Bayesian Factor Models with Dynamic Shrinkage for Time-Varying Correlation Matrices: Applications to Financial Crises

Daniel Andrew Coulson[1], David S. Matteson[1], Martin T. Wells[1]

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

Abstract

In this paper, we propose a novel approach to quantifying the risk associated with the co-movement of stocks in a portfolio over 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 and a multivariate factor stochastic volatility process for the observation error covariance matrices. To summarize the information in correlation matrices, we create a simple and intuitive scalar score. We prove a posterior concentration result to theoretically validate our modeling approach. Through a simulation study, we demonstrate that our estimation approach achieves superior performance in several key metrics and has the ability to rapidly adapt to changing market conditions, outperforming 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 that of 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 offers a

novel overall measure of correlation in a system, which can be utilized by both practitioners and researchers as a tool to quantify correlation risks in a portfolio and the degree of toxicity in a financial system.

Keywords: Bayesian time series, dynamic shrinkage prior, financial risk management, minimum variance portfolio, multivariate factor stochastic volatility, posterior concentration, uncertainty quantification

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 market instability is essential for investors to gain a fuller understanding of the risks associated with their portfolio.

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

Therefore, understanding the correlations within a portfolio is important to mitigate the impact of periods of greater market volatility. The benefits of diversification have been highlighted 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 during times of financial crisis.

The standard method for computing time-varying correlation matrices of a portfolio relies on rolling correlation estimators, which suffer from several problems. Lag effects induced by using a rolling window that contains several unrelated data points 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 due to over-parameterization, which naturally leads to estimators with higher variance than is usually desired. Rolling estimators also cannot estimate instantaneous correlation matrices, which is undesirable, especially when alternative methods exist that provide instantaneous estimates. Another popular approach to modeling time-varying correlation matrices is the Dynamic Conditional Correlation (DCC) model. To fit a DCC model, you first fit a univariate GARCH model to each of the time series to obtain the scaled residuals, and then estimate the correlation matrices from the residuals. However, due to the two-stage approach, if there is model misspecification in the first stage, this will 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. When making sensitive decisions, additional knowledge about the uncertainty of our correlation matrices could lead to more prudent decision-making, particularly regarding portfolio diversification.

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

To account for these difficulties, we propose a novel Bayesian approach to estimating correlation matrices from multivariate factor models. We utilize a Dynamic Shrinkage Prior (DSP) Process (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 that the state variables across assets are independent, with dependence across series through the unobserved observation error covariances. Through the DSP process, the model can adapt locally to changing market conditions and provide tighter credible intervals.

A problem with estimating time series of correlation matrices is how to summarize this information after the matrices have been estimated, so it can aid decision making. The most popular solution is to plot the estimated pairwise correlations through time; however, this can become cumbersome and ultimately uninformative in even moderate dimensions. Therefore, we propose a novel scalar score which ranges from -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 on 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 tracking only a scalar quantity rather than a matrix.

This paper will proceed as follows. In Section 2, we will introduce the methodology including the specification of priors and observation equations. Section 3 will discuss the estimation of correlation matrices and our correlation matrix summarization 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 the results of our proposed method, both in a simulation study and in 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 concludes and suggests 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 a focus on applications to asset returns.

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 utilize 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 US Treasury bills, are considered risk-free, although such assets are not truly risk-free.

Definition 2. *The excess market return is the return you can earn by investing in the market portfolio, which is the 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:*

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

Here, $t \in \mathbb{N}$, and $\boldsymbol{r}_t = (r_{1,t}, ..., r_{N,t})^{'}$ is the vector of N excess asset returns at time t. The vector of intercept terms is given by $\boldsymbol{\alpha}_t = (\alpha_{1,t}, ..., \alpha_{N,t})^{'}$. The factors in the model are $F_{1,t}, ..., F_{z,t}$, which, in this paper, are observed, and the vectors $\boldsymbol{\beta}_{1,t}, ..., \boldsymbol{\beta}_{z,t}$ 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 are built 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 that 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 priors.

DSP processes extend this idea to the four-parameter Z-distribution, which provides a natural extension. This is because the Z distribution can be written as normal mean-scale mixtures (Barndorff-Nielsen et al. [1982]), which can provide additional flexibility in the shape of the shrinkage; this 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.

Instead, Kowal et al. [2019] propose a prior process for the amount of shrinkage which has the advantage of having the shrinkage 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 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 Theorems 2 and 3 in Kowal et al. [2019]. By modeling the shrinkage through a prior stochastic process, DSP processes utilize previous observations to determine an appropriate amount of shrinkage, but they 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 that combines DSP with MFSV as DSP-MFSV. The observation equation and prior distribution specification are discussed in Sections 2.2.1 and 2.2.2, respectively. We discuss the computation of sampling 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 developed by 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 our portfolio of N assets at time $t \in \mathbb{N}$ is given by $r_{a,t}$, and $r_{M,t}$ is the excess market return at time t. Furthermore, $\epsilon_{a,t}$ is the idiosyncratic observation error, with $\alpha_{a,t}$ and $\beta_{a,t}$ being unobserved state variables. We assume that the state variables are independent across assets and we 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 that the process of time-varying covariance matrices is driven by a small number of latent factors. In particular,

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

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 that we may write the covariance matrix of $\boldsymbol{\epsilon}_t$ as $\boldsymbol{\Lambda} \tilde{\boldsymbol{\Sigma}}_{\mathbf{t}} \boldsymbol{\Lambda}^{'} + \bar{\boldsymbol{\Sigma}}_{\mathbf{t}}$.

2.2.2 Priors

In this section, we discuss our prior distribution specification for the observation equation in Section 2.2.1. We assume that the state variables of the assets in our model evolve independently of each other, and for each state variable in a given asset's observation equation, we further assume that these too are independent and evolve according to their own specific Dynamic Shrinkage Prior Process. These assumptions seem valid because α and β of one stock should not influence the α and β of another stock, since they could be considered to define the properties of a given stock. Then for a given asset a in our portfolio, we have:

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

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

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

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

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

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

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

$$\mu_M \sim N(0,100), \frac{\phi_M + 1}{2} \sim Beta(10,3), \sigma_M^2 \sim Ga(\frac{1}{2}, \frac{1}{2}). \tag{12}$$

The prior specification for the other state variables in a given observation equation is the same as above. By stating that our state variables evolve according to a Normal Random Walk (5) the problem becomes tractable and 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 $\tau_{a,0}^2$ determines the global level of shrinkage in all state variables and over all of time. Similarly, $\tau_{\beta_a}^2$ determines the amount of shrinkage over all time points of the parameter β of asset a in our portfolio, with the shrinkage implied by the half-Cauchy distribution. Finally, $\lambda_{\beta_a,t}^2$, the local shrinkage parameter, determines the amount of shrinkage of the β parameter of asset a in our portfolio 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 (7). We assume that the log-variance process evolves according to an autoregressive process of order (8), but with Z-distributed errors (9), rather than the typical normally distributed errors, due to the Z-distribution's ability to induce shrinkage. In particular, we utilize horseshoe-shaped shrinkage by specifying $Z(\frac{1}{2}, \frac{1}{2}, 0, 1)$ due to its symmetric level of shrinkage, which induces primarily strong or weak shrinkage. See Kowal et al. [2019] for a discussion of some of the other shrinkage types available. For the observation error covariance, we use

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

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

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

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

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

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

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

As can be seen in (13)-(15), the entries of the conditional covariance matrices from Section 2.2.1 follow independent stochastic volatility processes of order one, which allow the model to capture time-varying covariances of the observation errors of our observation equations. The prior utilizes normal-gamma shrinkage priors as in 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 that 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 that 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 for each observed time point. We then standardize these covariance matrices to obtain correlation matrices, and finally we use our proposed scalar score to summarize the correlation matrices in the series. After this procedure we will have posterior draws consisting of time series of the estimated scalar scores 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 an approximate low-rank structure. This low-rank structure, combined with dynamic shrinkage, provides low-variance covariance estimators.

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

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

$\implies var(\boldsymbol{r}_t) = var(\boldsymbol{\alpha}_t) + var(r_{M,t}\boldsymbol{\beta}_t) + var(\boldsymbol{\epsilon}_t)$, 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 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 level of correlation in a port-

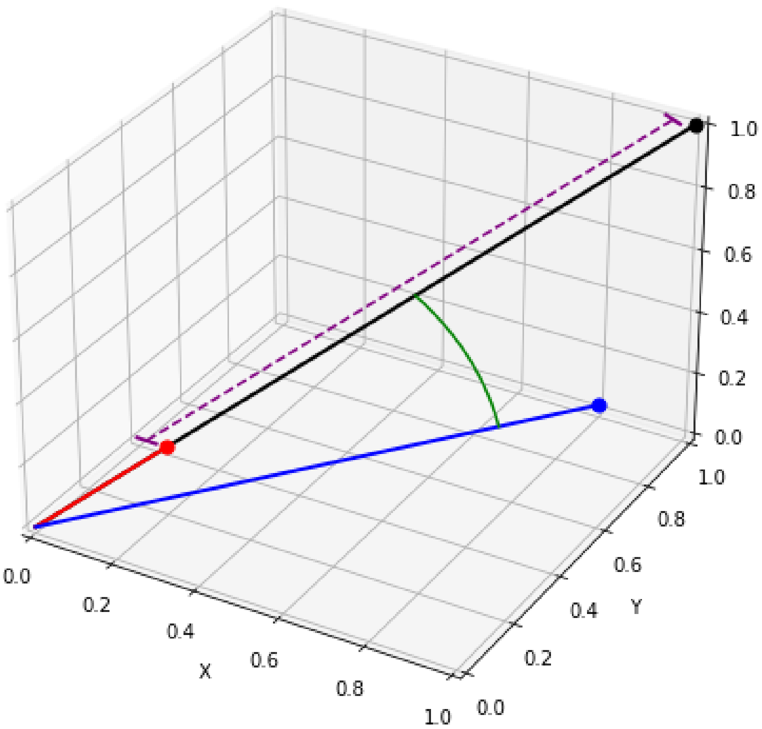

Figure 1: Plot of three vectors. The vector $(1, 1, 1)'$ is in black, the vector $(0.2, 0.2, 0.2)'$ is in red, and the vector $(0.9, 0.8, 0.2)'$ is in blue. The difference in length between $(0.2, 0.2, 0.2)'$ and $(1, 1, 1)'$ is represented by the purple dashed line, and the angle between the vectors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

folio. 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 $\boldsymbol{b}$ measures how much of the vector $\boldsymbol{a}$ is in the direction of the 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 the scalar projection takes into account both the length of the vector $\boldsymbol{a}$ and 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. This is 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 all-ones vectors. Therefore, to measure how close a matrix is to the all-ones matrix, we propose to measure how close each column vector is to the all-ones vector. In Figure 1, we can see that the $(0.2, 0.2, 0.2)'$ vector is perfectly aligned with the $(1, 1, 1)'$ vector, but it is noticeably shorter. Consequently, our proposed summary should take into account the length of the column vector. We also have the $(0.9, 0.8, 0.2)'$ vector. This vector is longer than the $(0.2, 0.2, 0.2)'$ vector, but it has a non-zero angle with the all-ones vector. Thus, to see how close a given column of a correlation matrix is to the all-ones vector, we also need to consider the angle between the column vector and $(1, 1, 1)'$. A quantity that accounts for both aspects is the scalar projection of the respective column of the correlation matrix onto the all-ones vector. Based on 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*

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

This score is the sum of the scalar projections of the columns of the correlation matrix onto the all-ones vector, scaled to the range $[-1, 1]$. This allows the score to not depend on the dimension of the matrix. This score builds upon the notion that correlation matrices can be seen as matrices whose entries are cosines of angles between vectors. 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, $score(\boldsymbol{A}) = \frac{1}{N(N-1)} \sum\limits_{j=1}^{N} [c(N-1)+1] - N = \frac{1}{N(N-1)}(cN(N-1) + N - N) = c$.*

Property 2. *The $score(\cdot)$ is invariant with respect 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, $\frac{1}{N(N-1)}(\sum\limits_{j=1}^{N}[c(N-1)+1] - N) = c$. Then if we add an additional row and column where all the pairwise correlations are c we get,*

$score(\boldsymbol{B}) = \frac{1}{N(N+1)}(\sum_{j=1}^{N+1}[cN+1] - (N+1)) = c.$

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 equal to 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

In this section, we state our main posterior concentration result and discuss its implications for this work. Posterior contraction is an important notion of Bayesian optimality, since it describes how the posterior distribution focuses around the true parameter with increasing data. This process ensures that Bayesian methods can identify the correct model over time. The contraction result theoretically validates DSP-MFSV, demonstrating that it can perform a consistent estimation, allowing for reliable inference such as the construction of valid credible intervals. We provide the proof strategy in this section and the full proof with all technical details in the Appendix.

Theorem 1. *Let $\epsilon_n = C_\epsilon n^{-\frac{1}{2+C_\theta}}$, where $C_\epsilon > 0, C_\theta \in \mathbb{N}$ are constants. Define the set Θ_n as the set of all $\theta \in \Theta$ such that the following hold:*

1. $|\mu_M| \leq log(n)$,
2. $|\phi_M| \leq \sqrt{1 - \frac{1}{log(log(n))}}$,
3. $\sigma_M^2 \leq log(n)^{0.5}$,
4. $||\Lambda||_F \leq log(n)$,
5. $\max_{1 \leq k \leq r} |\tilde{\mu}_k| \leq log(n)$,

6. $\max_{1\le k\le r} |\tilde{\phi}_k| \le \sqrt{1-\frac{1}{log(log(n))}}$,

7. $\max_{1\le k\le r} \tilde{\sigma}_k^2 \le log(n)^{0.5}$,

8. $\max_{1\le i\le N} |\bar{\mu}_i| \le log(n)$,

9. $\max_{1\le i\le N} |\bar{\phi}_i| \le \sqrt{1-\frac{1}{log(log(n))}}$,

10. $\max_{1\le i\le N} \bar{\sigma}_i^2 \le log(n)^{0.5}$,

11. $\max_{1\le i\le N} |\mu_{\alpha_i}| \le log(n^4)$,

12. $\max_{1\le i\le N} |\mu_{\beta_i}| \le log(n^4)$,

13. $\max_{1\le i\le N} |\phi_{\beta_i}| \le \sqrt{1-\frac{1}{log(log(n))}}$,

14. $\max_{1\le i\le N} |\phi_{\alpha_i}| \le \sqrt{1-\frac{1}{log(log(n))}}$.

Then, for every $M_n \to \infty$, we have

$$\mathbb{P}_{\theta_0}^{(n)} \Pi_n(\theta \in \Theta_n : ||R_\theta - R_{\theta_0}||_F \ge M_n \epsilon_n | X^{(n)}) \to 0,$$

where $R_\theta - R_{\theta_0} = \frac{1}{n}\sum_{t=1}^{n}(R_{\theta,t} - R_{\theta_0,t})$, with $R_{\theta,t}$ and $R_{\theta_0,t}$ denoting the estimated and true correlation matrices at time t, respectively.

Theorem 1 implies the following posterior concentration result about $score(R_\theta)$.

Corollary 2. *Under the conditions of Theorem 1, it follows that*

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

Theorem 1 and Corollary 2 establish posterior contraction results for DSP-MFSV CAPM. These results theoretically guarantee that as we observe more data over time, DSP-MFSV CAPM can reliably recover the true correlation dynamics of our stock portfolio. In addition to the empirical results in Section 6, this theoretically justifies the use of DSP-MFSV CAPM to model a financial system and risk management tasks.

To prove this result, we rely on Theorem 1 of Ghosal and van der Vaart [2007], which requires us to show that our model satisfies three conditions before we can apply the theorem. The first condition is that we can construct consistent tests, that is, tests with exponentially decaying Type 1 and Type 2 error probabilities. The second condition is an upper bound on the metric entropy of the set Θ_n with respect to the Frobenius norm of the difference between the average estimated and true correlation matrices, which bounds the growth of the size of the parameter space near the true parameter set. The third and final condition is a prior mass result, which requires that the prior mass does not increase too quickly in areas far away from the true parameter set, relative to the amount of prior mass near the true parameter set.

To prove the first condition, we first deduce the mixing properties of the prior correlation matrix process, which allow us to invoke Theorem 1 of Merlevède et al. [2009] to establish a Bernstein-type inequality on an arbitrary entry of the matrix $\frac{1}{n}\sum_{t=1}^{n}(R_{\theta,t} - R_{\theta_0,t})$. This Bernstein-type inequality then allows us to construct a consistent test which consists of a threshold on the largest entry in terms of the absolute value of the matrix $\frac{1}{n}\sum_{t=1}^{n}(R_{\theta,t} - R_{\theta_0,t})$.

To prove the metric entropy bound, we first derive bounds on the probability of each of the parameters in DSP-MFSV CAPM lying outside of the sieve, with each bound tending to zero in the limit. Now that we have shown that the probability of the sieve tends to one, we need to derive an upper bound on the metric entropy. To do this, we use the upper bounds on the parameters contained within the sieve to upper bound the entry-wise partial derivatives of a covariance matrix at some arbitrary time t with respect to each of the parameters in DSP-MFSV CAPM. Recognizing that Θ_n is a compact set and the fact that the mapping from the parameter space to some arbitrary entry in the covariance matrix is continuous on $\mathbb{R}^{C_1 n + C_2}$, for some $C_1, C_2 > 0$, we can invoke the Extreme Value Theorem. We can then utilize our partial derivative upper bounds, in combination with the Mean Value Theorem, to bound the Frobenius norm of $\frac{1}{n}\sum_{t=1}^{n}(R_{\theta,t} - R_{\theta_0,t})$. We then construct a superset of the Frobenius norm ball of our sieve by using our upper bound on $\frac{1}{n}\sum_{t=1}^{n}(R_{\theta,t} - R_{\theta_0,t})$. From this, we can then use a standard covering number upper bound applied to a ball in $\mathbb{R}^{C_1+C_2n}$, which, after some algebra, allows us to establish the upper bound on the metric entropy.

To prove the third condition, we trivially upper bound the prior mass on areas of the parameter space too far away from the true parameter set by 1. Therefore, it remains to lower bound the prior mass on parameters that are close to the true parameter. This closeness is defined by the set of all parameters which satisfy an upper bound on the Kullback-Leibler (KL) divergence between the true parameter space and the parameter values of interest, and an upper bound on the variance of the log-likelihood ratio between the true parameter set and some parameter set satisfying this condition. To lower bound the prior mass of such a set, we first construct a simpler subset. Then we lower bound the probability of this subset, which lower bounds the superset by monotonicity of measure. To construct the subset, we first derive the log-likelihood function of our vector of returns at a given time t, and utilize a second-order Taylor expansion of the KL divergence, which results in a quadratic form involving a Hessian matrix. By using Weyl's inequality and the Extreme Value Theorem, we can bound the Hessian matrix in Loewner order, which results in an upper bound on the KL divergence. To upper bound the variance of the log-likelihood, we first derive an upper bound in terms of some constant multiplied by the expectation under the true parameter set of some arbitrary moment of the absolute value of the log-likelihood ratio. We then exploit properties of normal log-likelihood ratios to obtain an upper bound on the variance of the log-likelihood ratio. We can then use these upper bounds to justify the construction of a subset. To lower bound the probability of this subset, we first state and prove a theorem which shows that the definite integral over a symmetric interval of a continuous function which is strictly positive at the center of the interval is lower bounded by some constant. We then invoke this theorem, in combination with our prior distributions with strictly positive density, to lower bound the probability, which gives the desired result.

