \frac{2b}{\pi}\int\limits_t^{\infty}\frac{1}{u^2}du = \frac{2b}{\pi}[-\frac{1}{u}]_t^{\infty} = \lim_{A\to\infty}\frac{2b}{\pi}[-\frac{1}{a}+\frac{1}{t}] = \frac{2b}{\pi t}$. Similarly, $\mathbb{P}(\tau > t) = \int\limits_0^t \frac{2}{\pi b}\frac{1}{1+(\frac{u}{b})^2}du = \frac{2}{\pi b}\int\limits_0^t\frac{1}{1+(\frac{u}{b})^2}$. Note that for $u \ge 0, 1+(\frac{u}{b})^2 \ge 1$ implies $1 \ge \frac{1}{1+(\frac{u}{b})^2}$. This implies $\mathbb{P}(\tau < t) \le \frac{2}{\pi b}\int\limits_0^t 1du = \frac{2}{\pi b}t$. Therefore, $\mathbb{P}(\tau_0^2\tau_1^2 > n^4) \le \frac{2}{\pi n^{\frac{3}{2}}} + \frac{2}{\pi n} = \frac{2}{\pi}(n^{-\frac{3}{2}}+n^{-1}) = \frac{2}{\pi}(n^{-\frac{1}{2}}+1)n^{-1}$. Similarly, $\mathbb{P}(\tau_0^2\tau_1^2 < \frac{1}{n^2}) \le \frac{2}{\pi}(n^{\frac{1}{2}}+1)n^{-1}$. This implies $\mathbb{P}(|\mu| > log(n)) \le \frac{2}{\pi}(n^{-\frac{1}{2}}+n^{\frac{1}{2}}+2)n^{-1} = \frac{2}{\pi}(n^{-\frac{3}{2}}+n^{-\frac{1}{2}}+2n^{-1}) \le \frac{8}{\pi}n^{-\frac{1}{2}}$.By the union bound we have $\mathbb{P}(\max_{1\le i\le N}|\mu_{\alpha_i}| > log(n^4)) \le \frac{8N}{\pi}n^{-\frac{1}{2}} \to 0$

For the other mean parameters in the model we have $\mu \sim N(0,10)$ or $\mu \sim N(0,100)$. We immediately know from the normal tail bound that $\mathbb{P}(|\mu| > t) \ge 2exp\{-\frac{t^2}{2\sigma^2}\}$.

Therefore, $\mathbb{P}(|\mu| > log(n)) \leq 2exp\{-\frac{(log(n))^2}{2\sigma^2}\} = 2n^{-\frac{log(n)}{2\sigma^2}}$. For $n \geq e$ we know $log(n) \geq 1 \iff log(n) \geq \frac{2\sigma^2}{2\sigma^2} \iff \frac{log(n)}{2\sigma^2} \geq \frac{1}{2\sigma^2} \iff -\frac{log(n)}{2\sigma^2} \leq -\frac{1}{2\sigma^2} \iff n^{-\frac{log(n)}{2\sigma^2}} \leq n^{-\frac{1}{2\sigma^2}}$. Therefore, $\mathbb{P}(|\mu| > log(n)) \leq n^{-\frac{1}{2\sigma^2}}$, where $\sigma^2 = 10$ or 100. Then, by the union bound we have $\mathbb{P}(|\mu_M| > log(n)) \leq n^{-\frac{1}{2\sigma^2}} \to 0, \mathbb{P}(\max_{1\leq k\leq r} |\tilde{\mu}_k| > log(n)) \leq rn^{-\frac{1}{2\sigma^2}} \to 0$,and $\mathbb{P}(\max_{1\leq i\leq N} |\bar{\mu}_i| < log(n)) \leq Nn^{-\frac{1}{2\sigma^2}} \to 0$

For all the persistence parameters, ϕ, they have prior distributions of the form $\frac{\phi+1}{2} \sim Beta(a,b)$ where $a > b$. Therefore, $\mathbb{P}(|\phi| > \sqrt{1-\frac{1}{log(log(n))}}) = \mathbb{P}(\phi > \sqrt{1-\frac{1}{log(log(n))}}) + \mathbb{P}(\phi < -\sqrt{1-\frac{1}{log(log(n))}}) = \mathbb{P}(\phi+1 > \sqrt{1-\frac{1}{log(log(n))}}+1) + \mathbb{P}(\phi+1 < -\sqrt{1-\frac{1}{log(log(n))}}+1) = \mathbb{P}(\frac{\phi+1}{2} > \frac{\sqrt{1-\frac{1}{log(log(n))}}+1}{2}) + \mathbb{P}(\frac{\phi+1}{2} < \frac{-\sqrt{1-\frac{1}{log(log(n))}}+1}{2}) = \mathbb{P}(X > 1-\epsilon) + \mathbb{P}(X < \epsilon)$, where $\epsilon = \frac{1-\sqrt{1-\frac{1}{log(log(n))}}}{2}$ and $X \sim Beta(a,b)$. Then, $\mathbb{P}(X < \epsilon) = \frac{1}{B(a,b)}\int_0^\epsilon x^{a-1}(1-x)^{b-1}dx$. For $x \in [0,1]$, $1-x \leq 1$ which implies $(1-x)^{b-1} \leq 1^{b-1} = 1$. Therefore, $\mathbb{P}(X \leq \epsilon) \leq \frac{1}{B(a,b)}\int_0^\epsilon x^{a-1}dx = \frac{1}{B(a,b)}[\frac{x^a}{a}]_0^\epsilon = \frac{1}{B(a,b)}[\frac{\epsilon^a}{a} - 0] = \frac{\epsilon^a}{aB(a,b)}$. Similarly, $\mathbb{P}(X > 1-\epsilon) = \frac{1}{B(a,b)}\int_{1-\epsilon}^1 x^{a-1}(1-x)^{b-1}dx$. For $x \in [1-\epsilon, 1], x \leq 1 \implies x^{a-1} \leq 1^{a-1} = 1$. Therefore, $\mathbb{P}(X > 1-\epsilon) \leq \frac{1}{B(a,b)}\int_{1-\epsilon}^1 (1-x)^{b-1}dx = \frac{1}{B(a,b)}[-\frac{(1-x)^b}{b}]_{1-\epsilon}^1 = \frac{\epsilon^b}{bB(a,b)}$. This implies $\mathbb{P}(|\phi| > \sqrt{1-\frac{1}{log(log(n))}}) \leq \frac{1}{aB(a,b)}(\frac{1-\sqrt{1-\frac{1}{log(log(n))}}}{2})^a + \frac{1}{bB(a,b)}(\frac{1-\sqrt{1-\frac{1}{log(log(n))}}}{2})^b \leq \frac{2}{bB(a,b)}(\frac{1-\sqrt{1-\frac{1}{log(log(n))}}}{2})^a$,with the upper bound resulting from the fact that $a > b$ for all of the persistence parameter priors in our model. This implies, $\mathbb{P}(\max_{1\leq k\leq r} |\tilde{\phi}_k| \geq \sqrt{1-\frac{1}{log(log(n))}}) \leq \frac{2r}{bB(a,b)}(\frac{1-\sqrt{1-\frac{1}{log(log(n))}}}{2})^a \to 0$ and $\mathbb{P}(\max_{1\leq i\leq N} |\bar{\phi}_i| \geq \sqrt{1-\frac{1}{log(log(n))}}) \leq \frac{2N}{bB(a,b)}(\frac{1-\sqrt{1-\frac{1}{log(log(n))}}}{2})^a$

Now, the variance parameters in the model have $Ga(\frac{1}{2},\frac{1}{2})$ distributions. Then, $\mathbb{P}(\sigma^2 > t) = \int_t^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}dx \leq \frac{\lambda^\alpha}{\Gamma(\alpha)}t^{\alpha-1}\int_t^\infty e^{-\lambda x}dx$ for $a < 1$, since $x \geq t$ implies $1 \geq \frac{t}{x}$ which means $\frac{1}{t} \geq \frac{1}{x}$ which implies $t^{a-1} \geq x^{a-1}$ for a<1. Thus, $\mathbb{P}(\sigma^2 > t) = \frac{\lambda^\alpha}{\Gamma(\alpha)}t^{\alpha-1}[-\frac{1}{\lambda}e^{-\lambda x}]_t^\infty = \lim_{A\to\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)}t^{\alpha-1}[-\frac{1}{\lambda}e^{-\lambda A} + \frac{1}{\lambda}e^{-\lambda t}] = \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}t^{\alpha-1}e^{-\lambda t}$. Therefore, in our prior specification $\mathbb{P}(\sigma^2 > log(n)) \leq \sqrt{\frac{2}{\pi}}(log(n))^{-\frac{1}{2}}e^{-\frac{1}{2}log(n)} = \sqrt{\frac{2}{\pi}}(log(n))^{-\frac{1}{2}}n^{-\frac{1}{2}} \leq \sqrt{\frac{2}{\pi}}n^{-\frac{1}{2}}$, for $e \leq n$. This implies $\mathbb{P}(\max_{1\leq k\leq r} \tilde{\sigma}_k^2 > log(n)^{0.5}) \leq r\sqrt{\frac{2}{\pi}}n^{-\frac{1}{2}} \to 0$ and

$\mathbb{P}(\max_{1\leq i\leq N} \bar{\sigma}_i^2 \geq log(n)^{0.5}) \leq N\sqrt{\frac{2}{\pi}}n^{-\frac{1}{2}} \rightarrow 0.$

Now bound the Frobenius norm of Λ. Firstly, $\mathbb{P}(|\Lambda_{ij}| > x|\lambda) = \mathbb{E}[\mathbf{1}\{|\Lambda_{ij} > x|\}|\lambda] = \mathbb{E}[\mathbb{E}[\mathbf{1}\{|\Lambda_{ij}| > x|\tau,\lambda\}]|\lambda]$, by the law of total expectation. This then gives $\mathbb{P}(|\Lambda_{ij}| > x|\lambda) = \mathbb{E}[\mathbb{P}(|\Lambda_{ij}| > x|\tau,\lambda)|\lambda] = \mathbb{E}[\mathbb{P}(|\Lambda_{ij}| > x|\tau)|\lambda] \leq \mathbb{E}[exp\{2exp\{-\frac{x^2}{2\tau^2}\}|\lambda] = \mathbb{E}[exp\{2exp\{-\frac{x^2}{2\tau^2}\}\mathbf{1}\{\tau^2 \leq L\}|\lambda]+\mathbb{E}[exp\{2exp\{-\frac{x^2}{2\tau^2}\}\mathbf{1}\{\tau^2 > L\}|\lambda] \leq 2exp-\frac{x^2}{2L}\mathbb{E}[\mathbf{1}\{\tau^2 \leq L\}|\lambda]+\mathbb{E}[exp\{2exp\{-\frac{x^2}{2\tau^2}\}\mathbf{1}\{\tau^2 > L\}|\lambda] = 2exp\{-\frac{x^2}{2L}\}\mathbb{P}(\tau^2 \leq L|\Lambda)+\mathbb{E}[exp\{2exp\{-\frac{x^2}{2\tau^2}\}\mathbf{1}\{\tau^2 > L\}|\lambda] \leq 2exp-\frac{x^2}{2L}+\mathbb{E}[exp\{2exp\{-\frac{x^2}{2\tau^2}\}\mathbf{1}\{\tau^2 > L\}|\lambda] \leq 2exp\{-\frac{x^2}{2L}\}+E[2\mathbf{1}\{\tau^2 > L\}|\lambda] = 2exp-\frac{x^2}{2L} + 2\mathbb{P}(\tau^2 > L|\lambda) = 2exp-\frac{x^2}{2L} + 2\mathbb{P}(e^{\gamma\tau^2} > e^{\gamma L}) \leq 2exp\{-\frac{x^2}{2L} + 2\mathbb{E}[e^{\gamma\tau^2}]e^{-\gamma L}$, by Markov's inequality. Then, $\mathbb{P}(|\Lambda_{ij}| > x|\lambda) \leq 2exp\{-\frac{x^2}{2L}\} + 2(1-\frac{\gamma}{\beta})^{-\alpha}e^{-\gamma L}$, using the formula for the moment generating function of a $Gamma(\alpha,\beta)$ distributed random variable. Letting $\gamma = \frac{1}{2}\beta$ gives $\mathbb{P}(|\Lambda_{ij}| > x|\lambda) \leq 2exp\{-\frac{x^2}{2L}\} + 2(\frac{1}{2})^{-\alpha}e^{-\gamma L} = 2exp\{-\frac{x^2}{2L}\} + 22^{\alpha}e^{-\gamma L} \leq 2exp\{-\frac{x^2}{2L}\} + 4e^{-0.025L\lambda^2}$, since $\alpha < 1$. Since we are doing computation on a computer there will be a smallest possible representable number and by letting L = x gives $\mathbb{P}(|\Lambda_{ij}| > x|\lambda \geq \lambda_0) \leq 2exp\{-\frac{x}{2}\} + 4exp\{-0.025\lambda_0^2 x\}$. Let $C_0 = min\{\frac{1}{2}, 0.025\lambda_0^2\}$, then $\mathbb{P}(|\Lambda_{ij}| > x) = \mathbb{E}[\mathbf{1}\{|\Lambda_{ij}| > x\}] = \mathbb{E}[\mathbb{P}(|\Lambda_{ij}| > x)|\lambda] \leq 6exp\{-C_0 x\}$, by using the law of total expectation. Now, $||\Lambda||_F^2 = \sum_{i,j}|\Lambda_{i,j}|^2 \leq \sum_{i,j}\max_{i,j}|\Lambda_{i,j}|^2 = Nr\max_{i,j}|\Lambda_{i,j}|^2$ which implies $||\Lambda||_F \leq \sqrt{Nr}\max_{i,j}|\Lambda_{i,j}|$. Therefore, $\mathbb{P}(||\Lambda||_F > x) \leq \mathbb{P}(\sqrt{Nr}\max_{i,j}|\Lambda_{ij}| > x) = \mathbb{P}(\max_{i,j}|\Lambda_{i,j}| > \frac{x}{\sqrt{Nr}}) = \mathbb{P}(\bigcup_{i,j}|\Lambda_{i,j}| > \frac{x}{\sqrt{Nr}}) \leq \sum_{i,j}\mathbb{P}(|\Lambda_{ij}| > \frac{x}{\sqrt{Nr}}) \leq \sum_{i,j}6exp\{-C_0(\frac{x}{\sqrt{Nr}})\} = 6Nrexp\{-\frac{C_0 x}{\sqrt{Nr}}\} = C1exp\{-C_2 x\}$, where $C_1 = 6Nr$, and $C_2 = \frac{C_0}{\sqrt{Nr}}$. This implies that $\mathbb{P}(||\Lambda||_F > log(n)) \leq C_1 n^{-C_2} \rightarrow 0$. Therefore, by the above bounds we have $\mathbb{P}(\Theta_n^c) \rightarrow 0$.

