DECONVOLUTION OF TRANSCRIPT PROFILING DATA AND ASYMPTOTIC INFERENCE OF CROSS CORRELATION IN L^∞

BY DIE SUN

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics and Biostatistics

Written under the direction of

Dr. Han Xiao

and approved by

New Brunswick, New Jersey

January, 2017

ABSTRACT OF THE DISSERTATION

Deconvolution of Transcript Profiling Data and Asymptotic Inference of Cross Correlation in L^{∞}

by Die Sun

Dissertation Director: Dr. Han Xiao

In this dissertation, we consider two research projects: deconvolution of transcript profiling data; and inferences for multivariate time series based on cross correlations, especially under high dimensionality.

The study of transcript profiling data such as macro-arrays or deep sequencing, has wide application in gene expression studies. A typical objective of gene expression study is to identify genes that are differentially expressed between groups of samples, such as normal vs. tumor tissue. However, most of the biological samples in scientific researches are heterogeneous: for the samples with identical cellular types, they may have very different proportions. Such variance in proportion will lead to confounding effects (Shen-Orr and Gaujoux, 2013). For example, the reflected gene expression variations are simply caused by the differences in proportions of cell subsets instead of the characteristic condition of a sample (e.g. disease). In order to eliminate the confounding effect, one solution might be to focus on the single cell subset. The isolation procedure, however, is limited by sample materials and financial budgets. Therefore, statistical deconvolution, which does not require any isolation, becomes necessary and practical. In the first project, we develop the Iterated Least Square (ILS) algorithms to estimate the cell specific signature and proportion matrix in complete blind case under homoscedasticity assumption, and theoretically justify the consistency of signature matrix estimate. We also find that the ILS estimate is equivalent to moment under homoscedasticity assumptions, and establish the central limit theorems for the moment estimates. In the heteroscedastic case, the ILS is no longer asymptotically unbiased. Thus, we propose to use the moment estimate, and develop the asymptotics of signature expression estimates. Both numerical examples and real data analysis are employed to illustrate the estimation methods and their asymptotic properties.

Cross correlations are of fundamental importance in multivariate time series analysis. We consider tests for independence of component series based on sample cross correlations. We start with a study of cross correlations between two time series. We derive the central limit theorems for sample cross correlations at large lags, establish convergence rates for maximum sample cross correlations, and demonstrate how they can be used to identify the lead lag relationship for a bivariate time series. We also propose a window sum approach to reduce the computational cost when the series is long. As a second problem, we consider tests for independence of components series under high dimensionality. We propose to use the maximum sample cross correlation over a large range of lags as the test statistic. We also consider an extension to Ljung-Box type statistics. We show that the limiting distributions of the test statistics are extreme value distribution of type I. Our results allow both the number of series, and the range of lags to grow as powers of the sample size, and reveal that how large they can be is determined by the dependence condition and moment condition. We also propose to use the moving blocks bootstrap to improve the finite sample performance of these test procedures.

Acknowledgements

First, I would to greatly and sincerely thank my advisor, Dr. Han Xiao, for his patient guidance, support and encouragement. Dr. Xiao has been a great mentor, role model, and friend. He is always able to convey complex problem in plain language, and I benefit a lot from each meeting we had. His excellent guidance has made this a wonderful and fruitful journey. I am always amazed by his largely enriched experience, deep insights and great devotion in statistics, all of which will serve as the direction for me to follow in my career. The experience of working with him will definitely become the most invaluable treasure of my life.

I also want to thank Dr.Yuefeng Lu for his brilliant ideas and generous advice during my my CO-OP in Sanofi. Dr. Lu has been a great supervisor and I am truly fortunate to have had opportunity to work with him. He spent a lot of time and energy during my internship. Without him, I cannot extend the deconvolution project as one part of my thesis.

I would like to sincerely thank Professor John Kolassa, our graduate advisor, for his devotion and effort to help all graduate students, including but not limited to course selection, financial sources, TA/GA appointments and so on.

Many thanks also go to my rest committee members: Dr. Rong Chen, and Dr. Xiaodong Lin, for their friendly guidance, thought-provoking suggestions, and invaluable input.

I also want to express my appreciation to my peer colleagues for the enlightening discussions and encouragement. Thank all of them to make my PhD life memorable.

Finally, special thanks to my wonderful husband, Bochao Shen, for showing me the true happiness in life. I can face anything because he is by my side.

Dedication

To my family

Table of Contents

Chapter 1

Introduction

As stated by Trevor et al. (2001): "The field of Statistics is constantly challenged by the problems that science and industry brings to its door." In this dissertation, we consider new theoretical results to address two specific problems of recent statistical research:

The first project studies the transcript profiling data such as macroarrays or deep sequencing, and has wide application in gene expression studies. A typical objective of gene expression study is to identify genes that are differentially expressed between groups of samples, such as normal vs. tumor tissue. However, most of the biological samples in scientific researches are heterogeneous: for the samples with identical cellular types, they may have very different proportions. For example, in the context of immunology study, the whole blood samples are used to compare the differential expression between groups of patients. As most of the genes are expressed in each cell type, therefore, the reflected blood-level differential expression between groups might be simply caused by the the differences in proportions of cell types instead of the characteristic condition of a group (e.g. disease). In order to eliminate the confounding effect, one solution might be to focus on the single cell subset. The isolation procedure, however, is limited by sample materials and financial budges. As an alternative approach, one can estimate the cell specific expression of each gene, as well as cell type proportions in disease and normal groups. Since both cell specific expression levels and cell type proportions are not observed, this approach has been described as a deconvolution problem in the literature Shen-Orr et al. (2010). Although it is known that the relationship between the expression levels of pure and mixed samples are not strictly linear, the linearity assumption is reasonable as shown in previous research work Shen-Orr et al. (2010). That is the global expression value of gene j in sample i is the sum of its expressions in the r cell types:

$$g_{ij} = \sum_{k=1}^{r} w_{ik}h_{kj} + \epsilon_{ij}, \text{ and } 1 \leq i \leq n, 1 \leq j \leq p \tag{1.1}$$

where w_{ik} is the proportion of cell type k in sample i, which sum to 1, i.e. $\sum_{k=1}^{r} w_{ik} = 1$. h_{kj} is the specific gene j's expression in cell type k, whose value are non-negative, i.e. $h_{kj} \geq 0$. ϵ_{ij} is the random error term, which is independent across all sample i and gene j. We also require $p \geq r$.

Using matrices, we can rewrite Eq (1.1) into the following approximate matrix decomposition problem:

$$G_{np} = W_{nk}H_{kp} + E_{np}$$

where G_{np} is the $n \times p$ global gene expression matrix, W_{nk} is the $n \times k$ fraction matrix, H_{kp} is the $k \times p$ cell-specific matrix, and E_{np} is the $n \times p$ error matrix. And it is required that $p \geq r$.

Existing work focused on the estimation methods and applications (Shen-Orr et al., 2010; Shen-Orr and Gaujoux, 2013; Yang et al., 2012). However, the statistical properties have not been discussed extensively. In the first project, we develop the Iterated Least Square (ILS) algorithms to estimate both H (with constraints) and W under homeostatic assumption, theoretically justify the associated statistical properties. As a side product, we find that ILS is equivalent to moment estimate under homeostatic assumption, and derive the joint limiting distribution of H in specific case when only two cell types and two gene markers are considered. In the heteroscedastic case, the ILS is no longer asymptotically unbiased. Thus, we propose to use the moment estimate, and study its associated statistical properties. The underlying theory supporting our limiting distribution are Multivariate Lindeberg-Levy CLT Greene (2002) and Multivariate delta Method van der Vaart (1998). We will illustrate the above estimation method and statistical properties by using numerical example and real data analysis.

The second project considered the lead lag relationship among component time series. Lead lag relationship is an important type of cross correlations and temporal dependence. It is the cross correlation between time series shifted in time relative to one another. Suppose there is a p-dimensional stationary time series x_{jt}, $1 \leq j \leq p$ and $1 \leq t \leq T$. The cross covariance between j-th and k-th time series at lag s is defined as $\gamma_{jk}(s) = \mathrm{Cov}(x_{jt}, x_{k,t+s})$. In particular, $\gamma_{jj}(\cdot)$ gives the autocovariance function of the j-th component series. The cross correlation is then given by $\rho_{jk}(s) = \gamma_{jk}(s)/\sqrt{\gamma_{jj}(0)\gamma_{kk}(0)}$. Lead-lag correlation has been widely used in many engineering and basic science fields, including electrical, acoustic, geophysical applications and economics (Nelson-Wong et al., 2009), (Duffy and Hughes-Clarke, 2005), (Basappa and Lakdawala, 2000), (Cohen, 1981). For example, Berndt and Ostrovnaya (2007) indicated that investors absorb information revealed in the CDS market into option prices within a few days, i.e. CDS market lead the option market. Ideally if $\gamma_{jk}(s)$ is zero for all positive s, and nonzero for some negative s, then there is a unidirectional relationship from the j-th series to the k-th series.

More comprehensive relationship among the p series can be modeled by autoregressive (VAR) models. However, when p is large, fitting a VAR over all series is not computationally or statistically feasible. The problem becomes easier if the p series can be partitioned into smaller groups, where the between groups dependence is weak or negligible, and VAR models can be built with each group. Cross correlations may serve as a proxy of the distance or closeness, since they measure the linear relationship between any two series. The problem can be viewed as a clustering problem, where it may be assumed that different groups are not correlated. A closely related problem is to test whether these p series are correlated at all. This can also be translated to the following testing problem

$$\mathrm{H}_0: \quad \gamma_{jk}(s) = 0, \quad \forall\, j \neq k, s \in \mathbb{Z}. \quad \text{vs} \quad \mathrm{H}_1: \quad \gamma_{jk}(s) \quad \text{for some } j \neq k,$$

under the "large T, large p" paradigm, where the dimension p may be comparable to, or even larger than the sample size T. The cross covariances can be estimated by the sample version:

$$\hat{\gamma}_{jk}(s) = \frac{1}{T} \sum_{1 \le t, t+s \le T} (x_{jt} - \bar{x}_j)(x_{k,t+s} - \bar{x}_k)$$

where $\bar{x}_j$ is the sample mean of the j-th series. We consider the maximum type test statistic:

$$\tilde{M}_1 = \max_{|s| \le s_T,\, 1 \le j < k \le p} \hat{\gamma}_{jk}(s).$$

Since the correlation between two series may exist at some unknown but very large lag s, here we allow the range s to expand with the sample size, i.e. s_T is allowed to approach infinity as T increases. Sometimes the cross correlation between two series may exist at many adjacent lags, but is weak at each of them. In this case, the following test statistic can have larger power.

$$\tilde{M}_m = \max_{|s| \le s_T,\, 1 \le j < k \le p} \tilde{Q}_{jk}(s), \tag{1.2}$$

where

$$\tilde{Q}_{jk}(s) = \sum_{l=s+1}^{s+m} \hat{\gamma}_{jk}^2(s).$$

We will show that the test statistics converge to extreme value distribution of type I (also called Gumbel distribution) after proper normalizations. Due to the existence of temporal dependence, we carry out theoretical analysis under the framework of causal representation and physical dependence measures (Wu, 2005). Our proof makes use of the Gaussian approximation result Zaitsev (1987).

On the other hand, it is well know that the Gumbel type convergence is usually slow. In order to evaluate the finite sample test performance, we propose to use bootstrap method to improve the finite sample performance. More specifically, we use the moving blocks bootstrap of Liu and Singh (1992). Recently, Hill and Motegi (2016) and Zhang and Cheng (2014) also considered bootstrap methods for the maximum type statistics under the time series context.

We shall show two main results. The first theoretical results considers the simplest scenario when there are only two time series. In section 3.3.1, we will show that under mild dependence condition and moment constraint (4th order moment) for bi-variate stationary process, the maximum cross correlation can asymptotically identify the true lead (or lag) even if it increases at a rate that is slower than the sample size. This results provides mathematical explanation how maximum cross correlation works and why it could be used to explore the lead-lag relationship. Furthermore, in order to reduce computational cost, we also propose window sum approach, and theoretically confirm its feasibility and accuracy. The second results considers high dimensional time series. Under mild condition, we establish the the Gumbel convergence of maximum cross correlation over wide rage of lags.

The rest of the thesis is organized as following. In Chapter 2, we consider the statistical models for the deconvolution of transcript profiling data, and show the theoretical results. In Chapter 3, we show the asymptotic results for maximum cross correlation among bivariate and high dimensional series.

Chapter 2

Deconvolution of Transcript Profiling Data

2.1 Introduction

A typical objective of gene expression study is to identify genes that are differentially expressed between groups of samples, such as normal vs. tumor tissue. However, most of the biological samples in scientific researches are heterogeneous. The main reason for the heterogeneity is the variation in cell proportions: for the samples with identical cellular types, they may have very different proportions. Such variance in proportion will lead to confounding effects. For example, it might provide misleading information Shen-Orr and Gaujoux (2013): The reflected gene expression variations are simply caused by the differences in proportions of cell subsets instead of the characteristic condition of a sample (e.g. disease). A second example is that it might cause signal strength dilution: A gene that is differentially expressed in a cell subset presented in low proportion in a sample might be masked by the signal from the same gene expressed in a prevalent cell subset (Cobb et al., 2005). Therefore, in order to obtain a detailed understanding of the role of each cell subset, the cell subset level measurement and interpretation of phenotypic changes between specific conditions is critically important.

The heterogeneity problem mentioned above has been acknowledged for a long time by Davey and Kell (1996). Experimental technique to isolate the mixed cell population is one of the solutions. However, the isolation procedure might introduce bias and entails a loss of a systems perspective (i.e. biologically meaningful changes happen in multiple cell subsets and between them) (Whitney et al., 2003). And the variance between expression values from different isolation methods is also recognized (Cobb et al., 2005). Furthermore, it is limited by sample materials and financial budget.

Computational deconvolution becomes an attractive way to deal with these concerns: it is capable of extracting cell-specific information from heterogeneous samples. Since the pioneering work of Venet et al. (2001), a variety of methods such as Repsilber et al. (2010); Lähdesmäki et al. (2005); Kuhn et al. (2011); Bolen et al. (2011) have been proposed by many researchers, casting insights into how to estimate the cell-specific signatures and/or relative cell type proportions from global gene expression measurements, such as micro-array and RNA-seq.

Nonetheless, the statistical inference on these estimates is not extensively discussed, especially in the complete blind case where both cell-specific and relative cell type proportions are unknown. In this chapter, we develop a new iterative least square (ILS) algorithm in the homoeostatic case, and derive the asymptotic statistical properties of the estimates under mild assumptions. In the heteroscedastic case, we find that the ILS is no longer asymptotic unbiased. Therefore we propose to use the moment estimate, and consider the associated statistical inference.

Although the relationship between the expression levels of pure and mixed samples is not strictly linear, the linearity assumption is reasonable as shown in previous research Shen-Orr et al. (2010). That is the global expression value of gene j in sample i can be characterized as the sum of its expressions in the r cell types,

$$g_{ij} = \sum_{k=1}^{r} w_{ik} h_{kj} + \epsilon_{ij}, \text{ and } 1 \leq i \leq n, 1 \leq j \leq p \tag{2.1}$$

where w_{ik} is the proportion of cell type k in sample i, which sum to 1, i.e. $\sum_{k=1}^{r} w_{ik} = 1$. h_{kj} is the specific gene j's expression in cell type k, whose value are non-negative, i.e. $h_{kj} \geq 0$. ϵ_{ij} is the random error term, which is independent across all sample i and gene j. We also require $p \geq r$.

Using matrices, we can rewrite Eq (2.1) into the following approximate matrix decomposition problem:

$$G_{np} = W_{nk} H_{kp} + E_{np} \tag{2.2}$$

where G_{np} is the $n \times p$ global gene expression matrix, W_{nk} is the $n \times k$ fraction matrix, H_{kp} is the $k \times p$ cell-specific matrix, and E_{np} is the $n \times p$ error matrix. And it is required that $p \geq r$. We will suppress the subscript in the latter.

There are various methods developed under the linear model framework of $G = WH + E$. With respect to the input data requirement, they can be classified into two major types Gaujoux and Seoighe (2013): (1) Partial Deconvolution: either the fraction matrix W or the cell-specific matrix H is known; (2) Complete Deconvolution: both fraction matrix W and cell-specific matrix H are needed to be estimated from the global gene expression data of the heterogeneous samples. It's clearly that the complexity of the latter increases due to more unknown parameters. More specifically, we will described each type in details in the following:

1. *Partial Deconvolution*: Either the fraction matrix W or the cell-specific matrix H is known.

    1.1. *The fraction matrix W is known*: This is an over determined problem provided that there are more samples than the number of cell types. Shen-Orr et al. (2010) proposed csSAM method. They regressed the global gene expression on the proportions using standard linear regression . Differential expression analysis was proposed with error estimates cooperated into test statistic. The approach was applied to whole-blood gene expression datasets from post-transplant kidney transplant recipients (9 were stable and 15 were experiencing acute rejection), identifying hundreds of genes that were differentially expressed. Other examples could be also found in Erkkilä et al. (2010); Stuart et al. (2004).

    1.2. *The cell-specific matrix H is known*: This is also an over determined estimation problem, since for each sample, generally we have more observations (genes) than unknown parameters(proportion for each cell type), which yields accurate estimate. Abbas et al. (2009) defined an optimised set of signatures for 17 different immune cell types, and proposed a heuristic algorithm based on standard linear regression with nonnegative constrain. The estimated proportion were scaled to sum-up to one after fitting. The method was applied to white blood cell samples from a cohort of Systemic Lupus Erythematosus (SLE) and healthy patients, and found increase in proportions (NK, T and monocytes) that were consistent with SLEDAI (Systemic Lupus Erythematosus Disease Activity Index). Gong et al. (2011) proposed an alternative algorithm that incorporates the sum-up to one constraint on the proportions within the fitting process . Wang et al. (2006) used similar approaches to analyze yeast cell cycle expression patterns from global mammary gene expression data .

2. *Complete Deconvolution*: Both fraction matrix W and cell-specific matrix H are needed to be estimated from the global gene expression data of the heterogeneous samples. Venet et al. (2001) pioneered the study of complete gene expression deconvolution . They proposed an alternative nonnegative least-squares approach by using a heuristic to limit the correlations between the estimated cell type signatures. Repsilber et al. (2010) proposed an Nonnegative Matrix Factorization (NMF) algorithm (deconf) . It corresponds exactly to Venet Venet et al. (2001) algorithm, but drops the correlation constraints. Although it is called 'complete' deconvolution, some prior knowledge is still needed. Such information is often in the form of marker gene sets, which are essentially expressed by the specific cell type.

2.2 Complete Deconvolution Model in Homoscedastic Case

Following the same notation as in Section 2.1, we will describe the constraints that we adopt on the model in the complete blind case, together with the method to compute the coefficients. Consider the model in formula (2.2), i.e.

$$G = WH + E, \text{ and } p \geq r$$

where G, W and H are defined as before, and $E = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p)$, where $\boldsymbol{e}_j = (e_{1j}, e_{2j}, \ldots, e_{nj})$. Furthermore, we assume $\boldsymbol{e}_j$ are independent across all genes j, which implies that $\text{cov}(\boldsymbol{e}_{j_1}, \boldsymbol{e}_{j_2}) = 0,$ for any $j_1 \neq j_2$. For each $j, (j = 1, \ldots, p)$, $\boldsymbol{e}_j$ have mean $\mathbf{0}_n$, and covariance matrix as $\sigma^2 I_n$.

2.2.1 Model Constraints

It is obvious that there might be a lot of such factorizations without any constraint. For example, if W^* and H^* is one solution, for any orthogonal matrix Q, W^*Q , $Q^{'}H^*$ is also one of the factorizations that satisfying Eq. (2.2). Therefore, in order to make the model both identifiable and meaningful, constraints are required. We adopt the same constraint on signature matrix as proposed by Gaujoux and Seoighe (2012). They used set of marker genes: which was known to be almost exclusively expressed by just one specific cell type. Therefore, the column of the signature matrix H contains all 0 except the one that corresponding to the cell-type specific marker gene. Table 2.1 shows a specific example of matrix H where CD3G and CD3E are markers for T-cell. Then the value of the corresponding columns will be 0 except for the row of T cells. The other columns should be read in the same way.

	CD3G	CD3E	.....	CD247	CD28
T-cell	$*$	$*$	...	0	0
B-cell	0	0	...	0	0
Granulocytes	0	0	...	0	0
Macrophage	0	0	...	$*$	$*$

Table 2.1: Signature Matrix with Gene Marker pre-specified

For the fraction matrix, the i-th row records the cell proportions for i-th sample (e.g. patient). Naturally the value of each element should be between $(0, 1)$, and the sum of each row is 1.