As the three conditions are satisfied, we can simply invoke Theorem 1 of Ghosal and van der Vaart [2007] and apply it to our model of the correlation matrix process. This result trivially holds for our scalar score of the correlation matrices, which is proved by the use of the triangle inequality and the Cauchy-Schwarz inequality.

5 MCMC sampler

All of our code is written in the R programming language (R Core Team [2023]), using and building on the code from 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 that 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 in Hosszejni and Kastner [2021] uses the Ancillarity Interweaving Strategy proposed in Yu and Meng [2011]. The application of this sampling technique to stochastic volatility models was discussed in Kastner and Frühwirth-Schnatter [2014]. To improve computational efficiency, Hosszejni and Kastner [2021] interfaces R with 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 are discussed in the next paragraph. For sampling the state variables, we use the sampler of Kowal et al. [2019]. However, we make some important changes. First, since we have several assets that have a joint model (through the observation error), we sample several time series simultaneously. Second, 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 to improve 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 efficiency due to the computational ease of sampling from Pólya-gamma distributions. This is based on 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 $\boldsymbol{\beta} = (\alpha_1, \beta_1, \ldots, \alpha_T, \beta_T)'$, $\boldsymbol{X} = \text{blockdiag}\big((1, r_{M,t})_{t=1}^T\big)$, $\boldsymbol{\Sigma}_\omega = \text{diag}\Big(\tau_0^2\tau_\alpha^2\lambda_{\alpha,1}^2,\ \tau_0^2\tau_\beta^2\lambda_{\beta,1}^2,\ \ldots,\ \tau_0^2\tau_\alpha^2\lambda_{\alpha,T}^2,\ \tau_0^2\tau_\beta^2\lambda_{\beta,T}^2\Big)$, $\boldsymbol{\Sigma}_\epsilon = \text{diag}\Big(\{\sigma_t^2\}_{t=1}^T\Big)$, $\boldsymbol{\ell}_\beta = \boldsymbol{X}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{y} = [\frac{y_1}{\sigma_1^2}(1, r_{M,1}), \ldots, \frac{y_T}{\sigma_T^2}(1, r_{M,T})]'$, let 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 δ_{ij} is the Kronecker delta, $\mathbf{1}$ denote an indicator function, and $\boldsymbol{I}_2$ denotes the 2x2 identity matrix. The posterior distribution of the state variables of a single asset is given by $\boldsymbol{\beta} \sim N(\boldsymbol{Q}_\beta^{-1}\boldsymbol{\ell}_\beta, \boldsymbol{Q}_\beta^{-1})$, where $\boldsymbol{Q}_\beta = \boldsymbol{X}'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{X} + (D_2' \bigotimes \boldsymbol{I}_2)\boldsymbol{\Sigma}_\omega^{-1}(D_2 \bigotimes \boldsymbol{I}_2)$. 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 $\boldsymbol{Q}_\beta E[\boldsymbol{\beta}] = \boldsymbol{\ell}_\beta$, which gives $E[\boldsymbol{\beta}] = \boldsymbol{Q}_\beta^{-1}\boldsymbol{\ell}_\beta$. 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 yields the variance of $\boldsymbol{\beta}$. We then use $var(\boldsymbol{\beta}) + E[\boldsymbol{\beta}]E[\boldsymbol{\beta}]' = E[\boldsymbol{\beta}\boldsymbol{\beta}']$ 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 MFSV model assumes that the covariances are driven by a small number of latent factors. The sampling of the idiosyncratic variances utilizes the same sampling procedure as univariate stochastic volatility processes. The factorstochvol package also utilizes the Ancillarity Interweaving Strategy of Yu and Meng [2011], and offers 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 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 $1 \leq t \leq T$. This 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 evaluate our proposed methodology, we perform a simulation study in Section 6.1. We also apply our methodology to two real-world examples of financial crises in Section 6.2 to observe the impact, if any, that portfolio diversification has on mitigating such risk. Additionally, we construct minimum variance portfolios and assess how the dynamically estimated portfolios compare 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 based on 100 simulations. For a given simulation, we fixed the length of all time series to 1000 time points and the number of simulated assets to 30. We set all pairwise correlations to be equal. We then construct the time series of the scores. We use the same time series of scores in each simulation as displayed in Figure 8 in the Appendix. Since at a particular time point all pairwise correlations are equal, the associated correlation matrix has a score equal to any given entry. We then construct the associated model-based covariance matrices by fixing the variances of the 30 excess asset returns to two. The remainder of our simulation construction is based on the time-varying parameter CAPM.

We simulate the time series of the αs and βs for each asset from independent multivariate normal distributions with mean vector $(0,1)'$ and diagonal covariance matrix with entries equal to 0.1. The time series of the excess market returns is 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 discussed in the previous paragraph. Following this, we simulate the observation errors from a multivariate normal distribution with mean vector equal to the zero vector and covariance matrix computed at the respective time point as discussed in the previous paragraph.

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

For each simulation, we computed the Root Mean Squared Error (RMSE) for each of the fitted models. For Bayesian models, we computed the empirical coverage and the mean empirical credible interval width. The RMSE 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 that contains the observations of the true time series, and $\hat{\boldsymbol{x}}$ is the estimate of the time series $\boldsymbol{x}$, and T denotes the length of the time series. The RMSE measures the accuracy of estimates, with a value of zero corresponding to perfect estimation.

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

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

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

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 gives the average width of the estimated HDIs in 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. DSP-MFSV CAPM also attains 95% coverage as desired, with noticeably tighter HDIs when compared to the normal approximation from the MFSV model, whose HDIs 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. First, we show that in two major financial crises in the 21st century, the U.S. subprime mortgage crisis (Sections 6.2.1-6.2.3) and the 2020 COVID-19 pandemic (Sections 6.2.4-6.2.6), stock portfolio diversification does not protect investors from correlation risks induced by such high volatility events. Second, 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 real-world examples.

6.2.1 U.S. subprime mortgage crisis

The U.S. subprime mortgage crisis occurred from 2007 to 2010 and originated from the U.S. housing bubble. In particular, the securitization of mortgages, including the infamous collateralized debt obligations (CDOs), played a crucial role in the crisis, as these instruments, although highly rated, were of vastly higher risk than advertised. We consider two portfolios, each 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 stocks from a range of industries within the S&P 500 stock index. We compute the excess asset returns using data downloaded from Yahoo Finance and the Fama-French data library (Fama [2023]). We then fit the DSP-MFSV CAPM to daily adjusted closing price data from 4 January 2006 to 31 December 2009. From this we obtain 3,000 saved posterior samples from 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 and shared behavior of these price time series, such as the sharp decrease in adjusted closing prices in late 2008.

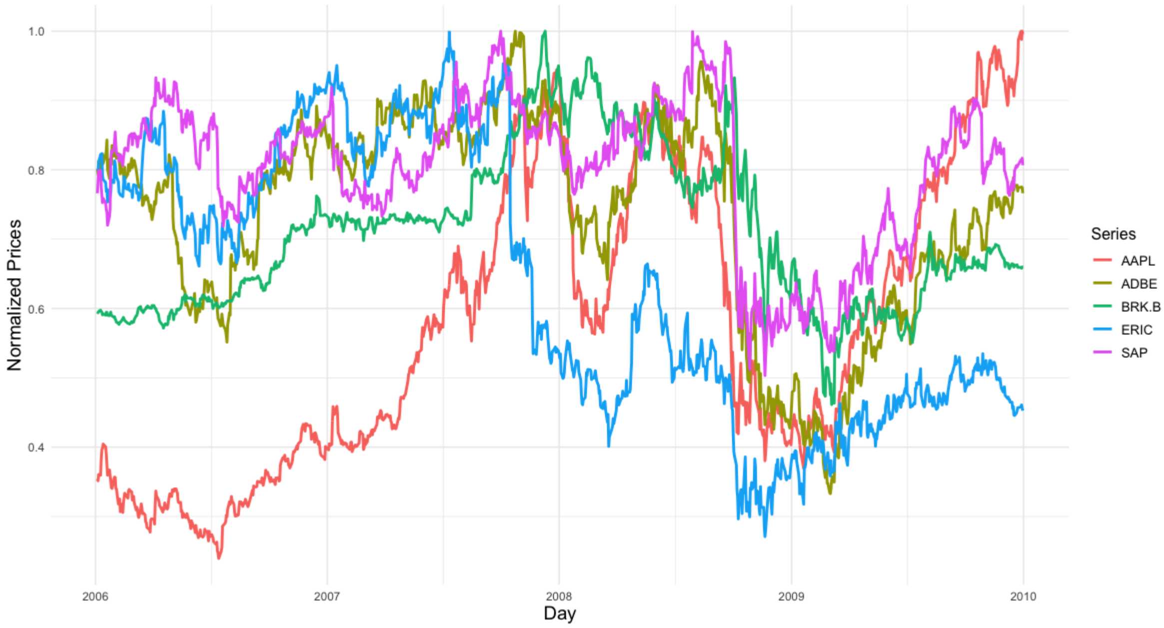

Figure 2: Plot of the daily normalized adjusted closing prices of five stocks from 4 January 2006 to 31 December 2009.

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 generally has a larger overall level of correlation than 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 than 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. They culminate in the large spikes in the correlations of both portfolios, which reached their peak on 12 November 2008,

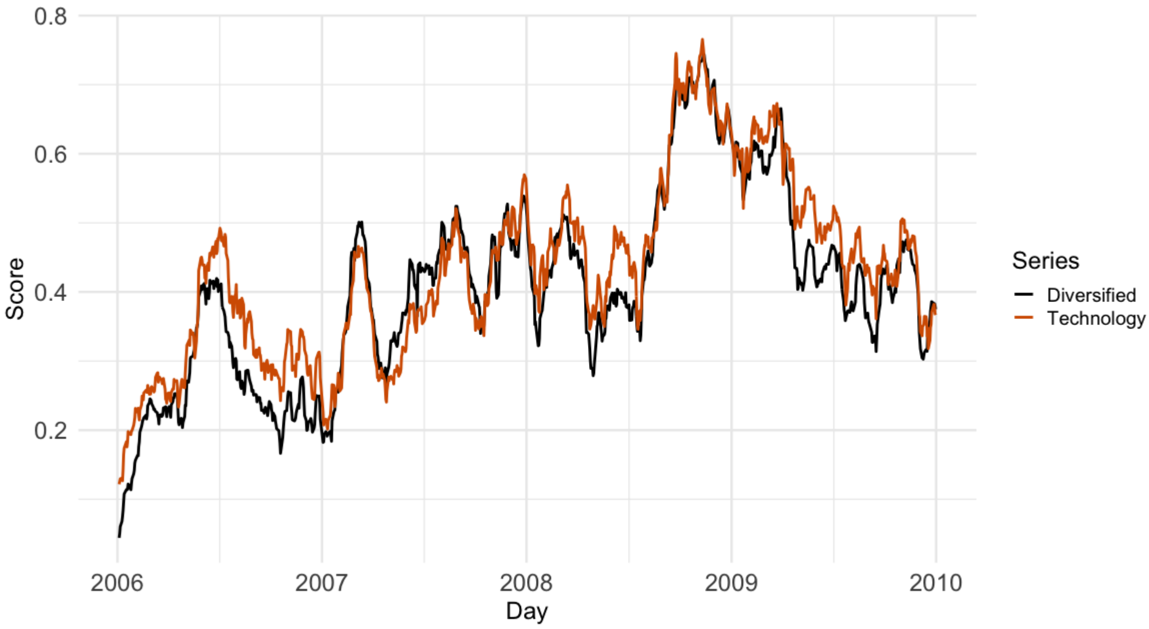

Figure 3: Plot of the estimated posterior mean score time series for the technology portfolio in vermilion and our estimated posterior mean score time series for the diversified portfolio in black.

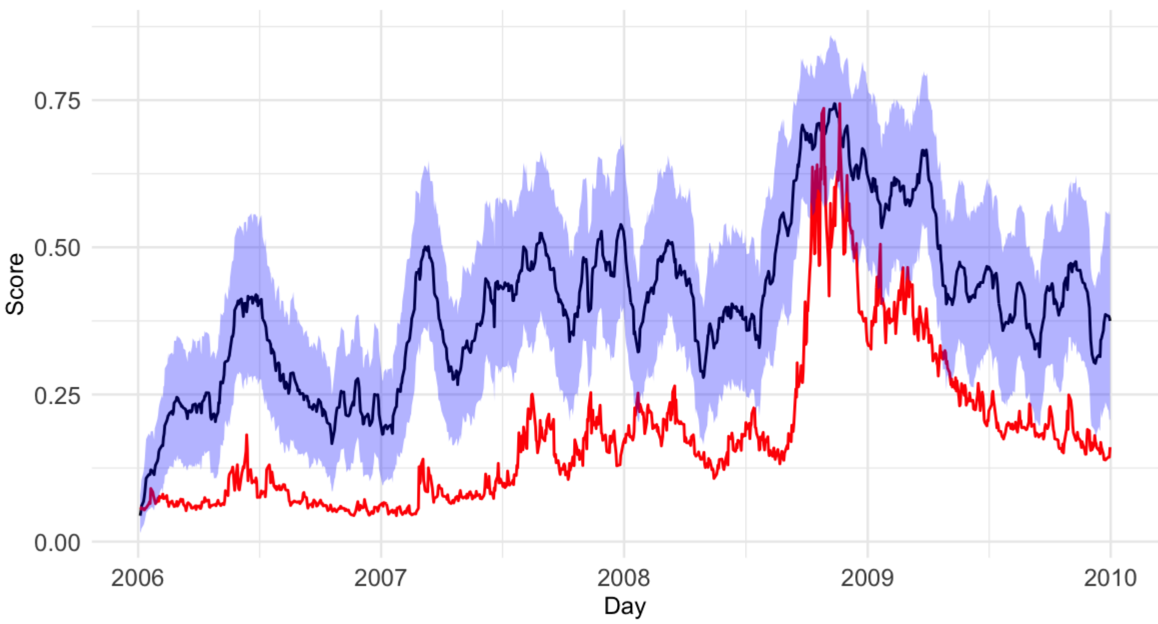

Figure 4: Plot of the estimated posterior score time series for our diversified portfolio in black and the VIX index in red. The 95% HDIs of our scores are represented by the boundary of the purple area. The VIX index is scaled to be between the smallest and largest posterior means of the estimated score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

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 are still affected by shocks in both portfolios, and during periods of the most intense economic stress, both portfolios see a similarly 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 emerging 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 may have been triggered by 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 2007 and continuing into 2008. The increase in correlation starts at approximately 0.329 on 22 July 2008. The correlation 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 Lehman 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. Figure 4 also illustrates that there is additional information beyond measures such as the VIX index, which provides some measure of the toxicity of the economy. Furthermore, Figure 4 provides a more accurate timeline of developments in the financial markets compared to the VIX, despite the fact that the VIX is constructed from a broader collection of assets, namely S&P 500 index options.

6.2.3 Minimum variance portfolio

The minimum variance portfolio is the portfolio that solves the problem $\min_{\boldsymbol{\omega}} \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}$, where $\boldsymbol{\omega}$ is a vector that sums to one and represents the proportion of each stock in which we should hold a long or short position, 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 that minimizes volatility, irrespective of expected return. Usually, this is estimated statically either by the covariance matrix derived from a static CAPM or by the sample covariance matrix. We compare the static global minimum variance portfolios with the dynamic global minimum variance portfolio, with respect to our diversified portfolio, at the time of maximum posterior mean correlation matrix score identified by 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. 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. On the other hand, the dynamically estimated portfolio has more extreme weights, 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 it. Interestingly, the dynamic portfolio places more than double the weights on WMT and MCD stocks compared to the static portfolios. 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 and McDonald's, which could make it advantageous to hold shares of their stocks in times of economic downturn.

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 estimated posterior mean covariance matrix from the fitted DSP-MFSV CAPM on 12 November 2008.

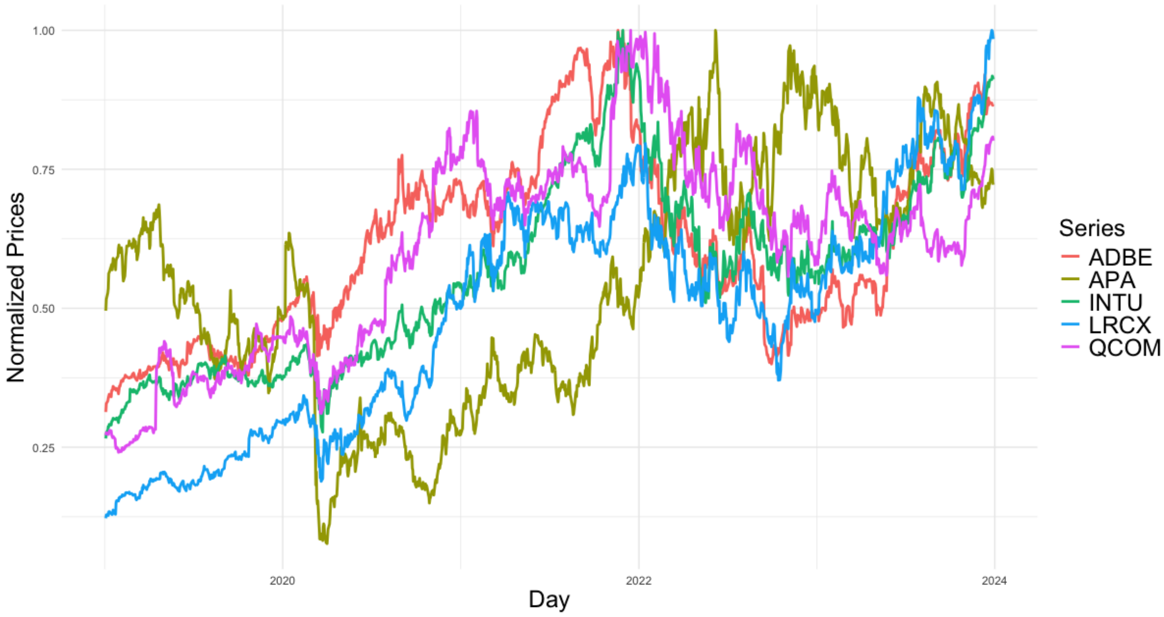

Figure 5: Plot of the normalized daily adjusted closing prices of five of the stocks in our diversified portfolio.

6.2.4 2020 COVID-19 pandemic

The COVID-19 pandemic started in Wuhan, China in December 2019. It quickly spread, triggering governments around the world to issue nationwide lockdowns to slow down the spread of the virus, which had a large negative economic impact. We consider two portfolios, each with 30 stocks. The first portfolio consists of 30 large technology companies listed on the NASDAQ. The second portfolio is a diversified portfolio which includes 10 stocks from our first portfolio and the rest are stocks from other industries included in the S&P 500. We compute excess asset returns 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 3,000 posterior samples of the score time series. Figure 5 shows the observed time series of the adjusted daily 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 that most of the stocks experience a noticeable decrease in their daily adjusted closing price.