To establish the metric entropy bound we will utilize the bounds on the parameters within the sieve to upper bound the partial derivatives of the covariance matrix with respect to each parameter. From this and by use of the extreme value theorem and mean value theorem we can upper bound the Frobenius norm of the difference between two different correlation matrices.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial h_{\alpha_{i,t-1}}} = \delta_{ij}exp\{h_{\alpha_{i,t-1}}\} \leq \delta_{ij}exp\{|\mu_{\alpha_i}| + \frac{clog(n)+log(2k(s))}{s}\}$. On Θ_n, $\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial h_{\alpha_{i,t-1}}} \leq \delta_{ij}exp\{log(n^4) + \frac{clog(n)+log(2k(s))}{s}\} = \delta_{ij}n^{4+\frac{c}{s}}C_1 < \infty$ for fixed n, where $C_1 = [2k(s)]^{\frac{2}{s}}$.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial h_{M,t-1}} = \beta_{i,t-1}\beta_{j,t-1}exp\{\mu_M - \phi_M\mu_M + \frac{1}{2}\sigma_M^2\}\frac{\partial}{\partial h_{M,t-1}}exp\{\phi_M h_{M,t-1}\} = \beta_{i,t-1}\beta_{j,t-1}exp\{\mu_M - \phi_M\mu_M + \frac{1}{2}\sigma_M^2\}\phi_M exp\{\phi_M h_{M,t-1}\} \leq (\kappa+1)C_0 n^\alpha log(n)exp\{log(n)\}exp\{\sqrt{1-\frac{1}{log(log(n))}}log(n)\}$ $exp\{\frac{1}{2}log(n)^{0.5}\}\sqrt{1-\frac{1}{log(log(n))}}exp\{\sqrt{1-\frac{1}{log(log(n))}}(log(n)+plog(n))\} \leq (\kappa+1)C_0 n^\alpha log(n)log(n)$ $exp\{log(n)\}exp\{log(n^{\frac{1}{2}})\}exp\{log(n)+plog(n)\}$ for $n \geq e$. This gives, $\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial h_{M,t-1}} \leq (\kappa+1)C_0 log(n)^2 n^{\alpha+\frac{5}{2}+p}$.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \mu_M} = (1-\phi_M)\beta_{i,t-1}\beta_{j,t-1}exp\{\phi_M h_{M,t-1} + \frac{1}{2}\sigma_M^2\}exp\{\mu_m - \phi\mu_m\} \leq (1-\sqrt{1-\frac{1}{log(log(n))}})\beta_{i,t-1}\beta_{j,t-1}exp\{\sqrt{1-\frac{1}{log(log(n))}}(log(n^4)+log(n^{\frac{c}{s}})+log([2k(s)])) + \frac{1}{2}log(n)^{0.5}\}exp\{log(n) - \sqrt{1-\frac{1}{log(log(n))}}log(n)\} \leq \beta_{i,t-1}\beta_{j,t-1}exp\{log(n^4)+log(n^{\frac{c}{s}}) + log([2k(s)]^{\frac{1}{s}}) + \frac{1}{2}log(n)^{0.5}\}$, for $n \geq e^e$. Therefore, $\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \mu_M} \leq \beta_{i,t-1}\beta_{j,t-1}n^4 n^{\frac{c}{s}}C_0 n^{\frac{1}{2}} = C_0\beta_{i,t-1}\beta_{j,t-1}n^{\frac{9}{2}+\frac{c}{s}}$. Now we need to bound β_t. $\mathbb{P}(\max_{1\leq a\leq N, 1\leq t\leq n}|\beta_{a,t}| > M) = \mathbb{P}(\{\max_{1\leq a\leq N, 1\leq t\leq n}|\beta_{a,t}| > M\}\cap\Theta_n) + \mathbb{P}(\{\max_{1\leq a\leq N, 1\leq t\leq n}|\beta_{a,t}| > M\}\cap\Theta_n^c) \leq \mathbb{P}(\{\max_{1\leq a\leq N, 1\leq t\leq n}|\beta_{a,t}| > M\}\cap\Theta_n) + \mathbb{P}(\theta_n^c) \leq 2nNnexp\{-\frac{M^2}{2nexp\{|mu|+\frac{clog(n)+log(2k(s))}{s}\}}\} + \mathbb{P}(\theta_n^c) \leq 2nNexp\{-\frac{M^2}{2nexp\{log(n)+\frac{clog(n)+log(2k(s))}{s}\}}\} + \mathbb{P}(\theta_n^c) = 2nexp\{-\frac{M^2}{C_0 n^{5+\frac{c}{s}}}\} + \mathbb{P}(\theta_n^c)$, where $C_0 = 2[2k(s)]^{\frac{1}{s}}$.Let $\alpha = \frac{c}{s}+5$. Then, $\mathbb{P}(\max_{1\leq a\leq N, 1\leq t\leq n}|\beta_{a,t}| > M) \leq 2nNexp\{-\frac{M^2}{C_0 n^\alpha}\} + \mathbb{P}(\Theta_n^c)$. Suppose $\frac{M^2}{C_0 n^\alpha} = (\kappa+1)log(n)$ for some $\kappa > 0$. Then, $M = \sqrt{(\kappa+1)C_0 n^\alpha log(n)} = C_1\sqrt{n^\alpha log(n)}$, where $C_1 := \sqrt{(\kappa+1)C_0}$. That is, $M = C_1 n^{\frac{\alpha}{2}}log(n)^{\frac{1}{2}}$. This implies, $\mathbb{P}(\max_{1\leq a\leq N, 1\leq t\leq n}|\beta_{a,t}| > C_1 n^{\frac{1}{2}(\frac{c}{s}+5)}log(n)^{\frac{1}{2}}) \leq 2nNexp\{-\frac{(\kappa+1)C_0 n^\alpha log(n)}{C_0 n^\alpha}\} + \mathbb{P}(\Theta_n^c) = 2nNexp\{log(n^{-(\kappa+1)})\} + \mathbb{P}(\Theta_n^c) = 2Nnn^{-\kappa-1} + \mathbb{P}(\Theta_n^c) = 2Nn^{-\kappa} + \mathbb{P}(\Theta_n^c) \to 0$. This implies, $\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \mu_M} \leq M^2 n^{\frac{1}{n}+\frac{1}{2}+\frac{c}{s}}[2k(s)]^{\frac{1}{s}} = C_2 n^{\frac{9}{2}+\frac{c}{s}+\frac{\alpha}{2}}log(n)^{\frac{1}{2}}$.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \phi_M} = \beta_i\beta_j exp\{\mu_M + \frac{1}{2}\sigma_M^2\}(h_M - \mu_M)exp\{\phi_M(h_M-\mu_M)\} \leq M^2 exp\{log(n) + \frac{1}{2}log(n)^{0.5}\}(|\bar{\mu}_m| + plog(n) + |\bar{\mu}_m|)exp\{\sqrt{1-\frac{1}{log(log(n))}}(2log(n)+plog(n))\} \leq M^2 exp\{log(n) + \frac{1}{2}log(n)\}(p+2)log(n)exp\{plog(n)+2log(n)\} = M^2 n^{\frac{3}{2}}(p+2)log(n)n^{p+2} = M^2 n^{\frac{3}{2}+p+2}(p+2)log(n) = C_2 n^{\frac{3}{2}+p+\alpha}log(n)^2$.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \sigma_M^2} = \beta_i\beta_j exp\{\mu_M + \phi_m(h_M-\mu_M)\}exp\{\sigma_M^2\} \leq M^2 exp\{log(n) + \sqrt{1-\frac{1}{log(log(n))}}(2+p)log(n)\}exp\{log(n)\} \leq M^2 exp\{log(n) + (2+p)log(n)\}exp\{log(n)\} = M^2 n^{3+p}n = M^2 n^{4+p} = C_2 n^{4+p+\alpha}log(n)$.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \beta_i} = \delta_{ij}\beta_j exp\{\mu_M + \phi_M(h_{M,t-1}-\mu_M) + \frac{1}{2}\sigma_M^2\} \leq Mexp\{log(n) + \sqrt{1-\frac{1}{log(log(n))}}(plog(n) + 2log(n)) + \frac{1}{2}log(n)\} \leq Mexp\{log(n) + log(n^{2+p}) + log(n^{\frac{1}{2}})\} = Mn^{\frac{7}{2}+p} = C_1 n^{\frac{7}{2}+p+\frac{\alpha}{2}}log(n)^{\frac{1}{2}}$.Similarly,

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \beta_j} \leq C_1 n^{\frac{7}{2}+p+\frac{\alpha}{2}} log(n)^{\frac{1}{2}}$

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \Lambda_{ik}} = \sum_{k=1}^{r} exp\{\tilde{\mu}_k + \tilde{\phi}_k(\tilde{h}_k - \tilde{\mu}_k) + \frac{1}{2}\tilde{\sigma}_k^2\}\Lambda_{jk} = \sum_{k=1}^{r} exp\{\tilde{\mu}_k\}exp\{\tilde{\phi}_k(\tilde{h}_k - \tilde{\mu}_k)\}exp\{\frac{1}{2}\tilde{\sigma}_k^2\}\Lambda_{jk} \leq \sum_{k=1}^{r} exp\{log(n)\}exp\{\sqrt{1-\frac{1}{log(log(n))}}(p+2)log(n)\}exp\{\frac{1}{2}log(n)^{0.5}\}log(n) \leq rnn^{2+p}n^{\frac{1}{2}}log(n) = rn^{\frac{7}{2}+p}log(n)$.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \tilde{h}_{k,t-1}} = \sum_{k=1}^{r} \Lambda_{ik}\Lambda_{jk}exp\{\tilde{\mu}_k\}exp\{-\tilde{\phi}_k\tilde{\mu}_k\}exp\{\frac{1}{2}\tilde{\sigma}_k^2\}\frac{\partial}{\partial \tilde{h}_{k,t-1}}exp\{\tilde{\phi}_k\tilde{h}_{k,t-1}\}$
$= \sum_{k=1}^{r} \Lambda_{ik}\Lambda_{jk}exp\{\tilde{\mu}_k\}exp\{-\tilde{\phi}_k\tilde{\mu}_k\}exp\{\frac{1}{2}\tilde{\sigma}_k^2\}\tilde{\phi}_k exp\{\tilde{\phi}_k\tilde{h}_{k,t-1}\}$
$\leq \sum_{k=1}^{r} log(n)^2 exp\{log(n)\}exp\{\sqrt{1-\frac{1}{log(log(n))}}log(n)\}exp\{\frac{1}{2}log(n)^{0.5}\}\sqrt{1-\frac{1}{log(log(n))}}$
$exp\{\sqrt{1-\frac{1}{log(log(n))}}(log(n)+plog(n))\} \leq rlog(n)^2 n^{\frac{7}{2}+p}$ for $n \geq e^e$.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \tilde{\mu}_k} = \sum_{k=1}^{r} \Lambda_{ik}\Lambda_{jk}exp\{\tilde{\phi}_k\tilde{h}_{k,t-1}\}exp\{\frac{1}{2}\tilde{\sigma}_k^2\}(1-\tilde{\phi}_k)exp\{(1-\tilde{\phi}_k)\tilde{\mu}_k\}$
$\leq \sum_{k=1}^{r} log(n)^2 exp\{\sqrt{1-\frac{1}{log(log(n))}}(plog(n+log(n)))\}exp\{\frac{1}{2}log(n)^{0.5}\}(1-\sqrt{1-\frac{1}{log(log(n))}})exp\{(1-\sqrt{1-\frac{1}{log(log(n))}})log(n)\} \leq rlog(n)^2 exp\{log(n^{p+1})\}exp\{log(n^{\frac{1}{2}})\}exp\{log(n)\} = rlog(n)^2 n^{p+1}n^{\frac{1}{2}}n = rlog(n)^2 n^{p+\frac{5}{2}}$.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\tilde{\phi}_k} = \sum_{k=1}^{r} \Lambda_{ik}\Lambda_{jk}exp\{\tilde{\mu}_k + \frac{1}{2}\tilde{\sigma}_k^2\}(\tilde{h}_{k,t-1} - \tilde{\mu}_k)exp\{\tilde{\phi}_k(\tilde{h}_{k,t-1} - \tilde{\mu}_k)\} \leq \sum_{k=1}^{r}(log(n))^2 exp\{log(n) + \frac{1}{2}log(n)^{0.5}\}(plog(n)+2log(n))exp\{\sqrt{1-\frac{1}{log(log(n))}}(plog(n)+2log(n))\} \leq rlog(n)^2 exp\{log(n^{\frac{3}{2}})\}(p+2)log(n)exp\{log(n^{2+p})\} = rlog(n)^2 n^{\frac{3}{2}}(p+2)log(n)n^{2+p} = rlog(n)^3(p+2)n^{\frac{7}{2}+p}$.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \tilde{\sigma}_k^2} = \sum_{k=1}^{r} \Lambda_{ik}\Lambda_{jk}exp\{\tilde{\mu}_k + \tilde{\phi}(\tilde{h}_{k,t-1} - \tilde{\mu}_k)\}\frac{1}{2}exp\{\frac{1}{2}\tilde{\sigma}_k^2\} \leq \frac{1}{2}\sum_{k=1}^{r} log(n)^2 exp\{log(n) + \sqrt{1-\frac{1}{log(log(n))}}(plog(n)+2log(n))\}exp\{\frac{1}{2}log(n)^{0.5}\} \leq \frac{r}{2}log(n)^2 exp\{log(n^{3+p})\}exp\{log(n^{\frac{1}{2}})\} = \frac{r}{2}log(n)^2 n^{\frac{7}{2}+p}$.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \bar{h}_{i,t-1}} = \delta_{ij}exp\{\bar{\mu}_i\}exp\{-\bar{\phi}_i\bar{\mu}_i\}exp\{\frac{1}{2}\bar{\sigma}_i^2\}\frac{\partial}{\partial \bar{h}_{i,t-1}}exp\{\bar{\phi}_i\bar{h}_{i,t-1}\}$
$= \delta_{ij}exp\{\bar{\mu}_i\}exp\{\bar{\phi}_i\bar{\mu}_i\}exp\{\frac{1}{2}\bar{\sigma}_i^2\}\bar{\phi}_i exp\{\bar{\phi}_i\bar{h}_{i,t-1}\} \leq exp\{log(n)\}exp\{\sqrt{1-\frac{1}{log(log(n))}}log(n)\}$
$exp\{\frac{1}{2}log(n)^{0.5}\}\sqrt{1-\frac{1}{log(log(n))}}exp\{\sqrt{1-\frac{1}{log(log(n))}}(log(n)+plog(n))\} \leq n^{\frac{7}{2}+p}$ for $n \geq e$

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \bar{\mu}_i} = \delta_{ij}exp\{\bar{\phi}_i\bar{h}_{i,t-1} + \frac{1}{2}\bar{\sigma}_i^2\}(1-\bar{\phi}_i)exp\{\bar{\mu}_i(1-\bar{\phi}_i)\} \leq exp\{\sqrt{1-\frac{1}{log(log(n))}}(log(n)+plog(n)) + \frac{1}{2}log(n)^{0.5}\}(1-\sqrt{1-\frac{1}{log(log(n))}})exp\{log(n)(1-\sqrt{1-\frac{1}{log(log(n))}})\} \leq exp\{log(n^{1+p})\}exp\{log(n^{\frac{1}{2}})\}exp\{log(n)\} = n^{\frac{5}{2}+p}$

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \bar{\phi}_i} = \delta_{ij}exp\{\bar{\mu}_i + \frac{1}{2}\bar{\sigma}_i^2\}(\bar{h}_{i,t-1} - \bar{\mu}_i)exp\{\bar{\phi}_i(\bar{h}_{i,t-1} - \bar{\mu}_i)\} \leq exp\{log(n) + \frac{1}{2}log(n)^{0.5}\}(plog(n) + 2log(n))exp\{\sqrt{1-\frac{1}{log(log(n))}}(plog(n)+2log(n))\} \leq exp\{log(n^{\frac{3}{2}})(2+p)log(n)exp\{log(n^{p+2})\} =$

$(2+p)log(n)n^{\frac{7}{2}}$.

$\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \bar{\sigma}_i^2} = \delta_{ij} exp\{\bar{\mu}_i + \bar{\phi}_i(\bar{h}_{i,t-1} - \bar{\mu}_i)\}\frac{1}{2}exp\{\frac{1}{2}\bar{\sigma}_i^2\} \leq exp\{log(n) + \sqrt{1 - \frac{1}{log(log(n))}}(plog(n) + 2log(n))\}\frac{1}{2}exp\{\frac{1}{2}log(n)^{0.5}\} \leq \frac{1}{2}exp\{log(n^{3+p})\}exp\{log(n^{\frac{1}{2}})\} = \frac{1}{2}n^{\frac{7}{2}+p}$

Recall that $R(\mathbf{r}_t)_{ij} = cov(\mathbf{r_t})_{ij} cov(\mathbf{r_t})_{ii}^{-\frac{1}{2}} cov(\mathbf{r_t})_{jj}^{-\frac{1}{2}}$. Therefore for some parameter θ, $\frac{\partial R(\mathbf{r_t})_{ij}}{\partial \theta} = cov(\mathbf{r_t})_{ii}^{-\frac{1}{2}} cov(\mathbf{r_t})_{jj}^{-\frac{1}{2}} \frac{\partial}{\partial \theta} cov(\mathbf{r_t})_{ij} + cov(\mathbf{r_t})_{ij}[cov(\mathbf{r_t})_{jj}^{-\frac{1}{2}}(-\frac{1}{2}cov(\mathbf{r_t})_{ii}^{-\frac{3}{2}}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{ii}) + cov(\mathbf{r_t})_{ii}^{-\frac{1}{2}}(-\frac{1}{2}cov(\mathbf{r_t})_{ij}^{-\frac{3}{2}}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{jj})] = (cov(\mathbf{r_t})_{ii}cov(\mathbf{r_t})_{jj})^{-\frac{1}{2}}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{ij}$
$-\frac{1}{2}cov(\mathbf{r_t})_{ij}[cov(\mathbf{r_t})_{jj}^{-\frac{1}{2}}cov(\mathbf{r_t})_{jj}^{-\frac{3}{2}}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{ii} + cov(\mathbf{r_t})_{ii}^{-\frac{1}{2}}cov(\mathbf{r_t})_{jj}^{-\frac{3}{2}}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{ij}]$
$= (cov(\mathbf{r_t})_{ii}cov(\mathbf{r_t})_{jj})^{-\frac{1}{2}}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{ij} - \frac{1}{2}cov(\mathbf{r_t})_{ij}[(cov(\mathbf{r_t})_{ii}cov(\mathbf{r_t})_{jj})^{-\frac{1}{2}}(cov(\mathbf{r_t})_{ii}^{-1}\frac{\partial}{\partial \theta cov(\mathbf{r_t})_{ii}} + cov(\mathbf{r_t})_{jj}^{-1}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{jj})] = (cov(\mathbf{r_t})_{ii}cov(\mathbf{r_t})_{jj})^{-\frac{1}{2}}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{jj}$
$-\frac{1}{2}cov(\mathbf{r_t})_{ij}(cov(\mathbf{r_t})_{ii}cov(\mathbf{r_t})_{jj})^{-\frac{1}{2}}(cov(\mathbf{r_t})_{ii}^{-1}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{ii} + cov(\mathbf{r_t})_{jj}^{-1}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{jj}) = \frac{\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{ij}}{\sqrt{cov(\mathbf{r_t})_{ii}cov(\mathbf{r_t})_{jj}}} - \frac{cov(\mathbf{r_t})_{ij}}{2\sqrt{cov(\mathbf{r_t})_{ii}cov(\mathbf{r_t})_{jj}}}(\frac{\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{ii}}{cov(\mathbf{r_t})_{ii}} + \frac{\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{jj}}{cov(\mathbf{r_t})_{jj}}) = \frac{1}{\sqrt{cov(\mathbf{r_t})_{ii}cov(\mathbf{r_t})_{jj}}}(\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{jj} - \frac{1}{2}cov(\mathbf{r_t})_{ij}(\frac{1}{cov(\mathbf{r_t})_{ii}}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{ii} + \frac{1}{cov(\mathbf{r_t})_{jj}}\frac{\partial}{\partial \theta}cov(\mathbf{r_t})_{jj}))$. We will now show that the variance terms are bound from below and above.