For notational convenience, we denote Ω as the set of all the possible solution satisfying the constraints, i.e

$$\Omega = \left\{ (W,H) \;\middle|\; \begin{array}{l} \sum_{k=1}^{k=r} w_{ik} = 1, w_{ik} \geq 0 (i = 1, 2, \ldots, n.k = 1, 2, \ldots, r) \\ \text{H satisfies signature matrix constraint} \end{array} \right\}$$

2.2.2 Algorithm to compute the coefficient

W, H will be solved by minimizing the squared difference between the observed and the modeled global expression matrix with constraints specified in Section 2.2.1, i.e.

$$(\hat{W}, \hat{H}) = \arg_{(W,H)\in\Omega} \|G_{np} - W_{nr}H_{rp}\|_F^2 \tag{2.3}$$

where $\|.\|_F$ denotes the Frobenius norm

Iterated least square method is applied to solve Eq. (2.3):

Step 1: Given an initial value of W^0, get the least square estimate of H with marker gene pattern and nonnegative constraints, denote the solution by $H^{(1)}$.

Step 2: Given $H = H^{(1)}$, get the constrained estimate of W, denote the solution by $W^{(1)}$.

Step 3: Repeat **Step** 1 and **Step** 2 (with the estimate from last step serving as initial value), until the estimate numerically converges.

We notice that all the estimates are in matrix form. In order to make the standard quadratic procedure applicable, we make some appropriate transformation.

Details about implementation for step 1 and step 2 are described below:

Step 1: For given W, denote the column of G and W by

$$\boldsymbol{g}_j = (g_{1j}, g_{2j} ...., g_{nj})'$$

$$\boldsymbol{w}_k = (w_{1k}, w_{2k}, ...w_{nk})'$$

$$\boldsymbol{e}_j = (e_{1j}, e_{2j}, ...e_{nj})'$$

Suppose the cell type k has k_d gene markers. Let $M_k = \{k_1, \ldots, k_d\}$ be the collection of the gene markers for the cell type k , where $k = 1, \ldots, r$. Besides, suppose there are q genes that are expressed on all cell types, and denote the collection of these genes by

$A_j = \{j_1, j_2, \ldots, j_q\}$.

With all these notation, now the objective function could be written as:

$$\underset{h_{kj}>0}{\arg}\ \min \left\{ \sum_{k=1}^{r} \sum_{j \in M_k} \|\boldsymbol{g}_j - h_{kj}\boldsymbol{w}_k\|_F^2 + \sum_{j \in A_j} \|\boldsymbol{g}_j - \sum_{k=1}^{r} h_{kj}\boldsymbol{w}_j\|_F^2 \right\} \tag{2.4}$$

Objective function (2.4) is familiar and standard. Quandratic programming with non-negative constraints is applicable to solve h_{kj}, and then transform back to H (h_{kj} is in the kth row and jth column). Note that all the h_{jk} that does not appear in objective function (2.4) are set as 0 as assumed.

Step 2: For fixed H,

Denote the row of G, W and E by the following:

$$\begin{aligned}
\boldsymbol{g}_{i.} &= (g_{i,1}, g_{i,2} \ldots, g_{i,p})' \\
\boldsymbol{w}_{i.} &= (w_{i,1}, w_{i,2}, \ldots w_{i,r})' \\
\boldsymbol{e}_{i.} &= (e_{i,1}, e_{i,2}, \ldots e_{i,r})'
\end{aligned}$$

The linear model can be rewritten as following:

$$\begin{pmatrix} \boldsymbol{g}'_{1.} \\ . \\ . \\ \boldsymbol{g}'_{(n-1).} \\ \boldsymbol{g}'_{n.} \end{pmatrix} = \begin{pmatrix} \boldsymbol{w}'_{1.}H \\ . \\ . \\ \boldsymbol{w}'_{(n-1).}H \\ \boldsymbol{w}'_{n.}H \end{pmatrix} + \begin{pmatrix} \boldsymbol{e}'_{1.} \\ . \\ . \\ \boldsymbol{e}'_{(n-1).} \\ \boldsymbol{e}'_{n.} \end{pmatrix}$$

Now the objective function becomes:

$$\begin{aligned}
&\arg_{\boldsymbol{w}_{i.}, i=1,2,\ldots,n} \min \sum_{i=1}^{i=n} \|\boldsymbol{g}_{i.} - H\boldsymbol{w}_{i.}\|^2 \\
&\text{subject to} \sum_{j=1}^{r} w_{i,j} = 1, w_{i,j} \geq 0, (i = 1, 2, \ldots, n, j = 1, 2, \ldots, r)
\end{aligned} \tag{2.5}$$

Remark 1. *For Eq. (2.5), it's equivalent to do n independent optimizations for each sample, i.e for each i, $\hat{\boldsymbol{w}}_i$ could be obtained through*

$$\begin{aligned}\arg_{\boldsymbol{w}_i} \min ||\boldsymbol{g}_{i.} - H\boldsymbol{w}_{i.}||^2 &= \arg_{\boldsymbol{w}_{i.}} \min\{\boldsymbol{g}_{i.}^{'}\boldsymbol{g}_{i.} - 2\boldsymbol{g}_{i.}^{'}H^{'}\boldsymbol{w}_{i.} + \boldsymbol{w}_{i.}HH^{'}\boldsymbol{w}_{i.}\} \\ &= \arg_{\boldsymbol{w}_{i.}} \min\{-2\boldsymbol{g}_{i.}^{'}H^{'}\boldsymbol{w}_{i.} + \boldsymbol{w}_{i.}HH^{'}\boldsymbol{w}_{i.}\},\end{aligned}$$

with restrictions (only first equality holds) specified as below:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} w_{i1} \\ w_{i2} \\ . \\ . \\ w_{ik} \end{pmatrix} \geq \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

2.2.3 Asymptotic Properties of the Estimate when cell type number $r = 2$

There is a lot of literature on the estimation methods and applications of the deconvolution problem. However, the inference of these estimates are not extensively discussed, especially in the complete blind case (neither W or H is known). In this section, we will explore the statistical properties of the Iterated Least Square (ILS) estimate. To be more close to practice, we also assume that the frequency matrix (W) is random and from some unknown population. We start from the simplest case, where only two cell types and two marker genes are under consideration.

Consider the model:

$$\boldsymbol{g}_j = h_{j0}\boldsymbol{w}_{j0} + \boldsymbol{\epsilon}_j \tag{2.6}$$

where $\boldsymbol{\epsilon}_1$ is independent of $\boldsymbol{\epsilon}_2$, and $\mathrm{E}(\boldsymbol{\epsilon}_j) = \boldsymbol{0}_n$, $\mathrm{Var}(\boldsymbol{\epsilon}_j) = \sigma_j^2 I$ $(j = 1, 2)$

Here $\boldsymbol{w}_{10}$ and $\boldsymbol{w}_{20}$ are random vectors with entries between 0 and 1, and satisfy $\boldsymbol{w}_{10} + \boldsymbol{w}_{20} = \boldsymbol{1}$, which is the vector with all entries equal to one. The two covariate vector $\boldsymbol{w}_{i0}$ are not observed, and the question is to estimate h_{i0} and $\boldsymbol{w}_{i0}$.

Define $x_{10} = \mathbf{1}'\boldsymbol{w}_{10}/n, x_{20} = \sqrt{\boldsymbol{w}_{10}'\boldsymbol{w}_{10}/n - x_{10}^2}$, and assume there exists a constant c, such that when n is large enough,

$$\lim_{n\to\infty} x_{10} > c, x_{20} > c, \text{ a.s.} \tag{2.7}$$

We first solve the following optimization problem:

$$(\hat{h}_1, \hat{h}_2, \hat{\boldsymbol{w}}_1, \hat{\boldsymbol{w}}_2) = \arg_{h_1,h_2,\boldsymbol{w}_1,\boldsymbol{w}_2} \min \|\boldsymbol{g}_1 - h_1\boldsymbol{w}_1\|^2 + \|\boldsymbol{g}_2 - h_2\boldsymbol{w}_2\|^2 \tag{2.8}$$

The common variance σ^2 can be estimated by:

$$\hat{\sigma}^2 = \frac{1}{n}\left(\|\boldsymbol{g}_1 - \hat{h}_1\hat{\boldsymbol{w}}_1\|^2 + \|\boldsymbol{g}_2 - \hat{h}_2\hat{\boldsymbol{w}}_2\|^2\right) \tag{2.9}$$

For notational convenience, we define:

$$\mu(w) = \mathrm{E}(w_{i1}), \text{ and } \mu_k(w_{i1}) = \mathrm{E}(w_{i1} - \mu(w))^k, k = 1, 2, 3 \dots$$

Theorem 1. *Let $\epsilon_{1,1}, \epsilon_{2,1}, \dots, \epsilon_{n,1}$ are i.i.d with mean 0, and variance σ_1^2, and $\epsilon_{1,2}, \epsilon_{2,2}, \dots, \epsilon_{n,2}$ are i.i.d with mean 0, and variance σ_2^2. Assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and Eq. (2.7) holds, then*

1) $(\hat{h}_1, \hat{h}_2) \to (h_{10}, h_{20})$ a.s

2) $\hat{\sigma}^2 \to \sigma^2$ a.s

3) Assume $\boldsymbol{\epsilon}_j \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$, and $0 < \mu(w) < 1$, then

$$\sqrt{n}\left(\begin{pmatrix} \hat{h}_1 \\ \hat{h}_2 \end{pmatrix} - \begin{pmatrix} h_{10} \\ h_{20} \end{pmatrix}\right) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Sigma_1)$$

Where $\Sigma_1 = \triangledown f'(\boldsymbol{u}_1)\Sigma\triangledown f(\boldsymbol{u}_1)$, and $\triangledown f(\boldsymbol{u}_1)$, $\Sigma = (\sigma_{ij})$ are given by

$$\nabla f(\boldsymbol{u}_1) = \begin{pmatrix} 1 & \frac{h_{20}}{h_{10}} \\ \frac{h_{10}}{h_{20}} & 1 \\ \dfrac{(1-\mu(w))h_{10}}{(h_{10}^2+h_{20}^2)\mu_2(w)} & -\dfrac{\mu(w)h_{20}}{(h_{10}^2+h_{20}^2)\mu_2(w)} \\ -\dfrac{(1-\mu(w))h_{10}}{(h_{10}^2+h_{20}^2)\mu_2(w)} & \dfrac{\mu(w)h_{20}}{(h_{10}^2+h_{20}^2)\mu_2(w)} \\ \dfrac{(h_{10}^2-h_{20}^2)(1-\mu(w))}{h_{20}(h_{10}^2+h_{20}^2)\mu_2(w)} & \dfrac{(h_{20}^2-h_{10}^2)\mu(w)}{h_{10}(h_{10}^2+h_{20}^2)\mu_2(w)} \end{pmatrix}$$

$$\begin{aligned}
\sigma_{11} &= \sigma^2 + h_{10}^2\mu_2(w) \\
\sigma_{22} &= \sigma^2 + h_{20}^2\mu_2(w) \\
\sigma_{33} &= h_{10}^4(u_4(w) - \mu_2^2(w)) + 2\sigma^4 + 4h_{10}^2\mu_2(w)\sigma^2 \\
\sigma_{44} &= h_{20}^4(u_4(w) - \mu_2^2(w)) + 2\sigma^4 + 4h_{20}^2\mu_2(w)\sigma^2 \\
\sigma_{55} &= h_{10}^2h_{20}^2(u_4(w) - \mu_2^2(w)) + h_{10}^2\mu_2(w)\sigma^2 + h_{20}^2\mu_2(w)\sigma^2 + \sigma^4 \\
\sigma_{12} &= -h_{10}h_{20}\mu_2(w) \\
\sigma_{13} &= h_{10}^3\mu_3(w) \\
\sigma_{14} &= h_{10}h_{20}^2\mu_3(w) \\
\sigma_{15} &= -h_{10}^2h_{20}\mu_3(w) \\
\sigma_{23} &= -h_{10}^2h_{20}\mu_3(w) \\
\sigma_{24} &= -h_{20}^3\mu_3(w) \\
\sigma_{25} &= h_{10}h_{20}^2\mu_3(w) \\
\sigma_{34} &= h_{10}^2h_{20}^2(\mu_4(w) - \mu_2^2(w)) + 2\sigma^4 \\
\sigma_{35} &= -h_{10}^3h_{20}(\mu_4(w) - \mu_2^2(w)) - 2h_{10}h_{20}\mu_2(w)\sigma^2 \\
\sigma_{45} &= -h_{10}h_{20}^3(\mu_4(w) - \mu_2^2(w)) - 2h_{10}h_{20}\mu_2(w)\sigma^2
\end{aligned} \tag{2.10}$$

Remark 2. For the limiting distribution part, the calculation of the covariance matrix is tedious. Here we derived some useful results of the covariance between sample mean and sample central moments, as well as the covariance of sample central moments.

Define

$$\bar{w}_1 = \sum_{i=1}^{i=n} \frac{w_{i1}}{n} \text{ and } m_k(w) = \sum_{i=1}^{i=n} \frac{(w_{i1} - \bar{w}_1)^k}{n}, k = 2, 3, \ldots \tag{2.11}$$

Then:

$$
\begin{aligned}
\mathrm{Cov}(\bar{w}_1, m_2) &= \frac{\mu_3(w)}{n} \\
\mathrm{Cov}(\bar{w}_1, m_3) &= -\frac{3(-2+n)(-1+n)\mu_2^2(w)}{n^3} + \frac{(-2+n)(-1+n)\mu_4(w)}{n^3} \\
\mathrm{Var}(m_2) &= \frac{\mu_4(w)}{n} - \frac{\mu_2(w)^2(n-3)}{n(n-1)} \\
\mathrm{Cov}(m_2, m_3) &= \frac{(-2+n)(-1+n)^2\mu_5(w)}{n^4} - \frac{2(-2+n)(-1+n)(-5+2n)\mu_2(w)\mu_3(w)}{n^4} \\
\mathrm{Var}(m_3) &= \frac{3(-2+n)(-1+n)(20-12n+3n^2)\mu_2(w)^3}{n^5} - \\
&\quad \frac{(-10+n)(-2+n)^2(-1+n)\mu_3(w)^2}{n^5} - \frac{3(-2+n)^2(-1+n)(-5+2n)\mu_2(w)\mu_4(w)}{n^5} \\
&\quad + \frac{(-2+n)^2(-1+n)^2\mu_6(w)}{n^5}
\end{aligned}
$$

Remark 3. Sometimes, people treat $\boldsymbol{w}_j$ as deterministic. In this case, we have the following central limit theorem:

$$
\sqrt{n}(\begin{pmatrix} \hat{h}_1 \\ \hat{h}_2 \end{pmatrix} - \begin{pmatrix} h_{10} \\ h_{20} \end{pmatrix}) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Sigma_2)
$$

where $\Sigma_2 = \nabla f(\boldsymbol{u}_1)' \begin{pmatrix} \sigma_{11} & 0 & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 & \sigma_{35} \\ 0 & 0 & 0 & \sigma_{44} & \sigma_{4,5} \\ 0 & 0 & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{pmatrix} \nabla f(\boldsymbol{u}_1)$

With $f(\boldsymbol{u}_1)$ and σ_{ij} defined as following:

$$
\nabla f(\boldsymbol{u}_1) = \begin{pmatrix} 1 & \frac{h_{20}}{h_{10}} \\ \frac{h_{10}}{h_{20}} & 1 \\ \frac{(1-\bar{w}_1)h_{10}}{(h_{10}^2+h_{20}^2)m_2(w)} & -\frac{\bar{w}_1 h_{20}}{(h_{10}^2+h_{20}^2)m_2(w)} \\ -\frac{(1-\bar{w}_1)h_{10}}{(h_{10}^2+h_{20}^2)m_2(w)} & \frac{\bar{w}_1 h_{20}}{(h_{10}^2+h_{20}^2)m_2(w)} \\ \frac{(h_{10}^2-h_{20}^2)(1-\bar{w}_1)}{h_{20}(h_{10}^2+h_{20}^2)m_2(w)} & \frac{(h_{20}^2-h_{10}^2)\bar{w}_1}{h_{10}(h_{10}^2+h_{20}^2)m_2(w)} \end{pmatrix}
$$

$$\begin{cases} \sigma_{11} & = \sigma_{2,2} = \sigma^2 \\ \sigma_{33} & = 2\sigma^4 + 4h_{10}^2 m_2(w)\sigma^2 \\ \sigma_{44} & = 2\sigma^4 + 4h_{20}^4 m_2(w)\sigma^2 \\ \sigma_{55} & = (h_{10}^2 + h_{20}^2) m_2(w)\sigma^2 + \sigma^4 \\ \sigma_{35} & = \sigma_{4,5} = \dfrac{2h_{10}h_{20}\sigma^2 \boldsymbol{w}_1' \boldsymbol{w}_2}{n} - 2h_{10}h_{20}\sigma^2 \bar{w}_1(1-\bar{w}_1) \end{cases} \tag{2.12}$$

Note that $\bar{w}_1$ and $m_k(w)$ are defined in Eq. (2.11).

Remark 4. As we do not know the true parameters of $\boldsymbol{w}_1$ in neither Eq.(2.10) or Eq.(2.12), we plug in the LS estimate conditional on $\hat{h}_1, \hat{h}_2$. Solve:

$$\begin{aligned} \hat{\boldsymbol{w}}_1^{LS} &= \arg_{\boldsymbol{w}_1} \min \|\boldsymbol{g}_1 - \hat{h}_1 \boldsymbol{w}_1\|^2 + \|\boldsymbol{g}_2 - \hat{h}_2(\mathbf{1} - \boldsymbol{w}_1)\|^2 \\ &= \frac{(\boldsymbol{g}_1 \hat{h}_1 - \boldsymbol{g}_2 \hat{h}_2 + \hat{h}_2^2)}{(\hat{h}_1^2 + \hat{h}_2^2)} \end{aligned}$$

Plugging $\boldsymbol{g}_j = h_{j0}\boldsymbol{w}_1 + \boldsymbol{\epsilon}_j$ in the solution $\hat{\boldsymbol{w}}_1^{LS}$, as $(\hat{h}_1, \hat{h}_2)$ is consistent estimate of (h_{10}, h_{20}), it's not difficult to see that $\hat{u}_{2K}^{LS}(w) = \dfrac{\sum_{i=1}^{i=n}(\hat{w}_{i,1}^{LS} - \bar{\hat{w}}^{LS})^{2K}}{n}$ is biased estimate of $\mu_{2K}(w)$ (or $m_2(w)$ if $\boldsymbol{w}_1$ is deterministic). Yet, it can be corrected by removing the bias.

For example, for the 2nd and 4th moment, the unbiased estimates are as following.

$$\begin{aligned} \hat{u}_2(w) &= \hat{u}_2^{LS}(w) - \frac{\hat{\sigma}^2}{(\hat{h}_1^2 + \hat{h}_1^2)} \\ \hat{u}_4(w) &= \hat{u}_4^{LS}(w) - \frac{6\hat{u}_2(w)\hat{\sigma}^2}{(\hat{h}_1^2 + \hat{h}_1^2)} - \frac{3\hat{\sigma}^4}{(\hat{h}_1^2 + \hat{h}_1^2)^2} \end{aligned}$$

Remark 5. The individual $w_{i,1}$ might not be a reliable estimate, but we could still get a consistent estimate of the mean value $\bar{w}_1$. Plugging in the consistent estimate of h_{10},

h_{20}, the estimated proportion could be expressed

$$\hat{\boldsymbol{w}}_1 = \boldsymbol{w}_{10} + \frac{\hat{h}_1\boldsymbol{\epsilon}_1 - \hat{h}_2\boldsymbol{\epsilon}_2}{\hat{h}_1^2 + \hat{h}_2^2},$$

So $\hat{\mu}(w) = \dfrac{\mathbf{1}'\hat{\boldsymbol{w}}_1}{n}$ is consistent estimate of $\mu(w)$

2.2.4 Asymptotic Properties of the Estimate when cell type number $r > 2$

Theorem 1 summarizes the asymptotic unbiasedness of ILS in the case when there are only two cell types and two marker genes. In this section, we will generalize the results in a more common case when the cell type number $r > 2, p \geq r > 2$.