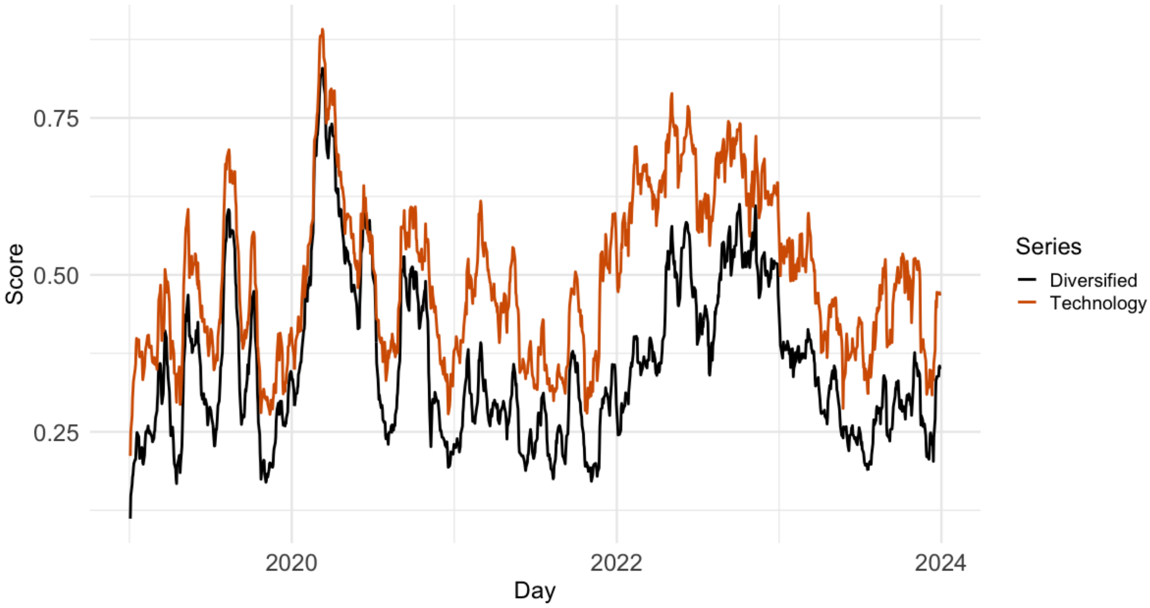

Figure 6: Plot of the estimated posterior score time series for our technology portfolio in vermilion and the estimated posterior score time series for our diversified portfolio in black. The posterior mean score time series is represented by the black line, and the 95% HDIs are represented by the boundary 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 protect 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 lockdowns across the globe starting soon afterwards, including the U.K., the 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 beginning of 2020, reaching their peak in early March. Therefore, despite diversification in our portfolio, we still suffer from the same shocks, demonstrating that

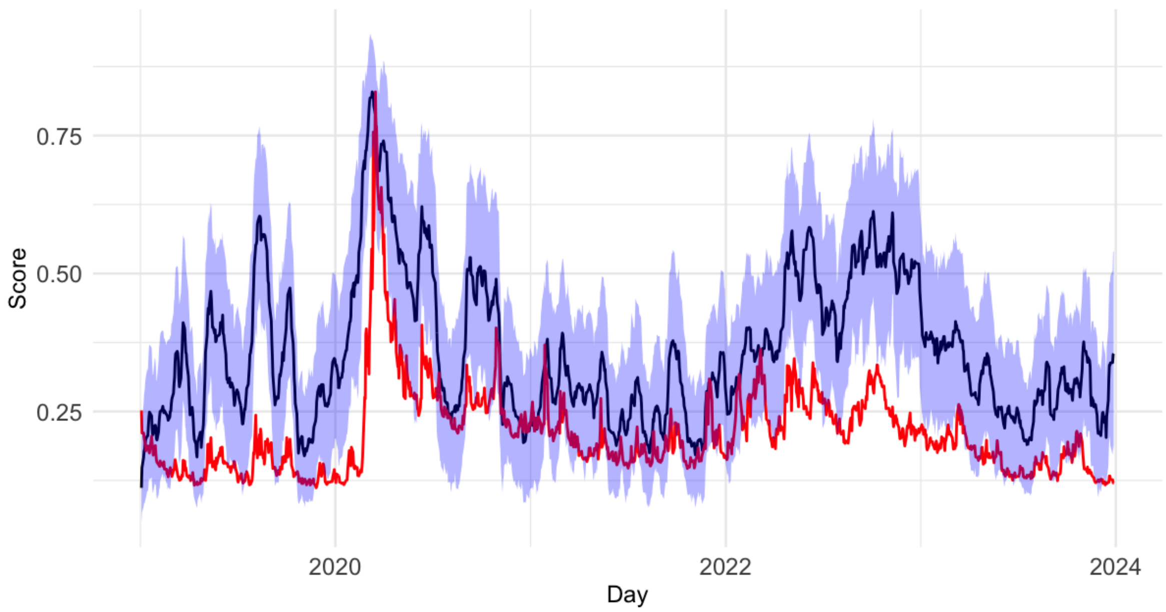

Figure 7: Plot of the estimated posterior score time series for our diversified portfolio in black and the VIX index in red. The 95% HDIs of our scores are represented by the boundary of the purple area. The VIX index is scaled to lie between the smallest and largest estimated posterior mean of the score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

there is no advantage to 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. This perhaps indicated 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 US Treasuries, which was occurring at the time. There were also other international stresses, such as the crash of the Argentinian stock market on 12 August 2019. The biggest spike in Figure 7 occurs on 9 March 2020 with a correlation of 0.83. We begin to see increases starting in early 2020 with reports of 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 2020 at which point the correlation level identified by our method is already 0.422. We then see several dramatic events happening in February, such as the increases in deaths internationally and a large decline in international stock markets from 24 to 28 February. The Federal Reserve then reduced 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 on 10 March 2020, the VIX index only peaked on 16 March 2020. This again shows the advantages of using our method to quantify the level of economic stress through the lens of modeling the correlation within an investor's portfolio and can serve as a quantitative tracker of the events of such economic downturns.

6.2.6 Minimum variance portfolio

See Section 6.2.3 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 10 March 2020, which is the day of the largest estimated posterior mean score of 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 stocks having weights with absolute values greater than one. The static sample covariance portfolio is more extreme, with 12 stocks having weights whose absolute values are greater than one, and the dynamic minimum variance portfolio is the most extreme, with 19 of its weights having absolute values greater than one. Interestingly, all three portfolios place a large weight on pharmaceutical stocks such as JNJ, MRK, and ABBV. The dynamic minimum variance portfolio places more than twice the weight on some of these stocks compared to the static portfolios. This makes intuitive sense, as with the outbreak of a global pandemic, there was an increase in demand for medical products and services, including calls for research into vaccines to fight COVID-19; therefore there would be an increase in the stock price of such companies.

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 estimated posterior mean covariance matrix from the fitted DSP-MFSV CAPM for 10 March 2020.

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 correlation matrices of even moderate dimension, we propose a scalar score to summarize a given correlation matrix. Through a posterior concentration result, we demonstrate theoretically the validity of our model. Through a simulation study, we show that our proposed model achieves desirable results in terms of root mean squared error and tight highest posterior density intervals when compared to competing methods. Through two real-world examples, we demonstrate the applicability of our model particularly in providing novel insights into an investor's portfolio, and establish that portfolio diversification does not alleviate the problems caused by financial crises. Finally, we compare the dynamically estimated minimum variance portfolio at the peak of each crisis with the traditional statically estimated minimum variance portfolios. Through this, we observe that the dynamically estimated minimum variance portfolio has more extreme weights in several 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

Section A

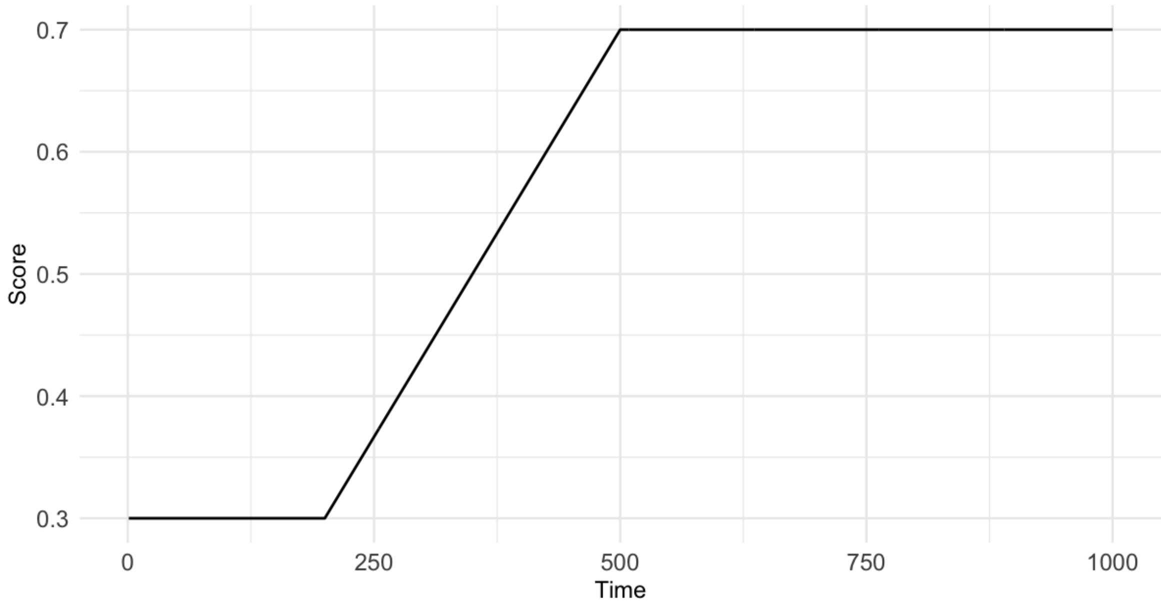

Figure 8: Plot of the correlation dynamics used in the simulation study of section 5.1

Section B

In this section we provide the full proof of our posterior concentration result.

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]

Lemma 3. *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. First, we define the σ-algebras we will be working with in view of Definition 5. Let

$$A_0^X = \sigma(X(t) : t \leq 0) \text{ and } A_k^X = \sigma(X(t) : t \geq k).$$

Then

$$\beta_X(k) = \mathbb{E}[\sup_{B\in A_k^X} |\mathbb{P}(B|A_0^X) - \mathbb{P}(B)|].$$

Now, consider $Y(t) = g(X(t))$. Then,

$$\begin{aligned} A_0^Y &= \sigma(Y(t) : t \leq 0) = \sigma(g(X(t)) : t \leq 0) \subseteq A_0^X \\ \text{and } A_k^Y &= \sigma(Y(t) : t \geq k) = \sigma(g(X(t)) : t \geq k) \subseteq A_k^X. \end{aligned}$$

Note that,

$$\begin{aligned} \mathbb{P}(B|A_0^Y) = \mathbb{E}[\mathbf{1}_B|A_0^Y] &= \mathbb{E}[\mathbb{E}[\mathbf{1}_B|A_0^X]|A_0^Y], \text{ by the law of total expectation,} \\ &= \mathbb{E}[\mathbb{P}(B|A_0^X)|A_0^Y]. \end{aligned}$$

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

$$\begin{aligned}\sup_{B\in A_k^Y} |\mathbb{P}(B|A_0^Y) - \mathbb{P}(B)| &\leq \sup_{B\in A_k^Y} \mathbb{E}[|\mathbb{P}(B|A_0^X) - \mathbb{P}(B)||A_0^Y] \\ &\leq \sup_{B\in A_k^X} \mathbb{E}[|\mathbb{P}(B|A_0^X) - \mathbb{P}(B)||A_0^Y] \leq \mathbb{E}[\sup_{B\in A_k^X} |\mathbb{P}(B|A_0^X) - \mathbb{P}(B)||A_0^Y].\end{aligned}$$

Which gives,

$$\begin{aligned}\beta_Y(k) := \mathbb{E}[\sup_{B\in A_k^Y} |\mathbb{P}(B|A_0^Y) - \mathbb{P}(B)|] &\leq \mathbb{E}[\mathbb{E}[\sup_{B\in A_k^X} |\mathbb{P}(B|A_0^X) - \mathbb{P}(B)||A_0^Y]] \\ &= \mathbb{E}[\sup_{B\in A_k^X} |\mathbb{P}(B|A_0^X) - \mathbb{P}(B)|], \text{ by the law of total expectation,} \\ &= \beta_X(k).\end{aligned}$$

Thus, $\beta_Y(k) \leq \beta_X(k) = O(\rho^k)$. Therefore, Y is also geometrically beta mixing. $\square$

Lemma 4. *Let $\mathcal{G} = \sigma(\beta_s : s \in \mathbb{Z})$. Then, the process*

$$\begin{aligned}cov(\mathbf{r}_t)_{ij}|\mathcal{F}_{t-1}, \mathcal{G} = \delta_{ij} \exp\{h_{\alpha_{i,t-1}}\} &+ \exp\{\mu_M + \phi_M(h_{M,t-1} - \mu_M) + \frac{1}{2}\sigma_M^2\}\beta_{i,t-1}\beta_{j,t-1} \\ &+ \sum_{k=1}^{r} \Lambda_{ik}\Lambda_{jk} \exp\{\tilde{\mu}_k + \tilde{\phi}_k(\tilde{h}_{k,t-1} - \tilde{\mu}_k) + \frac{1}{2}\tilde{\sigma}_k^2\} + \delta_{ij} \exp\{\bar{\mu}_i + \bar{\phi}_i(\bar{h}_{i,t-1} - \bar{\mu}_i) + \frac{1}{2}\bar{\sigma}_i^2\},\end{aligned}$$

where $\mathcal{F}_{t-1}$ is the information set at time t-1, is geometrically beta mixing.

Proof. Let,
$\boldsymbol{S}_t = (h_{M,t}, \{\tilde{h}_{k,t}\}_{k=1}^r, h_{\alpha_{i,t}}, \bar{h}_{i,t})'$, $\boldsymbol{\mu} = (\mu_M, \{\tilde{\mu}_k\}_{k=1}^r, \mu_{\alpha_i}, \bar{\mu}_i)$, $\boldsymbol{\phi} = diag(\phi_M, \{\tilde{\phi}_k\}_{k=1}^r, \phi_{\alpha_i}, \bar{\phi}_i)$
and $\boldsymbol{\epsilon}_t = (\epsilon_{M,t}, \{\eta_{k,t}\}_{k=1}^r, \eta_{\alpha_{i,t}}, \bar{\eta}_{i,t})$. Then,

$$\boldsymbol{S}_t = \boldsymbol{\mu} + \boldsymbol{\phi}(\boldsymbol{S}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t.$$

This can be written in the mean-centered form by defining $\boldsymbol{Z}_t := \boldsymbol{S}_t - \boldsymbol{\mu}$. Particularly,

$$\boldsymbol{Z}_t - \boldsymbol{\phi}\boldsymbol{Z}_{t-1} = \boldsymbol{\epsilon}_t.$$

In view of Mokkadem [1988] $B(0) = I_{3+r}, B(1) = -\boldsymbol{\phi}$, and $P(x) = \sum_{i=0}^{1} B(i)x^i$. The equation $detP(x) = 0 \iff (1-\phi_M x)(1-\tilde{\phi}_1 x)...(1-\tilde{\phi}_r x)(1-\phi_{\alpha_i}x)(1-\bar{\phi}_i x) = 0$ has solutions of the form $(1-\phi x) = 0 \iff x = \frac{1}{\phi}$. Also, recall that all the innovation distributions are absolutely continuous with respect to the Lebesgue measure and i.i.d. Since all the persistence parameters are less than 1 in absolute value, the solutions of the equation are all greater than 1 in absolute value. Therefore, by Theorem 1 of Mokkadem [1988] the process $(\boldsymbol{S}_t)_{t\in\mathbb{Z}}$ is geometric beta mixing.
Thus, by Lemma 3, the process $cov(\boldsymbol{r}_t)_{ij}|\mathcal{F}_{t-1}, \mathcal{G}$ is geometrically beta mixing. □ □

Lemma 5. *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 correlation matrix at time t under θ_0, in view of the subsequent posterior concentration result. Let $\mu = E_\theta[X_t]$. Then,*

$$\mathbb{P}_\theta(|\frac{1}{n}\sum_{t=1}^{n} X_t^{(i,j)}| > \epsilon_n + |\mu^{(i,j)}|) \le \exp\{-C_4 n\epsilon_n^2\},$$

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 to apply Theorem 1 of Merlevède et al. [2009]. Firstly,

$$|X_t^{(i,j)}| = |R_{t,\theta}^{(i,j)} - R_{t,\theta_0}^{(i,j)}| \le |R_{t,\theta}^{(i,j)}| + |R_{t,\theta_0}^{(i,j)}| \le 2 < \infty,$$

by the triangle inequality and the fact that the largest value of any entry in a correlation matrix is at most 1. This gives

$$|X_t^{(i,j)} - \mu^{(i,j)}| \le |X_t^{(i,j)}| + |\mu^{(i,j)}| \le 4 \le \infty, \text{by the triangle inequality.}$$

Secondly, by Lemma 4, 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

$$\alpha(n) \leq \kappa\rho^n = \exp\{log(\kappa\rho^n)\} = \exp\{log(\kappa) - n(-log(\rho))\}.$$

Choosing any C_1 such that $C_1 < -\log(\rho)$ we see $\alpha(n) \leq \exp\{log(\kappa) - C_1 n\}$. If $\kappa \leq 1$ then $log(\kappa) \leq 0$ which 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

$$\mathbb{P}(|\sum_{t=1}^{n} X_t^{(i,j)} - \mu^{(i,j)}| \geq x|\mathcal{G}) \leq \exp\{-\frac{C_3 x^2}{4^2 n + 4x(log(n))(log(log(n)))}\} \tag{26}$$

for $x \geq 0, n \geq 4$. Let

$$x = n\epsilon_n, \text{where } \epsilon_n = \frac{1}{\log(n)\log(\log(n))}.$$

Then,

$$\mathbb{P}(|\sum_{t=1}^{n} X_t^{(i,j)} - \mu^{(i,j)}| \geq n\epsilon_n|\mathcal{G}) \leq \exp\{-\frac{C_3 n^2 \epsilon_n^2}{16n + \frac{4n\log(n)\log(\log(n))}{\log(n)\log(\log(n))}}\} \text{ which implies}$$
$$\mathbb{P}(|\frac{1}{n}\sum_{t=1}^{n} X_t^{(i,j)} - \mu^{(i,j)}| \geq \epsilon_n|\mathcal{G}) \leq \exp\{-C_4 n\epsilon_n^2\}, \text{ where } C_4 := \frac{C_3}{20}.$$

Then, by the triangle inequality we obtain

$$\mathbb{P}(|\frac{1}{n}\sum_{t=1}^{n} X_t^{(i,j)}| \geq \epsilon_n + |\mu^{(i,j)}||\mathcal{G}) \leq \exp\{-C_4 n\epsilon_n^2\}$$

Finally, to remove the conditioning on $\mathcal{G}$,

$$
\begin{aligned}
\mathbb{P}(|\frac{1}{n}\sum_{t=1}^{n}X_t^{(i,j)}| \geq \epsilon_n + |\mu^{(i,j)}|) &= \mathbb{E}[\mathbf{1}\{|\frac{1}{n}\sum_{t=1}^{n}X_t^{(i,j)}| \geq \epsilon_n + |\mu^{(i,j)}|\}] \\
&= \mathbb{E}[\mathbb{E}[\mathbf{1}\{|\frac{1}{n}\sum_{t=1}^{n}X_t^{(i,j)}| \geq \epsilon_n + |\mu^{(i,j)}|\}|\mathcal{G}]], \text{ by law of total expectation} \\
&= \mathbb{E}[\mathbb{P}(|\frac{1}{n}\sum_{t=1}^{n}X_t^{(i,j)}| \geq \epsilon_n + |\mu^{(i,j)}||\mathcal{G})] \\
&\leq \exp\{-C_4 n\epsilon_n^2\}.\square
\end{aligned}
$$