Each of our parameters lies in a closed bounded interval. This means $\Theta_n = \prod\limits_{\#\theta} I_\theta$ where each $I_\theta \subset \mathbb{R}$ are closed and bounded intervals. Then by Tychonoff's theorem Θ_n is closed and bounded. We know that the function $\theta \to cov(\mathbf{r_t})_{ii}$ is continuous on $\mathbb{R}^{\#\theta}$. Therefore, it is continuous on Θ_n. Then by the extreme value theorem $cov(\mathbf{r_t})_{ii}$ attains its minimum and maximum value at least once on Θ_n. Let $c_n = \min\limits_{\theta \in \Theta_n} cov(\mathbf{r_t})_{ii}$ and $C_n = \max\limits_{\theta \in \Theta_n} cov(\mathbf{r_t})_{ii}$. Since $cov(\mathbf{r_t})_{ij}$ is strictly positive and Θ_n is bounded away from, ∞ $c_n > 0$ and $C_n < \infty$. Then, let $c = \min\limits_{1 \leq t \leq n, 1 \leq i \leq N} c_n(t,i)$ and $C = \max\limits_{1 \leq t \leq n, 1 \leq i \leq N} C_n$. Then, $0 < c \leq cov(\mathbf{r_t})_{ii} \leq C < \infty$.

Let $L_n = \max\limits_{0 \leq \ell \leq L} B_n^\ell$, that is the maximum of our covariance partial derivative bounds, where $L_n = O(n^\alpha (log(n))^\beta)$ for some constants α and β. Recall, $|cov(\mathbf{r_t})_{ij}| \leq \sqrt{cov(\mathbf{r_t})_{ii}cov(\mathbf{r_t})_{jj}}$. This implies, $|\frac{\partial R(\mathbf{r_t})_{ij}}{\partial \theta}| \leq \frac{L_n}{\sqrt{cc}} + \frac{C}{2\sqrt{cc}}(\frac{L_n}{c} + \frac{L_n}{c}) = \frac{L_n}{c} + \frac{C}{2c}(\frac{2L_n}{c}) = (\frac{1}{c} + \frac{C}{c^2})L_n$. This implies that $|\frac{\partial R(\mathbf{r_t})_{ij}}{\partial \theta}| \leq C' L_n$, where $C' = (\frac{1}{c} + \frac{C}{c^2})$. $||\nabla_\theta R(\mathbf{r_t})||_{OP} \leq ||\nabla_\theta R(\mathbf{r_t})|| = \sqrt{\sum\limits_{\ell=1}^{L}\sum\limits_{i,j=1}^{N}|\frac{\partial R(\mathbf{r_t})_{ij}}{\partial \ell}|^2} \leq \sqrt{LN^2C'^2L_n^2} = \sqrt{L}NC'L_n = C''L_n, C'' = \sqrt{L}NC'$. That is, $||\nabla_\theta R(\mathbf{r_t})||_{OP} \leq C''L_n$. By the mean value theorem, for $\theta, \theta' \in \Theta_n, \exists \epsilon_{t,i,j}$ where $\epsilon_{t,i,j} = \theta' + s_{t,i,j}(\theta - \theta')$ for $s_{t,i,j} \in (0,1)$ lying on the line segment from θ'

to θ such that $|R(\mathbf{r_t})_{ij}(\theta) - R(\mathbf{r_t})_{ij}(\theta^{'})| = |\nabla_\theta R(\mathbf{r_t})_{i,j}(\epsilon_{i,j,t})||(\theta - \theta^{'})|$ which implies $|R(\mathbf{r_t})_{ij}(\theta) - R(\mathbf{r_t})_{ij}(\theta^{'})| = |\nabla_\theta R(\mathbf{r_t})_{ij}(\epsilon_{t,i,j})^T(\theta - \theta^{'})| \leq ||\nabla_\theta R(\mathbf{r_t})_{ij}(\epsilon_{t,i,j})||_2||\theta - \theta^{'}||$, by the Cauchy-Schwarz inequality. Therefore, $|R(\mathbf{r_t})_{ij}(\theta) - R(\mathbf{r_t})_{ij}(\theta^{'})| \leq C^{''}L_n||\theta - \theta^{'}||_2$. Then, $||R(\mathbf{r_t}) - R(\mathbf{r_t})||_F^2 \leq \sum_{i=1}^{N}\sum_{j=1}^{N}|R(\mathbf{r_t})_{ij}(\theta) - R(\mathbf{r_t})_{ij}(\theta^{'})|^2 \leq N^2 C^{''2} L_n^2||\theta - \theta^{'}||_2^2$. This implies, $||R(\mathbf{r_t})(\theta) - R(\mathbf{r_t})(\theta^{'})||_F \leq C_1 L_n||\theta - \theta^{'}||_2$. Now, $||\frac{1}{n}\sum_{t=1}^{n} R(\mathbf{r_t})(\theta) - R(\mathbf{r_t})(\theta^{'})||_F = \frac{1}{n}||\sum_{t=1}^{n} R(\mathbf{r_t})(\theta) - R(\mathbf{r_t})(\theta^{'})||_F \leq \frac{1}{n}\sum_{t=1}^{n}||R(\boldsymbol{r}_t)(\theta) - R(\boldsymbol{r}_t)(\theta^{'})|| \leq C_1 L_n||\theta - \theta^{'}||_2$.

Recall that $\Theta_n \subset [-B_{max}(n), B_{max}(n)]^{\#\theta}$, where $B_{max} = O(n^{\frac{\alpha}{2}}\sqrt{log(n)})$ is the largest upper bound in the definition of Θ_n. Let $B_d(\theta_0, \epsilon) := \{\theta \in \Theta_n : d_n(\theta, \theta_0) < \epsilon\}$. Suppose, $||\theta - \theta_0^{'}|| < \delta$ which implies $d(\theta, \theta^{'}) \leq C_1 L_n||\theta - \theta^{'}||_2 \leq C_1 L_n\delta$. We want $d(\theta, \theta^{'}) < r$ for some radius r. Then, $C_1 L_n\delta < r \iff \delta < \frac{r}{C_1 L_n}$. Let $r = \frac{\epsilon\xi}{2}$ and $\delta < \frac{\frac{\epsilon\xi}{2}}{C_1 L_n}$. Then, $||\theta - \theta^{'}||_2 < \frac{\epsilon\xi}{2C_1 L_n}$ which implies $d(\theta, \theta^{'}) \leq \frac{\epsilon\xi}{2}$. We know that $N(\delta, [-B_{max}, B_{max}]^{\#\theta}, ||\cdot||_2) \leq (1 + \frac{2B_{max}}{\delta})^{\#\theta} \leq (\frac{3B_{max}}{\delta})^{\#\theta}$ which implies $log(N(\delta, [-B_{max}, B_{max}]^{\#\theta}, ||\cdot||2)) \leq dlog(\frac{3B_{max}}{\delta}) = \#\theta log(\frac{3B_{max}}{\delta}) = \#\theta[log(3) - \log(\delta) + log(O(n^{\frac{\alpha}{2}}\sqrt{log(n)}))] = \#\theta(\frac{\alpha}{2}log(n) + \frac{1}{2}log(log(n)) + log(\frac{1}{\delta}) + O(1))$. Since we previously shown that if $||\theta - \theta^{'}|| < \delta$ then $d(\theta, \theta^{'}) < r$ for $\delta < \frac{r}{C_1 L_n}$. That is, $\{\theta : \{||\theta - \theta^{'}||_2 < \delta\}\} \subset \{\theta : d_n(\theta, \theta^{'}) < r\}$ i.e. $B_2(\theta^{'}, \delta) \subset B_d(\theta^{'}, r)$, with $r = \frac{\epsilon\xi}{2}$ and $\delta = \frac{\epsilon\xi}{2C_1 L_n}$. Now, choose $\{\theta_1, ..., \theta_n\} \subset \Theta_n$ such that $\Theta_n \subset \bigcup_{j=1}^{N}\{\theta : ||\theta - \theta_j||_2 < \delta\}$. Now consider θ such that $d(\theta, \theta_0) < \epsilon$. Then by the previous argument θ must be in some $||\theta - \theta_j||_2 < \delta$. Then, $d(\theta, \theta_j) \leq C_1 L_n||\theta - \theta_j||_2 < C_1 L_n\delta = \frac{C_1 L_n\epsilon\xi}{2C_1 L_n} = \frac{\epsilon\xi}{2}$. Then, θ must be in the d_n ball of radius $\frac{\epsilon\xi}{2}$ centered around θ_j. Therefore, $B_d(\theta_0, \epsilon) \subset \bigcup_{j=1}^{N}\{\theta \in \Theta : d(\theta, \theta_j) < \frac{\epsilon\xi}{2}\}$. This implies $N(\frac{\epsilon\xi}{2}, B_d(\theta_0, \epsilon), d_n) \leq N(\delta, \Theta_n, ||\cdot||_2)$. That is, $N(\frac{\epsilon\xi}{2}, \{\theta \in \Theta_n : d(\theta, \theta_j) < \epsilon\}, d_n) \leq N(\delta, \Theta_n, ||\cdot||_2) \leq exp\{log(n^{\frac{\#\theta\alpha}{2}}) + log(log(n)^{\frac{\#\theta}{2}}) + log(\frac{1}{\delta}^{\#\theta}) + O(1)\}$. Recall, $\delta = \frac{\epsilon\xi}{2C_1 L_n}$, which means $log(\frac{1}{\delta}) = ln(\frac{2C_1 L_n}{\epsilon\xi}) = log(2C_1) + log(L_n) + log(\frac{1}{\xi\epsilon}) = O(log(log(n)) + log(\frac{1}{\epsilon\xi}))$. This implies, $log(N(\frac{\epsilon\xi}{2}, \{\theta \in \Theta_n : d(\theta, \theta_j < \epsilon)\}, d_n)) \leq \frac{\#\theta\alpha}{2}log(n) + \frac{\#\theta}{2}log(log(n)) + \#\theta log(L_n) + \#\theta log(\frac{1}{\xi\epsilon}) + O(1)$. Recall, $L_n = O(n^a, (log(n))^b)$. Then, $log(L_n) = log(C_1(n^a[log(n)]^b)) = alog(n) + blog(log(n)) + log(C_1)$. Let $\epsilon_n = \sqrt{\frac{\xi^2(a+b+\frac{1}{2}+\frac{\alpha}{2})\#\theta log(n)}{n}}$. Then, $log(\frac{1}{\xi\epsilon}) = log(\sqrt{\frac{n}{(a+b+\frac{1}{2}+\frac{\alpha}{2}+\#\theta log(n))}}) =$

$\frac{1}{2}log(\frac{n}{(a+\frac{3}{2}+\frac{\alpha}{2})\#\theta log(n)}) = \frac{1}{2}log(n) - \frac{1}{2}log((a+b+\frac{1}{2}+\frac{\alpha}{2})) - \frac{1}{2}log(\#\theta) - \frac{1}{2}log(log(n))$.

Therefore, $log(N(\frac{\epsilon\xi}{2}, \{\theta \in \Theta_n : d(\theta, \theta_j < \epsilon)\}, d_n)) \leq \frac{\#\theta\alpha}{2}log(n) + \frac{\#\theta}{2}log(log(n)) + \#\theta alog(n) + bdlog(n) + b\#\theta log(log(n)) + \frac{\#\theta}{2}log(n) - \frac{\#\theta}{2}log(a+b+\frac{1}{2}+\frac{\alpha}{2}) - \frac{\#\theta}{2}log(\#\theta) - \frac{\#\theta}{2}log(log(n))$. Which implies $log(N(\frac{\epsilon\xi}{2}, \{\theta \in \Theta_n : d(\theta, \theta_j < \epsilon)\}, d_n)) \leq \#\theta log(n)\{\frac{\alpha}{2} + a+b+\frac{1}{2}\} - \frac{\#\theta}{2}log(a+b+\frac{1}{2}+\frac{\alpha}{2}) + \#\theta log(C_1)$, using the fact that $loglog(n) \leq log(n)$ for $n > 1$. Let $B := \frac{\alpha}{2}+a+b+\frac{1}{2}$. Then we need, $dBlog(n) - \frac{\#\theta}{2}log(B) + \#\theta log(C_1) - \frac{\#\theta}{2}log(\#\theta) \leq \#\theta\epsilon_n^2 Blog(n) \iff \#\theta log(C_1) - \frac{\#\theta}{2}log(B) - \frac{\#\theta}{2}log(\#\theta) \leq \#\theta\epsilon_n^2 Blog(n) - \#\theta blog(n) = \#\theta b(\xi^2-1)log(n)$. This holds if $log(C_1) - \frac{1}{2}log(B) - \frac{1}{2}log(\#\theta) \leq B(\xi^2-1)log(n)$. Note that $\#\theta := c_2 n + c_0$, for some constants c_2, c_0. Then, the inequality holds if and only if $log(C_1) - \frac{1}{2}log(B) - \frac{1}{2}log(C_1 n + C_0) \leq B(\xi^2-1)log(n) \iff log(C_1) - \frac{1}{2}log(B) \leq B(\xi^2-1)log(n) + \frac{1}{2}log(C_2 n + C_0)$. Then for sufficiently large n, $log(C_1) - \frac{1}{2}log(B) \leq B(\xi^2-1)log(n) + \frac{1}{2}log(C_2 n) \iff log(C_1) - \frac{1}{2}log(B) \leq B(\xi^2-1)log(n) + \frac{1}{2}log(C_2 n) \iff log(C_1) - \frac{1}{2}log(B) - \frac{1}{2}log(C_2) \leq (B(\xi^2-1)+\frac{1}{2})log(n)$ which holds true for sufficiently large n and $\xi > 1$. Therefore, $log(N(\frac{\epsilon\xi}{2}, \{\theta \in \Theta_n : d(\theta, \theta_j < \epsilon)\}, d_n)) \leq n\epsilon_n^2$. $\square$

Theorem 6. *If f is continuous at x_0 with $f(x_0) > 0$ then $\exists\delta > 0$ such that $\forall r \leq \delta$, $\int\limits_{x_0-r}^{x_0+r} f(x)dx \geq f(x_0)r$.*

Proof. Let $\epsilon = \frac{1}{2}f(x_0) > 0$. Then, $\exists\delta > 0$ such that $|x - x_0| < \delta$ which implies $|f(x) - f(x_0)| < \frac{f(x_0)}{2}$, by the definition of a continuous function. Therefore, $-\frac{f(x_0)}{2} < f(x) - f(x_0) < \frac{f(x_0)}{2}$ which implies $f(x) > f(x_0) - \frac{f(x_0)}{2} = \frac{1}{2}f(x_0)$. Let $0 < r < \delta$. Then, for $x \in [x_0 - r, x_0 + r]$ we have $x_0 - r \leq x \leq x_0 + r$ which implies $-r \leq x - x_0 \leq r$. Therefore, $|x - x_0| \leq r$. Thus, $f(x) \geq \frac{f(x_0)}{2}$. This implies $\int\limits_{x_0-r}^{x_0+r} f(x)dx \geq \int\limits_{x_0-r}^{x_0+r} \frac{f(x_0)}{2}dx = \frac{f(x_0)}{2}[x]_{x_0-r}^{x_0+r} = \frac{f(x_0)}{2}(x_0 + r - x_0 + r) = rf(x_0)$. $\square$

Finally, we need to establish a prior ratio condition in order to apply the posterior concentration result in Theorem 1 of Ghosal and van der Vaart [2007].

Theorem 7. *$\frac{\Pi(\theta\in\Theta_n : j\epsilon_n < d_n(\theta,\theta_0) < 2j\epsilon_n)}{\Pi_n(B_n(\theta_0,\epsilon_n;k))} \leq exp\{\frac{1}{2}Kn\epsilon_n^2 j^2\}$, for sufficiently large $j \in \mathbb{N}$ and $K > 0$.*

Proof. We can trivially upper bound the numerator by one. Therefore, it remains to lower bound the denominator. To lower bound the denominator we will show that the set $B_n(\theta_0, \epsilon_n; k)$ is a superset of some other set, and then lower bound the denominator by the probability of this subset using monotonicity of measure. Let $X^{(n)} = (\mathbf{r}_1, ..., \mathbf{r}_n)$. $r_{a,t} = \alpha_{a,t} + r_{M,t}\beta_{a,t} + \epsilon_{a,t}$. Then, $X_t|\theta \sim N_N(\boldsymbol{\alpha}_t + r_{M,t}\boldsymbol{\beta}_t, \Sigma_t)$, where $\Sigma_t = \Lambda\tilde{\Sigma}_t\Lambda^T + \bar{\Sigma}_t$. Then, $\mathbb{P}_\theta^{(n)}(X^{(n)}, \theta^{(n)}) = P_{\theta_1}(\theta_1)\prod_{t=2}^{n} P(\theta_t|\theta_{t-1})\prod_{t=1}^{n} P(x_t|\theta_t)$. This implies, $\mathbb{P}_\theta^{(n)}(x^{(n)}) = \int P_{\theta_1}(\theta_1)\prod_{t=2}^{n} P(\theta_t|\theta_{t-1})\prod_{t=1}^{n} p(x_t|\theta_t)d\boldsymbol{\theta} = \prod_{t=1}^{n} P(x_t|\theta_t)$. This implies $\ell_n(\theta) = log(\prod_{t=1}^{n} P(x_t|\theta_t)) = \sum_{t=1}^{n} log(p(x_t|\theta_t)) = \sum_{t=1}^{n} \ell_t(\theta)$. Then, $\ell_t(\theta) = log[(2\pi)^{\frac{d}{2}}det(\Sigma_t)^{-\frac{1}{2}}exp\{-\frac{1}{2}(\mathbf{r}_t - \boldsymbol{\mu}_t)^T\Sigma_t^{-1}(\mathbf{r}_t - \boldsymbol{\mu}_t)\}] = -\frac{1}{2}(log((2\pi)^d) + log(det(\Sigma_t)) + (\boldsymbol{r}_t - \boldsymbol{\mu}_t)^T\Sigma_t^{-1}(\boldsymbol{r}_t - \boldsymbol{\mu}_t))$.