Consider the model:

$$\boldsymbol{g}_j = h_{j0}\boldsymbol{w}_{j0} + \boldsymbol{\epsilon}_j, (j = 1, 2, \ldots, r) \tag{2.13}$$

where $\boldsymbol{\epsilon}_{j_1}$ is independent of $\boldsymbol{\epsilon}_{j_2}$ if $j_1 \neq j_2$, and $\mathrm{E}(\boldsymbol{\epsilon}_j) = \mathbf{0}_n$, $\mathrm{Var}(\boldsymbol{\epsilon}_j) = \sigma^2 I$ ($\forall j = 1, \ldots r$

Similarly, we first solve the optimization problem:

$$(\hat{h}_1, \ldots, \hat{h}_k, \hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_k) = \arg_{h_1,\ldots,h_k,\boldsymbol{w}_1,\ldots,\boldsymbol{w}_k} \min \sum_{j=1}^{r} \|\boldsymbol{g}_j - h_j\boldsymbol{w}_j\|^2$$

And the common variance is estimated by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{r} \|\boldsymbol{g}_j - \hat{h}_j\hat{\boldsymbol{w}}_j\|^2$$

Theorem 2. *Let $\epsilon_{1,j}, \ldots, \epsilon_{n,j}$ are i.i.d with mean 0 and variance σ_j^2 $(j = 1, 2, \ldots, p)$. Consider model (2.13), assume $\sigma_j^2 = \sigma^2$, and $\mathrm{E}(w_{i,k}) > 0, \mathrm{Var}(w_{i,k}) > 0$ for $i =$*

$1, 2, \ldots, n, k = 1, 2, \ldots, r$. *Then*

$$\begin{cases} (\hat{h}_1, \ldots, \hat{h}_K) \to (h_{10}, \ldots, h_{K0}) \ a.s \\ \hat{\sigma}^2 \to \sigma^2 \ a.s \end{cases}$$

2.2.5 Simulation Study

Simulation study for consistency

For simplicity, we consider the case when there are only two cell types. Let $p = 2$, $h_{10} = 5, h_{20} = 10$. $\boldsymbol{w}_{10}$ is generated from i.i.d. U$(0, 1)$ distribution, and $\boldsymbol{w}_{20} = \mathbf{1} - \boldsymbol{w}_{10}$. The error terms $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ are independent on each other, and from some distribution with mean $\mathbf{0}$ and common covariance matrix $\sigma^2 I$. More specifically, for Table 2.3, and 2.2 both $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ are from standard Normal distribution, while for Table 2.5 and 2.4, $\boldsymbol{\epsilon}_1$ is generated from standard Normal distribution, and $\boldsymbol{\epsilon}_2$ from scaled t distribution ($df = 10$) with variance 1. $\boldsymbol{g}_j$ $j = 1, 2$ is generated from Model (2.6). We try two scenarios:

1. $\boldsymbol{w}_{j0}$ $(j = 1, 2)$ is refreshed in every repetition, i.e $\boldsymbol{w}_{10}$ is treated as random factor (See table 2.5 and 2.4).

2. $\boldsymbol{w}_{j0}$ $(j = 1, 2)$ is generated only once and then set constant over all repetitions (See table 2.3 and 2.2)

We repeat the simulation 1000 times, and summarize the results in Table 2.3, 2.2, 2.5 and 2.4. The simulation results are consistent with Theorem 1, regardless of how $\boldsymbol{w}_{j0}$ and $\boldsymbol{\epsilon}_1$ is generated, as long as the error terms share common variance, the average value of $\hat{h}_1, \hat{h}_2,$ and $\hat{\sigma}^2$ are close to the true values, and the average of squared deviation from true value is close to 0.

True h_1	5	True h_2	10
$\bar{h}_1$	5.015387490	$\bar{h}_2$	9.94706873
$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i,1} - h_{10})^2}{1000}$	0.002970813	$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i,2} - h_{20})^2}{1000}$	0.01353201
$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i,1} - \bar{h}_1)^2}{1000}$	0.002736775	$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i,2} - \bar{h}_2)^2}{1000}$	0.01013745
$\bar{\sigma}^2 = \sum_{i=1}^{i=1000} \frac{\hat{\sigma}_i^2}{1000}$	1.0003802743		
$\sum_{i=1}^{i=1000} \frac{(\hat{\sigma}_i^2 - \sigma^2)}{1000}$	0.0009686540		
$\sum_{i=1}^{i=1000} \frac{(\hat{\sigma}_i^2 - \bar{\sigma}^2)}{1000}$	0.0009685094		

Table 2.2: $n = 2000$, $\epsilon_1, \epsilon_2 \sim N(0, \sigma^2 = 1)$, $\boldsymbol{w}_{10}$ is regenerated in each repetitions. The results are based on 1000 repetitions

True h_1	5	True h_2	10
$\bar{h}_1$	5.017013194	$\bar{h}_2$	9.94575772
$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i1} - h_{10})^2}{1000}$	0.003056476	$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i2} - h_{20})^2}{1000}$	0.01306953
$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i1} - \bar{h}_1)^2}{1000}$	0.002769797	$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i2} - \bar{h}_2)^2}{1000}$	0.01013745
$\bar{\sigma}^2 = \sum_{i=1}^{i=1000} \frac{\hat{\sigma}_i^2}{1000}$	1.0002860431		
$\sum_{i=1}^{i=1000} \frac{(\hat{\sigma}_i^2 - \sigma^2)}{1000}$	0.0009624148		
$\sum_{i=1}^{i=1000} \frac{(\hat{\sigma}_i^2 - \bar{\sigma}^2)}{1000}$	0.0009623330		

Table 2.3: $n = 2000$, $\epsilon_1, \epsilon_2 \sim N(0, \sigma^2 = 1)$, $\boldsymbol{w}_{10}$ is constant over all repetitions. The results are based on 1000 repetitions

True h_{10}	5	True h_{20}	10
$\bar{h}_1$	5.019446510	$\bar{h}_2$	9.94229294
$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i1} - h_{10})^2}{1000}$	0.002997989	$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i2} - h_{20})^2}{1000}$	0.01345285
$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i1} - \bar{h}_1)^2}{1000}$	0.002619822	$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i2} - \bar{h}_2)^2}{1000}$	0.01012274
$\bar{\sigma}^2 = \sum_{i=1}^{i=1000} \frac{\hat{\sigma}_i^2}{1000}$	1.001095167		
$\sum_{i=1}^{i=1000} \frac{(\hat{\sigma}_i^2 - \sigma^2)^2}{1000}$	0.000966829		
$\sum_{i=1}^{i=1000} \frac{(\hat{\sigma}_i^2 - \bar{\sigma}^2)^2}{1000}$	0.000966829		

Table 2.4: $n = 2000$, $\epsilon_1 \sim N(0, \sigma^2 = 1)$, $\epsilon_2 \sim \sqrt{\frac{8}{10}} t(10)$, $\boldsymbol{w}_{10}$ is regenerated in each repetition. The results are based on 1000 repetitions

True h_{10}	5	True h_{20}	10
$\bar{h}_1$	5.018015363	$\bar{h}_2$	9.94678378
$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i1} - h_{10})^2}{1000}$	0.003139444	$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i2} - h_{20})^2}{1000}$	0.01341119
$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i1} - \bar{h}_1)^2}{1000}$	0.002814890	$\sum_{i=1}^{i=1000} \frac{(\hat{h}_{i2} - \bar{h}_2)^2}{1000}$	0.01057922
$\bar{\sigma}^2 = \sum_{i=1}^{i=1000} \frac{\hat{\sigma}_i^2}{1000}$	1.001095167		
$\sum_{i=1}^{i=1000} \frac{(\hat{\sigma}_i^2 - \sigma^2)^2}{1000}$	0.001015405		
$\sum_{i=1}^{i=1000} \frac{(\hat{\sigma}_i^2 - \bar{\sigma}^2)^2}{1000}$	0.001014206		

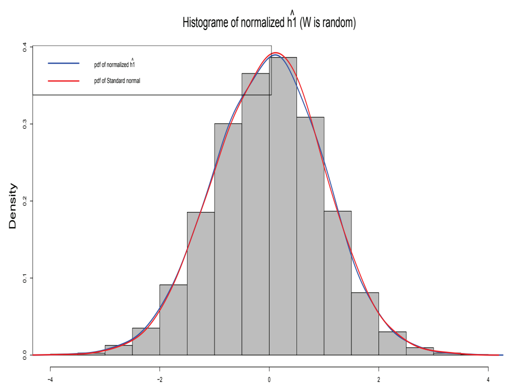
Table 2.5: $n = 2000$, $\epsilon_1 \sim N(0, \sigma^2 = 1)$, $\epsilon_2 \sim \sqrt{\frac{8}{10}} t(10)$, $\boldsymbol{w}_{10}$ is constant over all repetitions. The results are based on 1000 repetitions

Simulation study for limiting distribution

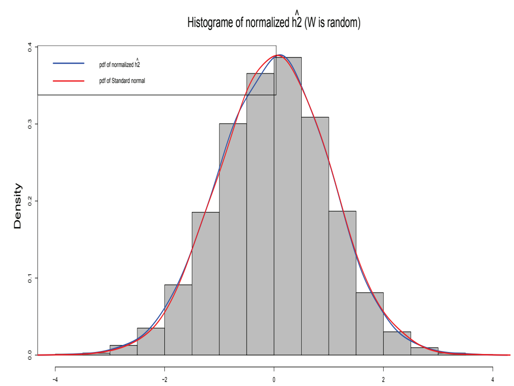
Similarly, we try two scenarios:

1. Random:Every repetition, we regenerate $\boldsymbol{w}_{10}$
2. Constant: $\boldsymbol{w}_{10}$ is generated only once and then set as constant over all repetitions

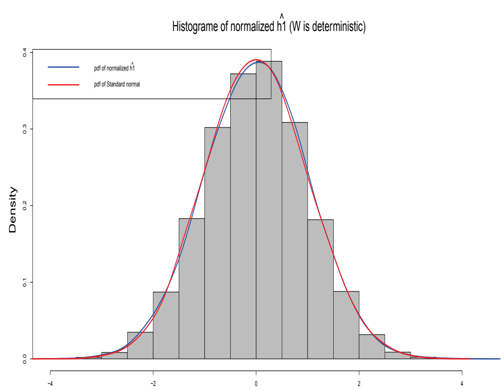
Let $h_{10} = 10, h_{20} = 8$ and sample size $n = 10^5$. $\boldsymbol{w}_{10}$ is generated from U$(0, 0.5)$, with $\boldsymbol{w}_{20}$ set as: $\boldsymbol{w}_{20} = \mathbf{1} - \boldsymbol{w}_{10}$. And the error terms $\boldsymbol{\epsilon}_1$, $\boldsymbol{\epsilon}_2$ are independent, from $\mathcal{N}(0, 0.7)$. $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ are generated through model (2.6). Afterwards, we can compute the LS estimate of $(\hat{h}_1^{(i)}, \hat{h}_2^{(i)})$. Meanwhile, the covariance matrix $\hat{\Sigma}^{(i)}$ can be calculated from Eq. (2.10), with estimates plugged in the formula. Thus, $(\hat{h}_1^{(i)}, \hat{h}_2^{(i)})$ can be normalized by left multiplying the square root of $\hat{\Sigma}^{(i)}$. The procedure is repeated 10^4 times. As mentioned in Remark 4, we make correction of the bias introduced by the even-order moments. Figure 2.1 shows the histogram of the normalized $\hat{h}_1^{(i)}, \hat{h}_2^{(i)}$ under both scenarios. The empirical probability function density lines (blue) of the normalized $\hat{h}_1^{(i)}$ and $\hat{h}_2^{(i)}, i = 1, \ldots, 10^4$ coincides with that of standard normal (red).



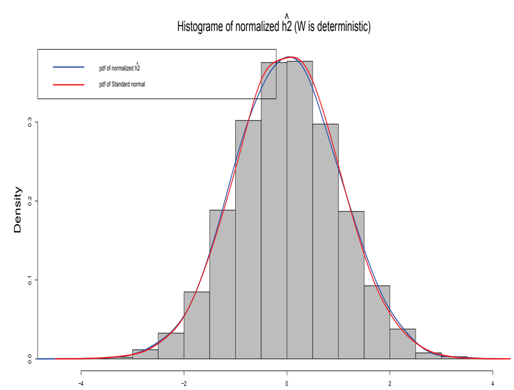
(a) pdf of $\hat{h}_1$ with $\boldsymbol{w}_1$ as random



(b) pdf of $\hat{h}_2$ with $\boldsymbol{w}_1$ as random



(c) pdf of $\hat{h}_1$ with $\boldsymbol{w}_1$ as deterministic



(d) pdf of $\hat{h}_2$ with $\boldsymbol{w}_1$ as deterministic

Figure 2.1: Empirical pdf of $\hat{h}_1$ and $\hat{h}_2$, the upper plots are for the case when $\boldsymbol{w}_{10}$ is refreshed every repetition, i.e. random , the bottom plots are for the case when $\boldsymbol{w}_{10}$ is deterministic

Besides, we conduct another simulation study with smaller sample size and larger noise variance. Now we decrease sample size to $n = 100$, and increase the noise variance to $\sigma^2 = 4$. The true signature expression values are: $h_1 = 10, h_2 = 5$. Other parameters are generated in the same way as the previous simulation study. Each time, based on the limiting distribution, we can calculate the confidence interval for the estimates $(\hat{h}_1^{(i)}, \hat{h}_2^{(i)})$. and checking if it covers the true parameter of h_{10}, h_{20}. Here, we define 'Exact CI' as the confidence interval calculated by plugging the true parameters in Eq. (2.10), and 'estimated CI' as the one calculated by plugging the estimates in Eq. (2.10). On the other hand, the confidence interval could be also constructed from bootstrap: we re-sample the 100 observations with replacement, and compute the moment estimates $(\hat{h}_1^b, \hat{h}_2^b)$ for each sampling $b = 1, \ldots, 1000$. Base on the quantile of these 1000 estimates, we could construct $1-\alpha$ confidence interval. We summarize the coverage probability and confidence interval length in Table 2.6 and 2.7. It is interesting to find that the coverage probability for both methods are close to the pre-specified level $(1-\alpha)$. However, the confidence interval length we derive from theorem is shorter than that from bootstrap, and more computationally efficient.

	Exact CI	estimated CI	Boostrap
99% Coverage Probability	0.973	0.977	0.987
99% CI Length	4.102	4.212	4.958
95% Coverage Probability	0.931	0.928	0.943
95% CI Length	3.128	3.212	3.594
90% Coverage Probability	0.877	0.874	0.885
90% CI Length	2.618	2.688	2.959

Table 2.6: Coverage Probability and CI length for $\hat{h}_1$,$n = 100, \sigma^2 = 4, h_1 = 10, h_2 = 5$, w_i generated from $Beta(a = 5, b = 5)$

	Exact CI	estimated CI	Boostrap
99% Coverage Probability	0.977	0.977	0.980
99% CI Length	2.051	2.039	2.181
95% Coverage Probability	0.933	0.932	0.936
95% CI Length	1.564	1.555	1.656
90% Coverage Probability	0.874	0.870	0.884
90% CI Length	1.309	1.301	1.385

Table 2.7: Coverage Probability and CI length for $\hat{h}_2$,$n = 100, \sigma^2 = 4, h_1 = 10, h_2 = 5$,w_i generated from $Beta(a = 5, b = 5)$

2.2.6 Real Data Analysis

We apply our method on the gene expression profiles of peripheral blood RNA, named as GSE33566. The data is from Yang et al. (2012). It consists of 123 samples. Among these 123 samples, 93 are pulmonary fibrosis (IPF) patients, while 30 are healthy, used as control group. Note that the data is on log2 scale and normalized already.

As we know that blood consists of several cell types, in this analysis, we decompose it into four categories: T cells, B cells, Granulocytes and others. Since it's very difficult to find the gene marker that are exclusively expressed in one specific cell type, instead of using the exact makers, we found eight approximate gene markers:

Makers for T cells: "CD3G","CD3D","CD28"

Makers for B cells: "CD19","CD22","CD79B"

Makers for Granulocytes: "CEACAM1","CEACAM3"

Assessment of the genes' quality as potential markers is important before estimate. Pairwise correlation is a good way to check whether the marker assumptions are reasonable or not. If there are multiple makers for one cell type, the correlation between these markers' global expression (tissue level, i.e. $\boldsymbol{g}_j$) will be large. The logic is simple. If M_1 and M_2 are marker genes for k-th cell type:

$$\boldsymbol{g}_{M_1} = h_{M_1}\boldsymbol{w}_k + \boldsymbol{\epsilon}_{M_1}, \boldsymbol{g}_{M_2} = h_{M_2}\boldsymbol{w}_k + \boldsymbol{\epsilon}_{M_2}.$$

Then $cor(\boldsymbol{g}_{M_1}, \boldsymbol{g}_{M_2}) = cor(h_{M_1}\boldsymbol{w}_k + \boldsymbol{\epsilon}_{M_1}, h_{M_2}\boldsymbol{w}_k + \boldsymbol{\epsilon}_{M_2}) = \dfrac{h_{M_1}h_{M_2}\mu_2(w)}{\sqrt{h_{M_1}^2\mu_2(w)+\sigma^2}\sqrt{h_{M_2}^2\mu_2(w)+\sigma^2}}$.

We notice that the correlation gets larger as the ratio of $\dfrac{\sigma^2}{\mu_2(w)}$ gets smaller. When $\sigma^2 = 0$, they are perfectly correlated with correlation equals to 1.

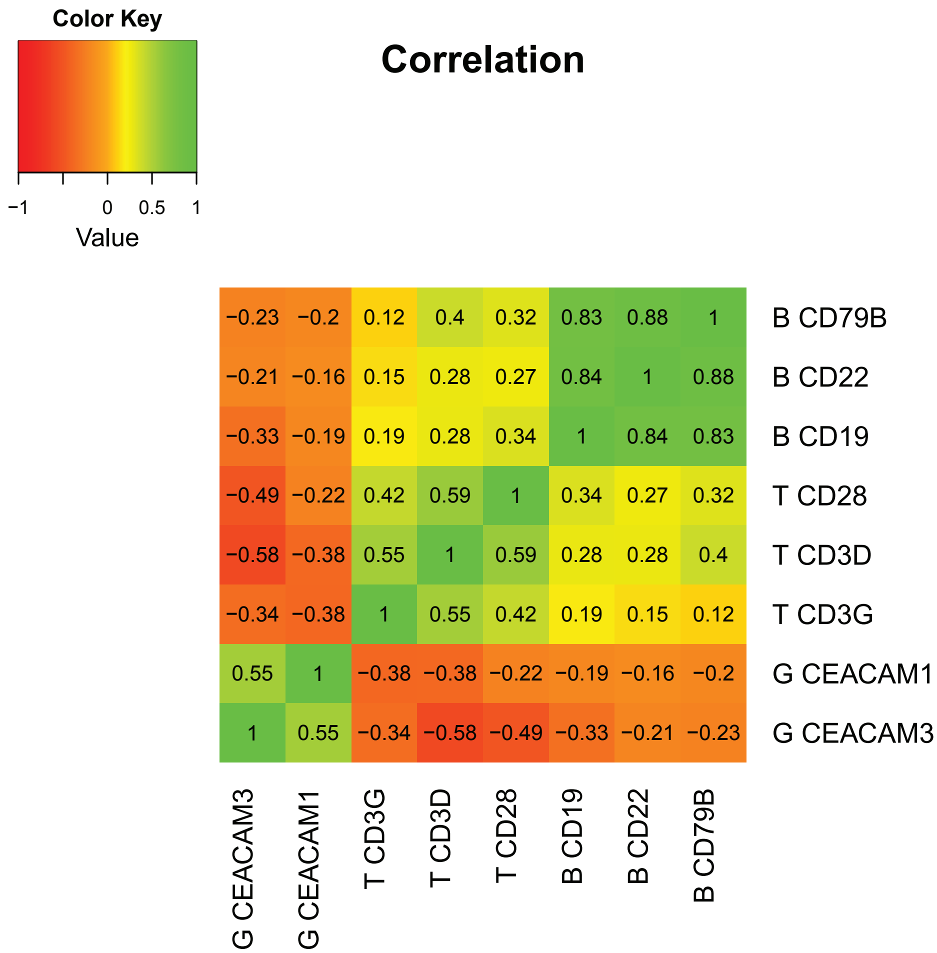


Figure 2.2: Heat map of the pairwise correlation

Fig. 2.2 is the heatmap of pairwise correlation. The row/column names start with the first letter of its specific expressed cell type. The gene markers exhibit clear block pattern(green square) across their specific cell.

Therefore, The marker information provided in Table 2.8 is reasonable in our data set. Besides the 8 markers, we also choose "HBA2" which have strong signal across all samples. So the cell specific expression matrix will be like below:

	CD3G	CD3D	CD28	CD19	CD22	CD79B	CEACAM1	CEACAM3	RHOA
T-cell	$*$	$*$	$*$	0	0	0	0	0	$*$
B-cell	0	0	0	$*$	$*$	$*$	0	0	$*$
Granulocytes	0	0	0	0	0	0	$*$	$*$	$*$
Others	0	0	0	0	0	0	0	0	$*$

Table 2.8: Cell Specific gene expression for GSE 33566

Iterative Least Square algorithm described in Section 2.2.2 is implemented.

The proportion estimates for both IPF and CTRL group are shown in Fig. 2.3. Table 2.9 summarizes the estimated cell specific expression matrix for both groups.