$\square$

Proposition 6. *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 Lemma 5.*

Proof. Consider the test

$$
\phi_n = \mathbf{1}\{\max_{1\leq i,j\leq N}|\bar{X}_n^{(i,j)}| > \tau\}, \text{ and } \tau = \frac{\epsilon_n}{2N}.
$$

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 Lemma 5, and the fact that the probability is taken with respect the null hypothesis we obtain

$$
\begin{aligned}
\mathbb{P}_{\theta_0}^{(n)}(\phi_n = 1) &\leq N^2\exp\{-C_4 n\tau^2\} = N^2\exp\{-C_4\frac{n\epsilon_n^2}{4N^2}\} = \exp\{\log(N^2) - C_5\frac{n\epsilon_n^2}{N^2}\}, C_5 := \frac{C_4}{4} \\
&\leq \exp\{-K_1 n\epsilon_n^2\}, \text{ where } \frac{C_5 n\epsilon_n^2}{N^2} - \log(N^2) \geq K_1 n\epsilon_n^2 \text{ iff } \frac{C_5}{N^2} - \frac{\log(N^2)}{n\epsilon_n^2} \geq K_1
\end{aligned}
$$

Now suppose we have an alternative hypothesis satisfying

$$||R_\theta - R_{\theta_0}||_F > \epsilon_n$$

That is,

$$\sqrt{\sum_{i,j=1}^{N} |R_\theta^{(i,j)} - R_{\theta_0}^{(i,j)}|^2} > \epsilon_n \text{ which implies } \exists (i^*, j^*) \text{ such that } |R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)}| \geq \frac{\epsilon_n}{N}.$$

Suppose to the contrary, that is

$$|R_\theta^{(i,j)} - R_{\theta_0}^{(i,j)}| < \frac{\epsilon_n}{N}, \forall 1 \leq i,j \leq N. \text{ Then,}$$

$$||R_\theta - R_{\theta_0}||_F = \sqrt{\sum_{i,j=1}^{N} |R_\theta^{(i,j)} - R_{\theta_0}^{(i,j)}|^2} < \sqrt{\sum_{i,j=1}^{N} |\frac{\epsilon_n}{N}|^2} = \epsilon_n.$$

That is, $||R_\theta - R_{\theta_0}||_F < \epsilon_n$ which is a contradiction. Furthermore, if $\phi_n = 0$ then

$$\max_{1 \leq i,j \leq N} |\bar{X}_n^{(i,j)}| \leq \tau.$$

This implies

$$|\bar{X}_n^{(i^*,j^*)}| \leq \frac{\epsilon_n}{2N}.$$

Then,

$$\mathbb{P}_\theta^{(n)}(1 - \phi_n) = \mathbb{P}_\theta(\phi_n = 0) = \mathbb{P}_\theta^{(n)}(\max_{1 \leq i,j \leq N} |\bar{X}_n^{(i,j)}| \leq \tau) \leq \mathbb{P}_\theta^{(n)}(|\bar{X}_n^{(i^*,j^*)}| \leq \tau)$$

Then,

$$\begin{aligned} |\bar{X}_n^{(i^*,j^*)} - (R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)})| &= |(R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)}) - \bar{X}_n^{(i^*,j^*)}| \\ &\geq ||R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)}| - |\bar{X}_n^{(i^*,j^*)}||, \end{aligned}$$

by the reverse triangle inequality. Fix $\theta_1 \in \Theta$ with $||R_{\theta_1} - R_{\theta_0}||_F > \epsilon_n$. For any $\theta \in \Theta$

with $||R_\theta - R_{\theta_1}||_F < \xi\epsilon_n$ we obtain

$$\begin{aligned}||R_\theta - R_{\theta_0}||_F &= ||(R_\theta - R_{\theta_1}) - (R_{\theta_0} - R_{\theta_1})||_F \\ &\geq |||R_\theta - R_{\theta_1}||_F - ||R_{\theta_0} - R_{\theta_1}||_F|, \text{ by the reverse triangle inequality} \\ &= |||R_{\theta_0} - R_{\theta_1}||_F - ||R_\theta - R_{\theta_1}||_F| \\ &> |\epsilon_n - \xi\epsilon_n| = |\epsilon_n(1-\xi)| = (1-\xi)\epsilon_n, \xi < 1.\end{aligned}$$

Hence, $\exists(i^*, j^*)$ such that $|R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)}| \geq \frac{(1-\xi)\epsilon_n}{N}$. Now,

$$|\bar{X}_n^{(i^*,j^*)} - (R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)})| \geq \frac{(1-\xi)\epsilon_n}{N} - \frac{\epsilon_n}{2N} = \frac{(\frac{1}{2}-\xi)\epsilon_n}{N}, 0 < \xi < \frac{1}{2}.$$

Let $A^*_{(i,j)}$ be the event that $\{|\bar{X}_n^{(i^*,j^*)}| \leq \tau\}$. On $A^*_{(i,j)}$ we have

$$|R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)}| \geq \frac{(1-\xi)\epsilon_n}{N} \text{ and } |\bar{X}_n^{(i^*,j^*)}| \leq \tau.$$

Therefore, on $A^*_{(i,j)}$ we have

$$|\bar{X}_n^{(i^*,j^*)} - (R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)})| \geq \frac{(\frac{1}{2}-\xi)\epsilon_n}{N}.$$

Thus,

$$A^*_{(i,j)} \subseteq \{|\bar{X}_n^{(i^*,j^*)} - (R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)})| \geq \frac{(\frac{1}{2}-\xi)\epsilon_n}{N}\}.$$

This implies,

$$\begin{aligned}\mathbb{P}_\theta^{(n)}(|\bar{X}_n^{(i^*,j^*)}| \leq \tau) &\leq \mathbb{P}_\theta^{(n)}(|\bar{X}_n^{(i^*,j^*)} - (R_\theta^{(i^*,j^*)} - R_{\theta_0}^{(i^*,j^*)})| \geq \frac{(\frac{1}{2}-\xi)\epsilon_n}{N}) \\ &= \mathbb{P}_\theta^{(n)}(|\bar{X}_n^{(i^*,j^*)} - \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_\theta[X_t^{(i^*,j^*)}]| \geq \frac{(\frac{1}{2}-\xi)\epsilon_n}{N})\end{aligned}$$

This then gives,

$$\mathbb{P}_\theta^{(n)}(1-\phi_n) \leq \exp\{-C_4 n(\frac{(\frac{1}{2}-\xi)\epsilon_n}{N})^2\}.$$

This implies,

$$\begin{aligned}\sup_{d_n(\theta,\theta_1)<\xi\epsilon_n} \mathbb{P}_\theta^{(n)}(1-\phi_n) &\leq \exp\{-\frac{C_4(\frac{1}{2}-\xi)^2 n\epsilon_n^2}{N^2}\} \\ &= \exp\{-K_2 n\epsilon_n^2\}, K_2 := \frac{C_4(\frac{1}{2}-\xi)^2}{N^2}\end{aligned}$$

Let $K = \min(K_1, K_2)$, then

$$\mathbb{P}_{\theta_0}^{(n)}(\phi_n = 1) \leq \exp\{-Kn\epsilon_n^2\} \text{ and } \sup_{d_n(\theta,\theta_1)<\xi\epsilon_n} \mathbb{P}_\theta^{(n)}(1-\phi_n) \leq \exp\{-Kn\epsilon_n^2\}.\square$$

$\square$

Proposition 7. $\sup_{\epsilon>\epsilon_n} \log N(\frac{1}{2}\epsilon\xi, \{\theta \in \Theta_n : ||R_n(\theta) - R_n(\theta_0)||_F < \epsilon\}, ||\cdot||_F) \leq n\epsilon_n^2$, *where* $\epsilon_n = C_\epsilon n^{-\frac{1}{2+C_\theta}}, C_\epsilon > 0, C_\theta \in \mathbb{N}$ *and* $\Theta_n = \{\theta \in \Theta : |\mu_M| \leq \log(n), |\phi_M| \leq \sqrt{1-\frac{1}{\log(\log(n))}}, \sigma_M^2 \leq \log(n)^{0.5}, ||\Lambda||_F \leq \log(n), \max_{1\leq k\leq r} |\tilde{\mu}_k| \leq \log(n),$ $\max_{1\leq k\leq r} |\tilde{\phi}_k| \leq \sqrt{1-\frac{1}{\log(\log(n))}}, \max_{1\leq k\leq r} \tilde{\sigma}_k^2 \leq \log(n)^{0.5}, \max_{1\leq i\leq N} |\bar{\mu}_i| \leq \log(n), \max_{1\leq i\leq N} |\bar{\phi}_i| \leq \sqrt{1-\frac{1}{\log(\log(n))}},$ $\max_{1\leq i\leq N} \bar{\sigma}_i^2 \leq \log(n)^{0.5}, \max_{1\leq i\leq N} |\mu_{\alpha_i}| \leq \log(n^4), \max_{1\leq i\leq N} |\mu_{\beta_i}| \leq log(n^4), \max_{1\leq i\leq N} |\phi_{\beta_i}| \leq \sqrt{1-\frac{1}{log(log(n))}},$ $\max_{1\leq i\leq N} |\phi_{\alpha_i}| \leq \sqrt{1-\frac{1}{log(log(n))}}\}$, $H_n = \{\theta \in \theta : \max_{1\leq i\leq N, 1\leq t\leq n-1} |h_{\alpha_{i,t-1}}| \leq |\mu_{\alpha_i}| + \frac{c\log(n)+\log(2k(s))}{s},$ $\max_{1\leq t\leq n} |h_{M,t-1}| \leq |\mu_M| + p\log(n), \max_{1\leq i\leq N, 1\leq t\leq n} |\beta_{i,t-1}| \leq \sqrt{(\kappa+1)C_0 n^\alpha \log(n)},$ $\max_{1\leq k\leq r, 1\leq t\leq n} |\tilde{h}_{k,t-1}| \leq |\tilde{\mu}_k| + p\log(n), \max_{1\leq i\leq N, 1\leq t\leq n} |\bar{h}_{i,t-1}| \leq |\bar{\mu}_i| + p\log(n)\}$ *for* $a, b, p, \alpha > 0$, *with* $\mathbb{P}(\Theta_n^c) \to 0$, $\mathbb{P}(H_n^c) \to 0$, *and* N *is the number of stocks in our portfolio.*

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

For the autoregressive processes of order 1 with normally distributed innovations

($\tilde{h}_t$,$\bar{h}_t$, and $h_{M,t}$) we have

$$
\begin{aligned}
h_t &= \sum_{j=0}^{\infty} \phi^j \eta_{t-j}, \text{ which implies} \\
M_{h_t}(s) &= \prod_{j=0}^{\infty} M_\eta(s\phi^j) = \prod_{j=0}^{\infty} \exp\{\frac{\sigma^2 s^2 \phi^{2j}}{2}\} = \exp\{\frac{1}{2}\sigma^2 s^2 \sum_{j=0}^{\infty} \phi^{2j}\} = \exp\{\frac{\sigma^2 s^2}{2(1-\phi^2)}\}.
\end{aligned}
$$

Then, the Chernoff bound gives us

$$
\mathbb{P}(h_t > u) \leq \exp\{\frac{\sigma^2 s^2}{2(1-\phi^2)} - su\} \text{ and } \mathbb{P}(h_t < -u) = \mathbb{P}(-h_t > u) \leq \exp\{\frac{\sigma^2 s^2}{2(1-\phi^2)} - su\}
$$

Therefore,

$$
\mathbb{P}(|h_t| > u) = \mathbb{P}(h_t > u) + \mathbb{P}(h_t < -u) \leq 2\exp\{\frac{\sigma^2 s^2}{2(1-\phi^2)} - su\}
$$

$$
\begin{aligned}
\text{Let } s &= \frac{(1-\phi^2)u}{\sigma^2}. Then \\
\mathbb{P}(|h_t| > u) &\leq 2\exp\{\frac{\sigma^2(1-\phi^2)^2 u^2}{2\sigma^4(1-\phi^2)} - \frac{(1-\phi^2)u^2}{\sigma^2}\} \\
&= 2\exp\{\frac{(1-\phi^2)u^2}{2\sigma^2} - \frac{(1-\phi^2)u^2}{\sigma^2}\} = 2\exp\{-\frac{(1-\phi^2)u^2}{2\sigma^2}\}.
\end{aligned}
$$

By the triangle inequality, $|h_t - \mu + \mu| \leq |h_t - \mu| + |\mu|$.This implies,

$$
\mathbb{P}(|h_t| > u) \leq \mathbb{P}(|h_t - \mu| + |\mu| > u) = \mathbb{P}(|h_t - \mu| > u - |\mu|)
$$

Now, let $u = |\mu| + p\log(n)$ for $p > 0$. Then,

$$
\mathbb{P}(|h_t| > u|\sigma^2, \phi) \leq 2\exp\{-\frac{(1-\phi^2)(|\mu| + p\log(n) - |\mu|)^2}{2\sigma^2}\} = 2\exp\{-\frac{(1-\phi^2)(p\log(n))^2}{2\sigma^2}\}.
$$

Then by using the union bound to bound the maximum we obtain

$$\mathbb{P}(\max_{1\le k\le r, 1\le t\le n-1} |\tilde{h}_{k,t-1}| \le |\tilde{\mu}_k| + p\log(n)|\tilde{\phi}, \tilde{\sigma}^2) \le 2r\exp\{-\frac{(1-\phi^2)p^2\log(n)^2}{2\sigma^2} + log(n)\}.$$

On the sieve,

$$|\phi| \le \sqrt{1 - \frac{1}{\log(\log(n))}} \text{ which implies } 1-\phi^2 \ge \frac{1}{\log(\log(n))}.$$

This implies,

$$\mathbb{P}(\max_{1\le k\le r, 1\le t\le n-1} |\tilde{h}_{k,t-1}| \le |\tilde{\mu}_k| + p\log(n)|\tilde{\phi}, \tilde{\sigma}^2) \le 2r\exp\{log(n) - \frac{p^2}{2}\frac{\log(n)^{1.5}}{\log(\log(n))}\} \to 0$$

Similarly for the other autoregressive processes of order 1 with normally distributed innovations, $\max_{1\le i\le N, 1\le t\le n} |\bar{h}_{i,t-1}|$ and $\max_{1\le t\le n} |h_{M,t-1}|$.

Before proceeding we first need to derive the the moment generating function of the four parameter Z-distribution. The moment generating function of $X \sim Z(\frac{1}{2}, \frac{1}{2}, 0, 1)$ is

$$M(t) = \mathbb{E}[e^{tx}] = \int_{-\infty}^{\infty} e^{tx} f(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{(t+\frac{1}{2})x}}{1+e^x}dx.$$

Let $u = e^x$ which implies

$$log(u) = x \text{ and } \frac{du}{dx} = e^x = u. \text{ This gives } \frac{du}{u} = dx.$$

Therefore,

$$M(t) = \frac{1}{\pi}\int_{0}^{\infty} \frac{u^{t+\frac{1}{2}}}{1+u}\frac{du}{u} = \frac{1}{\pi}\int_{0}^{\infty} \frac{u^{t-\frac{1}{2}}}{1+u}du.$$

$$\text{Letting } a = t + \frac{1}{2} \text{ implies } a - 1 = t - \frac{1}{2} \text{ which gives } M(t) = \frac{1}{\pi}\int_{0}^{\infty} \frac{u^{a-1}}{1+u}du.$$

Let,

$$v = \frac{u}{1+u} \text{ which implies } v(1+u) = u \text{ giving } u = \frac{v}{1-v}. \text{ This implies,}$$
$$\frac{du}{dv} = \frac{1}{(1-v)^2}, \text{ by the product rule. Subsequently, } du = \frac{dv}{(1-v)^2}.$$

Therefore,

$$\begin{aligned}
M(t) &= \frac{1}{\pi}\int_0^1 v^{a-1}(1-v)^{-(a-1)}(1-v)\frac{1}{(1-v)^2}dv = \frac{1}{\pi}\int_0^1 v^{a-1}(1-v)^{-(a-1)}(1-v)^{-1}dv \\
&= \frac{1}{\pi}\int_0^1 v^{a-1}(1-v)^{-a}dv = \frac{1}{\pi}\int_0^1 v^{a-1}(1-v)^{(1-a)-1}dv = \frac{1}{\pi}B(a, 1-a) \\
&= \frac{1}{\pi}B(t+\frac{1}{2}, \frac{1}{2}-t) = \frac{1}{\pi}\frac{\Gamma(t+\frac{1}{2})\Gamma(\frac{1}{2}-t)}{\Gamma(1)} = \frac{1}{\pi}\Gamma(t+\frac{1}{2})\Gamma(\frac{1}{2}-t)
\end{aligned}$$

Let $z = t + \frac{1}{2}$ which implies
$M(t) = \frac{1}{\pi}\Gamma(z)\Gamma(1-z) = \frac{1}{\pi}\frac{\pi}{sin(\pi z)} = \frac{1}{sin(\pi z)}$ by Euler's reflection formula.

Thus,

$$M(t) = \frac{1}{sin(\pi(t+\frac{1}{2}))} = \frac{1}{sin(\pi t + \frac{\pi}{2})} = \frac{1}{cos(\pi t)} = sec(\pi t).$$

For autoregressive process of order 1 with Z-distributed innovations

$$X_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} \text{ which implies } M_X(s) = \prod_{j=0}^{\infty} sec(\pi\phi^j s)$$

For fixed $s \in (0, 0.5)$ the infinite product converges. For such an s we have $|\pi\phi^j s| = \pi|\phi|^j s \le \pi s < \frac{\pi}{2}$.Therefore, $\pi\phi^j s \in (-\frac{\pi}{2}, \frac{\pi}{2})$.Consider $\sum_{j=1}^{\infty} \log(sec(\pi\phi^j s))$ where by Taylor expansion we see $\log(sec(x)) = \frac{x^2}{2} + O(x^4)$.This implies $\log(sec(\pi\phi^j s)) = \frac{\pi^2\phi^{2j}s^2}{2} + O(\phi^{4j})$.Observe $\lim_{j\to\infty} \frac{\frac{1}{2}\pi^2 s^2\phi^{2(j+1)} + O(\phi^{4j})}{\frac{1}{2}\pi^2 s^2\phi^{2j} + O(\phi^{4j})} = \phi^2$, 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,

$$\begin{aligned}
\mathbb{P}(|X_t| > u) &= \mathbb{P}(X_t > u) + \mathbb{P}(X_t \le -u) \\
&\le M_X(s)e^{-su} + M_X(-s)e^{-su} \\
&\le 2K(s)e^{-su}, K(s) = max\{M_X(s), M_X(-s)\}.
\end{aligned}$$

This implies

$\mathbb{P}(|h_t| > u) \le \mathbb{P}(|h_t| > u - |\mu|) \le 2K(s)\exp\{-s(u - |\mu|)\}$.