Let $U = \{\theta \in \Theta : ||\theta - \theta_0||_2 \leq \delta\}$ for some $\delta > 0$. Note that, $\lambda_{min}(\Sigma_t) = \lambda_{min}(\Lambda\tilde{\Sigma}_t\Lambda^T + \bar{\Sigma}_t) \geq \lambda_{min}(\Lambda\tilde{\Sigma}_t\Lambda^T) + \lambda_{min}(\bar{\Sigma}_t)$, by Weyl's inequality, since the matrices are real and symmetric. Then, $\lambda_{min}(\Sigma_t) \geq \lambda_{min}(\bar{\Sigma}_t) = \min_{1\leq i\leq m} e^{\bar{h}_{t,i}} > 0$, by the extreme value theorem and the fact that the log variance term cannot equal zero. Similarly, $\lambda_{max}(\Sigma_t) = \lambda_{max}(\Lambda\tilde{\Sigma}_t\Lambda^T) + \Lambda_{max}(\bar{\Sigma}_t) \leq C < \infty$, by using Weyl's inequality and the extreme value theorem. By second order Taylor expansion we can write the KL divergence as $KL(\theta_0, \theta) = \frac{1}{2}(\theta_0 - \theta)^T H_t(\theta)(\theta_0, \theta)$, where $H_t(\theta) = \nabla_\theta^2 KL(\theta_0, \theta) = -\mathbb{E}_{\theta_0}[-\nabla_\theta^2 log(\nabla_{\theta,t}(\boldsymbol{r}_t))]$. We know that the Hessian matrix is continuous. Since we are on U_n we may invoke the extreme value theorem. Therefore, the entries of the Hessian matrix are bounded. Thus, $\exists\lambda_{min}, \lambda_{max}$ such that $\infty < \lambda_{min} \leq \lambda(H_t(\theta)) \leq \lambda_{max} < \infty$. This implies $\lambda_{min}I \preceq H_t(\theta) \preceq \lambda_{max}I$. Thus, $\frac{1}{2}(\theta_0 - \theta)^T\lambda_{min}I(\theta_0 - \theta) \leq KL_t(\theta, \theta_0) \leq \frac{1}{2}(\theta_0 - \theta)^T\lambda_{max}I(\theta_0 - \theta)$.Therefore, $\frac{1}{2}\lambda_{min}||\theta_0 - \theta||_2^2 \leq KL_t(\theta_0, \theta) \leq \frac{1}{2}\lambda_{max}||\theta_0 - \theta||_2^2$ for $\theta \in U$. Finally, $\frac{1}{2}\lambda_{min}n||\theta_0 - \theta||_2^2 \leq KL_n(\theta_0, \theta) \leq \frac{1}{2}n\lambda_{max}||\theta_0 - \theta||_2^2$.

Now, $V_{k,0}(P_{\theta_0}, P_\theta) = \int P_{\theta_0}|log(\frac{P_{\theta_0}}{P_\theta}) - KL(P_{\theta_0}, P_\theta)|^k dx = \mathbb{E}_{\theta_0}[|log(\frac{P_{\theta_0}}{P_\theta}) - KL(P_{\theta_0}, P_\theta)|^k] = \mathbb{E}_{\theta_0}[|log(\frac{P_{\theta_0}}{P_\theta}) - \mathbb{E}_{\theta_0}[log(\frac{P_{\theta_0}}{P_\theta})]|^k \leq \mathbb{E}_{\theta_0}[(|log(\frac{P_{\theta_0}}{P_\theta})| + |\mathbb{E}_{\theta_0}[log(\frac{P_{\theta_0}}{P_\theta})]|)^k] \leq \mathbb{E}_{\theta_0}[2^{k-1}(|log(\frac{P_{\theta_0}}{P_\theta})|^k + |\mathbb{E}_{\theta_0}[log(\frac{P_{\theta_0}}{P_\theta})]|^k)]$, since $(a+b)^k \leq 2^{k-1}(a^k + b^k)$ for $a, b \geq 0, k \geq 1$. Then,$V_{k,0}(P_{\theta_0}, P_\theta) \leq \mathbb{E}_{\theta_0}[2^{k-1}(|log(\frac{P_{\theta_0}}{P_\theta})|^k + |\mathbb{E}_{\theta_0}[log(\frac{P_{\theta_0}}{P_\theta})]|^k)] \leq \mathbb{E}_{\theta_0}[2^{k-1}(|log(\frac{P_{\theta_0}}{P_\theta})|^k + \mathbb{E}_{\theta_0}[|log(\frac{P_{\theta_0}}{P_\theta})|^k])]$, by Jensen's inequality. Then, $V_{k,0}(P_{\theta_0}, P_\theta) \leq 2^{k-1}(\mathbb{E}_{\theta_0}[|log(\frac{P_{\theta_0}}{P_\theta})|^k] + \mathbb{E}_{\theta_0}[|log(\frac{P_{\theta_0}}{P_\theta})|^k]) = 2^k\mathbb{E}_{\theta_0}[|log(\frac{P_{\theta_0}}{P_\theta})|^k]$.

If $X \sim N(\boldsymbol{\mu}, \Sigma)$, then $\ell_\theta(x) = log\{(2\pi)^{-\frac{d}{2}} det(\Sigma)^{-\frac{1}{2}} exp\{-\frac{1}{2}(x-\boldsymbol{\mu})'\Sigma^{-1}(x-\boldsymbol{\mu})\}\} = -\frac{d}{2}log(2\pi) - \frac{1}{2}log(det(\Sigma)) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$. Then, $log(\frac{P_{\theta_0}}{P_\theta}) = log(P_{\theta_0}) - log(P_\theta) = -\frac{d}{2}log(2\pi) - \frac{1}{2}log(det(\Sigma_0)) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0) + \frac{d}{2}log(2\pi) + \frac{1}{2}log(det(\Sigma_\theta)) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\theta})'\Sigma_\theta^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma_\theta^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) + \frac{1}{2}log(det(\Sigma_\theta)) - \frac{1}{2}log(det(\Sigma_0)) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_0})\Sigma_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu_0}) = \frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_\theta)'\Sigma_\theta^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_\theta) - (\boldsymbol{x}-\boldsymbol{\mu}_0)'\Sigma_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\} + \frac{1}{2}log(\frac{|\Sigma_\theta|}{\Sigma_0})$. Let $\eta = X - \mu_0$. Then, $\ell_t(\theta) = \frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu_\theta})'\Sigma_\theta^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_\theta) - \eta^T\Sigma_0^{-1}\eta + log(\frac{|\Sigma_\theta|}{|\Sigma_0|})\}$. Since $\eta = X - \mu_0$, this implies $X = \eta + \mu_0$. Therefore, $(\boldsymbol{x}-\boldsymbol{\mu}_\theta) = \eta + \boldsymbol{\mu}_0 - \boldsymbol{\mu}_\theta = \eta - \Delta\boldsymbol{\mu}$, where $\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_\theta - \boldsymbol{\mu}_0$. This implies that $\ell_t(\theta) = \frac{1}{2}\{(\eta-\Delta\boldsymbol{\mu})'\Sigma_\theta^{-1}(\eta-\Delta\boldsymbol{\mu}) - \eta'\Sigma_0^{-1}\eta + log(\frac{|\Sigma_\theta|}{|\Sigma_0|})\}$. Now, $(\eta-\Delta\boldsymbol{\mu})^T\Sigma_\theta^{-1}(\eta-\Delta\boldsymbol{\mu}) = (\eta'-\Delta\boldsymbol{\mu}')\Sigma_\theta^{-1}(\eta = \Delta\mu) = (\eta'\Sigma_\theta^{-1} - \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1})(\eta-\Delta\boldsymbol{\mu}) = \eta'\Sigma_\theta^{-1}\eta - \eta'\Sigma_\theta^{-1}\Delta\boldsymbol{\mu} - \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta + \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\Delta\boldsymbol{\mu} = \eta'\Sigma_\theta^{-1}\eta - 2\Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta + \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\Delta\boldsymbol{\mu}$. This implies $\ell_t(\theta) = \frac{1}{2}\{log(\frac{|\Sigma_\theta|}{|\Sigma_0|}) + \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\Delta\boldsymbol{\mu}\} + \frac{1}{2}\{\eta'\Sigma_\theta^{-1}\eta - 2\Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta - \eta'\Sigma_0^{-1}\eta\} = C_\theta + \frac{1}{2}\eta'\Sigma_\theta^{-1}\eta - \Delta\mu^T\Sigma_\theta^{-1}\eta - \frac{1}{2}\eta'\Sigma_0^{-1}\eta = C_\theta + \frac{1}{2}\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta$, where $C_\theta = log(\frac{|\Sigma_\theta|}{|\Sigma_0|})$. Observe that, $\eta \sim N(0, \Sigma_0)$. Then, $\mathbb{E}_{\theta_0}[\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta] = \mathbb{E}_{\theta_0}[\eta'\Sigma_\theta^{-1}\eta] - \mathbb{E}_{\theta_0}[\eta'\Sigma_0^{-1}\eta] = tr(\Sigma_\theta^{-1}) - tr(\Sigma_0^{-1}\Sigma_0) = tr(\Sigma_0^{-1}\Sigma_0) - N$.

This implies $\mathbb{E}_{\theta_0}[\ell_t(\theta)] = C_0 + \frac{1}{2}tr(\Sigma_\theta^{-1}\Sigma_0) - N$. Let $Y_t = \ell_t(\theta) - \mathbb{E}_{\theta_0}[\ell_t(\theta)] = C_\theta + \frac{1}{2}\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta - C_\theta - \frac{1}{2}tr(\Sigma_\theta^{-1}\Sigma_0) + \frac{N}{2} = \frac{1}{2}\{\eta'(\Sigma_{\theta_0}^{-1} - \Sigma_0^{-1})\eta - tr(\Sigma_\theta^{-1}\Sigma_0)\} + \frac{N}{2} = \frac{1}{2}\{\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - tr(\Sigma_\theta^{-1}\Sigma_0) + tr(\Sigma_0^{-1}\Sigma_0)\} - \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta = \frac{1}{2}[\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - [tr(\Sigma_\theta^{-1}\Sigma_{\theta_0}) - tr(\Sigma_{\theta_0}^{-1}\Sigma_{\theta_0})]] - \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta = \frac{1}{2}[\eta'(\Sigma_\theta^{-1} - \Sigma_{\theta_0}^{-1})\eta - tr(\Sigma_\theta^{-1}\Sigma_0 - \Sigma_0^{-1}\Sigma_0)] - \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta$. Therefore, $\mathbb{E}_{\theta_0}[Y_t] = \frac{1}{2}[\mathbb{E}_{\theta_0}[\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta] - tr((\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0)] - \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\mathbb{E}_{\theta_0}[\eta] = \frac{1}{2}[tr((\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0 - tr((\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0)] - 0 = 0$. Now, $Var(Y_t) = Var(\frac{1}{2}\{\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - tr((\Sigma_\theta^{-1} - \Sigma_0)\Sigma_0)\} - \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta) = Var_{\theta_0}(\frac{1}{2}\{\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - tr((\Sigma_\theta^{-1} - \Sigma_0)\Sigma_0)\}) + Var_{\theta_0}(-\Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta) + 2cov(\frac{1}{2}\{\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - tr((\Sigma_\theta^{-1} - \Sigma_{\theta_0}^{-1})\Sigma_0\}, -\Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta) = \frac{1}{4}Var_{\theta_0}(\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta) + \Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}Var(\eta)\mu\Sigma_\theta^{-1} + 2cov(\frac{1}{2}\{\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - tr((\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0)\}, -\Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta) = \frac{1}{2}tr([(\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0]^2) + \Delta\boldsymbol{\mu}^T\Sigma_\theta^{-1}\Sigma_0\Delta\boldsymbol{\mu}\Sigma_\theta + 2cov(\frac{1}{2}\{\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - tr((\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0)\}, -\Delta\boldsymbol{\mu}'\Sigma_\theta^{-1}\eta)$. Observe that, $tr([(\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0]^2) = tr((\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0(\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0) \leq ||(\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0||_F^2 \leq ||(\Sigma_\theta^{-1} - \Sigma_0^{-1})||_F^2||\Sigma_0||_F^2$. By the mean value theorem we have that $\exists\epsilon_{ij}$ on the segment between θ and θ_0 such that $\Sigma_{ij}(\theta) - \Sigma_{ij}(\theta_0) = \nabla_\theta\Sigma_{ij}(\epsilon_{ij})'(\theta - \theta_0) \implies |\Sigma_{ij}(\theta) - \Sigma_{ij}(\theta_0)| \leq ||\nabla_\theta\Sigma_{ij}(\epsilon_{ij})^T||_2||\theta - \theta_0||_2$.

From previous results, we know that $||\nabla_\theta \Sigma_{ij}(\epsilon_{ij})^T||_2$ is bounded on a compact set. Therefore, $||\Sigma(\theta)-\Sigma(\theta_0)||_F^2 \leq L_1^2||\theta-\theta_0||_2^2$. Observe that $||\Sigma_\theta^{-1}-\Sigma_0^{-1}||_F = ||\Sigma_\theta^{-1}(\Sigma_0 - \Sigma_\theta)\Sigma_0^{-1}||_F \leq ||\Sigma_\theta^{-1}||_2||(\Sigma_0-\Sigma_\theta)\Sigma_0^{-1}||_F$, since $||AB||_F \leq ||A||_2||B||_F$ .Then, $||\Sigma_\theta^{-1} - \Sigma_0^{-1}||_F \leq ||\Sigma_\theta^{-1}||_2||\Sigma_0-\Sigma_\theta||_2||\Sigma_{\theta_0}^{-1}||_F \leq ||\Sigma_0^{-1}||_2||\Sigma_0-\Sigma_\theta||_F||\Sigma_\theta^{-1}||_F \leq ||\Sigma_\theta^{-1}||_F||\Sigma_\theta^{-1}||_F L_1||\theta-\theta_0||_2 \leq ||\Sigma_{\theta_0}^{-1}||_F||\Sigma_\theta^{-1}||_F L_1||\theta-\theta_0||_2 = \sqrt{tr(\Sigma_{\theta_0}^{-1}\Sigma_{\theta_0}^{-1})tr(\Sigma_\theta^{-1}\Sigma_\theta^{-1})}L_1||\theta-\theta_0||_2 = \sqrt{tr(\Sigma_0^{-2})tr(\sigma_\theta^{-2})}L_1||\theta-\theta_0||_2$
$= \sqrt{\sum_{i=1}^N \lambda_i(\Sigma_\theta^{-1})\sum_{i=1}^n \lambda_i(\Sigma_0^{-1})}L_1||\theta-\theta_0||_2 = \sqrt{\frac{1}{\sum_{i=1}^N \lambda_i(\Sigma_0^2)\sum_{i=1}^N \lambda_i(\Sigma_\theta^2)}}L_1||\theta-\theta_0||_2 \leq \sqrt{\frac{N^2}{\lambda_{\theta_0}^2\lambda_\theta^2}}L_1||\theta-\theta_0||_2 = \frac{N}{\lambda_{\theta_0}^2\lambda_\theta^2}L_1||\theta-\theta_0||_2$.

Let $\delta = \frac{\epsilon_n}{\sqrt{max\{C_1,C_2\}}}$. Then, $U_n = \{\theta \in \Theta : ||\theta-\theta_0||_2 \leq \frac{\epsilon_n}{\sqrt{max\{C_1,C_2\}}}\}$. On U_n we have shown that $KL_n(\theta_0,\theta) \leq n\epsilon_n^2$ and $V_{2,0}(\theta_0,\theta) \leq n\epsilon_n^2$. Therefore, $U_n \subseteq B_n(\theta_0,\epsilon_n,2)$ which implies $\mathbb{P}(B_n(\theta_0,\epsilon_n,2)) \geq \mathbb{P}(U_n)$. Therefore, to lower bound the denominator, we can lower bound the probability of U_n.

The required probability is given by $\mathbb{P}(U_n) = \mathbb{P}(\theta \in \Theta : ||\theta-\theta_0||_2 \leq \frac{\epsilon_n}{C_3})$. Note that, $\bigcap_{j=1}^{\#\theta}\{|\theta^{(j)}-\theta_0^{(j)}| \leq \frac{\epsilon_n}{C_3}\} \supseteq \{||\theta-\theta_0||_2 \leq \frac{\epsilon_n}{C_3}\}$. This implies $\mathbb{P}(||\theta-\theta_0||_2 \leq \frac{\epsilon_n}{C_3}) \geq \mathbb{P}(\bigcap_{j=1}^{\#\theta}\{|\theta^{(j)}-\theta_0^{(j)}| \leq \frac{\epsilon_n}{C_3}\}) = \prod_{j=1}^{\#\theta}\mathbb{P}(|\theta^{(j)}-\theta_0^{(j)}| \leq \frac{\epsilon_n}{C_3}) = \prod_{k=1}^{\#\theta}\int_{\theta_0-\frac{\epsilon_n}{C_3}}^{\theta_0+\frac{\epsilon_n}{C_3}} f\theta_j(x)dx$.Now we just need to compute or lower bound these integrals.