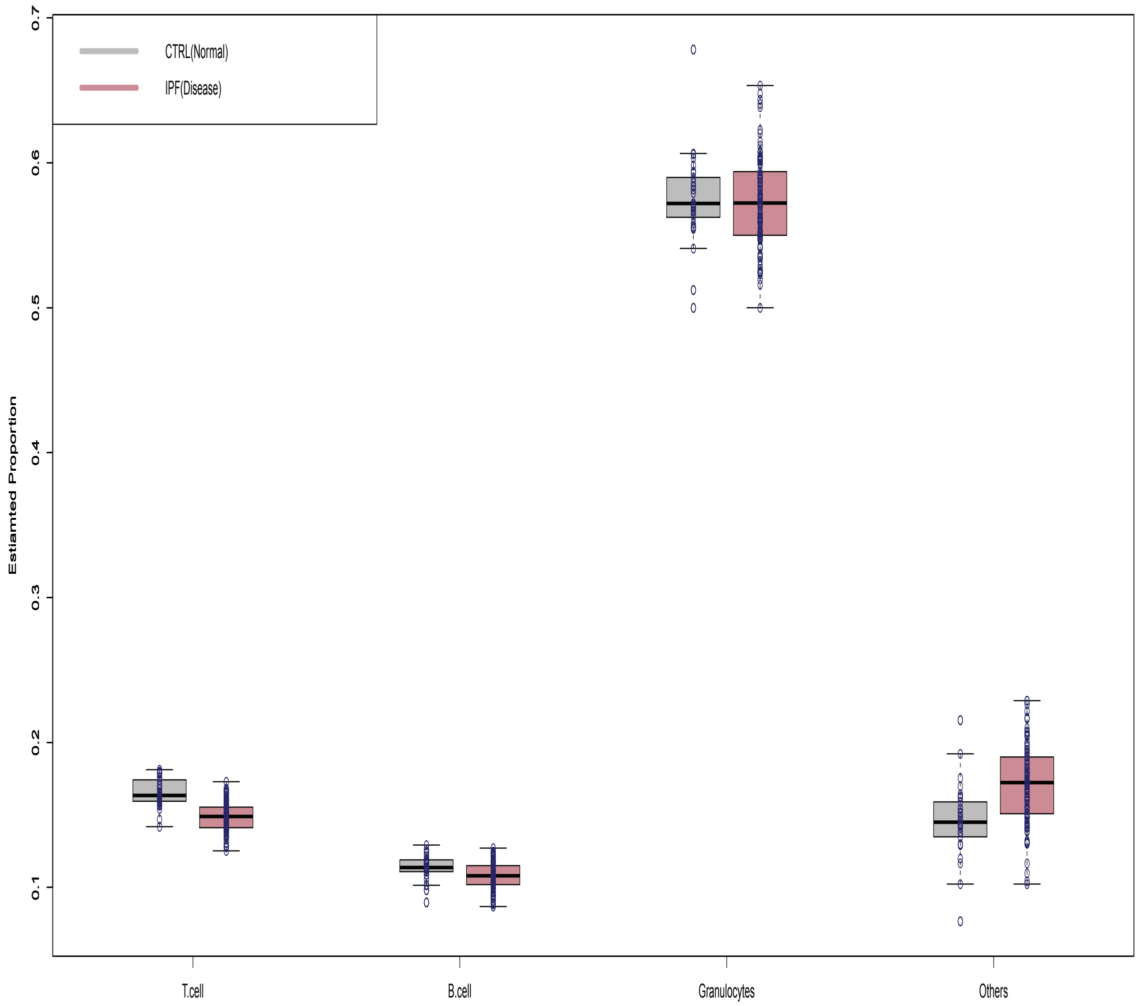


Figure 2.3: Estimated Proportions

	Control				IPF			
	T-cell	B-cell	Granulocytes	Others	T-cell	B-cell	Granulocytes	Others
CD3G	72.8	0.0	0.0	0.0	74.5	0.0	0.0	0.0
CD3D	61.9	0.0	0.0	0.0	60.5	0.0	0.0	0.0
CD28	45.4	0.0	0.0	0.0	44.8	0.0	0.0	0.0
CD19	0.0	84.4	0.0	0.0	0.0	90.1	0.0	0.0
CD22	0.0	56.3	0.0	0.0	0.0	61.6	0.0	0.0
CD79B	0.0	68.9	0.0	0.0	0.0	72.6	0.0	0.0
CEACAM1	0.0	0.0	14.8	0.0	0.0	0.0	15.0	0.0
CEACAM3	0.0	0.0	13.4	0.0	0.0	0.0	14.1	0.0
HBA2	15.4	8.3	13.8	5.9	24.0	0.0	14.5	3.2

Table 2.9: Estimated Signature matrix $\hat{H}$

We see that there is no big changes for the gene markers between the two groups. Yet, It's interesting to find that the proportion for T cell and B cell from IPF patients are smaller than those of normal samples. Furthermore, we conduct permutation test to see if such downward bias is statistically significant. Every time, we permute the labels (labels that indicate whether the sample is IPF or healthy), and calculate the marginal proportional difference between the 'new' groups (after permutation). We repeat the permutation 1000 times, based on which, we can compute the pvalues. The test results are summarized in table 2.10, p-values for T cell and B cells suggests that such downward bias is significant.

	Mean for IPF	Mean for CTRL	P value (Permutation Test)
T-cell	0.149	0.157	0.002
B-cell	0.107	0.122	0
Granulocytes	0.573	0.575	0.745

Table 2.10: Test of the estimated proportion

2.3 Equivalence of Moment Estimate and ILS in Homoscedastic Case

In this section, we will discuss the equivalence of ILS estimates and moment estimate under homoscedastic assumption. We will also show the the limiting distribution of ILS estimate of $(\hat{h}_1, \hat{h}_2)$ by the equivalence.

Consider the same Model (2.6), i.e.

$$\begin{cases} \mathbf{g}_1 &= h_{10}\boldsymbol{w}_1 + \boldsymbol{\epsilon}_1 \\ \mathbf{g}_2 &= h_{20}(1-\boldsymbol{w}_1) + \boldsymbol{\epsilon}_2 \end{cases}$$

Where $\boldsymbol{g}_j = \begin{pmatrix} g_{1j} \\ g_{2j} \\ \dots \\ g_{nj} \end{pmatrix}$, $\boldsymbol{\epsilon}_j = \begin{pmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \dots \\ \epsilon_{nj} \end{pmatrix}$, $\boldsymbol{\epsilon}_j \sim N(0, \sigma^2 I)$, $j = 1, 2$

By first and second moment, we have

$$\begin{cases} \mathrm{E}(\bar{g}_1) &= h_{10}\mu_1(w) \\ \mathrm{E}(\bar{g}_2) &= h_{20}(1-\mu_1(w)) \\ \mathrm{E}(V\hat{a}r(g_1)) &= h_1^2\mu_2(w) + \sigma^2 \\ \mathrm{E}(V\hat{a}r(g_2)) &= h_2^2\mu_2(w) + \sigma^2 \\ \mathrm{E}(C\hat{o}v(g_1, g_2)) &= -h_1h_2\mu_2(w) \end{cases}$$

where $\hat{\mathrm{Var}}(g_1), \hat{\mathrm{Var}}(g_2)$ is sample variance, $\bar{g}_1, \bar{g}_2$ is sample mean, and $\hat{\mathrm{Cov}}(g_1, g_2)$ is sample covariance. By solving the equalities, we have:

$$\begin{cases} \hat{h}_1 &= \bar{g}_1 + k\bar{g}_2 \\ \hat{h}_2 &= \bar{g}_2 + \dfrac{1}{k}\bar{g}_1 \end{cases} \tag{2.14}$$

where $k = \dfrac{c + \sqrt{c^2+4}}{2}$, and $c = \dfrac{\hat{\mathrm{Var}}(g_2) - \hat{\mathrm{Var}}(g_1)}{\hat{\mathrm{Cov}}(g_1, g_2)}$

On the other hand, we derive the explicit solution of ILS in the proof of (1) in Theorem 1. For more details, please refer to section 2.5.1. Compared the solution in (2.14) to the LS solution in Eq. (2.29), they are exact the same. Therefore, the equivalence is shown.

Next, we will show the limiting distribution of $(\hat{h}_1^{LS}, \hat{h}_2^{LS})$. Because of the equivalence, it's sufficient to show the limiting distribution of moment estimate:

Let $\boldsymbol{u}_1 = \begin{pmatrix} h_{10}\mu(w) \\ h_{20}(1-\mu(w)) \\ h_{10}^2\mu_2(w) - \sigma^2 \\ h_{20}^2\mu_2(w) - \sigma^2 \\ -h_{10}h_{20}\mu_2(w) \end{pmatrix}$, $\boldsymbol{v}_1 = \begin{pmatrix} \bar{g}_1 \\ \bar{g}_2 \\ \hat{\mathrm{Var}}(g_1) \\ \hat{\mathrm{Var}}(g_2) \\ \hat{\mathrm{Cov}}(g_1, g_2) \end{pmatrix}$

By Multivariate Lindeberg-Levy CLT Greene (2002), the joint distribution of the sample statistics is:

$$\sqrt{n}(\boldsymbol{v}_1 - \boldsymbol{u}_1) \xRightarrow{D} \mathcal{N}(0, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ & & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ & & & \sigma_{44} & \sigma_{45} \\ & & & & \sigma_{55} \end{pmatrix})$$

As $(\hat{h}_1, \hat{h}_2)$ is function of $\boldsymbol{v}_1$

$$f_1(\boldsymbol{v}_1) = \begin{pmatrix} \hat{h}_1 \\ \hat{h}_2 \end{pmatrix} = \begin{pmatrix} \bar{g}_1 + k\bar{g}_2 \\ \bar{g}_2 + \frac{1}{k}\bar{g}_1 \end{pmatrix}$$

By Multivariate delta Method van der Vaart (1998):

$$
\begin{aligned}
&\sqrt{n}(f(\boldsymbol{v}_1) - f_1(\boldsymbol{u}_1)) \\
&\stackrel{D}{\Longrightarrow}\sqrt{n}\nabla f(\boldsymbol{u}_1)^T\{\begin{pmatrix} \bar{g}_1 \\ \bar{g}_2 \\ \hat{\mathrm{Var}}(g_1) \\ \hat{\mathrm{Var}}(g_2) \\ \hat{\mathrm{Cov}}(g_1, g_2) \end{pmatrix} - \begin{pmatrix} h_{10}\mu(w) \\ h_{20}(1-\mu(w)) \\ h_{10}^2\mu_2(w) - \sigma^2 \\ h_{20}^2\mu_2(w) - \sigma^2 \\ -h_{10}h_{20}\mu_2(w) \end{pmatrix}\} \\
&\stackrel{D}{\Longrightarrow}\nabla f_1(\boldsymbol{u}_1)^T\mathcal{N}(0, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ & & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ & & & \sigma_{44} & \sigma_{45} \\ & & & & \sigma_{55} \end{pmatrix})
\end{aligned}
\tag{2.15}
$$

Where $\nabla f(\boldsymbol{u}_1)$ is derivative of f w.r.t $\boldsymbol{u}_1$. The closed form of $\nabla f_1(\boldsymbol{u}_1)$ and σ_{ij} are summarized in Eq.(2.10).

Note that the calculation of the covariance matrix $\Sigma = (\sigma_{ij})$ is standard, though tedious, thus we omit the details. And the results in remark 2 is useful for the calculation. All the above results are based on the assumption that $\boldsymbol{w}_{10}$ is random. For the special case when $\boldsymbol{w}_{10}$ is deterministic, the proof is almost the same, except for the calculation of $\Sigma = (\sigma_{ij})$. The calculation of Σ under the deterministic assumption will be simpler. The closed form of Σ is given in Eq. (2.12)

2.4 Complete Deconvolution Model in Heteroscedastic Case

Section 2.4 summarizes both algorithm and asymptotic unbiasedness of the estimate when the error term share common variance across all genes. Figure (2.6) in the proof section (2.27) explains the rationale from geometric view. However, if the homoscedasticity is violated, iterated LS will be biased and not appropriate. We will illustrate the biasness by a simple simulation study. With very similar generating scheme as described in section 2.2.5, instead of common error variance, we set them differently: ϵ_1

is from standard normal distribution, while ϵ_2 is from normal distribution with mean 0 and variance 0.5. The results are summarized in the table below based on 1000 repetitions. The average deviation of $\hat{h}_1$ and $\hat{h}_2$ from true values are 0.7945747 and 0.1518678 respectively.

True h_1	10	True h_2	15
Average Estimate $\bar{h}_1$	10.8798	Average Estimate $\bar{h}_2$	14.61578
$\sum_{i=1}^{i=n}\frac{(\hat{h}_{i1}-h_1)^2}{n}$	0.7945747	$\sum_{i=1}^{i=n}\frac{(\hat{h}_{i1}-h_1)^2}{n}$	0.1518567
$\sum_{i=1}^{i=n}\frac{(\hat{h}_{i1}-\bar{h}_1)^2}{n}$	0.0205309	$\sum_{i=1}^{i=n}\frac{(\hat{h}_{i1}-\bar{h}_1)^2}{n}$	0.004235257

Table 2.11: $n = 2000$,$\epsilon_1 \sim N(0, \sigma = 1)$, $\epsilon_2 \sim N(0, \sigma = .5)$, based on 1000 repetitions

Due to the limitation of ILS in heteroscedastic case mentioned above, we propose to use moment estimate. Later, we will also show the joint limiting distribution of $(\hat{h}_1, \hat{h}_2)$.

Moment estimate is algebraically complicated if there are too many unknown parameters. Therefore, for simplicity, we will investigate on the case when there are only two cell types and two signatures, i.e. $r = 2, p = 2$.

Consider the model:

$$\begin{cases} \mathbf{g}_1 & = h_{10}\boldsymbol{w}_{10} + \boldsymbol{\epsilon}_1 \\ \mathbf{g}_2 & = h_{20}(\mathbf{1} - \boldsymbol{w}_{10}) + \boldsymbol{\epsilon}_2 \end{cases} \tag{2.16}$$

where $\boldsymbol{g}_j = \begin{pmatrix} g_{1j} \\ g_{2j} \\ \dots \\ g_{nj} \end{pmatrix}$, $\boldsymbol{\epsilon}_j = \begin{pmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \dots \\ \epsilon_{nj} \end{pmatrix}$, and $\boldsymbol{\epsilon}_j \sim N(0, \sigma_j^2 I), j = 1, 2, \sigma_1^2 \neq \sigma_2^2$

2.4.1 Moment Estimate when W is asymmetric

Assume $\boldsymbol{w}_{10}$ is from some unknown population, and for $j = 1, 2$

$$\begin{aligned} &\mu(w) = \mathrm{E}(w_{ij}) > c, \text{ for some } c > 0 \\ &\mu_3(w) = \mathrm{E}((w_{ij} - \mu(w))^3) \neq 0 \end{aligned} \tag{2.17}$$

Note that the nonzero assumption of $\mu_3(w)$ requires that $\boldsymbol{w}_{10}$ be asymmetric. Otherwise, we need to consider higher order moments.

Solution of the moment estimate

Define:

$$\begin{cases} \bar{g}_1 &= \sum_{i=1}^{n} \dfrac{g_{i1}}{n} \\ \bar{g}_2 &= \sum_{i=1}^{n} \dfrac{g_{i2}}{n} \\ \hat{\mu}_{11} &= \dfrac{\sum_{i=1}^{i=n}(g_{i1} - \bar{g}_1)^2}{n} \\ \hat{\mu}_{22} &= \dfrac{\sum_{i=1}^{i=n}(g_{i2} - \bar{g}_2)^2}{n} \\ \hat{\mu}_{12} &= \dfrac{\sum_{i=1}^{i=n}(g_{i1} - \bar{g}_1)(g_{i2} - \bar{g}_2)}{n} \\ \hat{\mu}_{122} &= \dfrac{\sum_{i=1}^{i=n}(g_{i1} - \bar{g}_1)(g_{i2} - \bar{g}_2)^2}{n} \\ \hat{\mu}_{112} &= \dfrac{\sum_{i=1}^{i=n}(g_{i1} - \bar{g}_1)^2(g_{i2} - \bar{g}_2)}{n} \end{cases} \tag{2.18}$$

We first calculate

$$\begin{cases} \mathrm{E}(\bar{g}_1) &= h_{10}\mu(w) \\ \mathrm{E}(\bar{g}_2) &= h_{20}(1 - \mu(w)) \\ \mathrm{E}(\hat{\mu}_{11}) &= h_{10}^2\mu_2(w) + \sigma_1^2 \\ \mathrm{E}(\hat{\mu}_{22}) = &= h_{20}^2\mu_2(w) + \sigma_2^2 \\ \mathrm{E}(\hat{\mu}_{12}) &= -h_{10}h_{20}\mu_2(w) \\ \mathrm{E}(\hat{\mu}_{122}) = &= h_{10}h_{20}^2\mu_3(w) \\ \mathrm{E}(\hat{\mu}_{112}) &= -h_{10}^2h_{20}\mu_3(w) \end{cases} \tag{2.19}$$

Where $\bar{g}_1, \bar{g}_2, \hat{\mu}_{11}, \hat{\mu}_{22}, \hat{\mu}_{12}, \hat{\mu}_{112}, \hat{\mu}_{122}$ are defined in Eq (2.18). From the last two equality, we obtain the ratio of $\dfrac{\hat{h}_1}{\hat{h}_2}$. Together with the first two equalities, we are able to solve $\hat{h}_1, \hat{h}_1,$ and $\hat{\mu}(w)$. Plug in these estimate in the left equalities, we could get the solutions below:

$$\begin{cases} \hat{h}_1 = \bar{g}_1 + \lambda \bar{g}_2 \\ \hat{h}_2 = \bar{g}_2 + \dfrac{1}{\lambda}\bar{g}_1 \\ \hat{\mu}(w) = \dfrac{\bar{g}_1}{\hat{h}_1} \\ \hat{\mu}_2(w) = \dfrac{\hat{\mu}_{12}}{\hat{h}_1\hat{h}_2} \\ \hat{\sigma}_1^2 = \hat{\mu}_{11} - \hat{h}_1^2\hat{\mu}_2(w) \\ \hat{\sigma}_2^2 = \hat{\mu}_{22} - \hat{h}_2^2\hat{\mu}_2(w) \\ \hat{u}_3(w) = \dfrac{\hat{u}_{122}}{\hat{h}_1\hat{h}_2^2} \end{cases} \tag{2.20}$$

where $\lambda = -\dfrac{-\hat{\mu}_{112}}{\hat{\mu}_{122}}$

Remark 6. We notice that only the first, second, sixth and seventh equalities are used in calculating $\hat{h}_1$ and $\hat{h}_2$.