Let $u_n = |\mu| + \frac{c\log(n) + \log(2k(s))}{s}$, for some constant $c > 0$. This implies

$$\begin{aligned}
\mathbb{P}(|h_t| > |\mu| + \frac{c\log(n) + \log(2k(s))}{s}) &\le 2K(s)\exp\{-s(|\mu| + \frac{c\log(n) + \log(2K(s))}{s} - |\mu|)\} \\
&= 2K(s)\exp\{\log(\frac{n^{-c}}{2K(s)})\} = 2K(s)\frac{n^{-c}}{2K(s)} = n^{-c}.
\end{aligned}$$

This implies,

$$\mathbb{P}(\max_{1\le a\le N, 1\le t\le n} |h_{a,t}| > |\mu| + \frac{c\log(n) + \log(2k(s))}{s}) \le Nn^{-C_1+1} = Nn^{-C_2} \to 0, \text{ for some } C_1, C_2 > 0.$$

Now, we can bound the tail probability of the mean parameters.

$$\begin{aligned}\mathbb{P}(|\mu| > \log(n^4)) &= \mathbb{P}(\mu > \log(n^4)) + \mathbb{P}(\mu < -\log(n^4)) \\ &= \mathbb{P}(\log(\tau_0^2\tau_1^2) > \log(n^4)) + \mathbb{P}(\log(\tau_0^2\tau_1^2) < \log(\frac{1}{n^4})) \\ &= \mathbb{P}(\tau_0^2\tau_1^2 > n^4) + \mathbb{P}(\tau_0^2\tau_1^2 < \frac{1}{n^4}).\end{aligned}$$

Observe that 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 $\tau_0^2 < n^{-2}$ or $\tau_1^2 < n^{-2}$. 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 $\tau \sim C^+(0,b)$,

$$\begin{aligned}\mathbb{P}(\tau > t) &= \int_t^\infty \frac{2}{\pi b}\frac{1}{1+\frac{u^2}{b^2}}du = \frac{2}{\pi b}\int_t^\infty \frac{1}{1+\frac{u^2}{b^2}}du = \frac{2}{\pi b}\int_t^\infty \frac{b^2}{b^2+u^2}du \\ &= \frac{2b}{\pi}\int_t^\infty \frac{1}{b^2+u^2}du \le \frac{2b}{\pi}\int_t^\infty \frac{1}{u^2}du = \frac{2b}{\pi}[-\frac{1}{u}]_t^\infty = \lim_{a\to\infty}\frac{2b}{\pi}[-\frac{1}{a}+\frac{1}{t}] = \frac{2b}{\pi t}\end{aligned}$$

Similarly,

$$\mathbb{P}(\tau < t) = \int_0^t \frac{2}{\pi b}\frac{1}{1+(\frac{u}{b})^2}du = \frac{2}{\pi b}\int_0^t \frac{1}{1+(\frac{u}{b})^2}du.$$

Note that for $u \ge 0, 1+(\frac{u}{b})^2 \ge 1$ which implies $1 \ge \frac{1}{1+(\frac{u}{b})^2}$.

This implies $\mathbb{P}(\tau < t) \le \frac{2}{\pi b}\int_0^t 1du = \frac{2}{\pi b}t$. 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,

$$\mathbb{P}(|\mu| > \log(n^4)) \leq \frac{2}{\pi}(n^{-\frac{1}{2}} + n^{\frac{1}{2}} + 2)n^{-1} = \frac{2}{\pi}(n^{-\frac{3}{2}} + n^{-\frac{1}{2}} + 2n^{-1}) \leq \frac{8}{\pi}n^{-\frac{1}{2}}.$$

By the union bound , $\mathbb{P}(\max_{1\leq i\leq N} |\mu_{\alpha_i}| > \log(n^4)) \leq \frac{8N}{\pi}n^{-\frac{1}{2}} \to 0.$

Similarly, for $\max_{1\leq i\leq N} |\mu_{\beta_i}|$.

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

$\mathbb{P}(|\mu| > t) \leq 2\exp\{-\frac{t^2}{2\sigma^2}\}$.Therefore, $\mathbb{P}(|\mu| > \log(n)) \leq 2\exp\{-\frac{(\log(n))^2}{2\sigma^2}\} = 2n^{-\frac{\log(n)}{2\sigma^2}}$.

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,

$$\mathbb{P}(|\mu| > \log(n)) \leq 2n^{-\frac{1}{2\sigma^2}} \text{ where } \sigma^2 = 10 \text{ or } 100.$$

Then,

$\mathbb{P}(|\mu_M| > \log(n)) \leq 2n^{-\frac{1}{2\sigma^2}} \to 0$.By the union bound,

$\mathbb{P}(\max_{1\leq k\leq r} |\tilde{\mu}_k| > \log(n)) \leq 2rn^{-\frac{1}{2\sigma^2}} \to 0$ and $\mathbb{P}(\max_{1\leq i\leq N} |\bar{\mu}_i| < \log(n)) \leq 2Nn^{-\frac{1}{2\sigma^2}} \to 0.$

For all the persistence parameters, ϕ, they have prior distributions of the form

$$\frac{\phi + 1}{2} \sim Beta(a, b)\text{where } a > b > 0.$$

Therefore,

$$\begin{aligned}
\mathbb{P}(|\phi| > \sqrt{1-\frac{1}{\log(\log(n))}}) &= \mathbb{P}(\phi > \sqrt{1-\frac{1}{\log(\log(n))}}) + \mathbb{P}(\phi < -\sqrt{1-\frac{1}{\log(\log(n))}}) \\
&= \mathbb{P}(\phi + 1 > \sqrt{1-\frac{1}{\log(\log(n))}} + 1) \\
&+ \mathbb{P}(\phi + 1 < -\sqrt{1-\frac{1}{\log(\log(n))}} + 1) \\
&= \mathbb{P}(\frac{\phi+1}{2} > \frac{\sqrt{1-\frac{1}{\log(\log(n))}}+1}{2}) \\
&+ \mathbb{P}(\frac{\phi+1}{2} < \frac{-\sqrt{1-\frac{1}{\log(\log(n))}}+1}{2}) \\
&= \mathbb{P}(X > 1-\epsilon) + \mathbb{P}(X < \epsilon) \\
&\text{where } \epsilon = \frac{1-\sqrt{1-\frac{1}{\log(\log(n))}}}{2} \text{ and } X \sim Beta(a,b).
\end{aligned}$$

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,

$\mathbb{P}(X < \epsilon) \leq \frac{1}{B(a,b)} \int_0^{\epsilon} x^{a-1} dx = \frac{1}{B(a,b)}[\frac{x^a}{a}]_0^{\epsilon} = \frac{1}{B(a,b)}[\frac{\epsilon^a}{a} - 0] = \frac{\epsilon^a}{aB(a,b)}$.Similarly,

$$\mathbb{P}(X > 1-\epsilon) = \frac{1}{B(a,b)} \int_{1-\epsilon}^{1} x^{a-1}(1-x)^{b-1} dx.$$

For $x \in [1-\epsilon, 1], x \leq 1$ which 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,

$$\mathbb{P}(|\phi| > \sqrt{1 - \frac{1}{\log(\log(n))}}) \leq \frac{1}{aB(a,b)}(\frac{1 - \sqrt{1 - \frac{1}{\log(\log(n))}}}{2})^a + \frac{1}{bB(a,b)}(\frac{1 - \sqrt{1 - \frac{1}{\log(\log(n))}}}{2})^b$$
$$\leq \frac{2}{bB(a,b)}(\frac{1 - \sqrt{1 - \frac{1}{\log(\log(n))}}}{2})^a, \text{since } a > b.$$

This implies,

$$\mathbb{P}(\max_{1 \leq k \leq r} |\tilde{\phi}_k| \geq \sqrt{1 - \frac{1}{\log(\log(n))}}) \leq \frac{2r}{bB(a,b)}(\frac{1 - \sqrt{1 - \frac{1}{\log(\log(n))}}}{2})^a \to 0.$$

Similarly, for $|\phi_M|$, $\max_{1 \leq i \leq N} |\bar{\phi}_i|$, $\max_{1 \leq i \leq N} |\phi_{\alpha_i}|$, and $\max_{1 \leq i \leq N} |\phi_{\beta_i}|$.
Now, the variance parameters in the model have $Ga(\frac{1}{2}, \frac{1}{2})$ distributions. Then,

$$\mathbb{P}(\sigma^2 > t) = \int_t^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \leq \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \int_t^\infty e^{-\lambda x} dx \text{ for } a < 1, \text{ since}$$

$x \geq t$ 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,

$$\mathbb{P}(\sigma^2 > t) \leq \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} [-\frac{1}{\lambda} e^{-\lambda x}]_t^\infty = \lim_{A \to \infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} [-\frac{1}{\lambda} e^{-\lambda A} + \frac{1}{\lambda} e^{-\lambda t}] = \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}.$$

In our prior specification

$$\mathbb{P}(\sigma^2 > \log(n)) \leq \sqrt{\frac{2}{\pi}} (\log(n))^{-\frac{1}{2}} e^{-\frac{1}{2}\log(n)} = \sqrt{\frac{2}{\pi}} (\log(n))^{-\frac{1}{2}} n^{-\frac{1}{2}} \leq \sqrt{\frac{2}{\pi}} n^{-\frac{1}{2}}, \text{for } e \leq n.$$

This implies $\mathbb{P}(\max_{1 \leq k \leq r} \tilde{\sigma}_k^2 > \log(n)^{0.5}) \leq r\sqrt{\frac{2}{\pi}} n^{-\frac{1}{2}} \to 0$, similarly for $\max_{1 \leq i \leq N} \bar{\sigma}_i^2$ and σ_M^2.

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 , $\mathbb{P}(|\Lambda_{ij}| > x|\lambda) = \mathbb{E}[\mathbb{P}(|\Lambda_{ij}| > x|\tau, \lambda)|\lambda] = \mathbb{E}[\mathbb{P}(|\Lambda_{ij}| > x|\tau)|\lambda]$

$$
\begin{aligned}
&\leq \mathbb{E}[2\exp\{-\frac{x^2}{2\tau^2}\}|\lambda] \\
&= \mathbb{E}[2\exp\{-\frac{x^2}{2\tau^2}\}\mathbf{1}\{\tau^2 \leq L\}|\lambda] + \mathbb{E}[2\exp\{-\frac{x^2}{2\tau^2}\}\mathbf{1}\{\tau^2 > L\}|\lambda] \\
&\leq 2\exp\{-\frac{x^2}{2L}\}\mathbb{E}[\mathbf{1}\{\tau^2 \leq L\}|\lambda] + \mathbb{E}[2\exp\{-\frac{x^2}{2\tau^2}\}\mathbf{1}\{\tau^2 > L\}|\lambda] \\
&= 2\exp\{-\frac{x^2}{2L}\}\mathbb{P}(\tau^2 \leq L|\lambda) + \mathbb{E}[2\exp\{-\frac{x^2}{2\tau^2}\}\mathbf{1}\{\tau^2 > L\}|\lambda] \\
&\leq 2\exp\{-\frac{x^2}{2L}\} + \mathbb{E}[2\exp\{-\frac{x^2}{2\tau^2}\}\mathbf{1}\{\tau^2 > L\}|\lambda] \\
&\leq 2\exp\{-\frac{x^2}{2L}\} + E[2\mathbf{1}\{\tau^2 > L\}|\lambda] \\
&= 2\exp\{-\frac{x^2}{2L}\} + 2\mathbb{P}(\tau^2 > L|\lambda) = 2\exp\{-\frac{x^2}{2L}\} + 2\mathbb{P}(e^{\gamma\tau^2} > e^{\gamma L}) \\
&\leq 2\exp\{-\frac{x^2}{2L}\} + 2\mathbb{E}[e^{\gamma\tau^2}]e^{-\gamma L}, \text{ by Markov's inequality.}
\end{aligned}
$$

Then using the formula for the moment generating function of a $Gamma(\alpha, \beta)$ distributed random variable

$$\mathbb{P}(|\Lambda_{ij}| > x|\lambda) \leq 2\exp\{-\frac{x^2}{2L}\} + 2(1 - \frac{\gamma}{\beta})^{-\alpha}e^{-\gamma L}$$

Letting $\gamma = \frac{1}{2}\beta$ gives

$$
\begin{aligned}
\mathbb{P}(|\Lambda_{ij}| > x|\lambda) &\leq 2\exp\{-\frac{x^2}{2L}\} + 2(\frac{1}{2})^{-\alpha}e^{-\gamma L} = 2\exp\{-\frac{x^2}{2L}\} + 22^{\alpha}e^{-\gamma L} \\
&\leq 2\exp\{-\frac{x^2}{2L}\} + 4e^{-0.025L\lambda^2}, \text{ since } \alpha < 1.
\end{aligned}
$$

Since we are doing computation on a computer, there will be a smallest possible representable number, say λ_0, and letting L = x gives

$$\mathbb{P}(|\Lambda_{ij}| > x|\lambda \geq \lambda_0) \leq 2\exp\{-\frac{x}{2}\} + 4\exp\{-0.025\lambda_0^2 x\}.$$

Let $C_0 = min\{\frac{1}{2}, 0.025\lambda_0^2\}$, then

$$\mathbb{P}(|\Lambda_{ij}| > x) = \mathbb{E}[\mathbf{1}\{|\Lambda_{ij}| > x\}] = \mathbb{E}[\mathbb{P}(|\Lambda_{ij}| > x)|\lambda] \leq 6\exp\{-C_0 x\},$$

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, $\mathbb{P}(||\Lambda||_F > x) \leq \mathbb{P}(\sqrt{Nr} \max_{i,j} |\Lambda_{ij}| > x)$

$$= \mathbb{P}(\max_{i,j} |\Lambda_{i,j}| > \frac{x}{\sqrt{Nr}}) = \mathbb{P}(\bigcup_{i,j} |\Lambda_{i,j}| > \frac{x}{\sqrt{Nr}}) \leq \sum_{i,j} \mathbb{P}(|\Lambda_{ij}| > \frac{x}{\sqrt{Nr}})$$

$$\leq \sum_{i,j} 6\exp\{-C_0(\frac{x}{\sqrt{Nr}})\} = 6Nr\exp\{-\frac{C_0 x}{\sqrt{Nr}}\} = C_1\exp\{-C_2 x\},$$

where $C_1 = 6Nr,$ and $C_2 = \frac{C_0}{\sqrt{Nr}}$.

This implies $\mathbb{P}(||\Lambda||_F > \log(n)) \leq C_1 n^{-C_2} \to 0.$

Now we need to bound β_t.

$$\begin{aligned}
\mathbb{P}(\max_{1\leq a\leq N, 1\leq t\leq n} |\beta_{a,t}| > M) &= \mathbb{P}(\{\max_{1\leq a\leq N, 1\leq t\leq n} |\beta_{a,t}| > M\} \cap \Theta_n / \{\beta_{a,t}\}_{1\leq a\leq N, 1\leq t\leq n}) \\
&+ \mathbb{P}(\{\max_{1\leq a\leq N, 1\leq t\leq n} |\beta_{a,t}| > M\} \cap \Theta_n^c / \{\beta_{a,t}\}_{1\leq a\leq N, 1\leq t\leq n}) \\
&\leq \mathbb{P}(\{\max_{1\leq a\leq N, 1\leq t\leq n} |\beta_{a,t}| > M\} \cap \Theta_n / \{\beta_{a,t}\}_{1\leq a\leq N, 1\leq t\leq n}) \\
&+ \mathbb{P}(\theta_n^c / \{\beta_{a,t}\}_{1\leq a\leq N, 1\leq t\leq n}) \\
&\leq 2nN \exp\{-\frac{M^2}{2\exp\{|\mu| + \frac{c\log(n)+\log(2k(s))}{s}\}}\} + \mathbb{P}(\theta_n^c / \{\beta_{a,t}\}_{1\leq a\leq N, 1\leq t\leq n}) \\
&\leq 2nN \exp\{-\frac{M^2}{2\exp\{4\log(n) + \frac{c\log(n)+\log(2k(s))}{s}\}}\} + \mathbb{P}(\theta_n^c / \{\beta_{a,t}\}_{1\leq a\leq N, 1\leq t\leq n}) \\
&= 2nN \exp\{-\frac{M^2}{C_0 n^{4+\frac{c}{s}}}\} + \mathbb{P}(\theta_n^c / \{\beta_{a,t}\}_{1\leq a\leq N, 1\leq t\leq n}), \text{ where } C_0 = 2[2k(s)]^{\frac{1}{s}}.
\end{aligned}$$

Let $\alpha = \dfrac{c}{s} + 4$.Then ,$\mathbb{P}(\max_{1\leq a\leq N, 1\leq t\leq n} |\beta_{a,t}| > M) \leq 2nN \exp\{-\dfrac{M^2}{C_0 n^\alpha}\} + \mathbb{P}(\theta_n^c / \{\beta_{a,t}\})$.
Suppose $\dfrac{M^2}{C_0 n^\alpha} = (\kappa+1)\log(n)$ 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,
$\mathbb{P}(\max_{1\leq a\leq N, 1\leq t\leq n} |\beta_{a,t}| > C_1 n^{\frac{1}{2}(\frac{c}{s}+5)} \log(n)^{\frac{1}{2}}) \leq 2nN \exp\{-\dfrac{(\kappa+1)C_0 n^\alpha \log(n)}{C_0 n^\alpha}\} + \mathbb{P}(\theta_n^c / \{\beta_{a,t}\})$
$= 2nN \exp\{\log(n^{-(\kappa+1)})\} + \mathbb{P}(\theta_n^c / \{\beta_{a,t}\}) = 2Nnn^{-\kappa-1} + \mathbb{P}(\theta_n^c / \{\beta_{a,t}\}) = 2Nn^{-\kappa} + \mathbb{P}(\theta_n^c / \{\beta_{a,t}\}) \to 0.$

Therefore by the bounds above,

$$\mathbb{P}(\Theta_n^c) \to 0 \text{ and } \mathbb{P}(H_n^c) \to 0$$

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.