Recall, $X_{t+1} = \mu + \phi(X_t-\mu)+\epsilon_t$, with $f_\epsilon(x) = \frac{1}{\pi}\frac{e^{\frac{x}{2}}}{1-e^x}$. Let $Y_t = X_t - \mu$. This implies $Y_{t+1} = \phi Y_t + \epsilon_t$. Then, $\phi_Y(u) = \mathbb{E}[e^{iuY_t}] = \mathbb{E}[exp\{iu(\phi Y_{t-1}+\epsilon_t)\}] = \mathbb{E}[exp\{iu\phi Y_{t-1}\}exp\{iu\epsilon_t\}] = \mathbb{E}[exp\{iu\phi Y_{t-1}\}]\mathbb{E}[exp\{iu\epsilon_t\}] = \phi_Y(\phi\mu)\phi_\epsilon(u) = \prod_{k=0}^N \phi_\epsilon(\phi^k u)\phi_Y(\phi^{N+1}u)$. As $N\to\infty, \phi^{N+1}\to 0$ because $|\phi|<1$. This implies $\phi_Y(u) = \prod_{k=0}^\infty \phi_\epsilon(\phi^k u)$. $\phi_\epsilon(u) = \mathbb{E}[e^{iu\epsilon}]$. Observe that $cosh(x) = \frac{e^x+e^{-x}}{2}$ which implies $2cosh(\frac{x}{2}) = e^{\frac{x}{2}}+e^{-\frac{x}{2}}$. Furthermore, $\frac{e^{\frac{x}{2}}}{1+e^x} = \frac{e^{\frac{x}{2}}}{e^0+e^x} = \frac{e^{\frac{x}{2}}}{e^{-\frac{x}{2}}e^{\frac{x}{2}}+e^{\frac{x}{2}}e^{\frac{x}{2}}} = \frac{e^{\frac{x}{2}}}{e^{\frac{x}{2}}(e^{-\frac{x}{2}}+e^{\frac{x}{2}})} = \frac{1}{e^{-\frac{x}{2}}+e^{\frac{x}{2}}} = \frac{1}{2cosh(\frac{x}{2})}$. This implies $f_\epsilon(x) = \frac{1}{2\pi cosh(\frac{x}{2})}$. This results in, $\phi_\epsilon(u) = \int_{-\infty}^\infty e^{iux}\frac{1}{2\pi cosh(\frac{x}{2})}dx = \frac{1}{2\pi}\int_{-\infty}^\infty e^{iux}sech(\frac{x}{2}) = sech(\pi u)$. Therefore, $\phi_Y(u) = \prod_{k=0}^\infty sech(\pi\phi^k u)$. Therefore, $f_Y(y) = \frac{1}{2\pi}\int_{-\infty}^\infty e^{-iuy}\prod_{k=0}^\infty sech(\pi\phi^k u)du$. Then, $f_X(x) = f_Y(y-\mu) = \frac{1}{2\pi}\int_{-\infty}^\infty e^{-iu(x-\mu)}\prod_{k=0}^\infty sech(\pi\phi^k u)$.

Now we will apply Theorem 6 to each of the parameters blocks. Firstly, look at the DSP priors. $f_X(x_0) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-iu(x_0-\mu)} \prod\limits_{k=1}^{\infty} sech(\pi\phi^k u)du = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-iu(x_0-\mu)}\phi_Y(u)du$. The function $\phi_Y(u)$ is integrable. Then, by the dominated convergence theorem the map $x \to \int\limits_{-\infty}^{\infty} e^{-iu(x-\mu)}\phi_Y(u)du$ is continuous. If $x \to x_0$ then for each fixed $u, e^{-iu(x-u)} \to e^{-iu(x_0-u)}$ we have $|e^{-iu(x-u)}\phi_Y(u)| \leq |\phi_Y(u)|$ which is integrable. Therefore, $\lim\limits_{x\to x_0} f(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-iu(x_0-u)}\phi_Y(u)du = f(x_0)$. Therefore, f is continuous at x_0 and $f(x_0) > 0$. Then by Theorem 6, $\int\limits_{\theta_0-\frac{\epsilon_n}{C_3}}^{\theta_0+\frac{\epsilon_n}{C_3}} f_h(x)dx \geq f(h_{0,t-1})C_3\epsilon_n$.

For the mean parameters $\mu = ln(\tau_0^2\tau_1^2), \tau_0 \sim C^+(0, \frac{1}{\sqrt{n}}), \tau_1 \sim C^+(0,1)$. The probability density function of a half cauchy distribution with scale σ^2 is $\frac{2}{\pi\sigma}\frac{1}{1\frac{y^2}{\sigma^2}} = \frac{2}{\pi\sigma(1+\frac{y^2}{\sigma^2})} = \frac{2\sigma}{\pi\sigma^2(1+\frac{y^2}{\sigma^2})} = \frac{2\sigma}{\pi(\sigma^2+y^2)}$. Let $X = \tau^2$. Then, $f_X(x) = f_\tau(\sqrt{x})\frac{d}{dx}(\sqrt{x}) = \frac{2\sigma}{\pi(\sigma^2+x)}\frac{1}{2\sqrt{x}} = \frac{\sigma}{\sqrt{x}(\sigma^2+x)\pi}$. Now, let $U = ln(X)$ which implies $X = e^u$ which leads to $dx = e^u du$. Therefore, $f_U(u) = f_X(e^u)e^u = \frac{\sigma e^u}{\sqrt{e^u}(\sigma^2+e^u)\pi} = \frac{\sigma e^{\frac{1}{2}u}}{(\sigma^2+e^u)\pi}$. Let $\mu = U_0 + U_1$. Then, $f_\mu(m) = \int\limits_{-\infty}^{\infty} f_{U_0}(t)f_{U_1}(m-t)dt = \int\limits_{-\infty}^{\infty} \frac{S_0}{\pi}\frac{e^{\frac{t}{2}}}{(S_0^2+e^t)}\frac{S_1}{\pi}\frac{e^{\frac{(m-t)}{2}}}{e^{m-t}+S_1^2}dt = \frac{S_0S_1e^{\frac{m}{2}}}{\pi^2}\int\limits_{\infty}^{\infty}\frac{1}{(S_0^2+e^t)(e^{m-t}+S_1^2)}dt$. Let $x = e^t$ which implies $dt = \frac{1}{t}dx$. Therefore, $\int\limits_{-\infty}^{\infty}\frac{1}{(e^t+s_0^2)(e^{m-t}+S_1^2)}dt = \int\limits_{0}^{\infty}\frac{1}{x(x+S_0^2)(\frac{e^m}{x}+s_1^2)}dx = \int\limits_{0}^{\infty}\frac{1}{(x+s_0^2)(e^m+s_1^2x)}dx$. Note that, $\frac{1}{(x+a)(b+cx)} = \frac{1}{b-ac}(\frac{1}{x+a} - \frac{c}{b+cx}])$. Therefore, the integral with such an integrand is equal to $\frac{1}{b-ac}[ln(x+a) - ln(x+cx)]_0^\infty = \frac{1}{b-ac}ln(\frac{b}{ac})$. Therefore, $\int\limits_{0}^{\infty}\frac{1}{(x+s_0^2)(e^m+s_1^2x)}dx = \frac{m+ln(n)}{e^m-\frac{1}{n}}$. This implies $f_\mu(m) = \frac{S_0S_1e^{\frac{m}{2}}}{\pi^2}\frac{m+ln(n)}{e^m-\frac{1}{n}} = \frac{e^{\frac{m}{2}}}{\sqrt{n}\pi^2}\frac{m+ln(n)}{e^m-\frac{1}{2}}$. Suppose, $\mu_0 \neq -ln(n)$. Then the factor $\frac{e^{\frac{m}{2}}}{\sqrt{n}\pi^2}$ is strictly positive and continuous everywhere on $\mathbb{R}$. Then, $\frac{m+ln(n)}{e^m-\frac{1}{n}}$ is the quotient of two functions that are continuous at μ_0 and have non-zero values when $\mu_0 \neq -ln(n)$. Therefore, the quotient is continuous at μ_0. We also know that $sign(m+ln(n)) = sign(e^m - \frac{1}{n})$. Therefore, $\frac{m+ln(n)}{e^m-\frac{1}{n}} > 0$. Therefore, $\lim\limits_{m\to\mu_0} f_\mu(m) = f_\mu(\mu_0) > 0$. When $\mu_0 = -ln(n)$ we get $\frac{0}{0}$ which is undefined. However, f_μ extends continuously and positively through this point. Let $g(m) = m + ln(n)$ and $h(m) = e^m - \frac{1}{n}$, and $B(m) = \frac{g(m)}{h(m)}$. Now, apply L'Hôpital's rule, $\lim\limits_{m\to-ln(n)} B(m) = \lim\limits_{m\to-ln(n)}\frac{g'(m)}{h'(m)} = e^{ln(n)} = n$. Further, $\lim\limits_{m\to-ln(n)} A(m) = \frac{e^{-\frac{ln(n)}{2}}}{\sqrt{n}\pi} =$

$\frac{n^{-\frac{1}{2}}}{\sqrt{n}\pi^2} = \frac{1}{\pi^2}$. This implies $\lim_{m \to -ln(n)} f_\mu(n) = \lim_{m \to -ln(n)} A(m)B(m) = \frac{n}{\pi^2}$. So, defining $f_\mu(-ln(n)) := \frac{1}{\pi^2}$ makes f_μ continuous and positive at $m = -ln(n)$. Therefore, by Theorem 6, $\int_{\theta_0 - \frac{\epsilon_n}{C_3}}^{\theta_0 + \frac{C_3}{\epsilon_n}} f_\mu(x)dx \geq f_\mu(\mu_0)C_3\epsilon_n$.

Let $Y = \frac{\phi+1}{2}$. Then, $f_Y(y) = \frac{1}{B(a,b)} y^{a-1}(1-y)^{b-1}$. Observe that, $\phi = g(Y) = 2Y - 1$. Then, $Y = g^{-1}(\phi) = \frac{\phi+1}{2}$. Then, $\frac{d}{dx}(g^{-1}(\phi)) = \frac{d}{dx}(\frac{x+1}{2}) = \frac{1}{2}$. This implies, $f_\phi(\phi) = f_Y(g^{-1}(x))|\frac{d}{dx}g^{-1}(x)| = f_Y(\frac{\phi+1}{2})\frac{1}{2}$, since $y = \frac{x+1}{2}$. We set, $f_X(x) = \frac{1}{2B(a,b)}(\frac{\phi+1}{2})^{a-1}(1 - \frac{\phi+1}{2})^{b-1} = \frac{1}{2^{a+b-1}B(a,b)}(\phi+1)^{a-1}(1-\phi)^{b-1}$, where $|\phi| < 1$. On $(-1,1), (\phi+1)^{a-1}$ and $(1-\phi)^{b-1}$ are both continuous. Therefore, $f_X(x)$ is continuous $\forall \phi \in (-1,1)$. For $\phi_0 \in (-1,1), \phi_0 + 1 > 0$ and $1 - \phi_0 > 0$. Therefore, $(x_0+1)^{a-1}$ and $(1-x_0)^{b-1}$ are both greater than zero. The normalization constant is also positive. Therefore, $f_\phi(\phi_0) > 0$. By Theorem 6, this implies $\int_{\phi_0 - \frac{\epsilon_n}{C_3}}^{\phi_0 + \frac{\epsilon_n}{C_3}} f_\phi(x)dx \geq f_\phi(\phi_0)C_3\epsilon_n$.

Note that $\beta_{a,t-1}$,$\bar{\mu}_a$ for each asset a, and $\tilde{\mu}_j$ are realized from normal distributions which is continuous everywhere and has strictly positive density on $\mathbb{R}$. Therefore we may invoke Theorem 6 to bound the integrals. Similarly, for $\bar{\sigma}^2$ and $\tilde{\sigma}^2$, they both follow gamma densities which are continuous everywhere and strictly positive, which allows us to attain similar bounds. Furthermore, the conditional distribution of each $\Lambda_{i,j}$ is also normal, which again allows us to get a lower bound. For the remaining log variance processes with normally distribution innovations $f_h(x)\sqrt{\frac{1-\phi^2}{2\pi\sigma^2}}exp\{-(\frac{1-\phi^2}{2\sigma^2}x^2)\}$ which is continuous and positive at x_0.Therefore, $\int_{h_{0,t-1} - \frac{\epsilon_n}{C_3}}^{h_{0,t-1} + \frac{\epsilon_n}{C_3}} f_{h,t-1}(x)dx \geq f_{h,t-1}(h_{0,t-1})C_3\epsilon_n$, by Theorem 6.

Therefore $\Pi_n(B_n(\theta_0, \epsilon_n, k)) \geq m^{\#\theta}\epsilon_n^{\#\theta}$, where $m := minC_3 f_{\theta_j}(\theta_{j,0})$ and $\#\theta = C_d n + C_0$. Thus, we just need to verify that $\frac{1}{(m\epsilon_n)^{\#\theta}} \leq e^{\frac{Kn\epsilon_n^2 j^2}{2}} \iff -\#\theta log(m\epsilon_n) \leq \frac{Kn\epsilon_n^2 j^2}{2}$. Note that, $\epsilon_n = \frac{C_{\epsilon_n}\#\theta log(n)}{n}$, and $\#\theta = C_d n$ for sufficiently large n, which implies $\epsilon_n^2 = C_{\epsilon_n}C_d log(n)$ for sufficiently large n. This implies we need to verify, $-C_d n(log(m) + log(\epsilon_n)) \leq \frac{Kn\epsilon_n^2 j^2}{2} \iff -C_d(log(m) + log(\sqrt{C_{\epsilon_n}C_d log(n)})) \leq \frac{K\epsilon_n^2 j^2}{2} \iff -C_d(log(m) + \frac{1}{2}log(C_{\epsilon_n}C_d log(n))) \leq \frac{K\epsilon_n^2 j^2}{2} \iff -C_3(log(m) + \frac{1}{2}log(C_4 log(n))) \leq \epsilon_n^2 j^2 \iff$

$-C_3log(m) - \frac{C_3}{2}log(C_4log(C_4log(n))) \leq \epsilon_n^2 j^2 \iff -C_5 - C_6log(C_4log(n)) \leq \epsilon_n^2 j^2 \iff \frac{-C_5 - C_6log(C_4log(n))}{\epsilon_n^2} \leq j^2 \iff \frac{-C_5 - C_6log(C_4log(n))}{\frac{C_8 nlog(n)}{n}} \leq j^2 \iff -(\frac{C_5 + C_6log(C_4log(n))}{C_8log(n)}) \leq j^2$ which holds for sufficiently large j. □

We can now invoke Theorem 1 of Ghosal and van der Vaart [2007] to establish a posterior concentration results for our model for time varying correlation matrices.

Theorem 8. $\mathbb{P}_{\theta_0}^{(n)}\Pi_n(\theta \in \Theta_n : ||R_\theta - R_{\theta_0}||_F \geq M_n\epsilon_n|X^{(n)}) \to 0$ *for every* $M_n \to \infty$

Proof. We have proved that all the conditions of Theorem 1 of Ghosal and van der Vaart [2007]. Therefore, by application of their theorem 1 we prove theorem 9. □

We can then extend this further to verify the posterior concentration of our scalar summary of correlation matrices.

Theorem 9. $\mathbb{P}_{\theta_0}^{(n)}\Pi_n(\theta \in \Theta_n : |Score(R_\theta) - Score(R_{\theta_0})| \geq \frac{M\epsilon_n}{N-1}|X_n) \to 0$

Proof. $score(A) - score(B) = \frac{\sum_{j=1}^{N} \boldsymbol{x}_j^T\mathbf{1} - N}{N(N-1)} - \frac{\sum_{j=1}^{N} \boldsymbol{y}_j^T\mathbf{1} - N}{N(N-1)} = \frac{1}{N(N-1)}\sum_{j=1}^{N}(\boldsymbol{x}_j^T\mathbf{1} - \boldsymbol{y}_j^T\mathbf{1}) = \frac{1}{N(N-1)}\sum_{j=1}^{N}(\boldsymbol{x}_j - \boldsymbol{y}_j)^T\mathbf{1}$. Observe that $|(\boldsymbol{x}_j - \boldsymbol{y}_j)^T\mathbf{1}| \leq ||(\boldsymbol{x}_j - \boldsymbol{y}_j)^T||_2||\mathbf{1}||_2 = \sqrt{N}||\boldsymbol{x}_j - \boldsymbol{y}_j||_2$, by the Cauchy-Schwarz inequality. This implies $|score(A) - score(B)| = |\frac{1}{N-1}\sum_{j=1}^{N}(\boldsymbol{x}_j - \boldsymbol{y}_j)^T\mathbf{1}| = \frac{1}{N(N-1)}|\sum_{j=1}^{N}(\boldsymbol{x}_j - \boldsymbol{y}_j)^T\mathbf{1}| \leq \frac{1}{N(N-1)}\sum_{j=1}^{N}|(\boldsymbol{x})_j - \boldsymbol{y}_j)^T\mathbf{1}|$, by the triangle inequality. Then, $|score(A) - score(B)| \leq \frac{1}{N(N-1)}\sum_{j=1}^{N}\sqrt{N}||\boldsymbol{x}_j - \boldsymbol{y}_j||_2 = \frac{\sqrt{N}}{N(N-1)}\sum_{j=1}^{N}||\boldsymbol{x}_j - \boldsymbol{y}_j||_2$. Now, $\sum_{j=1}^{N}||\boldsymbol{x}_j - \boldsymbol{y}_j||_2 = \sum_{j=1}^{n} 1 \cdot ||\boldsymbol{x}_j - \boldsymbol{y}_j||_2 = a'b =< a, b >= | < a, b > |$, since the euclidean norm is non-negative. Then, the Cauchy-Schwartz inequality tells us that $||a||_2 = \sqrt{1^2 + ... + 1^2} = \sqrt{N}, ||b||_2 = \sqrt{\sum_{j=1}^{N}||\boldsymbol{x}_j^T - \boldsymbol{y}_j^2||_2^2} = ||A - B||_F$. This implies $\sum_{j=1}^{N}||\boldsymbol{x}_j - \boldsymbol{y}_j||_2 \leq \sqrt{N}||A - B||_F$. This implies $|score(A) - Score(B)| \leq \frac{\sqrt{N}}{N(N-1)}\sum_{j=1}^{N}||\boldsymbol{x}_j - \boldsymbol{y}_j||_2 \leq \frac{\sqrt{N}}{N(N-1)}\sqrt{N}||A - B||_F = \frac{1}{N-1}||A - B||_F$. Therefore, $\mathbb{P}_{\theta_0}^{(n)}\Pi_n(\theta \in \Theta_n : |socre(R_\theta) - score(R_{\theta_0})| \geq \frac{M\epsilon_n}{N-1}|X^{(n)}) \leq \mathbb{P}_{\theta_0}^{(n)}\Pi_n(\theta \in \Theta_n : ||R_\theta - R_{\theta_0}||_F > M_n\epsilon_n|X^{(n)}) \to 0$. □

5 MCMC sampler

All our computation is through the R programming language (R Core Team [2023]) using and building upon the code of Kowal et al. [2019] and Hosszejni and Kastner [2021].