Let $\boldsymbol{v}_2 = \begin{pmatrix} \bar{g}_1 \\ \bar{g}_2 \\ \hat{u}_{112} \\ \hat{u}_{122} \end{pmatrix}$, then $\begin{pmatrix} \hat{h}_1 \\ \hat{h}_2 \end{pmatrix} = f(\boldsymbol{v}_2) = \begin{pmatrix} \bar{g}_1 + \lambda\bar{g}_2 \\ \bar{g}_2 + \frac{1}{\lambda}\bar{g}_1 \end{pmatrix}$

Asymptotic properties of the moment estimate

Theorem 3. *Let $\epsilon_{1,j}, \epsilon_{2,j}, \ldots, \epsilon_{n,j}$ be i.i.d, with mean 0 and variance σ_j^2 $(j = 1, 2)$. Consider the model in (2.16), where $\sigma_1^2 \neq \sigma_2^2$. Assume $\boldsymbol{w}_1$ is from some unknown random population, with $0 < \mathrm{E}(w_i) < 1$, and $\mu_3(w) \neq 0$, for $i = 1, 2, \ldots, n$ Then*

$$\sqrt{n}(\begin{pmatrix} \hat{h}_1 \\ \hat{h}_2 \end{pmatrix} - \begin{pmatrix} h_{10} \\ h_{20} \end{pmatrix}) \overset{D}{\Longrightarrow} \mathcal{N}(\mathbf{0}, \Sigma_3)$$

where $\Sigma_3 = \triangledown f(\boldsymbol{u}_2)'\Sigma \triangledown f(\boldsymbol{u}_2)$, *and* $\triangledown f(\boldsymbol{u}_2)$, $\Sigma = (\sigma_{ij})$ *are given by:*

$$
\begin{aligned}
\sigma_{11} =& h_{10}^2\mu_2(w) + \sigma_{10}^2 \\
\sigma_{22} =& h_{20}^2\mu_2(w) + \sigma_{20}^2 \\
\sigma_{33} =& h_{10}^4h_{20}^2(\mu_6(w) - 6\mu_2(w)\mu_4(w) - \mu_3^2(w) + 9\mu_2^3(w)) + h_{10}^4\sigma_{20}^2(\mu_4(w) - \mu_2^2(w)) + \\
& 3\sigma_{10}^4\sigma_{20}^2 + 4h_{10}^2h_{20}^2(\mu_4(w) - \mu_2^2(w))\sigma_{10}^2 + 3h_{20}^2\mu_2(w)\sigma_{10}^4 + \\
& h_{10}^2h_{20}^2\sigma_{10}^2(\mu_4(w) - 3\sigma_w^4) + 4h_{10}^2\mu_2(w)\sigma_{10}^2\sigma_{20}^2 \\
\sigma_{44} =& h_{10}^2h_{20}^4(\mu_6(w) - 6\mu_2(w)\mu_4(w) - \mu_3^2(w) + 9\mu_2^3(w)) + h_{20}^4\sigma_{10}^2(u_4(w) - \mu_2^2(w)) + \\
& 3\sigma_{10}^2\sigma_{20}^4 + 4h_{10}^2h_{20}^2(\mu_4(w) - \mu_2^2(w))\sigma_{20}^2 + 3h_{10}^2\mu_2(w)\sigma_{20}^4 + \\
& h_{10}^2h_{20}^2\sigma_{20}^2(\mu_4(w) - 3\mu_2^2(w)) + 4h_{20}^2\mu_2(w)\sigma_{10}^2\sigma_{20}^2 \\
\sigma_{12} =& -h_{10}h_{20}\mu_2(w) \\
\sigma_{13} =& -h_{10}^3h_{20}(\mu_4(w) - 3\mu_2^2(w)) \\
\sigma_{14} =& h_{10}^2h_{20}^2(\mu_4(w) - 3\mu_2^2(w)) \\
\sigma_{23} =& h_{10}^2h_{20}^2(\mu_4(w) - 3\mu_2^2(w)) \\
\sigma_{24} =& -h_{10}h_{20}^3(\mu_4(w) - 3\mu_2^2(w)) \\
\sigma_{34} =& -h_{10}^3h_{20}^3(\mu_6(w) - 6\sigma_w^2\mu_4(w) - \mu_3^2(w) + 9\mu_2^3(w)) - h_{10}^3h_{20}\sigma_{20}^2(\mu_4(w) - \\
& 3\mu_2^2(w)) - 2h_{10}^3h_{20}\sigma_{20}^2(\mu_4(w) - \mu_2(w)^4) - 2h_{10}h_{20}^3\sigma_{10}^2(\mu_4(w) - \mu_2(w)^4) - \\
& 9h_{10}h_{20}\mu_2(w)\sigma_{10}^4\sigma_{20}^4 - 4h_{10}h_{20}\mu_2(w)\sigma_{10}^2\sigma_{20}^2
\end{aligned}
\tag{2.21}
$$

$$
\triangledown f(\boldsymbol{u}_2) = \begin{pmatrix} 1 & \frac{h_{20}}{h_{10}} \\ \frac{h_{10}}{h_{20}} & 1 \\ \dfrac{-(1-\mu(w))}{h_{10}h_{20}\mu_3(w)} & \dfrac{\mu(w)}{h_{10}^2\mu_3(w)} \\ \dfrac{-(1-\mu(w))}{h_{20}^2\mu_3(w)} & \dfrac{\mu(w)}{h_{10}h_{20}\mu_3(w)} \end{pmatrix} \tag{2.22}
$$

Remark 7. Under the same condition of Theorem 3, in a special case, let $\boldsymbol{w}_{10}$ be deterministic, then

$$\sqrt{n}(\begin{pmatrix} \hat{h}_1 \\ \hat{h}_2 \end{pmatrix} - \begin{pmatrix} h_{10} \\ h_{20} \end{pmatrix}) \overset{D}{\Longrightarrow} \mathcal{N}(\mathbf{0}, \Sigma_4)$$

where $\Sigma_4 = \nabla f(\boldsymbol{u}_2)^{'} \Sigma f(\boldsymbol{u}_2)$, where $f(\boldsymbol{u}_2)$ is defined as before with $\mu_k(w)$ replaced by $m_k(w)$, and $\Sigma = (\sigma_{ij})$ is defined as below. Note that all the σ_{ij} not listed below are 0.

$$\begin{aligned}
\sigma_{11} =& \sigma_1^2 \\
\sigma_{22} =& \sigma_2^2 \\
\sigma_{33} =& 2m_2(w)h_{20}^2\sigma_1^4 + 4h_{10}^2h_{20}^2m_4(w)\sigma_1^2 - 4h_{10}^2h_2^2\sigma_1^2m_2^2(w) + h_{10}^4m_4(w)\sigma_2^2 \\
& - \sigma_2^2h_{10}^4m_2^2(w) + 2\sigma_2^2\sigma_1^4 + 4h_{10}^2m_2(w)\sigma_2^2\sigma_1^2 \\
\sigma_{44} =& 2m_2(w)h_{10}^2\sigma_2^4 + 4h_{10}^2h_{20}^2m_4(w)\sigma_2^2 - 4h_{10}^2h_{20}^2\sigma_2^2m_2^2(w) + h_{20}^4m_4(w)\sigma_1^2 \\
& - \sigma_1^2h_{20}^4m_2^2(w) + 2\sigma_1^2\sigma_2^4 + 4h_{20}^2m_2(w)\sigma_1^2\sigma_2^2 \\
\sigma_{34} =& - 2m_4(w)h_{10}^3h_{20}\sigma_2^2 + 2h_{10}^3h_{20}\sigma_2^2\mu_2^2(w) - 2h_{10}h_2^3\sigma_1^2m_4(w) + \\
& 2h_{10}h_{20}^3\sigma_1^2m_2^2(w) - 4m_2(w)h_{10}h_{20}\sigma_1^2\sigma_2^2
\end{aligned} \tag{2.23}$$

Remark 8. Let $\boldsymbol{v}_{20} = \begin{pmatrix} \bar{g}_1 \\ \bar{g}_2 \\ \hat{\mu}_{11} \\ \hat{\mu}_{22} \\ \hat{\mu}_{12} \\ \hat{u}_{112} \\ \hat{u}_{122} \end{pmatrix}$, the covariance matrix of $\boldsymbol{v}_{20}$ is also calculated, and the results are given in formula (2.32) (when W is random), and (2.33) (when W is deterministic).

Remark 9. The covariance matrix in the limiting distribution involve 4th order moment estimate of $\mu_4(w)$. One may estimate through higher order of moments. As a simpler way, it can also be solved by LS method conditioning on the moment estimate of $\hat{h}_1, \hat{h}_2$. Similar as remark 4, the even order moment estimate of $\mu_{2K}(w)$ is biased, and can be corrected by removing the bias. For example , the second and fourth moment with bias corrected are:

$$\hat{u}_2(w) = \hat{u}_2^{LS}(w) - \frac{(\hat{h}_1^2\hat{\sigma}_1^2 + \hat{h}_2^2\hat{\sigma}_2^2)}{(\hat{h}_1^2 + \hat{h}_1^2)^2}$$
$$\hat{u}_4(w) = \hat{u}_4^{LS}(w) - \frac{6\hat{u}_2(w)(\hat{h}_1^2\hat{\sigma}_1^2 + \hat{h}_2^2\hat{\sigma}_2^2)}{(\hat{h}_1^2 + \hat{h}_1^2)^2} - \frac{3(\hat{h}_1^2\hat{\sigma}_1^2 + \hat{h}_2^2\hat{\sigma}_2^2)^2}{(\hat{h}_1^2 + \hat{h}_1^2)^4}$$

Simulation study for limiting distribution

Let $h_{10} = 10, h_{20} = 5$ and sample size $n = 10^5$ ($n = 10^6$ for the case when $\boldsymbol{w}_{10}$ is random). $\boldsymbol{w}_{10}$ is generated from B$(2, 5)$, and $\boldsymbol{w}_{20}$ is set as: $\boldsymbol{w}_{20} = \mathbf{1} - \boldsymbol{w}_{10}$. The distribution of of $\boldsymbol{w}_{10}$ is shown in figure 2.4 below, which is asymmetric.

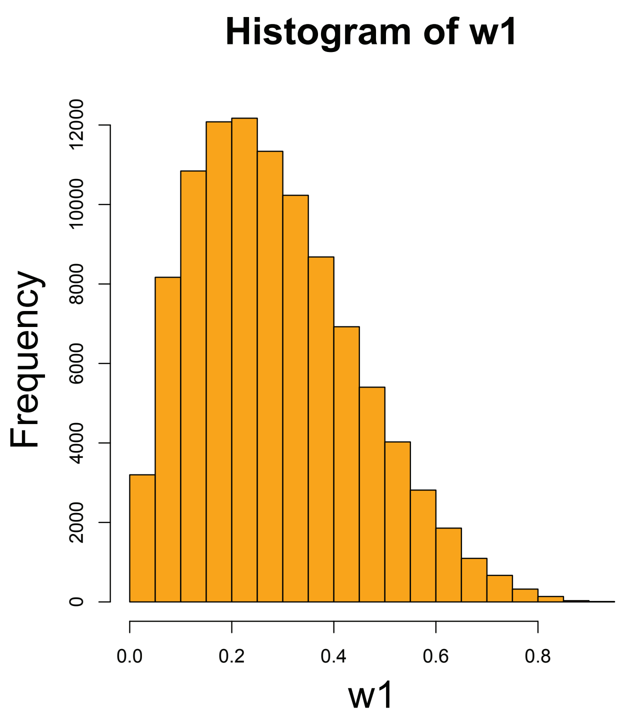
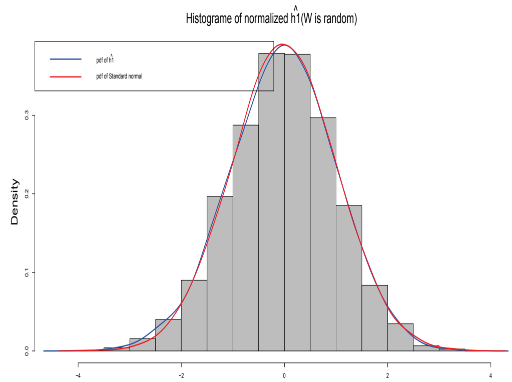


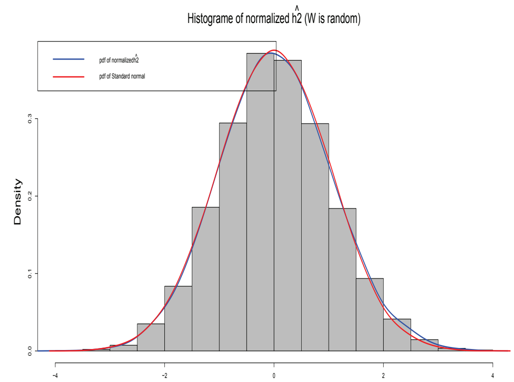
Figure 2.4: Histogram of $\boldsymbol{w}_1$

Similar as the previous simulation, we try two scenario: first, in each repetition, we regenerate $\boldsymbol{w}_{10}$ and $\boldsymbol{w}_{20}$. Second,we generate $\boldsymbol{w}_{10}$ and $\boldsymbol{w}_{20}$ only once and keep it as constant over all repetition.

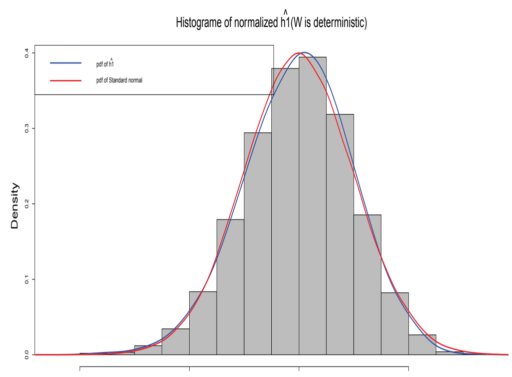
Each time, $\boldsymbol{\epsilon}_1 \sim \mathcal{N}(0, 0.7)$, $\boldsymbol{\epsilon}_2 \sim \mathcal{N}(0, 1)$, and $\boldsymbol{g}_1$, $\boldsymbol{g}_2$ are generated through formula (2.16). By Eq. (2.20) we can estimate $(\hat{h}_1^{(i)}, \hat{h}_2^{(i)})$. Meanwhile, we can calculate covariance matrix $\hat{\Sigma}^{(i)}$ from theorem 3. Note that if $\boldsymbol{w}_{10}$ is considered as deterministic, we use the Eq. (2.23) in Remark 7. Thus we can normalized $(\hat{h}_1^{(i)}, \hat{h}_2^{(i)})$ by by left multiplying the square root of $\hat{\Sigma}^{(i)}$. One thing we need to be cautious about is that the covariance matrix formula involves the forth central moment, i.e. $\mu_4(w)$, which is biased if we directly plug in $\hat{h}_1, \hat{h}_2$ and solve the least square of $\boldsymbol{w}_{10}$. For details please see remark 9. We repeat the procedure 10^4 times. Fig. 2.5 shows the histogram of the normalized $\hat{h}_1^{(i)}$ and $\hat{h}_2^{(i)}, i = 1, \ldots, 10^4$ under both scenarios. The empirical pdf of the normalized $\hat{h}_{i1}$,$\hat{h}_{i2}$ (blue) are close to that of standard normal (red).



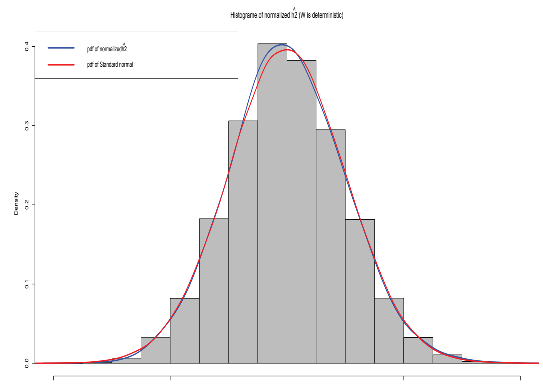
(a) pdf of $\hat{h}_1$ when $\boldsymbol{w}_1$ is random



(b) pdf of $\hat{h}_2$ when $\boldsymbol{w}_1$ is random



(c) pdf of $\hat{h}_1$ when $\boldsymbol{w}_1$ is deterministic



(d) pdf of $\hat{h}_2$ when $\boldsymbol{w}_1$ is deterministic

Figure 2.5: Empirical pdf of $\hat{h}_1$ and $\hat{h}_2$, the upper plots are for the case when $\boldsymbol{w}_{10}$ is random and regenerated in every repetition. The bottom plots are for the case when $\boldsymbol{w}_{10}$ is deterministic

Besides, we conduct another simulation study with smaller sample size $n = 10^3$, and larger noise variance $\sigma_1 = .9, \sigma_2 = 1.2$. The signature expression values are $h_{10} = 30, h_{20} = 50$, and other parameters are generated the same way as the previous study. Every time, we can compute moment estimate of $(\hat{h}_1^{(i)}, \hat{h}_2^{(i)})$, together with the covariance matrix $\hat{\Sigma}^{(i)}$ from the limiting distribution of Theorem 3. Thus, we are able to construct the $(1-\alpha)$ confidence interval for both $\hat{h}_1^{(i)}$ and $\hat{h}_2^{(i)}$. On the other hand, bootstrap is also conducted, based on which, we can obtain another confidence interval. The coverage probability and confidence interval length for each approach are summarized in Table 2.12 and 2.13: The coverage probability for both methods are quite close to the pre-specified level $(1-\alpha)$. However, the confidence interval length we derive from Theorem 3 is shorter than that from bootstrap.

	Exact CI	estimated CI	Boostrap
99% Coverage Probability	0.971	0.967	0.983
99% CI Length	2.651	2.611	2.946
95% Coverage Probability	0.912	0.910	0.938
95% CI Length	2.021	1.991	2.252
90% Coverage Probability	0.844	0.839	0.886
90% CI Length	1.691	1.666	1.891

Table 2.12: Coverage Probability and CI length for $\hat{h}_1$, $n = 1000$, $\sigma_{10} = 0.9, \sigma_{10} = 1.2, h_{10} = 30, h_{20} = 50$

	Exact CI	estimated CI	Boostrap
99% Coverage Probability	0.972	0.962	0.982
99% CI Length	2.525	2.501	2.812
95% Coverage Probability	0.909	0.911	0.936
95% CI Length	1.926	1.907	2.151
90% Coverage Probability	0.842	0.847	0.886
90% CI Length	1.611	1.596	1.804

Table 2.13: Coverage Probability and CI length for $\hat{h}_2$,n $= 1000$,$\sigma_{10} = 0.9, \sigma_{10} = 1.2, h_{10} = 30, h_{20} = 50$

2.4.2 Moment Estimate when W is unknown constant and symmetric

In section 2.4.1, we have studied the asymptotic properties when $\boldsymbol{w}_{10}$ is asymmetric. In this section, we will investigate in the case where $\boldsymbol{w}_{10}$ is symmetric. $\boldsymbol{w}_{10}$ is considered as constant over this subsection 2.4.2

Consider the same model as Model (2.16). Assuming that $\boldsymbol{w}_{j0}, j = 1, 2$ is unknown constant vector, and if

$$\begin{aligned} &\lim_{n\to\infty} m_3(w) = \frac{\sum_{i=1}^{i=n}(w_{i(10)} - \bar{w}_1)^3}{n} = 0 a.s \\ &\lim_{n\to\infty} \bar{w}_{j0} > 0, j = 1, 2 a.s \\ &\lim_{n\to\infty} m_4(w) \neq 0 a.s \end{aligned} \tag{2.24}$$

Note that the zero assumption of $m_3(w)$ makes the third order moment in Equation (2.19) invalid as both sides of the equality are 0. Thus we need to consider higher order moment.

Solution of the moment estimate

Based on the first, second and fourth moment, we could get the following equalities:

$$\begin{cases} \mathrm{E}(\bar{g}_1) = h_{10}\bar{w}_1 \\ \mathrm{E}(\bar{g}_2) = h_{20}(1-\bar{w}_1) \\ \mathrm{E}(u_{11}) = \mathrm{E}(\hat{\mathrm{var}}(g_1)) = h_{10}^2 m_2(w) + \sigma_1^2 \\ \mathrm{E}(u_{22}) = \mathrm{E}(\hat{\mathrm{var}}(g_2)) = h_{20}^2 m_2(w) + \sigma_2^2 \\ \mathrm{E}(u_{12}) = \mathrm{E}(\hat{\mathrm{cov}}(g_1, g_2)) = -h_{10}h_{20}m_2(w)^2 \\ \mathrm{E}(u_{1112}) = \mathrm{E}(\frac{\sum_{i=1}^{i=n}(g_{i,1}-\bar{g}_1)^3(g_{i,2}-\bar{g}_2)}{n}) = -h_{10}^3 h_2 m_4(w) - 3h_{10}h_{20}m_2(w)\sigma_1^2 \\ \mathrm{E}(u_{1222}) = \mathrm{E}(\frac{\sum_{i=1}^{i=n}(g_{i,1}-\bar{g}_1)(g_{i,2}-\bar{g}_2)^3}{n}) = -h_{10}h_{20}^3 m_4(w) - 3h_{10}h_{20}m_2(w)\sigma_2^2 \end{cases} \tag{2.25}$$

It's not difficult to solve the equations:

$$\begin{cases} \hat{h}_1 & = \bar{g}_1 + \lambda_2\bar{g}_2 \\ \hat{h}_2 & = \bar{g}_2 + \frac{1}{\lambda_2}\bar{g}_1 \end{cases}$$

where $\lambda_2 = \sqrt{\frac{\hat{u}_{1112} - 3\hat{u}_{12}\hat{u}_{11}}{\hat{u}_{1222} - 3\hat{u}_{12}\hat{u}_{22}}}$

Asymptotic properties of the moment estimate

Theorem 4. *Consider the model in (2.16), assume (2.24) holds. In the special case when $\boldsymbol{w}_{10}$ is deterministic. Then*

$$\sqrt{n}(\begin{pmatrix} \hat{h}_1 \\ \hat{h}_2 \end{pmatrix} - \begin{pmatrix} h_{10} \\ h_{20} \end{pmatrix}) \overset{D}{\Longrightarrow} \mathcal{N}(\mathbf{0}, \Sigma_5)$$