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

for fixed n, where $C_1 = [2k(s)]^{\frac{1}{s}}$.

$$\begin{aligned}
\frac{\partial cov(\boldsymbol{r}_t)_{ij}}{\partial h_{M,t-1}} &= \beta_{i,t-1}\beta_{j,t-1}\exp\{\mu_M - \phi_M\mu_M + \frac{1}{2}\sigma_M^2\}\frac{\partial}{\partial h_{M,t-1}}\exp\{\phi_M h_{M,t-1}\} \\
&= \beta_{i,t-1}\beta_{j,t-1}\exp\{\mu_M - \phi_M\mu_M + \frac{1}{2}\sigma_M^2\}\phi_M\exp\{\phi_M h_{M,t-1}\} \\
&\leq (\kappa+1)C_0 n^\alpha \log(n)\exp\{\log(n)\}\exp\{\sqrt{1-\frac{1}{\log(\log(n))}}\log(n)\} \\
&\times \exp\{\frac{1}{2}\log(n)^{0.5}\}\sqrt{1-\frac{1}{\log(\log(n))}}\exp\{\sqrt{1-\frac{1}{\log(\log(n))}}(\log(n)+p\log(n))\} \\
&\leq (\kappa+1)C_0 n^\alpha \log(n) n\exp\{\log(n)\}\exp\{\log(n^{\frac{1}{2}})\}\exp\{\log(n)+p\log(n)\} \text{ for } n \geq e^e.
\end{aligned}$$

This gives, $\frac{\partial cov(\boldsymbol{r}_t)_{ij}}{\partial h_{M,t-1}} \leq (\kappa+1)C_0\log(n)^2 n^{\alpha+\frac{7}{2}+p}$.

$$\begin{aligned}
\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \mu_M} &= (1-\phi_M)\beta_{i,t-1}\beta_{j,t-1}\exp\{\phi_M h_{M,t-1} + \frac{1}{2}\sigma_M^2\}\exp\{\mu_M - \phi\mu_m\} \\
&\leq (1-\sqrt{1-\frac{1}{\log(\log(n))}})\beta_{i,t-1}\beta_{j,t-1}\exp\{\sqrt{1-\frac{1}{\log(\log(n))}}(\log(n)+\log(n^p)) \\
&+ \frac{1}{2}\log(n)^{0.5}+\}\exp\{\log(n) - \sqrt{1-\frac{1}{\log(\log(n))}}\log(n)\} \\
&\leq \beta_{i,t-1}\beta_{j,t-1}\exp\{\log(n) + plog(n) + \log(n^{\frac{1}{2}})\}\exp\{\log(n)\}, \text{ for } n \geq e^e. \\
&\leq M^2 n^{\frac{5}{2}+p} \\
&= C_2 n^{\frac{5}{2}+\alpha+p}\log(n).
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \phi_M} &= \beta_i\beta_j \exp\{\mu_M + \frac{1}{2}\sigma_M^2\}(h_M - \mu_M)\exp\{\phi_M(h_M - \mu_M)\} \\
&\leq M^2 \exp\{\log(n) + \frac{1}{2}\log(n)^{0.5}\}(|\mu_M| + p\log(n) + |\mu_M|) \\
&\times \exp\{\sqrt{1 - \frac{1}{\log(\log(n))}}(2\log(n) + p\log(n))\} \\
&\leq M^2 \exp\{\log(n) + \frac{1}{2}\log(n)\}(p+2)\log(n)\exp\{p\log(n) + 2\log(n)\} \\
&= M^2 n^{\frac{3}{2}}(p+2)\log(n) n^{p+2} \\
&= M^2 n^{\frac{3}{2}+p+2}(p+2)\log(n) = C_2 n^{\frac{7}{2}+p+\alpha}\log(n)^2.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \sigma_M^2} &= \frac{1}{2}\beta_i\beta_j \exp\{\mu_M + \phi_m(h_M - \mu_M)\}\exp\{\frac{1}{2}\sigma_M^2\} \\
&\leq \frac{1}{2}M^2 \exp\{\log(n) + \sqrt{1 - \frac{1}{\log(\log(n))}}(2+p)\log(n)\}\exp\{\frac{1}{2}\log(n)^{0.5}\} \\
&\leq \frac{1}{2}M^2 \exp\{(3+p)\log(n)\}\exp\{\log(n^{\frac{1}{2}})\}, n \geq e^e \\
&= M^2 n^{3+p} n = \frac{1}{2}M^2 n^{\frac{7}{2}+p} = \frac{1}{2}C_2 n^{\frac{7}{2}+p+\alpha}\log(n).
\end{aligned}
$$

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

$$\begin{aligned}\frac{\partial cov(\mathbf{r}_t)_{ij}}{\partial \Lambda_{ik}} &= \exp\{\tilde{\mu}_k + \tilde{\phi}_k(\tilde{h}_k - \tilde{\mu}_k) + \frac{1}{2}\tilde{\sigma}_k^2\}\Lambda_{jk} = \exp\{\tilde{\mu}_k\}\exp\{\tilde{\phi}_k(\tilde{h}_k - \tilde{\mu}_k)\}\exp\{\frac{1}{2}\tilde{\sigma}_k^2\}\Lambda_{jk} \\ &\leq \exp\{\log(n)\}\exp\{\sqrt{1 - \frac{1}{\log(\log(n))}}(p+2)\log(n)\}\exp\{\frac{1}{2}\log(n)^{0.5}\}\log(n) \\ &\leq nn^{2+p}n^{\frac{1}{2}}\log(n) = n^{\frac{7}{2}+p}\log(n), \text{with a factor of 2 when i = j, } n \geq e^e\end{aligned}$$

$$\begin{aligned}\frac{\partial cov(\boldsymbol{r}_t)_{ij}}{\partial \tilde{h}_{k,t-1}} &= \Lambda_{ik}\Lambda_{jk}\exp\{\tilde{\mu}_k\}\exp\{-\tilde{\phi}_k\tilde{\mu}_k\}\exp\{\frac{1}{2}\tilde{\sigma}_k^2\}\frac{\partial}{\partial \tilde{h}_{k,t-1}}\exp\{\tilde{\phi}_k\tilde{h}_{k,t-1}\} \\ &= \Lambda_{ik}\Lambda_{jk}\exp\{\tilde{\mu}_k\}\exp\{-\tilde{\phi}_k\tilde{\mu}_k\}\exp\{\frac{1}{2}\tilde{\sigma}_k^2\}\tilde{\phi}_k\exp\{\tilde{\phi}_k\tilde{h}_{k,t-1}\} \\ &\leq \log(n)^2\exp\{\log(n)\}\exp\{\sqrt{1 - \frac{1}{\log(\log(n))}}\log(n)\}\exp\{\frac{1}{2}\log(n)^{0.5}\} \\ &\times \sqrt{1 - \frac{1}{\log(\log(n))}}\exp\{\sqrt{1 - \frac{1}{\log(\log(n))}}(\log(n) + p\log(n))\} \\ &\leq \log(n)^2 n^{\frac{7}{2}+p} \text{ for } n \geq e^e\end{aligned}$$

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

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

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

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

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

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

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

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,

$$\begin{aligned}
\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})_{jj}^{-\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})_{ii}^{-\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})_{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})_{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})_{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}) \\
&= \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})_{ij} \\
&- \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})).
\end{aligned}$$

We will now show that the variance terms are bounded from below and above. Each of our parameters lies in a closed and bounded interval. We have that Θ_n is closed and bounded and $\#\theta < \infty$. Therefore, by the Heine-Borel theorem it is compact. 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 on Θ_n. Let $c_n = \min_{\theta \in \Theta_n} cov(\mathbf{r_t})_{ii}$ and $C_n = \max_{\theta \in \Theta_n} cov(\mathbf{r_t})_{ii}$. Since $cov(\mathbf{r_t})_{ii}$ is strictly positive and Θ_n is compact, $c_n > 0$ and $C_n < \infty$. Then, let $c = \min_{1 \le t \le n, 1 \le i \le N} c_n(t, i)$ and $C = \max_{1 \le t \le n, 1 \le i \le N} C_n$. Then, $0 < c \le cov(\mathbf{r_t})_{ii} \le C < \infty$. Let $L_n = \max_{0 \le \ell \le L} B_n^\ell$, 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,
$|\frac{\partial R(\mathbf{r_t})_{ij}}{\partial \theta}| \leq \frac{L_n}{\sqrt{cc}} + \frac{C}{2\sqrt{cc}}(\frac{L_n}{c} + \frac{L_n}{c}) = \frac{L_n}{c} + \frac{C}{2c}(\frac{2L_n}{c}) = (\frac{1}{c} + \frac{C}{c^2})L_n$.Therefore,
$|\frac{\partial R(\mathbf{r_t})_{ij}}{\partial \theta}| \leq C^{'}L_n,$ where $C^{'} = (\frac{1}{c} + \frac{C}{c^2})$.Now,
$||\nabla_\theta R(\mathbf{r_t})||_{OP} \leq ||\nabla_\theta R(\mathbf{r_t})||_F = \sqrt{\sum_{\ell=1}^{L}\sum_{i,j=1}^{N}|\frac{\partial R(\mathbf{r_t})_{ij}}{\partial \ell}|^2} \leq \sqrt{LN^2C'^2L_n^2} = \sqrt{L}NC^{'}L_n$
$= C^{''}L_n, C^{''} = \sqrt{L}NC^{'}$.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 some $s_{t,i,j} \in (0,1),$ with $\epsilon_{t,i,j}$ lying on the line segment from θ' to $\theta,$ such that $R(\mathbf{r_t})_{ij}(\theta) - R(\mathbf{r_t})_{ij}(\theta^{'}) = \nabla_\theta R(\mathbf{r_t})_{ij}(\epsilon_{t,i,j})^T(\theta - \theta^{'})$.

This implies,

$$
\begin{aligned}
|R(\mathbf{r_t})_{ij}(\theta) - R(\mathbf{r_t})_{ij}(\theta^{'})| &= |\nabla_\theta R(\mathbf{r_t})_{ij}(\epsilon_{t,i,j})^T(\theta - \theta^{'})| \\
&\leq ||\nabla_\theta R(\mathbf{r_t})_{ij}(\epsilon_{t,i,j})||_2||\theta - \theta^{'}||_2, \text{by the Cauchy-Schwarz inequality.}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
|R(\mathbf{r_t})_{ij}(\theta) - R(\mathbf{r_t})_{ij}(\theta^{'})| &\leq C^{''} L_n ||\theta - \theta^{'}||_2 .\text{Then,} \\
||R(\mathbf{r_t}) - R(\mathbf{r_t})||_F^2 &= \sum_{i=1}^{N}\sum_{j=1}^{N} |R(\mathbf{r_t})_{ij}(\theta) - R(\mathbf{r_t})_{ij}(\theta^{'})|^2 \\
&\leq N^2 C^{''2} L_n^2 ||\theta - \theta^{'}||_2^2 .\text{This implies ,} \\
||R(\mathbf{r_t})(\theta) - R(\mathbf{r_t})(\theta^{'})||_F &\leq C_1 L_n ||\theta - \theta^{'}||_2, C_1 = NC^{''}. Now, \\
||\frac{1}{n}\sum_{t=1}^{n} R(\mathbf{r_t})(\theta) - R(\mathbf{r_t})(\theta^{'})||_F &= \frac{1}{n}||\sum_{t=1}^{n} R(\mathbf{r_t})(\theta) - R(\mathbf{r_t})(\theta^{'})||_F \leq \frac{1}{n}\sum_{t=1}^{n} ||R(\boldsymbol{r}_t)(\theta) - R(\boldsymbol{r}_t)(\theta^{'})||_F \\
&\leq C_1 L_n ||\theta - \theta^{'}||_2 .
\end{aligned}
$$

Recall that $\Theta_n \subset [-B_{max}, B_{max}]^{C_\theta}$, where $B_{max} = 4log(n)$ is the largest upper bound in the definition of Θ_n and $C_\theta \in \mathbb{N}$ is a constant. Let $B_d(\theta^{'}, \epsilon) := \{\theta \in \Theta_n : d_n(\theta, \theta^{'}) < \epsilon\}$.

Suppose $||\theta - \theta^{'}|| < \delta$ which implies $d_n(\theta, \theta^{'}) \leq C_1 L_n ||\theta - \theta^{'}||_2 \leq C_1 L_n \delta$. We want $d_n(\theta, \theta^{'}) < r$ 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 $d_n(\theta, \theta^{'}) < \frac{\epsilon\xi}{2}$.

We know that

$$
\begin{aligned}
N(\delta, [-B_{max}, B_{max}]^{C_\theta}, ||\cdot||_2) &\leq (1 + \frac{2B_{max}}{\delta})^{C_\theta} \leq (\frac{3B_{max}}{\delta})^{C_\theta} \text{ which implies,} \\
\log(N(\delta, [-B_{max}, B_{max}]^{C_\theta}, ||\cdot||_2)) &\leq C_\theta \log(\frac{3B_{max}}{\delta}) \\
&= C_\theta [log(3) - log(\delta) + log(4log(n))] \\
&= C_\theta [log(4log(n)) + log(3) + log(\frac{1}{\delta})].
\end{aligned}
$$

We have 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

$\delta = \frac{\epsilon\xi}{2C_1L_n}$.

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 $d_n(\theta, \theta_0) < \epsilon$. Since, the balls cover Θ_n,θ must be in some $||\theta - \theta_j||_2 < \delta$. Then, $d_n(\theta, \theta_j) \leq C_1L_n||\theta - \theta_j||_2 < C_1L_n\delta = \frac{C_1L_n\epsilon\xi}{2C_1L_n} = \frac{\epsilon\xi}{2}$. Then, θ must be in the d_n ball of radius $\frac{\epsilon\xi}{2}$ centered around θ_j. Therefore, $B_d(\theta_0, \epsilon) \subset \bigcup_{j=1}^{N}\{\theta \in \Theta_n : d_n(\theta, \theta_j) < \frac{\epsilon\xi}{2}\}$. This implies $N(\frac{\epsilon\xi}{2}, B_d(\theta_0, \epsilon), d_n) \leq N(\delta, \Theta_n, || \cdot ||_2)$. That is,

$$\begin{aligned}N(\frac{\epsilon\xi}{2}, B_d(\theta_0, \epsilon), d_n) &\leq N(\delta, \Theta_n, || \cdot ||_2)\\ &\leq exp\{C_\theta[log(4log(n)) + log(3) + log(\frac{1}{\delta})]\end{aligned}$$

Recall, $\delta = \dfrac{\epsilon\xi}{2C_1L_n}$, which means

$\log(\dfrac{1}{\delta}) = \log(\dfrac{2C_1L_n}{\epsilon\xi}) = \log(2C_1) + \log(L_n) + \log(\dfrac{1}{\xi\epsilon})$.

This implies $\log(N(\dfrac{\epsilon\xi}{2}, B_d(\theta_0, \epsilon), d_n)) \leq C_\theta[log(4log(n)) + log(3) + log(2C_1) + log(L_n) + log(\dfrac{1}{\xi\epsilon})]$.

$\log(L_n) = log(C_2[n^a log^b(n)]) = alog(n) + blog(log(n)) + log(C_2)$.

Let $\epsilon = \epsilon_n = C_\epsilon n^{-\frac{1}{2+C_\theta}}$.Then,

$log(\dfrac{1}{\xi\epsilon_n}) = log(\dfrac{1}{\xi C_\epsilon n^{-\frac{1}{2+C_\theta}}}) = \dfrac{1}{2 + C_\theta}log(n) - log(\xi C_\epsilon), \epsilon_n \to 0,$ and $(n\epsilon_n^2)^{-1} = O(1)$.

Therefore,

$$\begin{aligned}\log(N(\frac{\epsilon\xi}{2}, \{\theta \in \Theta_n : d_n(\theta, \theta_j) < \epsilon\}, d_n)) \leq\ & C_\theta[log(4log(n)) + log(3) + log(2C_1) + alog(n)\\ &+ blog(log(n)) + log(C_2) + log(\frac{1}{\xi\epsilon})]\end{aligned}$$

Which implies

$$\begin{aligned}
\log(N(\frac{\epsilon\xi}{2}, \{\theta \in \Theta_n : d_n(\theta, \theta_j) < \epsilon\}, d_n)) &\leq C_\theta[log(4log(n)) + log(3) + log(2C_1) + alog(n) \\
&+ blog(log(n)) + log(C_2) + \frac{1}{2 + C_\theta} log(n) - log(\xi C_\epsilon)] \\
&= C_\theta[(a + \frac{1}{2 + C_\theta})log(n) + blog(log(n)) + log(4log(n)) + C_N] \\
&= Alog(n) + Blog(log(n)) + Clog(4log(n)) + D \\
&\leq Elog(n) + D, \text{ for } n \geq 9
\end{aligned}$$

Then,

$$Elog(n) + D \leq n\epsilon_n^2 \iff Elog(n) + D \leq C_\epsilon^2 n^{\frac{C_\theta}{2+C_\theta}},$$

which holds for sufficiently large n. Therefore,

$$\sup_{\epsilon > \epsilon_n} \log(N(\frac{\epsilon\xi}{2}, \{\theta \in \Theta_n : d_n(\theta, \theta_0) < \epsilon\}, d_n)) \leq n\epsilon_n^2,$$

since $N(\frac{\epsilon\xi}{2}, B_d(\theta_0, \epsilon), d_n) \leq N(\delta, \Theta_n, || \cdot ||_2)$ with the upper bound decreasing in ϵ since covering numbers are non-increasing in the radius. Therefore, the supremum is attained at ϵ_n $\square$

Lemma 8. *If f is continuous at x_0 with $f(x_0) > 0$ then $\exists \delta > 0$ such that $\forall r < \delta$, $\int_{x_0-r}^{x_0+r} f(x)dx \geq f(x_0)r$.*

Proof.

Let $\epsilon = \dfrac{1}{2}f(x_0) > 0$.

Then, $\exists \delta > 0$ such that $|x - x_0| < \delta$.This implies

$|f(x) - f(x_0)| < \dfrac{f(x_0)}{2}$, by the definition of a continuous function.

Therefore, $-\dfrac{f(x_0)}{2} < f(x) - f(x_0) < \dfrac{f(x_0)}{2}$, which implies

$f(x) > f(x_0) - \dfrac{f(x_0)}{2} = \dfrac{1}{2}f(x_0)$.

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 , $f(x) \geq \dfrac{f(x_0)}{2}$.This implies

$$\int\limits_{x_0-r}^{x_0+r} f(x)dx \geq \int\limits_{x_0-r}^{x_0+r} \frac{f(x_0)}{2}dx = \frac{f(x_0)}{2}[x]_{x_0-r}^{x_0+r} = \frac{f(x_0)}{2}(x_0 + r - x_0 + r) = rf(x_0).$$

$\square$

Proposition 9. *$\frac{\Pi(\theta\in\Theta_n : j\epsilon_n < d_n(\theta,\theta_0) < 2j\epsilon_n)}{\Pi_n(B_n(\theta_0,\epsilon_n;k))} \leq \exp\{\frac{1}{2}Kn\epsilon_n^2 j^2\}$, 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,$\mathbb{P}_\theta^{(n)}(x^{(n)}|\theta^{(n)}) = \prod_{t=1}^{n} P(x_t|\theta_t)$.The likelihood is give by

$$\ell_n(\theta) = \log(\prod_{t=1}^{n} P(x_t|\theta_t)) = \sum_{t=1}^{n} \log(P(x_t|\theta_t)) = \sum_{t=1}^{n} \ell_t(\theta).\text{This implies}$$

$$\ell_t(\theta) = \log[(2\pi)^{-\frac{N}{2}} det(\Sigma_t)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{r}_t - \boldsymbol{\mu}_t)^T \Sigma_t^{-1}(\mathbf{r}_t - \boldsymbol{\mu}_t)\}]$$
$$= -\frac{1}{2}(\log((2\pi)^N) + \log(det(\Sigma_t)) + (\boldsymbol{r}_t - \boldsymbol{\mu}_t)^T \Sigma_t^{-1}(\boldsymbol{r}_t - \boldsymbol{\mu}_t)).$$

Let $U = \{\theta \in \Theta : ||\theta - \theta_0||_2 \leq \delta\}$ for some $\delta > 0$.

On U we have the following bounds on Σ_t.

$\lambda_{min}(\Sigma_t) = \lambda_{min}(\Lambda\tilde{\Sigma}_t\Lambda^T + \bar{\Sigma}_t)$

$\geq \lambda_{min}(\Lambda\tilde{\Sigma}_t\Lambda^T) + \lambda_{min}(\bar{\Sigma}_t)$, by Weyl's inequality, since the matrices are real and symmetric.

Then , $\lambda_{min}(\Sigma_t) \geq \lambda_{min}(\bar{\Sigma}_t) = \min_{1 \leq i \leq m} e^{\bar{h}_{t,i}} > 0$, since $\tilde{\Sigma}_t \succeq 0$ and the extreme value theorem.