The first component of our MCMC algorithm is to draw samples of the posterior variance of $r_{M,t}$ in (2), which is independent of the other components of our MCMC sampling algorithm. We assume the excess market return follows a stochastic volatility process (Taylor [1982]) of order one. To obtain posterior samples we utilize the R package stochvol (Hosszejni and Kastner [2021]). The sampling algorithm of Hosszejni and Kastner [2021] uses the ancilitary interweaving strategy proposed in Yu and Meng [2011]. The application of this sampling technique to stochastic volatility models was then discussed in Kastner and Frühwirth-Schnatter [2014]. To improve computation time Hosszejni and Kastner [2021] interface R to C++.

The next step of our sampling algorithm is to sample the state variables in the linear factor models for each asset in the portfolio of interest. For example, in the time varying parameter CAPM for a given asset in a portfolio, this would be $\alpha_1, ..., \alpha_T$ and $\beta_1, ..., \beta_T$. We sample the time series of the state variables independently for each asset, with the only dependence across assets coming from the observation error covariances which is discussed in the next paragraph. For sampling the state variables, we utilize the sampler of Kowal et al. [2019]. However, we make some important changes. Firstly, since we have several assets which have a joint model (through the observation error) we sample several times series simultaneously. Secondly, since we assume the assets are dependent through the observation error, we replace the observation error variance for a single asset which was assumed to follow a stochastic volatility process of order one in the original DSP model, with the sampled observation error variance for the respective asset from the multivariate stochastic volatility process (discussed in the penultimate paragraph in this section). Kowal et al. [2019] also use parameter expansion (Liu and Wu [1999]) which is known in the literature for improving the efficiency in MCMC computation, particularly for Gibbs sampling algorithms. Specifically, Kowal et al. [2019] use a Pólya-gamma parameter expansion for sampling from the four parameter Z distribution which improves the computational efficiency due to the computational ease of sampling from Pólya-gamma distributions. This is based upon combining the work of Polson et al. [2013] with Barndorff-Nielsen et al. [1982] using the fact that a four parameter Z distribution can be written as a Normal mean scale mixture.

For a single asset in a portfolio let $\boldsymbol{\beta} = (\alpha_1, \beta_1, \ldots, \alpha_T, \beta_T)'$, $\boldsymbol{X} = \text{blockdiag}\big((1, r_{M,t})_{t=1}^{T}\big)$, $\Sigma_\omega = \text{diag}\Big(\tau_0^2\tau_\alpha^2\lambda_{\alpha,1}^2,\ \tau_0^2\tau_\beta^2\lambda_{\beta,1}^2,\ \ldots,\ \tau_0^2\tau_\alpha^2\lambda_{\alpha,T}^2,\ \tau_0^2\tau_\beta^2\lambda_{\beta,T}^2\Big)$, $\Sigma_\epsilon = \text{diag}\Big(\{\sigma_t^2\}_{t=1}^{T}\Big)$, $\boldsymbol{\ell}_\beta = \boldsymbol{X}'\Sigma_\epsilon^{-1}\boldsymbol{y} = [\frac{y_1}{\sigma_1^2}(1, r_{M,1}), ..., \frac{y_T}{\sigma_T^2}(1, r_{M,T})]'$, D_2 be a matrix with entries given by $d_{ij} = \delta_{ij} - 2\mathbf{1}\{i = j+1, i \geq 3\} + \mathbf{1}\{i = j+2\}$, where $\delta_{ij}$is the Kronecker delta and $\mathbf{1}$ denotes an indicator function, and I_2 denotes the 2x2 identity matrix. Then the posterior distribution of the state variables from a single asset is given by $\boldsymbol{\beta} \sim N(\boldsymbol{Q}_\beta^{-1}\boldsymbol{\ell}_\beta, \boldsymbol{Q}_\beta^{-1})$, where $\boldsymbol{Q}_\beta = X'\boldsymbol{\Sigma}_\epsilon^{-1}X + (D_2' \bigotimes I_2)\boldsymbol{\Sigma}_\omega^{-1}(D_2 \bigotimes I_2)$. Then, in a given MCMC sample in order to compute (20) we need to obtain a sample of the posterior variance of $var(\boldsymbol{\alpha}_t)$ and $E[\boldsymbol{\beta}_t\boldsymbol{\beta}_t^T]$. To compute the posterior mean we use a Cholesky decomposition of $\boldsymbol{Q}_\beta$ to solve the linear system $\boldsymbol{Q}_\beta E[\boldsymbol{\beta}] = \boldsymbol{\ell}_b$ which gives $E[\boldsymbol{\beta}] = Q_\beta^{-1}\boldsymbol{\ell}_\beta$. To compute the posterior variance we find the inverse of the Cholesky factor of $\boldsymbol{Q}_\beta$ and then compute the sum of squared row entries, which gives us the variance of $\boldsymbol{\beta}$. We then use $var(\boldsymbol{\beta}) + E[\boldsymbol{\beta}]E[\boldsymbol{\beta}]^T = E[\boldsymbol{\beta}\boldsymbol{\beta}^T]$ to also obtain an MCMC sample of the expected outer product.

For drawing samples of the observation error covariances from the MFSV model we utilize the R package factorstochvol (Hosszejni and Kastner [2021]). For computational tractability the authors assume that the covariances are driven by a small number of latent factors. The sampling of the idiosyncratic variances utilize the same sampling procedure as univariate stochastic volatility processes. The authors also utilize the ancillarity interweaving strategy of Yu and Meng [2011] and offer alternative interweaving strategies. For our sampling scheme we use deep interweaving for the largest absolute entries in each column of the factor loading matrix $\boldsymbol{\Lambda}$. We then feed the posterior sampled observation error variances into our sampling of the individual asset observation equations. Finally, we then have a posterior sample of $var(\boldsymbol{\alpha}_t), var(r_{M,t}), E[\boldsymbol{\beta}_t\boldsymbol{\beta}_t^{'}]$, and $var(\boldsymbol{\epsilon}_t)$ for $1 \leq t \leq T$, which allows us to use (20) to compute the posterior covariance matrix time series. We then standardize each of these matrices to obtain the correlation matrix, and then apply our score function, to obtain a posterior sample of the score time series.

6 Results

To assess our proposed methodology, we performed a simulation study in Section 6.1. We also apply our proposed methodology to two real world examples, in Section 6.2, of financial crises to observe the impact, if any, that portfolio diversification has on mitigating the impact of such crises. We also construct the minimum variance portfolios and assess how the dynamically estimated portfolio compares to the statically estimated portfolios.

6.1 Simulation study

In this Section we discuss our simulation study including the details of how we formed the simulations and the results from our simulation study based on 100 simulations. For a given simulation we fix the length of all time series to 1000 time points and the number of simulated assets to 30. We also fix all the pairwise correlations to be equal. We then construct the time series of the scores, where we use the same time series of scores in each simulation as displayed in Figure 8 in the appendix. Since all our pairwise correlations are equal they are equal to the constructed scalar score at the respective time point. We then construct the associated model-based covariance matrices by fixing the variances of the 30 excess asset returns to two. We build the rest of our simulation from the time varying parameter CAPM.

We simulate the time series of the α's and β's for each asset from independent multivariate normal distributions with mean vector $(0, 1)^T$ and diagonal covariance matrix with entries equal to 0.1. The time series of the excess market returns are simulated from a standard normal white noise process. Subsequently we construct the observation error covariance matrices such that the overall model-based covariance matrices are equal to those defined in the earlier step. Following this we simulate the observation errors from a multivariate normal distribution with mean vector equal to the zero vector, and covariance matrix given by the computed covariance matrix at the respective time point from the previous step.

Now we have simulated α, β from the CAPM, and observation error time series for each of the 30 assets. In addition, we also have the simulated excess market returns. We then combine these according to the CAPM to obtain simulated excess asset returns for 30 assets. Then we fit an exponentially weighted rolling correlation estimate, the proposed DSP-MFSV CAPM,DCC model, and an MFSV model.

For each simulation we computed the root mean squared error (RMSE) for each of the fitted models and for the Bayesian models the empirical coverage and mean empirical credible interval width. The root mean squared error is given by

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(x_t - \hat{x}_t)^2}{T}}. \tag{22}$$

Where, $\boldsymbol{x} = (x_1, ..., x_T)^T$ is the vector containing the observations of the true time series, and $\hat{\boldsymbol{x}}$ is the estimate of the time series $\boldsymbol{x}$, and T is the length of the time series . The RMSE gives a measure of the accuracy of estimates with a RMSE of zero corresponding to perfect estimation.

The empirical coverage and mean empirical credible interval width are given by:

$$\text{empirical coverage} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{\text{lower}(t), \text{upper}(t)\}(x(t)) \tag{23}$$

$$\text{mean credible interval width} = \frac{1}{T}\sum_{t=1}^{T} \text{lower}(t) - \text{upper}(t). \tag{24}$$

In (23) and (24), lower(t) and upper (t) refer to the lower and upper bounds of the estimated 95 % highest density interval (HDI) and $\mathbb{1}\{\text{lower}(t), \text{upper}(t)\}(x(t))$ is equal to 1 if the observed value of the time series is within the interval $[\text{lower}(t), \text{upper}(t)]$ and zero otherwise. The empirical coverage gives us the empirical probability that the estimated HDIs contain the true time series. The mean credible interval informs us the average width of the estimated HDIs across the time series.

The results of our 100 simulations are displayed in table 1, where we saved 3000 MCMC samples from the Bayesian models with a burn in period of 1500 samples and a thinning rate of 4.

Model	RMSE	mean empirical coverage	mean credible interval width
DSP-MFSV-CAPM	0.038	0.953	0.120
MFSV	0.050	0.999	0.334
Rolling	0.072	NA	NA
DCC	0.116	NA	NA

Table 1: Simulation study results

Our model achieves the best performance in terms of RMSE, and achieves 95% coverage as desired, and with noticeably tighter HDIs when compared to the normal approximation from the MFSV model which are 2.78 times wider.

6.2 Real world examples

To demonstrate the applicability of our proposed methodology to practitioners and researchers we investigate two problems. Firstly, we show that in two major financial crises this century, the U.S subprime mortgage crisis (Section 5.2.1-Section 5.2.3), and the 2020 COVID-19 pandemic (Section 5.2.4- Section 5.2.6), that stock portfolio diversification does not protect investors from correlation risks induced by such high volatility events. Secondly, we explore how the standard static minimum variance portfolio differs from the dynamically estimated minimum variance portfolio at the peak of these financial crises. In performing this analysis we draw 13,500 samples from the posterior distribution, with a burn-in period of 1500 samples, and a thinning rate of 4, which gives us a total of 3000 saved MCMC samples in both examples.

6.2.1 U.S. subprime mortgage crisis

The U.S. subprime mortgage crisis occurred from 2007 to 2010. This was a global financial crisis which originated from the U.S. housing bubble. Particularly, the securitization of mortgages including the infamous collateralized debt obligations (CDOs) which although highly rated were of a vastly higher risk than advertised. When the U.S housing bubble burst this triggered a global financial crisis. We consider two portfolios with 30 stocks. The first portfolio consists of 30 large technology stocks from the period. The second portfolio is a more diversified portfolio in which we include 10 technology stocks from the first portfolio and the remainder, consisting of large stocks from a range of industries, from the S&P 500 stock index. We compute the excess asset returns by using data downloaded from Yahoo Finance and the Fama-French data library (Fama [2023]).We then we fit the DSP-MFSV CAPM to daily adjusted closing price data from 4 January 2006 to the 31 December 2009. From this we obtain 3000 saved posterior samples of the score time series, where the score is the scalar summary of the estimated correlation matrix at a given time point (Definition 4). We also plot the daily normalized prices of five of the stocks in the diversified portfolio in Figure 2. We observe some interesting, shared behavior of these price time series, such as the sharp decrease in adjusted closing prices in late 2008.

6.2.2 Diversification risk

In this Section we discuss the extent to which portfolio diversification could have helped to shield an investor's portfolio from such a crisis. Having fitted our DSP-MFSV CAPM we plot the time series of the estimated scores along with 95% HDIs. In Figure 3 we plot the posterior score of both of our portfolios and the posterior score of the diversified portfolio with 95% highest density (posterior) intervals (HDIs) and the VIX index in figure 4. By inspecting Figure 3 we see that the technology portfolio gener-

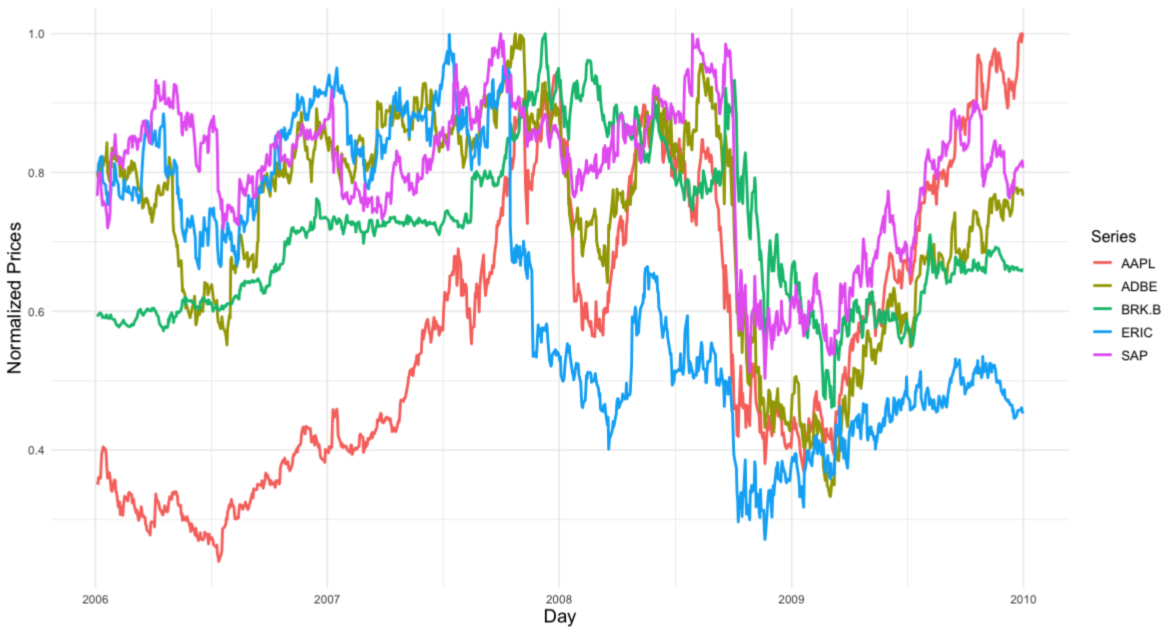

Figure 2: Plot of the daily normalized adjusted closing prices of 5 stocks from the 4th of January 2006 to the 31st of December 2009.

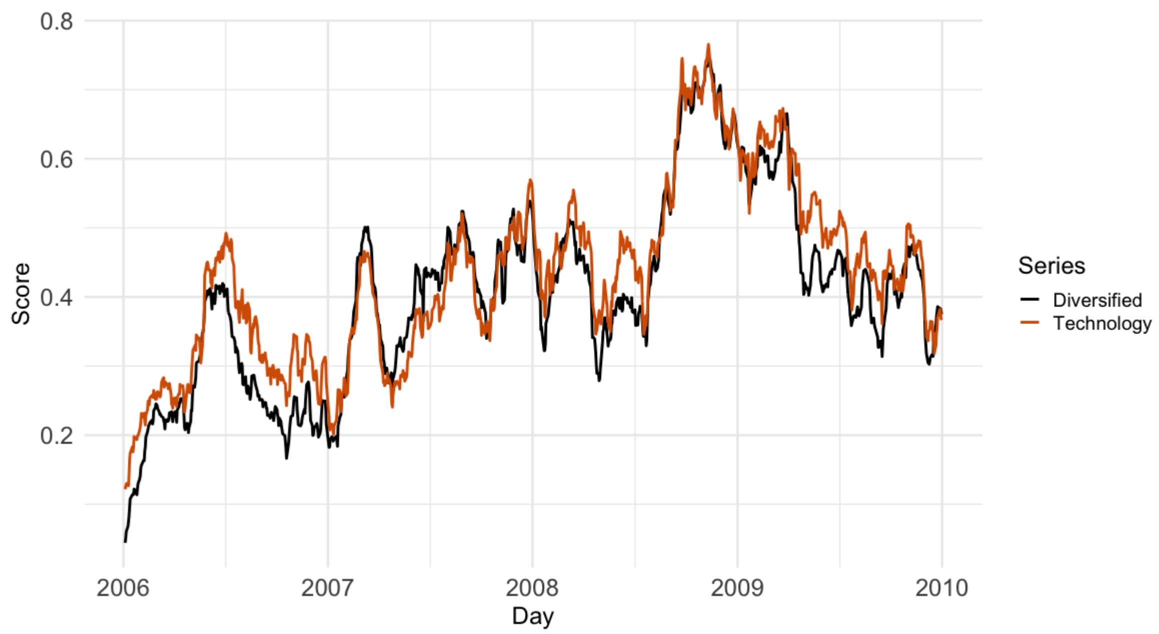

Figure 3: Plot of the estimated posterior mean score time series for the technology portfolio in vermillion and our the estimated posterior mean score time series for the diversified portfolio in Black.