Where $\Sigma_5 = \triangledown f'(\boldsymbol{u}_3)\Sigma\triangledown f(\boldsymbol{u}_3)$ with Σ and $\triangledown f(\boldsymbol{u}_3)$ defined as following:

$$\Sigma = \begin{pmatrix} \sigma_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ & \sigma_{22} & 0 & 0 & 0 & 0 & 0 \\ & & \sigma_{33} & 0 & \sigma_{35} & \sigma_{36}^{\star} & \sigma_{37}^{\star} \\ & & & \sigma_{44} & \sigma_{45} & \sigma_{46}^{\star} & \sigma_{47}^{\star} \\ & & & & \sigma_{55} & \sigma_{56}^{\star} & \sigma_{57}^{\star} \\ & & & & & \sigma_{66}^{\star} & \sigma_{67}^{\star} \\ & & & & & & \sigma_{77}^{\star} \end{pmatrix}$$

where σ_{ij} are same as listed in Eq. (2.33), except for those marked with $\star$. $\sigma_{ij}^{\star}$ are given as following

$$\begin{aligned}
\sigma_{66}^{\star} =& h_{10}^6 m_6(w)\sigma_2^2 + 15h_{20}^2 m_2(w)\sigma_1^6 + 15\sigma_1^6\sigma_2^2 + 9h_{10}^4 h_{20}^2 m_6(w)\sigma_1^2 + 9h_{10}^4 m_4(w)\sigma_1^2\sigma_2^2 + \\
& 18h_{10}^2 h_{20}^2 m_4(w)\sigma_1^4 + 27h_{10}^2 m_2(w)\sigma_1^4\sigma_2^2 + 3h_{10}^4 m_4(w)\sigma_1^2\sigma_2^2 + 9h_{10}^2 h_{20}^2 m_4(w)\sigma_1^2 + 9h_{10}^2 m_6(w)\sigma_1^4\sigma_2^2 \\
\sigma_{77}^{\star} =& h_{20}^6 m_6(w)\sigma_1^2 + 15h_{10}^2 m_6(w)\sigma_2^6 + 15\sigma_2^6\sigma_1^2 + 9h_{20}^4 h_{10}^2 m_6(w)\sigma_2^2 + 9h_{20}^4 m_4(w)\sigma_2^2\sigma_1^2 + \\
& 18h_{20}^2 h_{10}^2 m_4(w)\sigma_2^4 + 27h_{20}^2 m_2(w)\sigma_2^4\sigma_1^2 + 3h_{20}^4 m_4(w)\sigma_2^2\sigma_1^2 + 9h_{10}^2 h_{20}^2 m_4(w)\sigma_2^2 + 9h_{20}^2 m_2(w)\sigma_2^4\sigma_1^2 \\
\sigma_{36}^{\star} =& -12h_{10}h_{20}\sigma_1^4 m_2(w) - 6h_{10}^3 h_{20} m_4(w)\sigma_1^2 \\
\sigma_{37}^{\star} =& -2h_{10}h_{20}^3 m_4(w)\sigma_1^2 \\
\sigma_{46}^{\star} =& -2h_{10}^3 h_{20} m_4(w)\sigma_2^2 \\
\sigma_{47}^{\star} =& -12h_{10}h_{20}\sigma_2^4 m_2(w) - 6h_{20}^3 h_{10} m_4(w)\sigma_2^2 \\
\sigma_{56}^{\star} =& h_{10}^4 m_4(w)\sigma_2^2 + 3h_{10}^2 m_2(w)\sigma_1^2\sigma_2^2 + 3h_{20}^2 m_2(w)\sigma_1^4 + 3h_{20}^2 h_{10}^2 m_4(w)\sigma_1^2 + 3\sigma_2^2\sigma_1^4 + 3h_{10}^2 m_2(w)\sigma_2^2\sigma_1^2 \\
\sigma_{57}^{\star} =& h_{20}^4 m_4(w)\sigma_1^2 + 3h_{20}^2 m_2(w)\sigma_1^2\sigma_2^2 + 3h_{10}^2 m_2(w)\sigma_2^4 + 3h_{10}^2 h_{20}^2 m_4(w)\sigma_2^2 + 3\sigma_1^2\sigma_2^4 + 3h_{20}^2 m_2(w)\sigma_1^2\sigma_2^2 \\
\sigma_{67}^{\star} =& 3h_{10}^4 m_4(w)\sigma_2^4 + 3h_{10}^4 h_{20}^2 m_6(w)\sigma_2^2 + 3h_{20}^4 m_4(w)\sigma_1^4 + 9h_{20}^2 m_2(w)\sigma_1^4\sigma_2^2 + 9\sigma_1^4\sigma_2^4 + 9h_{20}^2 m_2(w)\sigma_1^4\sigma_2^2 + \\
& 3h_{10}^2 h_2^4 m_6(w)\sigma_1^2 + 9h_{10}^2 h_2^2 m_4(w)\sigma_1^2\sigma_2^2 + 9h_{10}^2 m_2(w)\sigma_1^2\sigma_2^4 + 9h_{10}^2 h_{20}^2 m_4(w)\sigma_1^2\sigma_2^2 + 9h_{10}^2 m_2(w)\sigma_1^2\sigma_2^4
\end{aligned}$$

And $\nabla f(\boldsymbol{u}_3)$ *is:*

$$\nabla f(\boldsymbol{u}_3) = \begin{pmatrix} 1 & \frac{h_{20}}{h_{10}} \\ \frac{h_{10}}{h_{20}} & 1 \\ \frac{3m_2(w)\bar{w}_2}{2h_{10}|-m_4(w)+3m_2^2(w)|} & -\frac{3m_2(w)h_{20}\bar{w}_1}{2h_{10}^2|-m_4(w)+3m_2^2(w)|} \\ -\frac{3\bar{w}_2 h_{10} m_2(w)}{2h_{20}^2(-m_4(w)+3m_2^2(w))} & \frac{3\bar{w}_1 m_2(w)}{2h_{20}(-m_4(w)+3m_2^2(w))} \\ \frac{3\bar{w}_2(h_{10}^2\sigma_{20}^2 - h_2^2\sigma_{10}^2)}{2h_{10}^2 h_2^3(3m_2(w)^2 - m_4(w))} & \frac{-3\bar{w}_1(h_{10}^2\sigma_{20}^2 - h_{20}^2\sigma_{10}^2)}{2h_{10}^3 h_{20}^2(3m_2^2(w) - m_4(w))} \\ \frac{\bar{w}_2}{2h_{10}^2 h_{20}|-m_4(w)+3m_2^2(w)|} & -\frac{\bar{w}_1}{2h_{10}^3|-m_4(w)+3m_2^2(w)|} \\ -\frac{\bar{w}_2}{2h_{20}^3(-m_4(w)+3m_2^2(w))} & \frac{\bar{w}_1}{2h_{10}h_{20}^2(-m_4(w)+3m_2^2(w))} \end{pmatrix} \tag{2.26}$$

2.5 Proof of the theorems

2.5.1 Proof of Theorem 1

Proof of (1) in Theorem 1

Proof. Let $\boldsymbol{w}_{10} = x_{10}\mathbf{1}_n + x_{20}\boldsymbol{\eta}_{10}$, where $\boldsymbol{\eta}_{10} \perp \mathbf{1}_n, \|\boldsymbol{\eta}_{10}\| = n$.

and the coefficients are

$$x_{10} = \frac{\mathbf{1}_n^{'}\boldsymbol{w}_{10}}{n}, x_{20} = \sqrt{\frac{\boldsymbol{w}_{10}^{'}\boldsymbol{w}_{10} - n||x_{10}||^2}{n}}.$$

Consider new orthogonal basis in R^n:

$$\{\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3, ..., \boldsymbol{z}_n\},$$

where $\boldsymbol{z}_1 = \frac{\mathbf{1}_n}{\sqrt{n}}, \boldsymbol{z}_2 = \frac{\boldsymbol{\eta}_{10}}{\sqrt{n}}$.

Let $\{(\boldsymbol{w}_1, \boldsymbol{w}_2) \mid \boldsymbol{w}_2 = \mathbf{1}_n - \boldsymbol{w}_1\}$ be the colloction of all possible solution that satisfy Eq. (2.7). Now we will project $\boldsymbol{w}_{10}, \boldsymbol{w}_{20}, \boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \boldsymbol{w}_1, \boldsymbol{w}_2$ on the new orthogonal basis.

$$\boldsymbol{w}_{10} = x_{10}\mathbf{1}_n + x_{20}\boldsymbol{\eta}_{10}$$

$$\boldsymbol{w}_{20} = (1 - x_{10})\mathbf{1}_n - x_{20}\boldsymbol{\eta}_{10}$$

$$\boldsymbol{\epsilon}_1 = e_{11}\mathbf{1}_n + e_{12}\boldsymbol{\eta}_{10} + \boldsymbol{\zeta}_1$$

$$\boldsymbol{\epsilon}_2 = e_{21}\mathbf{1}_n + e_{22}\boldsymbol{\eta}_{10} + \boldsymbol{\zeta}_2$$

$$\boldsymbol{w}_1 = x_1\mathbf{1}_n + x_2\boldsymbol{\eta}_{10} + \boldsymbol{\zeta}_w$$

$$\boldsymbol{w}_2 = (1 - x_1)\mathbf{1}_n - x_2\boldsymbol{\eta}_{10} + \boldsymbol{\zeta}_w$$

By Central Limit Theorem, there are some interesting observations on the coefficient. Proof of these observations are provided in Remark 11.

1. Observation 1: For any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|e_{11}| \leq \epsilon) = 1, \lim_{n\to\infty} P(|e_{21}| \leq \epsilon) = 1$$
$$\lim_{n\to\infty} P(|e_{12}| \leq \epsilon) = 1, \lim_{n\to\infty} P(|e_{22}| \leq \epsilon) = 1$$

2. Observation 2: For any $\epsilon > 0$,

$$\lim_{n\to\infty} P(\big|\|\boldsymbol{\zeta}_1\| - (n-2)\sigma^2\big| \leq \epsilon) = 1, \lim_{n\to\infty} P(\big|\|\boldsymbol{\zeta}_2\| - (n-2)\sigma^2\big| \leq \epsilon) = 1$$

3. Observation 3: For any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\frac{\boldsymbol{\zeta}_1'\boldsymbol{\zeta}_2}{(n-2)\sigma^2}| \leq \epsilon) = 1$$

Observation 2 and 3 state that if sample size n is large enough, the length of $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ will be $(n-2)\sigma^2$, and that they will be orthogonal to each other.

With the representations in new orthogonal space, the original problem (2.8) could be rewritten as

$$\underset{x_1,x_2,h_1,h_2,\boldsymbol{\zeta}_w}{\arg\min} \|h_{10}(x_{10}\mathbf{1}_n + x_{20}\boldsymbol{\eta}_{10}) + e_{11}\mathbf{1}_n + e_{12}\boldsymbol{\eta}_{10} + \boldsymbol{\zeta}_1 - h_1(x_1\mathbf{1}_n + x_2\boldsymbol{\eta}_{10} + \boldsymbol{\zeta}_w)\|^2$$

$$+ \|h_{20}\left((1 - x_{10})\mathbf{1}_n - x_{20}\boldsymbol{\eta}_{10}\right) + e_{21}\mathbf{1}_n + e_{22}\boldsymbol{\eta}_{10} + \boldsymbol{\zeta}_2 - h_2\left((1 - x_1)\mathbf{1}_n - x_2\boldsymbol{\eta}_{10} - \boldsymbol{\eta}_w\right)\|^2$$

$$\Longleftrightarrow$$

$$\underset{h_1,h_2}{\arg} \quad \underset{x_1,x_2,\boldsymbol{\zeta}_w}{\arg} \min\{||h_{10}x_{10}\mathbf{1}_n + e_{11}\mathbf{1}_n - h_1x_1\mathbf{1}_n||^2 + ||h_{10}x_{20}\boldsymbol{\eta}_{10} + e_{12}\boldsymbol{\eta}_{10} - h_1x_2\boldsymbol{\eta}_{10}||^2$$

$$+ ||\boldsymbol{\zeta}_1 - h_1\boldsymbol{\zeta}_w||^2 + ||h_{20}(1 - x_{10})\mathbf{1}_n + e_{21}\mathbf{1}_n - h_2(1 - x_1)\mathbf{1}_n||^2$$

$$+ || - h_{20}x_{20}\boldsymbol{\eta}_{10} + h_2x_2\boldsymbol{\eta}_{10} + e_{22}\boldsymbol{\eta}_{10}|| + ||\boldsymbol{\zeta}_2 + h_2\boldsymbol{\zeta}_w||^2\}$$

Now, we are going to show that:

$$\underset{h_1,h_2,\boldsymbol{\zeta}_w}{\arg} \min\{\|\boldsymbol{\zeta}_2 + h_2\boldsymbol{\zeta}_w\|^2 + \|\boldsymbol{\zeta}_1 - h_1\boldsymbol{\zeta}_w\|^2\} = (n - 2)\sigma^2$$

which is independent of the choice of h_1, h_2

For any fixed h_1, h_2, take derivative with respect of $\boldsymbol{\zeta}_w$, it's easy to get

$$\hat{\boldsymbol{\zeta}}_w = \frac{h_1\boldsymbol{\zeta}_1 - h_2\boldsymbol{\zeta}_2}{h_1^2 + h_2^2} \tag{2.27}$$

Plugging in $\hat{\boldsymbol{\zeta}}_w$, together with the Observation 2 and 3, for any $\epsilon > 0$, let $\epsilon^\star = \dfrac{h_1^2 + h_2^2}{(h_1 + h_2)^2}\epsilon$, we have:

$$\begin{aligned}
&\underset{h_1,h_2}{\arg} \ \underset{\boldsymbol{\zeta}_w}{\arg\min}\{||\boldsymbol{\zeta}_2 + h_2\boldsymbol{\zeta}_w||^2 + ||\boldsymbol{\zeta}_1 - h_1\boldsymbol{\zeta}_w||^2\}\\
=&\underset{h_1,h_2}{\arg} \min\{||\boldsymbol{\zeta}_2 + h_2\frac{h_1\boldsymbol{\zeta}_1 - h_2\boldsymbol{\zeta}_2}{h_1^2 + h_2^2}||^2 + ||\boldsymbol{\zeta}_1 - h_1\frac{h_1\boldsymbol{\zeta}_1 - h_2\boldsymbol{\zeta}_2}{h_1^2 + h_2^2}||^2\}\\
=&\underset{h_1,h_2}{\arg} \min\{\frac{(h_1^4 + h_1^2h_2^2)||\boldsymbol{\zeta}_2||^2 + (h_1^2h_2^2 + h_2^4)||\boldsymbol{\zeta}_1||^2 + (2h_1^3h_2 + 2h_1h_2^3)\boldsymbol{\zeta}_1^{'}\boldsymbol{\zeta}_2}{h_1^2 + h_2^2}\}\\
\geq&\frac{(h_1^4 + h_2^4 + 2h_1^2h_2^2)((n - 2)\sigma^2 - \epsilon^\star) + (2h_1^3h_2 + 2h_2^3h_1)(-\epsilon^\star)}{h_1^2 + h_2^2}\\
=&(n - 2)\sigma_2 + \frac{(h_1 + h_2)^2}{h_1^2 - h_2^2}\epsilon^\star\\
=&(n - 2)\sigma_2 - \epsilon
\end{aligned} \tag{2.28}$$

Since $\|\mathbf{1}_n\|^2 = \|\boldsymbol{\eta}_{10}\|^2 = n$, together with Observation 1 and Eq. (2.28), now the original problem could be further simplified to:

$$\arg\min_{x_1,x_2,h_1,h_2}\{n||h_{10}x_{10} - h_1x_1 + o_P(n)||^2 + n||h_{10}x_{20} - h_1x_2 + o_P(n)||^2 +$$
$$n||h_{20}(1 - x_{10}) - h_2(1 - x_1) + o_P(n)||^2 + n|| - h_{20}x_{20} + h_2x_2 + o_P(n)||^2\} + (n - 2)\sigma^2$$
$$\geq (n - 2)\sigma^2$$

The equality holds if and only if

$$h_{10}x_{10} = h_1x_1$$
$$h_{10}x_{20} = h_1x_2$$
$$h_{20}(1 - x_{10}) = h_2(1 - x_1)$$
$$h_{20}x_{20} = h_2x_2$$

Therefore, $(\hat{h}_1, \hat{h}_2) \stackrel{n\to\infty}{\longrightarrow} (h_{10}, h_{20})$ almost surely. □

Remark 10. *Fig. 2.6 provides an intuitive way to understand why Eq. (2.28) holds. For any* h_1, h_2*, the vector that minimized the* $\{\|\boldsymbol{\zeta}_2 + h_2\boldsymbol{\zeta}_w\|^2 + \|\boldsymbol{\zeta}_1 - h_1\boldsymbol{\zeta}_w\|^2\}$ *could be computed from Eq. (2.27). As we already observed that* $\boldsymbol{\zeta}_1 \perp \boldsymbol{\zeta}_2$ *(black solid line), it's easy to see that the red(or blue) triangles are congruent. Therefore the the minimum value is just the length of* $\|\boldsymbol{\zeta}_1\|^2$*(or* $\|\boldsymbol{\zeta}_2\|^2$*), which is* $(n - 2)\sigma^2$

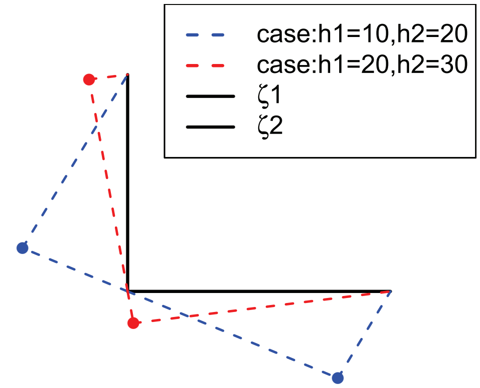


Figure 2.6: Geometric interpretation of $\arg_{h_1,h_2,\boldsymbol{\zeta}_w} \min\{||\boldsymbol{\zeta}_2 + h_2\boldsymbol{\zeta}_w||^2 + ||\boldsymbol{\zeta}_1 - h_1\boldsymbol{\zeta}_w||^2\}$, black solid line are $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_1$, red(blue) dashed line connected by solid circle is $\hat{\boldsymbol{\zeta}}_w$ from formula (2.27) by different h_1, h_2 values

Remark 11. *A short sketch of proof for Observation 1, 2 and 3 are provided below:*

First, let's discuss why Observation 1 holds.

$$\boldsymbol{\epsilon}_1 = \sqrt{n}e_{11}\frac{\mathbf{1}_n}{\sqrt{n}} + \sqrt{n}e_{21}\frac{\boldsymbol{\eta}_{10}}{\sqrt{n}} + e_{31}\boldsymbol{z}_3 + ... + e_{n1}\boldsymbol{z}_n,$$

where $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n\}$ *is the orthogonal basis defined above.*

Note that $\|\boldsymbol{\epsilon}_1\|^2 = n\sigma^2$*, by CLT*

$$\sqrt{n}e_{11} = \frac{\boldsymbol{\epsilon}_1^{'}\mathbf{1}_n}{\sqrt{n}} = \frac{\sum_{i=1}^{i=n}\epsilon_{i,1}}{\sqrt{n}} = \sqrt{n}\bar{\epsilon}_1 \sim N(0,\sigma^2)$$

Therefore, for any $\epsilon > 0$*,* $P(e_{11} > \epsilon) = P(\frac{\mathcal{N}(0,\sigma^2)}{\sqrt{n}} > \epsilon) \stackrel{n\to\infty}{\longrightarrow} 0$ *in probability.*

By the same logic, we could get $P(e_{21} > \epsilon) \stackrel{n\to\infty}{\longrightarrow} 0$

For e_{12} *and* e_{22}*, by CLT*

$$\sqrt{n}e_{21} = \frac{\boldsymbol{\epsilon}_1^{'}\boldsymbol{\eta}_{10}}{\sqrt{n}} = \frac{\sum_{i=1}^{i=n}\epsilon_{i,1}\eta_{10,i}}{\sqrt{n}} \sim \mathcal{N}(0,\sigma^2)$$

Therefore, for any $\epsilon > 0$*,* $P(e_{21} > \epsilon) = P(\frac{\mathcal{N}(0,\sigma^2)}{\sqrt{n}} > \epsilon) \stackrel{n\to\infty}{\longrightarrow} 0$.

With the same argument, we could get $P(e_{22} > \epsilon) \stackrel{n\to\infty}{\to} 0$.

As to Observation 2, it's easy to see $\|\boldsymbol{\zeta}_1\| = \|\boldsymbol{\zeta}_2\| = (n-2)\sigma^2$

And Observation 3 is implied by Chebyshev Inequality:

$$P(|\frac{\boldsymbol{\zeta}_1'\boldsymbol{\zeta}_2}{(n-2)\sigma^2}| > \epsilon) \leq \frac{E(\|\boldsymbol{\zeta}_1'\boldsymbol{\zeta}_2\|^2)}{(n-2)^2\sigma^4\epsilon^2} = 0$$

Thus, for individual $\hat{w}_{i,1}$, the estimate might not be reliable due to the two noise term. But by CLT, $\frac{\mathbf{1}'\hat{\boldsymbol{w}}_1}{n}$ is still consistent estimate for the population mean.