Similarly, $\lambda_{max}(\Sigma_t) \leq \lambda_{max}(\Lambda\tilde{\Sigma}_t\Lambda^T) + \lambda_{max}(\bar{\Sigma}_t)$

$\leq C < \infty$, by Weyl's inequality and the extreme value theorem.

We can now bound the KL divergence between P_θ and P_{θ_0}. Note that $KL_t(\theta_0, \theta_0) = 0$ and $\nabla_\theta KL_t(\theta_0, \theta_0) = 0$.Then, by Taylor's theorem, $KL_t(\theta, \theta_0) = 0 + 0 + \int_0^1 (1 - s)(\theta - \theta_0)' H(\theta_0 + s(\theta - \theta_0))(\theta - \theta_0)ds$. By the compactness of U, $\Sigma_t \succ 0$, and that we have a KL divergence between two Gaussians, we obtain

$$\lambda_{min} I \preceq H(\theta_0 + s(\theta - \theta_0)) \preceq \lambda_{max} I, \text{ which implies}$$
$$\lambda_{min}(\theta - \theta_0)'(\theta - \theta_0) \leq (\theta - \theta_0)' H(\theta_0 + s(\theta - \theta_0))(\theta - \theta_0) \leq \lambda_{max}(\theta - \theta_0)'(\theta - \theta_0).$$

This implies,

$$
\begin{aligned}
KL_t(\theta,\theta_0) &\leq \int_0^1 (1-s)\lambda_{max}||\theta-\theta_0||_2^2 ds \\
&= \lambda_{max}||\theta-\theta_0||_2^2 \int_0^1 (1-s)ds \\
&= \lambda_{max}||\theta-\theta_0||_2^2 [s-\frac{s^2}{2}]_0^1 \\
&= \frac{\lambda_{max}}{2}||\theta-\theta_0||_2^2 \\
&\leq C_1\delta^2, C_1 = \frac{\lambda_{max}}{2}.
\end{aligned}
$$

This implies, $KL_n(\theta,\theta_0) \leq C_1 n\delta^2$. Now,

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

$$
\begin{aligned}
\text{If } X \sim N(\boldsymbol{\mu}, \Sigma), \text{ then } \ell_\theta(x) &= \log\{(2\pi)^{-\frac{N}{2}} det(\Sigma)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(x-\boldsymbol{\mu})^{'}\Sigma^{-1}(x-\boldsymbol{\mu})\}\} \\
&= -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log(det(\Sigma)) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{'}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}). \\
\text{Then, } \log(\frac{P_{\theta_0}}{P_\theta}) &= \log(P_{\theta_0}) - \log(P_\theta) \\
&= -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log(det(\Sigma_0)) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^{'}\Sigma_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0) + \frac{N}{2}\log(2\pi) \\
&+ \frac{1}{2}\log(det(\Sigma_\theta)) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_\theta)^{'}\Sigma_\theta^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_\theta) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_\theta)^{'}\Sigma_\theta^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_\theta) \\
&+ \frac{1}{2}\log(det(\Sigma_\theta)) - \frac{1}{2}\log(det(\Sigma_0)) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_0})\Sigma_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu_0}) \\
&= \frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_\theta)^{'}\Sigma_\theta^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_\theta) - (\boldsymbol{x}-\boldsymbol{\mu}_0)^{'}\Sigma_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\} + \frac{1}{2}\log(\frac{|\Sigma_\theta|}{|\Sigma_0|}). \\
\text{Let } \eta &= X - \mu_0. \\
\text{Then }, \ell_t(\theta) &= \frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu_\theta})^{'}\Sigma_\theta^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_\theta) - \eta^{'}\Sigma_0^{-1}\eta + \log(\frac{|\Sigma_\theta|}{|\Sigma_0|})\}. \\
\text{Since } \eta &= X - \mu_0, \text{ this implies } X = \eta + \mu_0. \\
\text{Therefore }, (\boldsymbol{x}-\boldsymbol{\mu}_\theta) &= \eta + \boldsymbol{\mu}_0 - \boldsymbol{\mu}_\theta = \eta - \Delta\boldsymbol{\mu}, \text{ where } \Delta\boldsymbol{\mu} = \boldsymbol{\mu}_\theta - \boldsymbol{\mu}_0.
\end{aligned}
$$

This implies,

$$
\ell_t(\theta) = \frac{1}{2}\{(\eta-\Delta\boldsymbol{\mu})^{'}\Sigma_\theta^{-1}(\eta-\Delta\boldsymbol{\mu}) - \eta^{'}\Sigma_0^{-1}\eta + \log(\frac{|\Sigma_\theta|}{|\Sigma_0|})\}.
$$

$$
\begin{aligned}
\text{Now,} (\eta-\Delta\boldsymbol{\mu})^{'}\Sigma_\theta^{-1}(\eta-\Delta\boldsymbol{\mu}) &= (\eta^{'}-\Delta\boldsymbol{\mu}^{'})\Sigma_\theta^{-1}(\eta-\Delta\boldsymbol{\mu}) \\
&= (\eta^{'}\Sigma_\theta^{-1} - \Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1})(\eta-\Delta\boldsymbol{\mu}) \\
&= \eta^{'}\Sigma_\theta^{-1}\eta - \eta^{'}\Sigma_\theta^{-1}\Delta\boldsymbol{\mu} - \Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1}\eta + \Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1}\Delta\boldsymbol{\mu} \\
&= \eta^{'}\Sigma_\theta^{-1}\eta - 2\Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1}\eta + \Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1}\Delta\boldsymbol{\mu}. \\
\text{This implies } \ell_t(\theta) &= \frac{1}{2}\{\log(\frac{|\Sigma_\theta|}{|\Sigma_0|}) + \Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1}\Delta\boldsymbol{\mu}\} + \frac{1}{2}\{\eta^{'}\Sigma_\theta^{-1}\eta - 2\Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1}\eta - \eta^{'}\Sigma_0^{-1}\eta\} \\
&= C_\theta + \frac{1}{2}\eta^{'}\Sigma_\theta^{-1}\eta - \Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1}\eta - \frac{1}{2}\eta^{'}\Sigma_0^{-1}\eta \\
&= C_\theta + \frac{1}{2}\eta^{'}(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - \Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1}\eta, C_\theta = \frac{1}{2}\{\log(\frac{|\Sigma_\theta|}{|\Sigma_0|}) + \Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1}\Delta\boldsymbol{\mu}\}
\end{aligned}
$$

$$\text{Observe that,} \eta \sim N(0, \Sigma_0).$$

$$\begin{aligned}
\text{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}\Sigma_0) - tr(\Sigma_0^{-1}\Sigma_0) = tr(\Sigma_\theta^{-1}\Sigma_0) - N.\text{This implies} \\
\mathbb{E}_{\theta_0}[\ell_t(\theta)] &= C_\theta + \frac{1}{2}tr(\Sigma_\theta^{-1}\Sigma_0) - \frac{N}{2}. \\
\text{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^{-1} - \Sigma_0^{-1})\eta - tr(\Sigma_\theta^{-1}\Sigma_0)\} + \frac{N}{2} - \Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1}\eta \\
&= \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_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_0 - \Sigma_0^{-1}\Sigma_0)] - \Delta\boldsymbol{\mu}^{'}\Sigma_\theta^{-1}\eta. \\
\text{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. \\
\text{Now ,} Var(Y_t) &= Var(\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) \\
&= Var_{\theta_0}(\frac{1}{2}\{\eta^{'}(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - tr((\Sigma_\theta^{-1} - \Sigma_0^{-1})\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_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_{\theta_0}(\eta)\Sigma_\theta^{-1}\Delta\boldsymbol{\mu} \\
&+ 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}^{'}\Sigma_\theta^{-1}\Sigma_0\Sigma_\theta^{-1}\Delta\boldsymbol{\mu} \\
&+ 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). \\
\text{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.
\end{aligned}$$

By the Mean Value Theorem $\exists \epsilon_{ij} = \theta_0 + s_{ij}(\theta - \theta_0), s_{ij} \in (0,1)$ such that

$$\Sigma_{ij}(\theta) - \Sigma_{ij}(\theta_0) = \nabla_\theta \Sigma_{ij}(\epsilon_{ij})^{'}(\theta - \theta_0), \text{ which implies}$$
$$|\Sigma_{ij}(\theta) - \Sigma_{ij}(\theta_0)| \leq ||\nabla_\theta \Sigma_{ij}(\epsilon_{ij})||_2 ||\theta - \theta_0||_2.$$

From the proof of Proposition 7, we found that all the partial derivatives are continuous. Since U is a compact set, we have that $||\nabla_\theta \Sigma_{ij}(\epsilon')||_2$ is bounded by some constant, say $L_{ij} < \infty$, by the extreme value theorem. Therefore,

$$||\Sigma(\theta) - \Sigma(\theta_0)||_F^2 \leq L_1^2 ||\theta - \theta_0||_2^2, L_1 = (\sum_{i,j=1}^{N} L_{ij}^2)^{\frac{1}{2}}. \text{ Observe that}$$

$$\begin{aligned}
||\Sigma_\theta^{-1} - \Sigma_0^{-1}||_F &= ||\Sigma_\theta^{-1}(\Sigma_0 - \Sigma_\theta)\Sigma_0^{-1}||_F \\
&\leq ||\Sigma_\theta^{-1}||_F ||(\Sigma_0 - \Sigma_\theta)\Sigma_0^{-1}||_F, \text{ since } ||AB||_F \leq ||A||_F ||B||_F. \text{ Then },\\
||\Sigma_\theta^{-1} - \Sigma_0^{-1}||_F &\leq ||\Sigma_\theta^{-1}||_F ||\Sigma_0 - \Sigma_\theta||_F ||\Sigma_{\theta_0}^{-1}||_F \\
&\leq ||\Sigma_\theta^{-1}||_F ||\Sigma_0 - \Sigma_\theta||_F ||\Sigma_\theta^{-1}||_F \\
&\leq ||\Sigma_\theta^{-1}||_F ||\Sigma_\theta^{-1}||_F L_1 ||\theta - \theta_0||_2 \\
&\leq ||\Sigma_\theta^{-1}||_F ||\Sigma_\theta^{-1}||_F L_1 ||\theta - \theta_0||_2 \\
&= \sqrt{tr(\Sigma_0^{-1}\Sigma_0^{-1}) tr(\Sigma_\theta^{-1}\Sigma_\theta^{-1})} L_1 ||\theta - \theta_0||_2 \\
&= \sqrt{tr(\Sigma_0^{-2}) tr(\Sigma_\theta^{-2})} L_1 ||\theta - \theta_0||_2 \\
&= \sqrt{\sum_{i=1}^{N} \lambda_i(\Sigma_\theta^{-2}) \sum_{i=1}^{n} \lambda_i(\Sigma_0^{-2})} L_1 ||\theta - \theta_0||_2 \\
&\leq \sqrt{\frac{N^2}{\lambda_{0,min}^2 \lambda_{\theta,min}^2}} L_1 ||\theta - \theta_0||_2 \\
&= \frac{N}{\lambda_{0,min}\lambda_{\theta,min}} L_1 ||\theta - \theta_0||_2.
\end{aligned}$$

Therefore,

$$
\begin{aligned}
||\Sigma_\theta^{-1} - \Sigma_0^{-1}||_F &\leq C_2 L_1 ||\theta - \theta_0||_2. \text{ This implies,} \\
tr([(\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0]^2) &\leq C_2^2 L_1^2 ||\theta - \theta_0||_2^2 ||\Sigma_0||_F^2 \\
&\leq C_2^2 L_1^2 \delta^2 ||\Sigma_0||_F^2. \text{Therefore,} \\
Var(Y_t) &\leq \frac{1}{2} C_2^2 L_1^2 \delta^2 ||\Sigma_0||_F^2 + \Delta\mu' \Sigma_\theta^{-1} \Sigma_0 \Sigma_\theta^{-1} \Delta\mu \\
&+ 2cov(\frac{1}{2}\{\eta'(\Sigma_\theta^{-1} - \Sigma_0^{-1})\eta - tr((\Sigma_\theta^{-1} - \Sigma_0^{-1})\Sigma_0)\}, -\Delta\mu' \Sigma_\theta^{-1}\eta) \\
&= \frac{1}{2} C_2^2 L_1^2 \delta^2 ||\Sigma_0||_F^2 + \Delta\mu' \Sigma_\theta^{-1} \Sigma_0 \Sigma_\theta^{-1} \Delta\mu + 0
\end{aligned}
$$

$$
\begin{aligned}
\Delta\mu' \Sigma_\theta^{-1} \Sigma_0 \Sigma_\theta^{-1} \Delta\mu &\leq ||\Delta\mu||_2^2 ||\Sigma_\theta^{-1} \Sigma_0 \Sigma_\theta^{-1}||_2 \\
&\leq \delta^2 ||\Sigma_\theta^{-1} \Sigma_0 \Sigma_\theta^{-1}||_2 \\
&\leq \delta^2 ||\Sigma_\theta^{-1}||_F ||\Sigma_0||_F ||\Sigma_\theta^{-1}||_F \\
&= \delta^2 tr(\Sigma_\theta^{-2})^{\frac{1}{2}} tr(\Sigma_0^2)^{\frac{1}{2}} tr(\Sigma_\theta^{-2})^{\frac{1}{2}} \\
&= \delta^2 (\sum_{i=1}^{N} \frac{1}{\lambda_i(\Sigma_\theta)^2})(\sum_{i=1}^{N} \lambda_i(\Sigma_0)^2)^{\frac{1}{2}} \\
&\leq \delta^2 (\sum_{i=1}^{N} \frac{1}{\lambda_{min}(\Sigma_\theta)^2})(\sum_{i=1}^{N} \lambda_{max}(\Sigma_0)^2)^{\frac{1}{2}}, \text{ which exist and are finite on U} \\
&= \delta^2 \frac{N}{\lambda_{min}(\Sigma_\theta)^2} N^{\frac{1}{2}} \lambda_{max}(\Sigma_0) \\
&= \delta^2 C_3, \text{ where } C_3 = \frac{N}{\lambda_{min}(\Sigma_\theta)^2} N^{\frac{1}{2}} \lambda_{max}(\Sigma_0).
\end{aligned}
$$

Then,

$$
\begin{aligned}
Var(Y_t) &\leq C_4 \delta^2 + C_3 \delta^2, \text{ where } C_4 = \frac{1}{2} C_2^2 L_1^2 N^2 \lambda_{max}(\Sigma_0)^4 \\
&= C_5 \delta^2, \text{where } C_5 = C_4 + C_3. \text{Then,} \\
Var(Y_n) &\leq C_5 n \delta^2.
\end{aligned}
$$

Let

$$\delta = \frac{\epsilon_n}{\sqrt{max\{C_1, C_5\}}}.$$

$$\text{Then, } U_n = \{\theta \in \Theta : ||\theta - \theta_0||_2 \leq \frac{\epsilon_n}{\sqrt{max\{C_1, C_5\}}}\}$$

On U_n we have shown

$$KL_n(\theta_0, \theta) \leq n\epsilon_n^2 \text{ and } V_{2,0}(\theta_0, \theta) \leq n\epsilon_n^2 \text{.Therefore ,}$$

$$U_n \subseteq B_n(\theta_0, \epsilon_n, 2) \text{ which implies } \Pi(B_n(\theta_0, \epsilon_n, 2)) \geq \mathbb{P}(U_n).$$

Therefore, to lower bound the denominator, we can lower bound the probability of U_n. The required probability is given by

$$\mathbb{P}(U_n) = \mathbb{P}(\theta \in \Theta : ||\theta - \theta_0||_2 \leq \frac{\epsilon_n}{C_6}), \text{ where } C_6 = \frac{1}{\sqrt{max\{C_1, C_5\}}}.$$

$$\text{Note, } \bigcap_{j=1}^{C_\theta} \{|\theta^{(j)} - \theta_0^{(j)}| \leq \frac{\epsilon_n}{C_6}\} \subseteq \{||\theta - \theta_0||_2 \leq \frac{\epsilon_n}{C_6}\}. \text{ This implies}$$

$$\mathbb{P}(||\theta - \theta_0||_2 \leq \frac{\epsilon_n}{C_6}) \geq \mathbb{P}(\bigcap_{j=1}^{C_\theta} \{|\theta^{(j)} - \theta_0^{(j)}| \leq \frac{\epsilon_n}{C_6}\})$$

$$= \prod_{j=1}^{C_\theta} \mathbb{P}(|\theta^{(j)} - \theta_0^{(j)}| \leq \frac{\epsilon_n}{C_6}) = \prod_{k=1}^{C_\theta} \int_{\theta_0 - \frac{\epsilon_n}{C_6}}^{\theta_0 + \frac{\epsilon_n}{C_6}} f\theta_j(x)dx.$$

Now we just need to compute or lower bound these integrals. To start we will consider an arbitrary autoregressive process of order one with Z-distributed innovations.

Recall, $X_{t+1} = \mu + \phi(X_t - \mu) + \epsilon_t,$ with $f_\epsilon(x) = \dfrac{1}{\pi}\dfrac{e^{\frac{x}{2}}}{1+e^x}.$

Let $Y_t = X_t - \mu$. This implies $Y_{t+1} = \phi Y_t + \epsilon_t$.

$$
\begin{aligned}
\text{Then },\phi_Y(u) &= \mathbb{E}[e^{iuY_t}] \\
&= \mathbb{E}[\exp\{iu(\phi Y_{t-1} + \epsilon_{t-1})\}] \\
&= \mathbb{E}[\exp\{iu\phi Y_{t-1}\}\exp\{iu\epsilon_{t-1}\}] \\
&= \mathbb{E}[\exp\{iu\phi Y_{t-1}\}]\mathbb{E}[\exp\{iu\epsilon_{t-1}\}] \\
&= \phi_Y(\phi u)\phi_\epsilon(u) \\
&= \prod_{k=0}^{n}\phi_\epsilon(\phi^k u)\phi_Y(\phi^{n+1}u).
\end{aligned}
$$

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 $cosh(x) = \dfrac{e^x + e^{-x}}{2}$ which implies $2cosh(\dfrac{x}{2}) = e^{\frac{x}{2}} + e^{-\frac{x}{2}}$.

Furthermore , $\dfrac{e^{\frac{x}{2}}}{1+e^x} = \dfrac{e^{\frac{x}{2}}}{e^0+e^x}$

$$= \frac{e^{\frac{x}{2}}}{e^{-\frac{x}{2}}e^{\frac{x}{2}} + e^{\frac{x}{2}}e^{\frac{x}{2}}} = \frac{e^{\frac{x}{2}}}{e^{\frac{x}{2}}(e^{-\frac{x}{2}} + e^{\frac{x}{2}})} = \frac{1}{e^{-\frac{x}{2}} + e^{\frac{x}{2}}} = \frac{1}{2cosh(\frac{x}{2})}.$$

This implies $f_\epsilon(x) = \dfrac{1}{2\pi cosh(\frac{x}{2})}$.

This results in,

$$\phi_\epsilon(u) = \int_{-\infty}^{\infty} e^{iux}\frac{1}{2\pi cosh(\frac{x}{2})}dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{iux} sech(\frac{x}{2})dx = sech(\pi u), \text{by Fourier transform.}$$

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, $f_X(x) = f_Y(x-\mu) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iu(x-\mu)}\prod_{k=0}^{\infty} sech(\pi\phi^k u)du$.