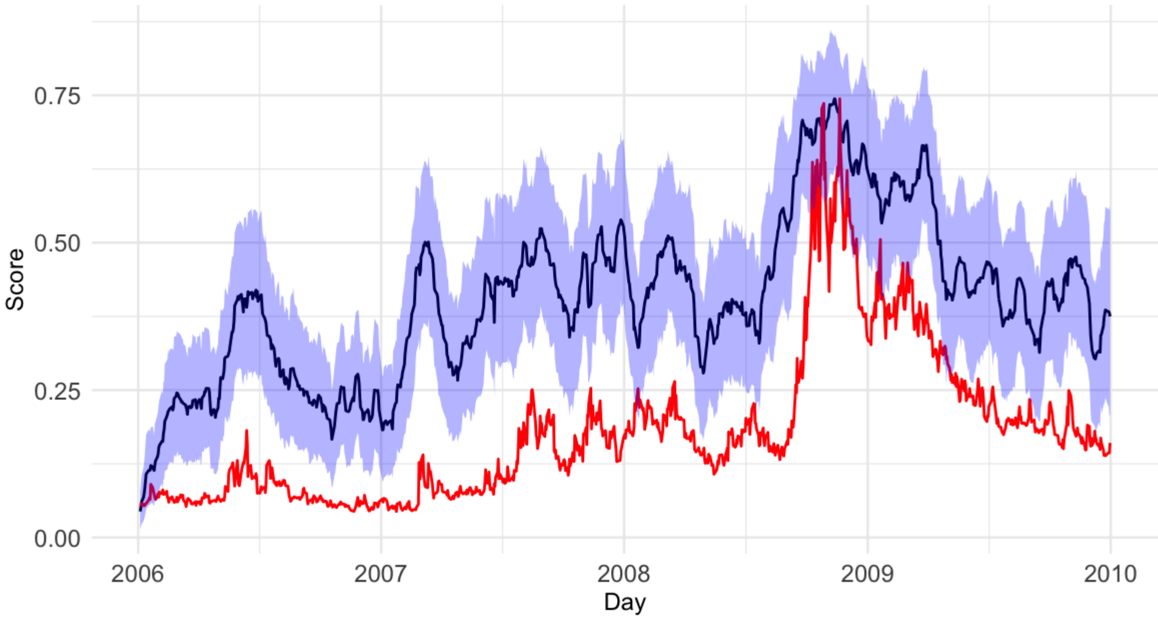

Figure 4: Plot of the estimated posterior score time series for our diversified portfolio in black and the VIX index in red. The 95% HDIs of our scores are represented by the boundary of the purple area. The VIX index is scaled to be between the smallest and largest posterior mean of the estimated score time series.

ally has a larger overall level of correlation compared to the diversified portfolio. For example, in mid-2006 we see that the correlation of the pure technology portfolio is larger than that of the diversified portfolio. As we proceed into 2007 this changes; for example in the first quarter of 2007, both portfolios see a spike in their correlations, with the diversified portfolio having a greater correlation compared to the pure technology portfolio. Overall, we can see that the shapes of the correlation dynamics are similar throughout the period of the U.S. sub-prime mortgage crisis, culminating in the large spikes in the correlations of both portfolios which reached their peak on 12 November 2008, coinciding with the large increase in the VIX index during the same period of time. Interestingly, our proposed correlation score begins to increase at the very start of mid-2008, prior to that of the VIX index. This highlights the additional information that can be provided by analyzing the correlations within a given investor's portfolio to detect a market downturn, and the ability of our model to monitor such movement. It also highlights, that despite diversification we still suffer from the shocks in both portfolios, and at the periods of the most intense economic stress, both portfolios see a similar large degree of correlation. In Figure 4 we see several spikes, such as in June-July 2006 when several events occurred, including an increase in the federal funds rate, and difficulties appearing in the CDO market, such as the growth of credit default swaps with regard to CDOs and the struggles of some institutions to sell their CDOs. We then see a noticeable spike in March 2007 which could stem from growing warnings about an impending crisis; for example a speech by Ben Bernanke discussing how Fannie Mae and Freddie Mac were causing a systemic risk to the U.S. economy. After this time period we observe a prolonged period of increased correlation starting around June and continuing into 2008. We begin to see the increase in correlation starting from approximately 0.329 on 22 July 2008. It increases rapidly during the most dramatic events of the crisis in September 2008, such as the government takeover of Fannie Mae and Freddie Mac and the bankruptcy of Lehmann brothers, reaching 0.520 on 3 September as identified by our method, and 0.708 by 25 September. We see some minor perturbations of the correlation although it remains quite high, reaching a peak of 0.744 on 12 November 2008 which reflects the aftermath of these dramatic events and the uncertainty around the government response to the crisis, such as loans and refinancing from the U.S. government. This example illustrates the use of our method in providing a novel tool for tracking the stability of an economic system over time and provides a quantitative approach to clearly identify key points of the crisis. It also illustrates that there is additional information beyond measures such as the VIX index, which provide some measure of the toxicity of the economy, and early warnings of what is to come compared to the VIX.

6.2.3 Minimum variance portfolio

The minimum variance portfolio is the portfolio which solves the problem $\min\limits_{\boldsymbol{\omega}} = \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}$, where $\boldsymbol{\omega}$ is a vector which sums to one and gives the proportion of each stock we should hold a long or short position in, and $\boldsymbol{\Sigma}$ is the covariance matrix of the excess asset returns (Tsay [2005]). The solution to this problem is given by:

$$\boldsymbol{\omega} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{25}$$

The minimum variance portfolio is the portfolio allocation which minimizes volatility irrespective of expected return. Usually this is estimated statically by either using the covariance derived from a static CAPM or using the sample covariance matrix. We compare how the static global minimum variance portfolios compare to the dynamic global minimum variance portfolio, with respect to our diversified portfolio, at the time of maximum correlation as identified by the largest value of the posterior mean of the score from the fitted DSP-MFSV CAPM.

In this case the maximum correlation point occurs on 12 November 2008. The minimum variance portfolios are given in Table two (see the appendix). For the statically estimated portfolios we see that most of the weights are quite small, leading to a mostly balanced portfolio with six stocks in both the static portfolios having a weight with an absolute value greater than 0.1; whereas the dynamically estimated portfolio has more extreme weights with the diversified portfolio containing 14 stocks with absolute values greater than 0.1. Both static portfolios place a large weight on JNJ stock, whereas the dynamic portfolio places only half of its weight on JNJ. Interestingly, the dynamic portfolio places more than double the weight of the static stocks on WMT and MCD stocks. This could be because the western economy was entering a period of severe economic recession, with large numbers of job losses and other forms of financial distress. As such consumers will go to cheaper places to buy goods and services, increasing the demand for discount retailers and restaurants such as Walmart or McDonalds, which could make it advantageous to hold their stock in times of economic downturn.

6.2.4 2020 COVID-19 pandemic

The COVID-19 pandemic started in Wuhan, China in December 2019. It quickly spread triggering governments across the world to issue nationwide lock downs to slow down

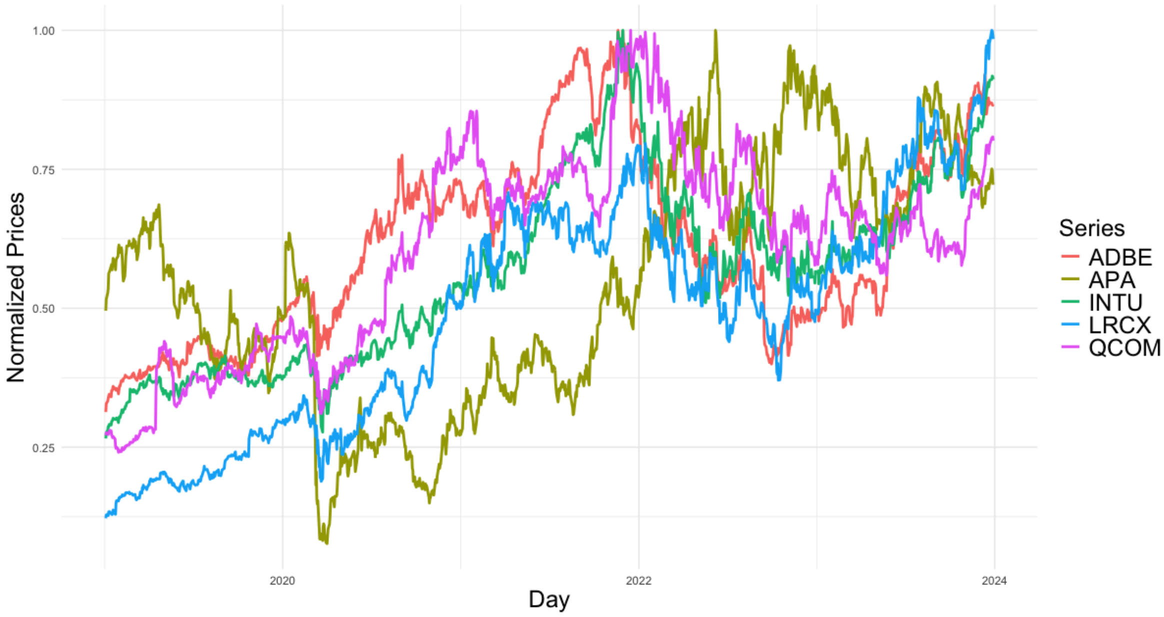

Figure 5: Plot of the normalized daily adjusted closing prices of five of the stocks in our diversified portfolio.

the spread of the virus, which had a large negative economic impact. We consider two portfolios with 30 stocks. The first portfolio consists of 30 large technology companies listed on the NASDAQ. The second portfolio is a diversified portfolio where we have 10 stocks from our first portfolio and the rest are large stocks from other industries included in the S&P 500. We compute the excess asset returns by using data downloaded from Yahoo Finance and the Fama-French data library (Fama [2023]). We consider daily data from 3 January 2019 to 29 December 2023. We then fit DSP-MFSV CAPM to the data and obtain 3000 posterior samples of the score time series. We can see in Figure 5 the observed time series of the daily adjusted closing price of five of the stocks in our diversified portfolio. The time series have some interesting commonalities, for example in early 2020 and late 2022 we see most of the stocks experience a noticeable decrease in their daily adjusted closing price.

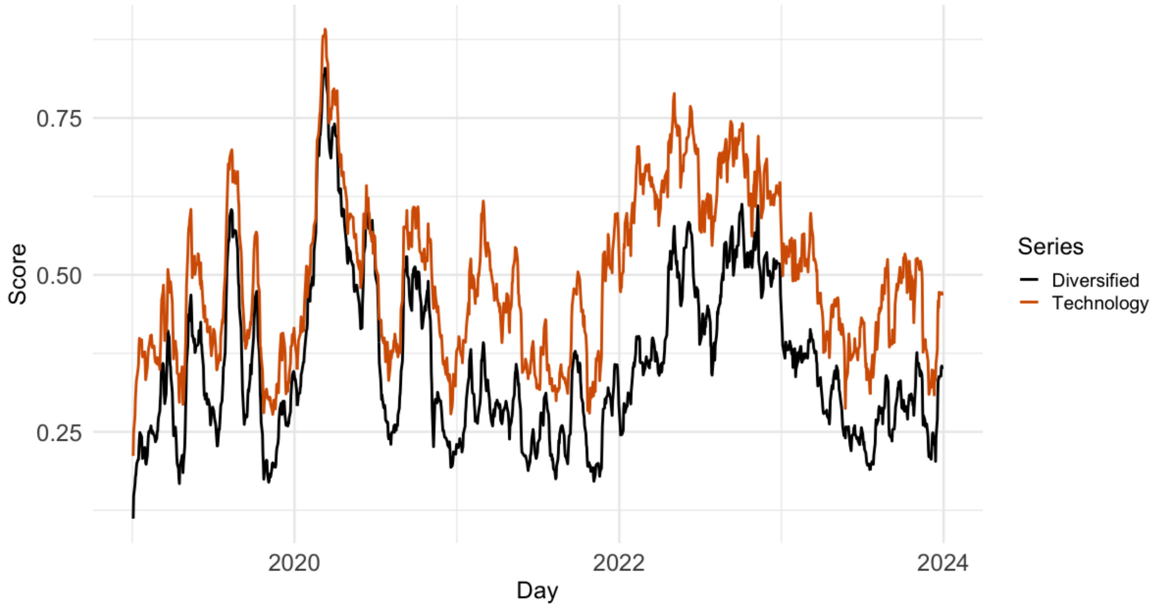

Figure 6: Plot of the estimated posterior score time series for our technology portfolio in vermillion and the estimated posterior score time series for our diversified portfolio in black. The posterior mean score time series is represented by the black line, and the 95% HDIs are represented by the boundary of the purple area.

6.2.5 Diversification risk

We now discuss the extent to which diversifying our first portfolio to obtain our second portfolio helped in protecting our portfolio from the economic impact of the COVID-19 pandemic. Having fitted our DSP-MFSV model we plot the time series of the estimated scores along with 95% HDIs. In Figure 6 we plot the posterior mean score time series of both portfolios and in Figure 7 we plot the posterior score of our diversified portfolio with the VIX index. In both Figure 6 and Figure 7 we see a large spike in our portfolios' correlations on 10 March 2020, with nationwide lock downs across the globe starting soon afterwards such as in the U.K, U.S, and Europe. We see that at the start of the time frame the diversified portfolio has a smaller correlation compared to the technology portfolio, but each series seems to share the same spikes in correlation. Furthermore, we can observe that both portfolios see an increase in their correlation at the start of 2020, reaching their peak in early March. Therefore, despite diversification in our portfolio we still suffer from the same shocks, demonstrating there is no advantage to

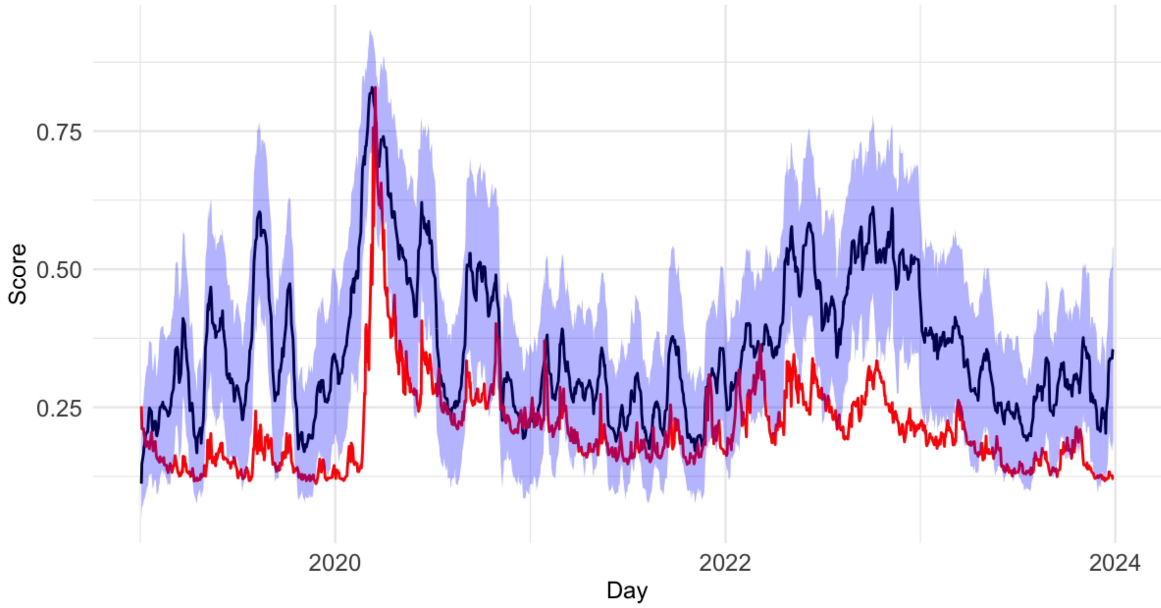

Figure 7: Plot of the estimated posterior score time series for our diversified portfolio in black and the VIX index in red. The 95% HDIs of our scores are represented by the boundary of the purple area. The VIX index is scaled to be between the smallest and largest posterior mean of the estimated score time series.

portfolio diversification in crisis periods. We observe several spikes in Figure 7. Firstly, we see a large spike, culminating in an overall correlation of 0.603 on 13 August 2019. This increase could be the result of several factors, such as the recently announced U.S. tariffs on Chinese goods and global signs of an economic downturn; these included the reduction of interest rates by the U.S. Federal Reserve announced at the end of July 2019, which was the first reduction in interest rates since 2008, perhaps indicating that the Fed believed an economic downturn would arrive soon and they were trying to encourage economic growth by making loans cheaper to encourage investment and spending, and thereby increase GDP. This was in combination with other signs of possible economic downturn, for example the inverted yield curve of U.S. treasures which was occurring at the time. There were also other international stresses such as the crashing of the Argentinian stock market on 12 August 2019. The biggest spike in Figure 7 occurs on the 9 March 2020 with a correlation of 0.83. We begin to see the increases starting in early 2020 with reports of the COVID-19 disease appearing in late 2019, with the World Health Organization declaring the COVID-19 outbreak a public health emergency of international concern on 30 January 2019 at which point the amount of the correlation level as identified by our method is already at 0.422. We then see several dramatic events happening in February, such as the increases in deaths internationally, and a large decline in international stocks markets from 24 to 28 February. The Federal Reserve then decreased interest rates on 3 March due to the growing concern regarding COVID-19, and the U.S. stock market saw a large decline on 9 March 2020. Despite our model peaking at 10 March 2020 the VIX index only peaked on 16 March 2020. This again shows the advantages of using our method to quantify the amount of economic stress through the lens of modeling the correlation within an investor's portfolio and can be used to give a quantitative tracker of the events of such economic downturns.

6.2.6 Minimum variance portfolio

See Section 5.2.3 for a brief introduction to minimum variance portfolios. We will now compare the minimum variance portfolios estimated by traditional static methods with a dynamically estimated minimum variance portfolio with respect to our diversified portfolio on the 10 March 2020, which is the day of the largest correlation in our diversified portfolio.