Alternative Proof of (1) in Theorem 1

Proof. We provide another way to prove Theorem 1 which concentrates on computation. As a side product, we can get the closed form of the ILS estimate. Let $\mathbf{g}_1 = \bar{g}_1\mathbf{1}_n + \boldsymbol{\theta}_1, \mathbf{g}_2 = \bar{g}_2\mathbf{1}_n + \boldsymbol{\theta}_2, \mathbf{w}_1 = \bar{w}_1\mathbf{1}_n + \boldsymbol{\gamma}_1$, where $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\gamma}_1$ is orthogonal to $\mathbf{1}_n$, thus the original problem could be rewritten as follows:

$$\begin{aligned}
&\underset{h_1,h_2,\mathbf{w}_1}{\arg}\ \min\{\|\mathbf{g}_1 - h_1\mathbf{w}_1\|^2 + \|\mathbf{g}_2 - h_2(\mathbf{1}_n - \mathbf{w}_1)\|^2\} \\
&= \underset{h_1,h_2,\boldsymbol{\gamma}_1}{\arg}\ \min\{||\bar{g}_1\mathbf{1}_n + \boldsymbol{\theta}_1 - h_1(\bar{w}_1\mathbf{1}_n + \boldsymbol{\gamma}_1)||^2 + \|\bar{g}_2\mathbf{1}_n + \boldsymbol{\theta}_2 - h_2((1-\bar{w}_1)\mathbf{1}_n - \boldsymbol{\gamma}_1)\|^2\} \\
&= \underset{h_1,h_2,\bar{w},\boldsymbol{\gamma}_1}{\arg}\ \min\{(\bar{g}_1 - h_1\bar{w}_1)^2 n + ||\boldsymbol{\theta}_1 - h_1\boldsymbol{\gamma}_1||^2 + ||g_2 - h_2(1-\bar{w}_1)||^2 n + ||\boldsymbol{\theta}_2 + h_2\boldsymbol{\gamma}_1||^2\} \\
&= \underbrace{\underset{h_1,h_2}{\arg}\{\arg_{\bar{w}_1}\min\{(\bar{g}_1 - h_1\bar{w})^2 n + (\bar{g}_2 - h_2(1-\bar{w}_1))^2 n\} +}_{\text{I}} \\
&\quad \underbrace{\underset{\boldsymbol{\gamma}_1}{\arg\min}\{\|\boldsymbol{\theta}_1 - h_1\boldsymbol{\gamma}_1\|^2 + \| - \boldsymbol{\theta}_2 - h_2\boldsymbol{\gamma}_1\|^2\}\}}_{\text{II}}
\end{aligned}$$

For any fixed h_1, h_2, the second term is minimized when $\hat{\boldsymbol{\gamma}}_1 = \frac{h_1\boldsymbol{\theta}_1 - h_2\boldsymbol{\theta}_2}{h_1^2 + h_2^2}$, plug in $\hat{\boldsymbol{\gamma}}_1$

$$\begin{aligned}
&\underset{h_1,h_2}{\arg}\ \underset{\boldsymbol{\gamma}_1}{\arg}\min\{||\boldsymbol{\theta}_1 - h_1\boldsymbol{\gamma}_1||^2 + || - \boldsymbol{\theta}_1 - h_2\boldsymbol{\gamma}_1||^2\} \\
&= \underset{h_1,h_2}{\arg}\ \min\{||\frac{h_1^2\boldsymbol{\theta}_1 + h_2^2\boldsymbol{\theta}_1 - h_1^2\boldsymbol{\theta}_1 + h_1h_2\boldsymbol{\theta}_2}{h_1^2 + h_2^2}||^2 + ||\frac{h_1^2\boldsymbol{\theta}_2 + h_2^2\boldsymbol{\theta}_2 - h_2^2\boldsymbol{\theta}_2 + h_1h_2\boldsymbol{\theta}_1}{h_1^2 + h_2^2}||^2\} \\
&= \arg\min_{h_1,h_2}\{\frac{h_2^4\boldsymbol{\theta}_1'\boldsymbol{\theta}_1 + h_1^2h_2^2\boldsymbol{\theta}_2'\boldsymbol{\theta}_2 + 2h_1h_2^3\boldsymbol{\theta}_2'\boldsymbol{\theta}_1 + h_1^4\boldsymbol{\theta}_2'\boldsymbol{\theta}_2 + h_1^2h_2^2\boldsymbol{\theta}_1'\boldsymbol{\theta}_1 + 2h_1^3h_2\boldsymbol{\theta}_1'\boldsymbol{\theta}_2}{(h_1^2 + h_2^2)^2}\} \\
&= \arg\min_{c=\frac{h1}{h2}}\{\frac{||\boldsymbol{\theta}_1 + c\boldsymbol{\theta}_2||^2}{1 + c^2}\}
\end{aligned}$$

Furthermore, for part I, we will show that its low bound is 0.

$$\begin{aligned}
&\arg_{h_1,h_2,\bar{w}_1} \min\{(\bar{g}_1 - h_1\bar{w})^2 n + (\bar{g}_2 - h_2(1-\bar{w}_1))^2 n\} \\
&= \arg_{c=\frac{h_1}{h_2}} \arg_{h_1,\bar{w}_1} \{(\bar{g}_1 - h_1\bar{w})^2 n + (\bar{g}_2 - \frac{h_1}{c}(1-\bar{w}_1))^2 n\}
\end{aligned}$$

For any fixed c, part I is minimized to 0 when

$$\begin{cases} \bar{w}_1 & = \dfrac{\bar{g}_1}{\bar{g}_1 + c\bar{g}_2} \\ \hat{h}_1 & = \bar{g}_1 + c\bar{g}_2 \end{cases}$$

Thus the original problem now becomes (it can be solved analytically):

$$\arg_{c=\frac{h1}{h2}} \min \frac{(\boldsymbol{\theta}_1 + c\boldsymbol{\theta}_2)^2}{1+c^2}$$

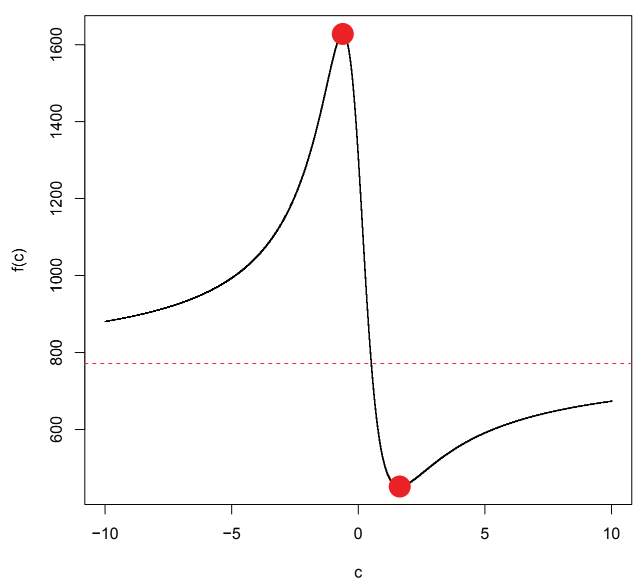


Figure 2.7: plot of $f(c) = \dfrac{(\theta_1 + c\theta_2)^2}{1+c^2}$,with Asymptote as $y = \theta_2'\theta_2$

$$\frac{\partial f(c)}{\partial c} = \frac{(2\boldsymbol{\theta}_2'\boldsymbol{\theta}_2 c + 2\boldsymbol{\theta}_1'\boldsymbol{\theta}_2)(1+c^2) - 2c(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1 + c^2\boldsymbol{\theta}_2'\boldsymbol{\theta}_2 + 2c\boldsymbol{\theta}_1'\boldsymbol{\theta}_2)}{(1+c^2)^2} = 0$$

$$c_{10} = \frac{b+\sqrt{b^2+4}}{2}, c_{20} = \frac{b-\sqrt{b^2+4}}{2},$$

where $b = \dfrac{\boldsymbol{\theta}_2'\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1'\boldsymbol{\theta}_1}{\boldsymbol{\theta}_1'\boldsymbol{\theta}_2}$, and c_{20} is maximum, while c_{10} is minimum.

As a side product, we get the explicit formula of the least square (LS) estimate:

$$\begin{cases} \bar{w}_1 &= \dfrac{\bar{g}_1}{\bar{g}_1 + c_{10}\bar{g}_2} \\ \hat{h}_1 &= \bar{g}_1 + c_{10}\bar{g}_2 \\ \hat{h}_2 &= \bar{g}_2 + \dfrac{\bar{g}_1}{c_{10}} \\ \hat{\mathbf{w}} &= \bar{w}_1\mathbf{1} + \hat{\boldsymbol{\gamma}}_1 = \dfrac{\bar{g}_1}{\bar{g}_1 + c_{10}\bar{g}_2}\mathbf{1} + \dfrac{\hat{h}_1\boldsymbol{\theta}_1 - \hat{h}_2\boldsymbol{\theta}_2}{\hat{h}_1^2 + \hat{h}_2^2} \end{cases} \tag{2.29}$$

Since $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ are observed(just centered g_1, g_2), the formula is explicit.

As we already shown in Section 2.3 that moment estimate is consistent. Therefore, the consistency of ILS of is shown.

Remark 12. *Compared to the moment estimate in Eq. 2.14, the formulas are exact the same. This means that ILS and moment estimate are equivalent when the error term share common variance. In addition, from the above formula, it's not difficult to see that the LS estimate is biased if the homoscedastic assumption is violated.*

□

Proof of (2) in Theorem 1

Proof. As we already show from Remark 5, $\hat{\boldsymbol{w}}_1 = \boldsymbol{w}_{10} + \dfrac{\hat{h}_1\boldsymbol{\epsilon}_1 - \hat{h}_2\boldsymbol{\epsilon}_2}{\hat{h}_1^2 + \hat{h}_2^2}$, plug $\hat{\boldsymbol{w}}$ in Eq (2.9), we get

$$
\begin{aligned}
\hat{\sigma}^2 =& \frac{1}{n}(\|h_{10}\boldsymbol{w}_{10} + \boldsymbol{\epsilon}_1 - \hat{h}_1(\boldsymbol{w}_{10} + \frac{\hat{h}_1\boldsymbol{\epsilon}_1 - \hat{h}_2\boldsymbol{\epsilon}_2}{\hat{h}_1^2 + \hat{h}_2^2})\|^2 \\
&+ \|h_{20}(\mathbf{1} - \boldsymbol{w}_{10}) + \boldsymbol{\epsilon}_2 - \hat{h}_2(\mathbf{1} - \boldsymbol{w}_{10} - \frac{\hat{h}_1\boldsymbol{\epsilon}_1 - \hat{h}_2\boldsymbol{\epsilon}_2}{\hat{h}_1^2 + \hat{h}_2^2})\|^2) \\
=& \frac{1}{n}(\|(h_{10} - \hat{h}_1)\boldsymbol{w}_{10} + \frac{\hat{h}_2^2\boldsymbol{\epsilon}_1 + \hat{h}_1\hat{h}_2\boldsymbol{\epsilon}_2}{\hat{h}_1^2 + \hat{h}_2^2}\|^2 + \\
& \| - (h_{20} - \hat{h}_2)\boldsymbol{w}_{10} + \frac{\hat{h}_1^2\boldsymbol{\epsilon}_2 + \hat{h}_1\hat{h}_2\boldsymbol{\epsilon}_1}{\hat{h}_1^2 + \hat{h}_2^2}\|^2) \\
=& \frac{1}{n}(\|(h_{10} - \hat{h}_1)\boldsymbol{w}_{10}\|^2 + \| - (h_{20} - \hat{h}_2)\boldsymbol{w}_{10}\|^2) + \sigma^2
\end{aligned}
$$

As we already show that for any $\epsilon > 0$,

$$
P(\|\hat{h}_1 - h_{10}\| < \epsilon/2/\mu(w)) = 1, P(\|\hat{h}_2 - h_{20}\| < \epsilon/2/\mu(w)) = 1
$$

Therefore,

$$
\begin{aligned}
P(\|\hat{\sigma}^2 - \sigma^2\| < \epsilon) &= P(\frac{1}{n}(\|(h_{10} - \hat{h}_1)\boldsymbol{w}_{10}\|^2 + \| - (h_{20} - \hat{h}_2)\boldsymbol{w}_{10}\|^2) < \epsilon) \\
&= P(\bar{w}_1\|h_{10} - \hat{h}_1\| + \bar{w}_1\|h_{20} - \hat{h}_2\| < \epsilon) \\
&= 1
\end{aligned}
$$

□

Proof of (3) in Theorem 1

Proof. The proof is shown by equivalence between moment estimate and LS. Details is provided in Section 2.3. □

2.5.2 Proof of Theorem 2

Proof. Similar as the proof of Theorem 1, let

$$\boldsymbol{w}_{10} = x_{10}\mathbf{1}_n + x_{11}\boldsymbol{\eta}_{10}, \text{ where } \boldsymbol{\eta}_{10} \perp \mathbf{1}_n \text{ and } \|\boldsymbol{\eta}_{10}\| = n$$

$$\boldsymbol{w}_{20} = x_{20}\mathbf{1}_n + x_{21}\boldsymbol{\eta}_{10} + x_{22}\boldsymbol{\eta}_{20}, \text{ where } \boldsymbol{\eta}_{20} \perp \text{Span}\{\mathbf{1}_n, \boldsymbol{\eta}_{10}\} \text{ and } \|\boldsymbol{\eta}_{20}\| = n$$

$$\ldots$$

$$\boldsymbol{w}_{(K-1)0} = x_{(K-1)0}\mathbf{1}_n + \sum_{i=1}^{K-1} x_{(K-1)i}\boldsymbol{\eta}_{i0}, \text{ where } \boldsymbol{\eta}_{(K-1)0} \perp \text{Span}\{\mathbf{1}_n, \ldots, \boldsymbol{\eta}_{(K-2)0}\} \text{ and } \|\boldsymbol{\eta}_{(K-1)0}\| = n$$

$$\boldsymbol{w}_{K0} = (1 - \sum_{i=1}^{i=K-1} x_i)\mathbf{1}_n - \sum_{i=1}^{K-1} x_{(K-1)i}\boldsymbol{\eta}_{i0}$$

$$\boldsymbol{\epsilon}_k = e_{k,0}\mathbf{1}_n + \sum_{j=1}^{k=K} e_{k,j}\boldsymbol{\eta}_{j0} + \boldsymbol{\zeta}_k, \text{ where } k = 1, \ldots, K$$

Let (ϵ_n) be a sequence of positive numbers which converges to 0 slowly, we will have similar observations:

$$P(|e_{i,k}| > \epsilon_n) = 0 \text{ where } i = 1, \ldots, K \text{ and } k = 1, \ldots, K$$

$$P(|\boldsymbol{\zeta}_i - (n-K)\sigma^2| > \epsilon_n) = 0 \text{ where } i = 1, \ldots, K$$

$$P(|\frac{\boldsymbol{\zeta}_i^{'}\boldsymbol{\zeta}_j^{'}}{(n-K)\sigma^2}| > \epsilon_n) = 0 \text{ where } i \neq j$$

Note that $\{\mathbf{1}_n, \boldsymbol{\eta}_{10}, \ldots, \boldsymbol{\eta}_{(K-1)0}\}$ is orthogonal space spanned by $\{\mathbf{1}_n, \boldsymbol{w}_{10}, \ldots, \boldsymbol{w}_{(K-1)0}\}$ with $\|\boldsymbol{\eta}_{j0}\| = n,$ where $j = 1, \ldots, K-1$

Represent $\boldsymbol{w}_i$ with the new basis $\{\mathbf{1}_n, \boldsymbol{\eta}_{10}, \ldots, \boldsymbol{\eta}_{(K-1)0}\}$:

$$\boldsymbol{w}_i = y_{i0}\mathbf{1}_n + \sum_{j=1}^{j=K-1} y_{ij}\boldsymbol{\eta}_{j0} + \boldsymbol{\eta}_{w_i}, \text{ where } i = 1, 2, \ldots, K-1$$

$$\boldsymbol{w}_K = (1 - y_{i0})\mathbf{1}_n - \sum_{j=1}^{j=K-1} y_{ij}\boldsymbol{\eta}_{j0} - \sum_{j=1}^{j=K-1} \boldsymbol{\eta}_{w_i}$$

Plug $\boldsymbol{w}_i, \boldsymbol{w}_{i0}, \boldsymbol{\epsilon}_i, i = 1, \ldots, K$ in the objective function (2.30):

$$
\begin{aligned}
&\arg_{(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_K,h_1,\ldots,h_K)} \min \sum_{i=1}^{k} \|\boldsymbol{g}_i - h_i \boldsymbol{w}_i\|^2 \\
\Longleftrightarrow &\arg_{(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_K,h_1,\ldots,h_K)} \min \sum_{i=1}^{k} \|h_{i0}\boldsymbol{w}_{i0} + \boldsymbol{\epsilon}_i - h_i \boldsymbol{w}_i\|^2
\end{aligned} \tag{2.30}
$$

With similar arguments, we can show that:

$$
\begin{aligned}
&\arg_{\boldsymbol{\eta}_{w_1},\ldots,\boldsymbol{\eta}_{w_{K-1}}} \min\{\sum_{i=1}^{i=K} \|\boldsymbol{\zeta}_i - h_i \boldsymbol{\eta}_{w_i}\|^2 + \|\boldsymbol{\zeta}_K + h_K \sum_{j=1}^{j=K-1} \boldsymbol{\eta}_{w_i}\|^2\} \\
&= (n-K)\sigma^2
\end{aligned}
$$

which is independent of the choice of $h_1, h_2, \ldots, h_K$. With exact same arguments, we could show that

$$
(\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_K) \overset{n\to\infty}{\to} (h_{10}, h_{20}, \ldots, h_{K0})
$$

And

$$
\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{K} \|\boldsymbol{g_j} - \hat{h}_j \hat{\boldsymbol{w}}_j\|^2
$$

is consistent estimate of σ^2

$\square$

2.5.3 Proof of Theorem 3

Proof. Define $\boldsymbol{v}_{20} = \begin{pmatrix} \bar{g}_1 \\ \bar{g}_2 \\ \hat{\mathrm{Var}}(g_1) \\ \hat{\mathrm{Var}}(g_2) \\ \hat{\mathrm{cov}}(g_1, g_2) \\ \hat{u}_{112} \\ \hat{u}_{122} \end{pmatrix}$ $\boldsymbol{u}_{20} = \begin{pmatrix} h_{10}u_w \\ h_{20}(1-u_w) \\ h_1^2\sigma_w^2 + \sigma_1^2 \\ h_{20}^2\sigma_w^2 + \sigma_2^2 \\ -h_{10}h_{20}\sigma_w^2 \\ u_3(3)h_{10}h_{20}^2 \\ -u_3(3)h_{10}^2h_{20} \end{pmatrix}$

We will divide the proof into two steps. In the first step, we will show the limiting distribution of $\boldsymbol{v}_{20}^T$ by multidimensional CLT Greene (2002). In the second step, we will show the asymptotic results of $(\hat{h}_1, \hat{h}_2)$ by multivariate delta method van der Vaart (1998).

Step 1:

By multidimensional CLT, we could get the joint distribution of $\boldsymbol{v}_{20}$. The computation of the covariance matrix is standard, thus we omit the details. The joint distribution of $\boldsymbol{v}_{20}$ is:

$$\sqrt{n}(\boldsymbol{v}_{20} - \boldsymbol{u}_{20}) \overset{D}{\Longrightarrow} \mathcal{N}(0, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} & \sigma_{17} \\ & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} & \sigma_{27} \\ & & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} & \sigma_{37} \\ & & & \sigma_{44} & \sigma_{45} & \sigma_{46} & \sigma_{47} \\ & & & & \sigma_{55} & \sigma_{56} & \sigma_{57} \\ & & & & & \sigma_{66} & \sigma_{67} \\ & & & & & & \sigma_{77} \end{pmatrix})$$

where σ_{ij} is defined in Eq. (2.21).