Now we will apply Lemma 8 to each of the parameter blocks. Firstly, look at the DSP priors.

$$f_X(x_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iu(x_0-\mu)}\prod_{k=0}^{\infty} sech(\pi\phi^k u)du = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iu(x_0-\mu)}\phi_Y(u)du.$$

The function $\phi_Y(u)$ is integrable.

Then, by the dominated convergence theorem the map $x \to \int_{-\infty}^{\infty} e^{-iu(x-\mu)}\phi_Y(u)du$ is continuous .

If $x \to x_0$ then for each fixed $u, e^{-iu(x-\mu)} \to e^{-iu(x_0-\mu)}$

we have $|e^{-iu(x-\mu)}\phi_Y(u)| \leq |\phi_Y(u)|$ which is integrable .

Therefore, $\lim_{x\to x_0} f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iu(x_0-\mu)}\phi_Y(u)du = f(x_0)$.

Therefore, f is continuous at x_0.

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iu(x-\mu)} sech(\pi u)\prod_{k=1}^{\infty} sech(\pi\phi^k u)du$$

Since $|\phi| < 1$, the infinite product of characteristic functions converge, which is itself a characteristic function.

Let $\{X_k\}_{k\geq 1}$ be i.i.d with hyperbolic secant distributions. Define $Y_k := \pi\phi^k X_k$. Then,

$$\phi_{Y_k}(u) = \mathbb{E}[e^{iuY_k}] = sech(\pi\phi^k u).$$

Then, for $n \geq 1$,

$$\phi_n(u) = \prod_{k=1}^{n} sech(\pi\phi^k u) = \prod_{k=1}^{n} \phi_{Y_k}(u) = \phi_{S_n}(u), \text{ which is a characteristic function.}$$

Note, $\mathbb{E}[X_k] = 0$ and $Var(X_k) = 1$. Then, $Var(Y_k) = \pi^2\phi^{2k}$. Therefore,

$$\sum_{k=1}^{\infty} Var(Y_k) = \pi^2 \sum_{k=1}^{\infty} \phi^{2k} = \frac{\pi^2\phi^2}{1-\phi^2} < \infty.$$

$$\begin{aligned}
\mathbb{E}[|S_m - S_n|^2] &= \mathbb{E}[(\sum_{k=n+1}^{m} Y_k)^2] \\
&= \mathbb{E}[\sum_{k=n+1}^{m} Y_k^2 + 2 \sum_{n+1\leq i,j\leq m} Y_iY_j] \\
&= \mathbb{E}[\sum_{k=n+1}^{m} Y_k^2] \\
&= \sum_{k=n+1}^{m} E[Y_k^2] \\
&= \sum_{k=n+1}^{m} Var(Y_k) \\
&= \pi^2 \sum_{k=n+1}^{m} \phi^{2k} \\
&\leq \pi^2 \sum_{k=n+1}^{\infty} \phi^{2k}.
\end{aligned}$$

Let $r = k - (n+1)$. This implies,

$$\begin{aligned}
\mathbb{E}[|S_m - S_n|^2] &\leq \pi^2 \sum_{r=0}^{\infty} \phi^{2(r+(n+1))} \\
&= \pi^2 \sum_{r=0}^{\infty} \phi^{2r} \phi^{2(n+1)} \\
&= \frac{\pi^2 \phi^{2(n+1)}}{1-\phi^2} \to 0, \text{ as } n \to \infty.
\end{aligned}$$

Therefore, $S_n \to S$ in probability. Now,

$$|e^{iuS_n} - e^{iuS}| \leq |u||S_n - S|.$$

Then,

$$\begin{aligned}
|\phi_n(u) - \mathbb{E}[e^{ius}]| &= |\mathbb{E}[e^{iuS_n}] - \mathbb{E}[e^{iuS}]| \\
&= |\mathbb{E}[e^{iuS_n} - e^{ius}]| \\
&\leq \mathbb{E}[|e^{iuS_n} - e^{ius}|] \\
&\leq \mathbb{E}[|u||S_n - S|] \\
&= |u|\mathbb{E}[|S_n - S|] \\
&\leq |u|(\mathbb{E}|S_n - S|^2)^{\frac{1}{2}} \to 0 \text{ as } n \to \infty.
\end{aligned}$$

Thus,

$$\phi(u) := \lim_{n\to\infty} \phi_n(u) = \mathbb{E}[e^{iuS}], \text{ where } \phi(u) = \prod_{k=1}^{\infty} sech(\pi\phi^k u).$$

Hence, there exists a probability measure H on $\mathbb{R}$ such that

$$R(u) = \prod_{k=1}^{\infty} sech(\pi\phi^k u) = \int_{-\infty}^{\infty} e^{iuy} H(dy).$$

Furthermore,

$$g(x) := \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-iux} sech(\pi u) du = \frac{1}{2\pi} sech(\frac{x}{2}) > 0, \forall x.$$

This implies,

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-iu(x-\mu)} sech(\pi u)(\int\limits_{-\infty}^{\infty} e^{iuy} H(dy))du \\ &= \int\limits_{-\infty}^{\infty} [\frac{1}{2\pi} e^{-iu(x-\mu-y)} sech(\pi u) du] H(dy), \end{aligned}$$

by the Fubini-Tonelli theorem since $|e^{-iu(x-\mu)} sech(\pi u) e^{iuy}| \leq sech(\pi u)$ with $\int\limits_{-\infty}^{\infty} sech(\pi u) du = 1$. Therefore,

$$f(x) = \int\limits_{-\infty}^{\infty} \frac{1}{2\pi} sech(\frac{x-\mu-y}{2}) H(dy) > 0,$$

since $\frac{1}{2\pi} sech(\frac{x-\mu-y}{2}) > 0$ everywhere and$H(y)$ is a probability measure.

Then by Lemma 8 for the log-variance processes we have

$$\int\limits_{h_{0,t-1}-\frac{\epsilon_n}{C_3}}^{h_{0,t-1}+\frac{\epsilon_n}{C_3}} f_h(x)dx \geq f_h(h_{0,t-1})\frac{1}{C_3}\epsilon_n, \text{ for sufficiently large n.}$$

For the mean parameters $\mu = log(\tau_0^2\tau_1^2), \tau_0 \sim C^+(0, \frac{1}{\sqrt{n}}), \tau_1 \sim C^+(0,1)$.

The probability density function of a half cauchy distribution with scale σis $\frac{2}{\pi\sigma}\frac{1}{1+\frac{y^2}{\sigma^2}} = \frac{2}{\pi\sigma(1+\frac{y^2}{\sigma^2})} = \frac{2\sigma}{\pi\sigma^2(1+\frac{y^2}{\sigma^2})} = \frac{2\sigma}{\pi(\sigma^2+y^2)}$.

Let $X = \tau^2$.

Then, $f_X(x) = f_\tau(\sqrt{x})\dfrac{d}{dx}(\sqrt{x}) = \dfrac{2\sigma}{\pi(\sigma^2 + x)}\dfrac{1}{2\sqrt{x}} = \dfrac{\sigma}{\sqrt{x}(\sigma^2 + x)\pi}$.

Now, let $U = log(X)$ which implies $X = e^u$ which leads to $dx = e^u du$.

Therefore, $f_U(u) = f_X(e^u)e^u = \dfrac{\sigma e^u}{\sqrt{e^u}(\sigma^2 + e^u)\pi} = \dfrac{\sigma e^{\frac{1}{2}u}}{(\sigma^2 + e^u)\pi}$.

Let $\mu = U_0 + U_1$.

$$\text{Then, } f_\mu(m) = \int\limits_{-\infty}^{\infty} f_{U_0}(t) f_{U_1}(m - t)dt$$

$$= \int\limits_{-\infty}^{\infty} \frac{S_0}{\pi}\frac{e^{\frac{t}{2}}}{(S_0^2 + e^t)}\frac{S_1}{\pi}\frac{e^{\frac{(m-t)}{2}}}{e^{m-t} + S_1^2}dt$$

$$= \frac{S_0 S_1 e^{\frac{m}{2}}}{\pi^2}\int\limits_{\infty}^{\infty}\frac{1}{(S_0^2 + e^t)(e^{m-t} + S_1^2)}dt.$$

Let $x = e^t$ which implies $dt = \dfrac{1}{x}dx$.

$$\text{Therefore, } \int\limits_{-\infty}^{\infty}\frac{1}{(e^t + s_0^2)(e^{m-t} + S_1^2)}dt = \int\limits_{0}^{\infty}\frac{1}{x(x + S_0^2)(\frac{e^m}{x} + s_1^2)}dx$$

$$= \int\limits_{0}^{\infty}\frac{1}{(x + s_0^2)(e^m + s_1^2 x)}dx.$$

Note that, $\dfrac{1}{(x + a)(b + cx)} = \dfrac{1}{b - ac}(\dfrac{1}{x + a} - \dfrac{c}{b + cx})$.

Therefore, the integral with such an integrand is equal to

$\dfrac{1}{b - ac}[log(x + a) - log(x + cx)]_0^\infty = \dfrac{1}{b - ac}log(\dfrac{b}{ac})$.

$$\text{Therefore, } \int\limits_{0}^{\infty}\frac{1}{(x + s_0^2)(e^m + s_1^2 x)}dx = \frac{m + log(n)}{e^m - \frac{1}{n}}.$$

This implies $f_\mu(m) = \dfrac{S_0 S_1 e^{\frac{m}{2}}}{\pi^2}\dfrac{m + log(n)}{e^m - \frac{1}{n}} = \dfrac{e^{\frac{m}{2}}}{\sqrt{n}\pi^2}\dfrac{m + log(n)}{e^m - \frac{1}{n}}$.

Suppose, $\mu_0 \neq -log(n)$.

Then the factor $\dfrac{e^{\frac{m}{2}}}{\sqrt{n}\pi^2}$ is strictly positive and continuous everywhere on $\mathbb{R}$.

Then, $\frac{m+log(n)}{e^m-\frac{1}{n}}$ is the quotient of two functions that are continuous at μ_0 and have non-zero values when $\mu_0 \neq -log(n)$.

Therefore, the quotient is continuous at μ_0.

We also know that $sign(m + log(n)) = sign(e^m - \frac{1}{n})$.Therefore, $\frac{m + log(n)}{e^m - \frac{1}{n}} > 0$.

Therefore, $\lim_{m\to\mu_0} f_\mu(m) = f_\mu(\mu_0) > 0$.

When $\mu_0 = -log(n)$we get $\frac{0}{0}$ which is undefined .

However, f_μ extends continuously and positively through this point.

Let $g(m) = m + log(n)$ and $h(m) = e^m - \frac{1}{n}, A(m) = \frac{e^{\frac{m}{2}}}{\sqrt{n}\pi^2}$, and $B(m) = \frac{g(m)}{h(m)}$.

Now, apply L'Hôpital's rule , $\lim_{m\to-log(n)} B(m) = \lim_{m\to-log(n)} \frac{g'(m)}{h'(m)} = e^{log(n)} = n$.

Further , $\lim_{m\to-log(n)} A(m) = \frac{e^{-\frac{log(n)}{2}}}{\sqrt{n}\pi^2} = \frac{n^{-\frac{1}{2}}}{\sqrt{n}\pi^2} = \frac{1}{n\pi^2}$.

This implies $\lim_{m\to-log(n)} f_\mu(n) = \lim_{m\to-log(n)} A(m)B(m) = \frac{1}{\pi^2}$.

So, defining $f_\mu(-log(n)) := \frac{1}{\pi^2}$ makes f_μ continuous and positive at $m = -log(n)$.

Therefore, by Lemma 8 , $\int_{\mu_0-\frac{\epsilon_n}{C_3}}^{\mu_0+\frac{C_3}{\epsilon_n}} f_\mu(x)dx \geq f_\mu(\mu_0)\frac{1}{C_3}\epsilon_n,$ for sufficiently large n.

$$
\begin{aligned}
\text{Let } Y &= \frac{\phi+1}{2}.\\
\text{Then }, f_Y(y) &= \frac{1}{B(a,b)} y^{a-1}(1-y)^{b-1}.\\
\text{Observe that } \phi &= g(Y) = 2Y-1.\\
\text{Then }, Y &= g^{-1}(\phi) = \frac{\phi+1}{2}.\\
\text{Then }, \frac{d}{d\phi}(g^{-1}(\phi)) &= \frac{d}{d\phi}(\frac{\phi+1}{2}) = \frac{1}{2}.\\
\text{This implies }, f_\phi(\phi) &= f_Y(g^{-1}(x))|\frac{d}{dx}g^{-1}(x)| = f_Y(\frac{\phi+1}{2})\frac{1}{2}, \text{ since } y = \frac{x+1}{2}.\\
\text{We set }, f_\phi(\phi) &= \frac{1}{2B(a,b)}(\frac{\phi+1}{2})^{a-1}(1-\frac{\phi+1}{2})^{b-1}\\
&= \frac{1}{2^{a+b-1}B(a,b)}(\phi+1)^{a-1}(1-\phi)^{b-1}, \text{ where } |\phi| < 1.
\end{aligned}
$$

On $(-1,1)$,$(\phi+1)^{a-1}$ and $(1-\phi)^{b-1}$ are both continuous.

Therefore ,$f_\phi(\phi)$ 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 Lemma 8, this implies $\displaystyle\int\limits_{\phi_0-\frac{\epsilon_n}{C_3}}^{\phi_0+\frac{\epsilon_n}{C_3}} f_\phi(x)dx \geq f_\phi(\phi_0)\frac{1}{C_3}\epsilon_n,$ for sufficiently large n.

Note that $\beta_{a,t-1}$,$\bar{\mu}_a$ for each asset a, and $\tilde{\mu}_j$ are realized from normal distributions that are continuous everywhere and have a strictly positive density in $\mathbb{R}$. Therefore, we may invoke Lemma 8 to bound the integrals. Similarly, for $\bar{\sigma}^2$ and $\tilde{\sigma}^2$, they both follow gamma densities that are continuous everywhere and strictly positive, which allows us to attain similar bounds for sufficiently large n. Furthermore, the conditional distribution of each $\Lambda_{i,j}$ is also normal, which again allows us to obtain a lower bound.

For the remaining log variance processes with normally distributed innovations

$$f_h(x) = \sqrt{\frac{1-\phi^2}{2\pi\sigma^2}}\exp\{-(\frac{1-\phi^2}{2\sigma^2}x^2)\}, |\phi| < 1, \sigma^2 >,$$

which is continuous and positive at x_0.Therefore,

$$\int_{h_{0,t-1}-\frac{\epsilon_n}{C_3}}^{h_{0,t-1}+\frac{\epsilon_n}{C_3}} f_h(x)dx \geq f_h(h_{0,t-1})\frac{1}{C_3}\epsilon_n, \text{by Lemma 8, for sufficiently large n}$$

Therefore,

$$\Pi_n(B_n(\theta_0, \epsilon_n, k)) \geq m^{C_\theta}\epsilon_n^{C_\theta}$$

where $m := minC_3 f_{\theta_j}(\theta_{j,0})$. Thus, we need to verify that

$$\begin{aligned}
\frac{1}{(m\epsilon_n)^{C_\theta}} &\leq e^{\frac{Kn\epsilon_n^2 j^2}{2}} \\
\iff \frac{1}{(mC_\epsilon n^{-\frac{1}{2+C_\theta}})^{C_\theta}} &\leq \frac{KC_\epsilon^2 n^{\frac{C_\theta}{2+C_\theta}} j^2}{2} \\
\iff \frac{n^{\frac{C_\theta}{2+C_\theta}}}{(mC_\epsilon)^{C_\theta}} &\leq \frac{KC_\epsilon^2 n^{\frac{C_\theta}{2+C_\theta}} j^2}{2} \\
\iff \frac{1}{(mC_\epsilon)^{C_\theta}} &\leq \frac{KC_\epsilon^2 j^2}{2} \\
\iff \frac{1}{m} &\leq \frac{KC_\epsilon^{2+C_\theta} j^2}{2} \\
\iff (\frac{2}{Kj^2 m})^{\frac{1}{2+C_\theta}} &\leq C_\epsilon
\end{aligned}$$

Therefore, the bound holds $\forall j \in \mathbb{N}$ for a suitably chosen C_ϵ. □

Theorem 10. $\mathbb{P}_{\theta_0}^{(n)}\Pi_n(\theta \in \Theta_n : ||R_\theta - R_{\theta_0}||_F \geq M_n\epsilon_n|X^{(n)}) \to 0$ *for every* $M_n \to \infty$

Proof. We have proved all the conditions of Theorem 1 of Ghosal and van der Vaart [2007]. Therefore, by application of their Theorem 1 we prove Theorem 10. □

Corollary 11. $\mathbb{P}_{\theta_0}^{(n)}\Pi_n(\theta \in \Theta_n : |Score(R_\theta) - Score(R_{\theta_0})| \geq \frac{M\epsilon_n}{N-1}|X_n) \to 0$

Proof.

$$\begin{aligned} score(A) - score(B) &= \frac{\sum_{j=1}^{N} \boldsymbol{x}_j^T \mathbf{1} - N}{N(N-1)} - \frac{\sum_{j=1}^{N} \boldsymbol{y}_j^T \mathbf{1} - N}{N(N-1)} \\ &= \frac{1}{N(N-1)} \sum_{j=1}^{N} (\boldsymbol{x}_j^T \mathbf{1} - \boldsymbol{y}_j^T \mathbf{1}) \\ &= \frac{1}{N(N-1)} \sum_{j=1}^{N} (\boldsymbol{x}_j - \boldsymbol{y}_j)^T \mathbf{1}. \end{aligned}$$

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,

$$\begin{aligned} |score(A) - score(B)| &= |\frac{1}{N-1} \sum_{j=1}^{N} (\boldsymbol{x}_j - \boldsymbol{y}_j)^T \mathbf{1}| \\ &= \frac{1}{N(N-1)} |\sum_{j=1}^{N} (\boldsymbol{x}_j - \boldsymbol{y}_j)^T \mathbf{1}| \\ &\leq \frac{1}{N(N-1)} \sum_{j=1}^{N} |(\boldsymbol{x})_j - \boldsymbol{y}_j)^T \mathbf{1}|, \text{ by the triangle inequality.} \end{aligned}$$

$$\begin{aligned} \text{Then }, |score(A) - score(B)| &\leq \frac{1}{N(N-1)} \sum_{j=1}^{N} \sqrt{N} ||\boldsymbol{x}_j - \boldsymbol{y}_j||_2 \\ &= \frac{\sqrt{N}}{N(N-1)} \sum_{j=1}^{N} ||\boldsymbol{x}_j - \boldsymbol{y}_j||_2. \end{aligned}$$

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 nonnegative. 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

$$\begin{aligned}
|score(A) - Score(B)| &\leq \frac{\sqrt{N}}{N(N-1)} \sum_{j=1}^{N} ||\boldsymbol{x}_j - \boldsymbol{y}_j||_2 \\
&\leq \frac{\sqrt{N}}{N(N-1)} \sqrt{N}||A - B||_F \\
&= \frac{1}{N-1}||A - B||_F.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{P}_{\theta_0}^{(n)} \Pi_n(\theta \in \Theta_n : |score(R_\theta) - score(R_{\theta_0})| \geq \frac{M\epsilon_n}{N-1}|X^{(n)}) \\
&\leq \mathbb{P}_{\theta_0}^{(n)} \Pi_n(\theta \in \Theta_n : ||R_\theta - R_{\theta_0}||_F > M_n\epsilon_n|X^{(n)}) \to 0.
\end{aligned}$$

$\square$