The minimum variance portfolios computed using two static methods and our DSP-MFSV CAPM are presented in Table three (see the appendix). The static CAPM covariance portfolio gives quite a balanced portfolio with only seven of the stocks having a weight with absolute value greater than one. The static sample covariance portfolio is more extreme with 12 stocks having an absolute value greater than one, and the dynamic minimum variance portfolio being the most extreme with 19 of its weights having an absolute value greater than one. Interestingly all three portfolios place a large weight on pharmaceutical stocks such as JNJ,MRK, and ABBV with the dynamic minimum variance portfolio placing more than twice the weight on some of these stocks compared to the static portfolios. This makes intuitive sense, as entering into a global pandemic there was an increased demand for medical goods and services including calls for research into vaccines fighting COVID-19, therefore there would be an increase in the stock price of such companies.

7 Conclusion

In summary we have proposed a novel approach for estimating time varying correlation matrices in a Bayesian fashion based upon dynamic shrinkage processes. To allow practitioners to derive meaningful information from time series consisting of even moderate dimension correlation matrices we propose a scalar score to summarize a given correlation matrix. Through a simulation study we show that our proposed model achieves desirable results in terms of RMSE and tight highest density posterior intervals when compared to the competing method. Through two real world examples we demonstrated the applicability of our model especially in providing novel insights into an investor's portfolio and established that portfolio diversification does not avail the problems caused by financial crises on an investor's portfolio. Finally, we compared the dynamically estimated minimum variance portfolio at the peak of each crisis with the traditional statically estimated minimum variance portfolios. Through this we observed that the dynamically estimated minimum variance portfolio has more extreme weights in a few companies compared to the more balanced statically estimated portfolios. Future work could include studying the theoretical properties of the proposed scalar score and expanding the framework to provide further insights into portfolio allocation.

References

Ole Barndorff-Nielsen, John Kent, and Michael Sørensen. Normal variance-mean mixtures and z distributions. *International Statistical Review/Revue Internationale de Statistique*, pages 145–159, 1982.

Carlos M Carvalho, Nicholas G Polson, and James G Scott. Handling sparsity via the horseshoe. In *Artificial Intelligence and Statistics*, pages 73–80. PMLR, 2009.

Yi-Ling Chen, Ming-Chun Wang, Jun-Biao Lin, and Ming-Chih Huang. How financial crises affect the relationship between idiosyncratic volatility and stock returns. *International Review of Economics & Finance*, 80:96–113, 2022.

Siddhartha Chib, Federico Nardari, and Neil Shephard. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341–371, 2006.

EF Fama. Production of us rm-rf, smb, and hml in the fama-french data library. *Chicago Both Paper*, (22-23), 2023.

Jennifer Foo and Dorota Witkowska. The 2020 covid-19 financial crisis impact on the european stock markets and economies. a preliminary analysis. *Folia Oeconomica Stetinensia*, 24(1):22–40, 2024.

Subhashis Ghosal and Aad van der Vaart. Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics*, pages 192–223, 2007.

Jim E Griffin and Philip J Brown. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188, 2010.

Darjus Hosszejni and Gregor Kastner. Modeling univariate and multivariate stochastic volatility in R with stochvol and factorstochvol. *Journal of Statistical Software*, 100 (12):1–34, 2021. doi: 10.18637/jss.v100.i12.

Leonidas Sandoval Junior and Italo De Paula Franca. Correlation of financial markets in times of crisis. *Physica A: Statistical Mechanics and its Applications*, 391(1-2): 187–208, 2012.

Gregor Kastner and Sylvia Frühwirth-Schnatter. Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408–423, 2014.

Daniel R Kowal, David S Matteson, and David Ruppert. Dynamic shrinkage processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81 (4):781–804, 2019.

Wen-Ling Lin, Robert F Engle, and Takatoshi Ito. Do bulls and bears move across borders? international transmission of stock returns and volatility. *Review of Financial Studies*, 7(3):507–538, 1994.

John Lintner. Security prices, risk, and maximal gains from diversification. *The Journal of Finance*, 20(4):587–615, 1965a.

John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1): 13–37, 1965b.

Jun S Liu and Ying Nian Wu. Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94(448):1264–1274, 1999.

Steve Lumby and Christopher Jones. *Investment Appraisal and Financial Decisions. 6th ed.* Cengage Learning Business Press, 1998.

Florence Merlevède, Magda Peligrad, and Emmanuel Rio. Bernstein inequality and moderate deviations under strong mixing conditions. In *High dimensional probability V: the Luminy volume*, volume 5, pages 273–293. Institute of Mathematical Statistics, 2009.

Abdelkader Mokkadem. Mixing properties of arma processes. *Stochastic processes and their applications*, 29(2):309–315, 1988.

Jan Mossin. Equilibrium in a capital asset market. *Econometrica: Journal of the Econometric Society*, pages 768–783, 1966.

Nicholas G Polson, James G Scott, and Jesse Windle. Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349, 2013.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL `https://www.R-project.org/`.

David Ruppert and David S Matteson. *Statistics and Data analysis for Financial Engineering*, volume 13. Springer, 2011.

William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.

Annastiina Silvennoinen and Timo Teräsvirta. Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations. Technical report, SSE/EFI Working Paper Series in Economics and Finance, 2005.

Bruno Solnik, Cyril Boucrelle, and Yann Le Fur. International market correlation and volatility. *Financial Analysts Journal*, 52(5):17–34, 1996.

Stephen John Taylor. Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time Series Analysis: Theory and Practice*, 1:203–226, 1982.

Ruey S Tsay. *Analysis of Financial Time Series*. John wiley & sons, 2005.

WH Wagner and SC Lau. The effect of diversification on risk. *Financial Analysts Journal*, 27:48–53, 1971.

Yaming Yu and Xiao-Li Meng. To center or not to center: That is not the questionan ancillarity–sufficiency interweaving strategy (asis) for boosting mcmc efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570, 2011.

Appendix

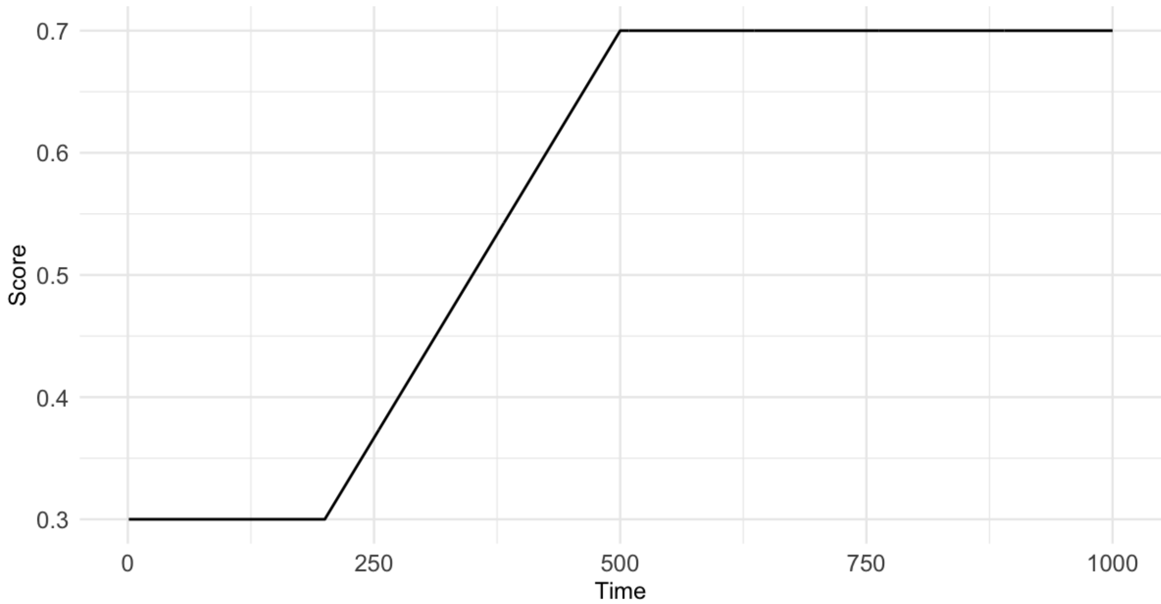

Figure 8: Plot of the correlation dynamics used in the simulation study of section 5.1

When the probability distribution of the innovations is $Z(\frac{1}{2}, \frac{1}{2}, 0, 1)$, then the density of η is given by $\frac{1}{\pi}\frac{e^{\frac{x}{2}}}{1+e^x}$. Then, $\mathbb{P}(\eta > u) = \int\limits_u^\infty \frac{1}{\pi}\frac{e^{\frac{x}{2}}}{1+e^x}dx$. Note that $1 + e^x \geq e^x$ which implies $\frac{1}{e^x} \geq \frac{1}{1+e^x}$ which leads to $\frac{e^{\frac{x}{2}}}{e^x} \geq \frac{e^{\frac{x}{2}}}{1+e^x}$ which finally gives, $e^{-\frac{x}{2}} \geq \frac{e^{\frac{x}{2}}}{1+e^x}$. This then gives $\mathbb{P}(\eta > u) \leq \int\limits_u^\infty \frac{1}{\pi}e^{-\frac{x}{2}}dx = \frac{1}{\pi}[-2e^{-\frac{x}{2}}]_u^\infty = \frac{2}{\pi}e^{-\frac{u}{2}}$. Similarly, $\mathbb{P}(\eta < -u) = \int\limits_{-\infty}^{-u} \frac{1}{\pi}\frac{e^{\frac{x}{2}}}{1+e^x}dx$. Note that $e^x \geq 0$ which implies $1 + e^x \geq 1$ which gives $1 \geq \frac{1}{1+e^x}$ which finally results in $e^{\frac{x}{2}} \geq \frac{e^{\frac{x}{2}}}{1+e^x}$. This then implies $\mathbb{P}(\eta < -u) \leq \frac{1}{\pi}\int\limits_{-\infty}^{u} e^{\frac{x}{2}}dx = \frac{1}{\pi}[2e^{\frac{x}{2}}]_{-\infty}^{-u} = \frac{2}{\pi}e^{-\frac{u}{2}}$. This then implies $\mathbb{P}(|\eta| > u) = \mathbb{P}(\eta > u) + \mathbb{P}(\eta < -u) \leq \frac{2}{\pi}e^{-\frac{u}{2}} + \frac{2}{\pi}e^{-\frac{u}{2}} = \frac{4}{\pi}e^{-\frac{u}{2}}$.

Note that when the innovations have an associated $N(0, \sigma^2)$ distribution this gives $var(h_t) = \frac{\sigma^2}{1-\phi^2} < \infty$. We have that $X_t = \phi X_{t-1} + \eta_t$, where $X_t := h_t - \mu$. This then tells us that $var(X_t) = \phi^2 var(X_{t-1}) + var(\eta_t)$.Note that $\mathbb{E}[\eta_t] = \int\limits_{-\infty}^{\infty} \frac{xe^{\frac{x}{2}}}{\pi(1+e^x)}dx$. Observe, $f(-x) = \frac{1}{\pi}\frac{e^{-\frac{x}{2}}}{(1+e^{-x})} = \frac{1}{\pi}\frac{e^{-\frac{x}{2}}}{e^{-x}(1+e^x)} = \frac{1}{\pi}\frac{e^{\frac{x}{2}}}{1+e^x} = f(x)$. Therefore the density is an even function which implies that the expectation is equal to zero and subsequently the variance of the innovation is given by the second moment of η_t. Since $|\phi| < 1$ the auto regressive processes are stationary therefore $var(X_t) - \phi^2 var(X_{t-1}) = E[\eta_t^2]$ implies $var(X_t) = \frac{E[\eta_t^2]}{1-\phi^2}$. Furthermore, $E[X_t] = E[h_t - \mu] = 0$. So we have $var(X_t) = E[X_t^2]$ which implies $E[X_t^2] = \frac{E[\eta_t^2]}{1-\phi^2}$ which implies $var(h_t) = \frac{E[\eta_t^2]}{1-\phi^2}$.

Note that the moment generating function of the $Z(\frac{1}{2}, \frac{1}{2}, 0, 1)$ innovations is $M(t) = \mathbb{E}[e^{tx}] = \int\limits_{-\infty}^{\infty} e^{tx} f(x)dx = \frac{1}{\pi} \int\limits_{-\infty}^{\infty} \frac{e^{(t+\frac{1}{2})x}}{1+e^x} dx$, Let $u = e^x$ which implies $ln(u) = x$ and $\frac{du}{dx} = xe^x$ which subsequently leads to $\frac{du}{u} = dx$. Therefore m $M(t) = \frac{1}{\pi} \int\limits_0^{\infty} \frac{u^{t+\frac{1}{2}}}{1+u} \frac{du}{u} = \frac{1}{\pi} \int\limits_0^{\infty} \frac{u^{t-\frac{1}{2}}}{1+u} du$. Now let $a = t + \frac{1}{2}$ which implies $a - 1 = t - \frac{1}{2}$. Then $M(t) = \frac{1}{\pi} \int\limits_0^{\infty} \frac{u^{a-1}}{1+u} du$. Now, let $v = \frac{u}{1+u} iff v(1+u) = u$ which implies $u = \frac{v}{1-v}$ which implies $\frac{du}{dv} = \frac{1}{(1-v)^2}$, by the product rule. This implies that $du = \frac{dv}{(1-v)^2}$. Therefore, $M(t) = \frac{1}{\pi} \int\limits_0^1 v^{a-1}(1 - v)^{-(a-1)}(1-v)\frac{1}{(1-v)^2} dv = \frac{1}{\pi} \int\limits_0^1 v^{a-1}(1-v)^{-(a-1)}(1-v^{-1}) dv = \frac{1}{\pi} \int\limits_0^1 v^{a-1}(1-v)^{-a} dv = \frac{1}{\pi} \int\limits_0^1 v^{a-1}(1-v)^{(1-a)-1} du = \frac{1}{\pi} B(a, 1-a) = \frac{1}{\pi} B(t+\frac{1}{2}, \frac{1}{2} - t) = \frac{1}{\pi} \frac{\Gamma(t+\frac{1}{2})\Gamma(\frac{1}{2}-t)}{\Gamma(1)} = \frac{1}{\pi}\Gamma(t + \frac{1}{2})\Gamma(\frac{1}{2} - t)$. Let $z = t + \frac{1}{2}$. Then, $M(t) = \frac{1}{\pi}\Gamma(z)\Gamma(1-z) = \frac{1}{\pi} \frac{\pi}{sin(\pi z)} = \frac{1}{sin(\pi z)}$, by Euler's reflection formula. Then this leads to $M(t) = \frac{1}{sin(\pi(t+\frac{1}{2}))} = \frac{1}{sin(\pi t + \frac{\pi}{2})} = \frac{1}{cos(\pi t)} = sec(\pi t)$.

Stock	Static CAPM covariance	Static Sample covariance	DSP-MFSV CAPM covariance
ADBE	-0.035	-0.021	-0.109
SAP	0.001	0.026	0.013
ERIC	-0.036	-0.023	-0.045
AAPL	-0.009	0.047	0.141
BRK-B	0.090	0.230	0.127
JNJ	0.317	0.367	0.162
PG	0.196	0.138	0.295
JPM	-0.071	-0.045	-0.037
XOM	-0.010	-0.045	0.081
BIDU	-0.013	-0.012	0.005
NOK	-0.026	0.013	-0.018
CVX	-0.038	-0.033	-0.133
PFE	0.062	0.030	-0.028
KO	0.173	0.067	0.167
DIS	-0.050	-0.087	-0.114
VZ	0.058	0.054	-0.004
LPL	-0.036	-0.056	-0.033
PEP	0.201	0.186	0.236
MRK	0.028	-0.059	-0.162
HD	-0.020	-0.036	-0.002
BAC	-0.049	-0.030	-0.065
UNH	-0.004	-0.038	-0.052
T	0.031	0.016	-0.069
CMCSA	-0.039	-0.054	-0.171
NVDA	-0.029	-0.009	-0.004
INTC	-0.032	0.015	-0.016
MCD	0.139	0.168	0.365
MMM	0.064	0.068	0.131
WMT	0.143	0.112	0.362
ORCL	-0.007	0.000	-0.025

Table 2: Global minimum variance portfolios using three different approaches. The first approach fits a static CAPM and then uses the model implied covariance matrix. The second approach uses the static sample covariance matrix. The third approach uses the model based covariance matrix from the 12th of November 2008 from fitting a DSP-MFSV CAPM model.

Stock	Static CAPM covariance	Static Sample covariance	DSP-MFSV CAPM covariance
ADBE	-0.047	-0.005	-0.061
QCOM	-0.039	0.022	-0.024
APA	-0.018	-0.022	-0.008
INTU	-0.078	-0.056	-0.016
LRCX	-0.070	-0.045	-0.102
ASML	-0.061	0.032	0.168
INFY	0.029	0.120	0.210
BRK-B	0.130	0.174	0.206
MMM	0.050	0.074	0.120
V	-0.007	0.109	-0.177
JPM	-0.017	0.033	-0.085
JNJ	0.206	0.201	0.392
UNH	0.036	-0.054	-0.088
PG	0.168	0.113	0.004
MA	-0.045	-0.142	-0.269
XOM	0.018	0.133	0.046
HD	0.001	-0.011	-0.132
PFE	0.090	0.026	0.154
ABBV	0.097	0.095	0.269
MRK	0.138	0.113	0.358
KO	0.154	0.128	-0.017
PEP	0.145	-0.101	-0.127
BAC	-0.038	-0.127	-0.155
WMT	0.158	0.220	0.229
NOW	-0.037	0.000	0.062
CSCO	0.025	-0.050	-0.180
CVX	0.000	-0.078	-0.045
INTC	-0.026	-0.028	-0.178
PANW	-0.003	0.085	0.240
CMCSA	0.041	0.042	0.207

Table 3: Global minimum variance portfolios using three different approaches. The first approach fits a static CAPM and then uses the model implied covariance matrix. The second approach uses the static sample covariance matrix. The third approach uses the model based covariance matrix from 10 March 2020 from fitting a DSP-MFSV CAPM model.