Step 2:

As shown in Eq. (2.19) $\hat{h}_1, \hat{h}_2$ are functions of $\boldsymbol{v}_2$. Specifically, they are:

$$f(\boldsymbol{v}_2) = \begin{pmatrix} \hat{h}_1 \\ \hat{h}_2 \end{pmatrix} = \begin{pmatrix} \bar{g}_1 + \lambda \bar{g}_2 \\ \bar{g}_2 + \frac{1}{\lambda}\bar{g}_1 \end{pmatrix}$$

By Multivariate Delta Method, we have:

$$\begin{aligned}
&\sqrt{n}(f(\boldsymbol{v}_{20}) - f(\boldsymbol{u}_{20})) \\
&\overset{D}{\Longrightarrow} \sqrt{n}\triangledown f(\boldsymbol{u}_{20})^T \mathcal{N}(0, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} & \sigma_{17} \\ & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} & \sigma_{27} \\ & & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} & \sigma_{37} \\ & & & \sigma_{44} & \sigma_{45} & \sigma_{46} & \sigma_{47} \\ & & & & \sigma_{55} & \sigma_{56} & \sigma_{57} \\ & & & & & \sigma_{66} & \sigma_{67} \\ & & & & & & \sigma_{77} \end{pmatrix})
\end{aligned} \tag{2.31}$$

where $\triangledown f(\boldsymbol{u}_{20})$ is the derivative of f with respect of $\boldsymbol{u}_{20}$, and are given by

$$\triangledown f(\boldsymbol{u}_{20}) = \begin{pmatrix} 1 & \frac{h_{20}}{h_{10}} \\ \frac{h_{10}}{h_{20}} & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \dfrac{-\bar{w}_2}{h_{10}h_{20}u_3(w)} & \dfrac{\bar{w}_1}{h_{10}^2 u_3(w)} \\ \dfrac{-\bar{w}_2}{h_{20}^2 u_3(w)} & \dfrac{\bar{w}_1}{h_1 h_{20} u_3(w)} \end{pmatrix}$$

$$
\begin{aligned}
\sigma_{11} =& h_{10}^2\mu_2(w) + \sigma_{10}^2 \\
\sigma_{22} =& h_{20}^2\mu_2(w) + \sigma_{20}^2 \\
\sigma_{33} =& h_{10}^4(\mu_4(w) - \mu_2(w)) + 2\sigma_{10}^4 + 4h_{10}^2\mu_2(w)\sigma_{10}^2 \\
\sigma_{44} =& h_{20}^4(\mu_4(w) - \mu_2(w)) + 2\sigma_{20}^4 + 4h_{20}^2\mu_2(w)\sigma_{20}^2 \\
\sigma_{55} =& h_{10}^2h_{20}^2(\mu_4(w) - \mu_2^2(w)) + h_{10}^2\mu_2(w)\sigma_{20}^2 + h_{20}^2\mu_2(w)\sigma_{10}^2 + \sigma_{10}^2\sigma_{20}^2 \\
\sigma_{66} =& h_{10}^4h_{20}^2(\mu_6(w) - 6\mu_2(w)\mu_4(w) - \mu_3^2(w) + 9\mu_2^3(w)) + h_{10}^4\sigma_{20}^2(\mu_4(w) - \mu_2^2(w)) + \\
& 3\sigma_{10}^4\sigma_{20}^2 + 4h_{10}^2h_{20}^2(\mu_4(w) - \mu_2^2(w))\sigma_{10}^2 + 3h_{20}^2\mu_2(w)\sigma_{10}^4 + \\
& h_{10}^2h_{20}^2\sigma_{10}^2(\mu_4(w) - 3\sigma_w^4) + 4h_{10}^2\mu_2(w)\sigma_{10}^2\sigma_{20}^2 \\
\sigma_{77} =& h_{10}^2h_{20}^4(\mu_6(w) - 6\mu_2(w)\mu_4(w) - \mu_3^2(w) + 9\mu_2^3(w)) + h_{20}^4\sigma_{10}^2(u_4(w) - \mu_2^2(w)) + \\
& 3\sigma_{10}^2\sigma_{20}^4 + 4h_{10}^2h_{20}^2(\mu_4(w) - \mu_2^2(w))\sigma_{20}^2 + 3h_{10}^2\mu_2(w)\sigma_{20}^4 + \\
& h_{10}^2h_{20}^2\sigma_{20}^2(\mu_4(w) - 3\mu_2^2(w)) + 4h_{20}^2\mu_2(w)\sigma_{10}^2\sigma_{20}^2 \\
\sigma_{12} =& -h_{10}h_{20}\mu_2(w) \\
\sigma_{13} =& h_{10}^3\mu_3(w) \\
\sigma_{14} =& h_{10}h_{20}^2\mu_3(w)
\end{aligned}
\tag{2.32}
$$

$$
\begin{aligned}
\sigma_{15} =& -h_{10}^2 h_{20}\mu_3(w)\\
\sigma_{16} =& -h_{10}^3 h_{20}(\mu_4(w) - 3\mu_2^2(w))\\
\sigma_{17} =& h_{10}^2 h_{20}^2(\mu_4(w) - 3\mu_2^2(w))\\
\sigma_{23} =& -h_{10}^2 h_{20}\mu_3(w)\\
\sigma_{24} =& -h_{20}^3\mu_3(w)\\
\sigma_{25} =& h_{10}h_{20}^2\mu_3(w)\\
\sigma_{26} =& h_{10}^2 h_{20}^2(\mu_4(w) - 3\mu_2^2(w))\\
\sigma_{27} =& -h_{10}h_{20}^3(\mu_4(w) - 3\mu_2^2(w))\\
\sigma_{34} =& h_{10}^2 h_{20}^2(\mu_4(w) - \mu_2^2(w)) + 2\sigma_{10}^2\sigma_{20}^2\\
\sigma_{35} =& -h_{10}^3 h_{20}(\mu_4(w) - \mu_2^2(w)) - 2h_{10}h_{20}\sigma_{10}^2\mu_2(w)\\
\sigma_{36} =& -h_{10}^4 h_{20}(\mu_5(w) - 4\mu_2(w)\mu_3(w)) - h_{10}^2 h_{20}\mu_3(w)\sigma_{10}^2 - 4h_{10}^2 h_{20}\mu_3(w)\sigma_{10}^2\\
\sigma_{37} =& h_{10}^3 h_{20}^2(\mu_5(w) - 4\mu_2(w)\mu_3(w)) + h_{10}^3\mu_3(w)\sigma_{20}^2 + 2h_{10}h_{20}^2\mu_3(w)\sigma_{10}^2\\
\sigma_{45} =& -h_{10}h_{20}^3(\mu_4(w) - \mu_2^2(w)) - 2h_{10}h_{20}\sigma_{20}^2\mu_2(w)\\
\sigma_{46} =& -h_{10}^2 h_{20}^3(\mu_5(w) - 4\mu_2(w)\mu_3(w)) - h_{20}^3\mu_3(w)\sigma_{10}^2 - 2h_{10}^2 h_{20}\mu_3(w)\sigma_{20}^2\\
\sigma_{47} =& h_{10}h_{20}^4(\mu_5(w) - 4\mu_2(w)\mu_3(w)) + h_{10}h_{20}^2\mu_3(w)\sigma_{20}^2 + 4h_{10}h_{20}^2\mu_3(w)\sigma_{20}^2\\
\sigma_{56} =& h_{10}^3 h_{20}^2(\mu_5(w) - 4\mu_2(w)\mu_3(w)) + h_{10}^3\mu_3(w)\sigma_{20}^2 + 2h_{10}h_{20}^2\mu_3(w)\sigma_{10}^2\\
\sigma_{57} =& -h_{10}^2 h_{20}^3(\mu_5(w) - 4\mu_2(w)\mu_3(w)) - 2h_{10}^3 h_{20}\mu_3(w)\sigma_{20}^2 - h_{20}^3\mu_3(w)\sigma_{10}^2\\
\sigma_{67} =& -h_{10}^3 h_{20}^3(\mu_6(w) - 6\sigma_w^2\mu_4(w) - \mu_3^2(w) + 9\mu_2^3(w)) - h_{10}^3 h_{20}\sigma_{20}^2(\mu_4(w) -\\
& 3\mu_2^2(w)) - 2h_{10}^3 h_{20}\sigma_{20}^2(\mu_4(w) - \mu_2(w)^4) - 2h_{10}h_{20}^3\sigma_{10}^2(\mu_4(w) - \mu_2(w)^4) -\\
& 9h_{10}h_{20}\mu_2(w)\sigma_{10}^4\sigma_{20}^4 - 4h_{10}h_{20}\mu_2(w)\sigma_{10}^2\sigma_{20}^2
\end{aligned}
$$

It follows immediately that theorem 3 holds as the third, fourth and fifth elements of $\boldsymbol{v}_{20}$ does not contribute to the calculation of $\hat{h}_1, \hat{h}_2$

Remark 13. *As a special case, when W is unknown constant, we can still get similar conculusion, with $\Sigma = (\sigma_{ij})$ given by:*

$$
\begin{aligned}
\sigma_{11} =&\sigma_1^2 \\
\sigma_{22} =&\sigma_2^2 \\
\sigma_{33} =&\sigma_1^4 + 4h_{10}^2 m_2(w)\sigma_1^2 \\
\sigma_{44} =&\sigma_2^4 + 4h_{20}^2 m_2(w)\sigma_2^2 \\
\sigma_{55} =&h_{10}^2 m_2(w)\sigma_2^2 + h_{20}^2 m_2(w)\sigma_1^2 + \sigma_1^2\sigma_2^2 \\
\sigma_{66} =&2m_2(w)h_{20}^2\sigma_1^4 + 4h_{10}^2 h_{20}^2 m_4(w)\sigma_1^2 - 4h_{10}^2 h_2^2 \sigma_1^2 m_2^2(w) + h_{10}^4 m_4(w)\sigma_2^2 \\
&- \sigma_2^2 h_{10}^4 m_2^2(w) + 2\sigma_2^2\sigma_1^4 + 4h_{10}^2 m_2(w)\sigma_2^2\sigma_1^2 \\
\sigma_{77} =&2m_2(w)h_{10}^2\sigma_2^4 + 4h_{10}^2 h_{20}^2 m_4(w)\sigma_2^2 - 4h_{10}^2 h_{20}^2 \sigma_2^2 m_2^2(w) + h_{20}^4 m_4(w)\sigma_1^2 \\
&- \sigma_1^2 h_{20}^4 m_2^2(w) + 2\sigma_1^2\sigma_2^4 + 4h_{20}^2 m_2(w)\sigma_1^2\sigma_2^2 \\
\sigma_{35} =&- 2h_{10}h_{20}\sigma_1^2\bar{w}_1(1-\bar{w}_1) + 2h_{10}h_{20}\sigma_1^2\frac{\boldsymbol{w}_1^{'}\boldsymbol{w}_2}{n} \qquad (2.33) \\
\sigma_{3,6} =&- 4m_3(w)h_{10}^2 h_{20}\sigma_1^2 - m_3(w)h_{10}h_{20}\sigma_1^2 \\
\sigma_{37} =&2m_3(w)h_{10}h_{20}^2\sigma_1^2 \\
\sigma_{45} =&- 2h_{10}h_{20}\sigma_2^2\bar{w}_1(1-\bar{w}_1) + 2h_{10}h_{20}\sigma_2^2\frac{\boldsymbol{w}_1^{'}\boldsymbol{w}_2}{n} \\
\sigma_{46} =&- 2m_3(w)h_{10}^2 h_2\sigma_2^2 \\
\sigma_{47} =&4m_3(w)h_{10}h_{20}^2\sigma_2^2 + m_3(w)h_{10}h_{20}\sigma_2^2 \\
\sigma_{56} =&m_3(w)h_{10}^3\sigma_2^2 + 2m_3(w)h_{10}h_{20}^2\sigma_1^2 \\
\sigma_{57} =&- m_3(w)h_{20}^3\sigma_1^2 - 2m_3(w)h_{20}h_{10}^2\sigma_2^2 \\
\sigma_{67} =&- 2m_4(w)h_{10}^3 h_{20}\sigma_2^2 + 2h_{10}^3 h_{20}\sigma_2^2\mu_2^2(w) - 2h_{10}h_2^3\sigma_1^2 m_4(w) + \\
&2h_{10}h_{20}^3\sigma_1^2 m_2^2(w) - 4m_2(w)h_{10}h_{20}\sigma_1^2\sigma_2^2
\end{aligned}
$$

□

2.5.4 Proof of Theorem 4

Proof. The arguments are quite similar as that of the proof of theorem 3. Thus we omit the details. □

Chapter 3

Asymptotic Inference of Maximum Cross Correlation of Stationary Process

3.1 Introduction

In the era of big data, there has been extensive research on the dependence among large number of variables (Bühlmann and Van De Geer, 2011; Bickel and Levina, 2008a,b; Cai et al., 2010). Statistical analysis are usually carried out based on independent samples. Recently multivariate analysis has also undergone rapid developments to study this type of cross sectional dependence among the variables, as well as the temporal dependence among the samples (Song and Bickel, 2011; Davis et al., 2015; Basu and Michailidis, 2015; Raskutti and Yuan, 2015).

Cross correlations play fundamental roles in measuring and analyzing cross-sectional and temporal dependence simultaneously. Suppose there is a p-dimensional stationary time series x_{jt}, $1 \leq j \leq p$ and $1 \leq t \leq T$. The cross covariance between j-th and k-th time series at lag s is defined as $\gamma_{jk}(s) = \mathrm{Cov}(x_{jt}, x_{k,t+s})$. In particular, $\gamma_{jj}(\cdot)$ gives the autocovariance function of the j-th component series. The cross correlation is then given by $\rho_{jk}(s) = \gamma_{jk}(s)/\sqrt{\gamma_{jj}(0)\gamma_{kk}(0)}$. One important type of cross-sectional and temporal dependence is the lead lag relationship among component series, i.e. observed values from one series may have an impact on another series a few time units later. Lead-lag relationship has been widely studied in many scientific fields, including economics, engineering, finance, geophysical sciences, and neuro-sciences (Nelson-Wong et al., 2009), (Duffy and Hughes-Clarke, 2005), (Basappa and Lakdawala, 2000), (Cohen, 1981). For example, Conover and Peterson (1999) found that before the passage of the Insider Trading Sanctions Act (ITSA) in 1984, the options market leads the stock market before negative surprises but that the stock market leads prior to positive surprises, while after the passage of ITSA there is no leading role for either market under positive or negative surprises. Berndt and Ostrovnaya (2007) provided a rigorous analysis on the relationship between credit market and option market. Their results indicated that investors absorb information revealed in the CDS market into option prices within a few days, i.e. CDS market lead the option market. Cross correlations may be used to infer such kind of relationship among different series. Ideally if $\gamma_{jk}(s)$ is zero for all negative s, and nonzero for some negative s, then there is a unidirectional relationship from the j-th series to the k-th series.

More comprehensive relationship among the p series may be modeled by vector autoregressive (VAR) models. However, when there are many series, i.e. p is large, fitting a VAR over all series is not computationally or statistically feasible. The problem becomes easier if the p series can be partitioned into smaller groups, where the between groups dependence is weak or negligible, and VAR models can be built with each group. Cross correlations can be used to measure the linear relationship between any two series, and may serve as a proxy of the distance or closeness between them.

The aforementioned problem can be viewed as a clustering problem, where it may be assumed that different groups are not correlated. A closely related problem is to test whether these p series are correlated at all. It is also a preliminary step before fitting a VAR to the data. This testing problem is related to, but different from the classical multivariate white noise test (see for example Chitturi, 1974). The most important distinction is that each individual series may have its own temporal dependence, and may not be a univariate white noise. However, similarly as the multivariate white noise test, cross correlations can be used to construct the test statistic.

Motivated by the preceding discussion, we consider the following testing problem

$$\mathrm{H}_0: \quad \gamma_{jk}(s) = 0, \quad \forall\, j \neq k, s \in \mathbb{Z}. \quad \text{vs} \quad \mathrm{H}_1: \quad \gamma_{jk}(s) \quad \text{for some } j \neq k, \tag{3.1}$$

under the "large T, large p" paradigm, where the dimension p may be comparable to, or even larger than the sample size T. The cross covariances can be estimated by the sample version:

$$\hat{\gamma}_{jk}(s) = \frac{1}{T} \sum_{1 \le t, t+s \le T} (x_{jt} - \bar{x}_j)(x_{k,t+s} - \bar{x}_k) \tag{3.2}$$

where $\bar{x}_j$ is the sample mean of the j-th series. We consider the maximum type test statistic:

$$\tilde{M}_1 = \max_{|s| \le s_T,\, 1 \le j < k \le p} \hat{\gamma}_{jk}(s). \tag{3.3}$$

Since the correlation between two series may exist at some unknown but very large lag s, here we allow the range s to expand with the sample size, i.e. s_T is allowed to approach infinity as T increases. Sometimes the cross correlation between two series may exist at many adjacent lags, but is weak at each of them. In this case, the following test statistic can have larger power.

$$\tilde{M}_{2m} = \max_{|s| \le s_T,\, 1 \le j < k \le p} \tilde{Q}_{jk}(s), \tag{3.4}$$

where

$$\tilde{Q}_{jk}(s) = \sum_{l=s+1}^{s+m} \hat{\gamma}_{jk}^2(s).$$

The testing problem is related to the classical white noise tests in time series analysis. The later is often used for diagnostics after a model being fitted to the data, see for example (Wikle and Hooten, 2010) for spatial-temporal modeling, and Tao et al. (2012) for a study of a large number of assets. Many classical tests have been invented for univariate time series, including Robinson (1991), Durbin and Watson (1950, 1951), Box and Pierce (1970), Durlauf (1991); Hong (1996), and many variants. A multivariate version of the Box and Pierce test was proposed by Chitturi (1974). Hosking (1980, 1981) gave several equivalent forms of this statistic, see also Ahn (1988); Escanciano et al. (2013); Mainassara (2011). Most of these tests are essentially based on sample autocovariances and cross covariances. Usually they involve a finite number of lags. Hong (1996) and Hong and Lee (2003) were the first to allow the number of lags to grow with the sample size. Xiao and Wu (2013) considered the maximum deviation of the sample autocovariances.

We will show that the test statistics converge to extreme value distribution of type

I (also called Gumbel distribution) after proper normalizations. Due to the existence of temporal dependence, we carry out theoretical analysis under the framework of causal representation and physical dependence measures (Wu, 2005). Our proof makes use of the Gaussian approximation result Zaitsev (1987).

On the other hand, it is well know that the Gumbel type convergence is usually slow. As a result, tests based on asymptotic limiting distribution may be distorted when the sample size is not large enough. We propose to use bootstrap method to improve the finite sample performance. More specifically, we use the moving blocks bootstrap of Liu and Singh (1992). Recently, Hill and Motegi (2016) and Zhang and Cheng (2014) also considered bootstrap methods for the maximum type statistics under the time series context.

The problem under consideration is also closely related to high dimensional covariance structure testing, which is of fundamental importance in high dimensional statistics. Let $\mathbf{X} = (x_{ij})_{1\le i\le n, 1\le j\le p}$ be the data matrix, whose n rows are independently and identically distributed, with mean vector $\boldsymbol{\mu}_n$ and covariance matrix $\Sigma_n = (\sigma_{ij})$. In many empirical studies, it is often assumed that $\Sigma_n = I_p$, where I_p is the $p \times p$ identity matrix. Therefore, it is important to test whether Σ is an identity or a diagonal matrix. Due to high dimensionality, the convectional LRT is drifted to infinity to when p is large (Bai et al., 2009). Chen et al. (2013) found that the empirical distance test Nagao (1973) is not consistent when both p and n are large, and proposed corrections to the empirical distance test. Assuming that the population distribution is Gaussian with mean $\boldsymbol{\mu}_n = 0$, Johnstone (2001) used the largest eigenvalue of the sample covariance matrix $X_n^T X_n$ as the test statistic, and proved that its limiting distribution follows the Trac-yWidom law Tracy and Widom (1994). His work was extended to the non-Gaussian case by Péché (2009); Soshnikov (2002). Other literature concerning on the second order properties among high dimensional data includes Cai et al. (2013); Tony Cai et al. (2014); Chen and Qin (2010). Let $\bar{x}_j = \frac{1}{n}\sum_{i=1}^n x_{ij}$, then the sample covariance between j-th and k-th column is $\hat{\sigma}_{jk} = \dfrac{1}{n}\sum_{1\le i\le T}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$. Jiang (2004) used $\max_{1\le i<j\le p} |\hat{\sigma}_{jk}|$ as the test statistic, and established the Gumbel type convergence under the assumption that the entries of X are independent and identically

distributed. His result was followed by Li et al. (2010); Zhou (2007) and Liu et al. (2008). In recent paper, Cai and Jiang (2011) extended the results in the way that each row of $\mathbf{X}$ is Gaussian and can be m-dependent. Xiao and Wu (2013) also showed the Gumbel convergence of a self-normalized version of $\max_{1\leq i<j\leq p} |\hat{\sigma}_{jk} - \sigma_{jk}|$, allowing Σ to be a general non-diagonal matrix.

The rest of this chapter is organized as follows. We first consider the cross covariance between two series in section 3.3.1, and show that under mild dependence and moment conditions, the maximum deviation of the sample cross covariance converges to extreme value distribution. This result can help to identify the true lead (or lag) if there is a underlying lead lag relationship. Furthermore, in order to reduce the computational cost, we also propose a window sum approach, where the lead (or lag) window can be identified. We then study the cross correlations among high dimensional time series in section 3.5. Under mild conditions, we establish the the Gumbel convergence of maximum sample cross correlation. We also propose to use the moving blocks bootstrap to improve the finite sample performance in section 3.5.2.

3.2 Physical Dependence Measurement

To develop an asymptotic results for the times series, it's necessary to impose suitable measures of dependence. We consider our theory in the general physical dependence of Wu (2005). Assume that (X_i) is a stationary causal process of the form:

$$X_i = \mathbf{g}(\ldots, \epsilon_{i-1}, \epsilon_i),$$

where $\mathbf{g}$ is a measurable function for which X_i is a properly defined random variables.

For notational simplicity, we define the operator:

$\Omega_k(X) := \mathbf{g}(\epsilon_j, ..., \epsilon_{k+1}, \epsilon_k^{'}, \epsilon_{k-1}, \ldots)$, where $(\epsilon_k^{'})_{k\in\mathbf{Z}}$ is an i.i.d copy of $(\epsilon_k^{'})$. Namely ϵ_k in X is replaced by $\epsilon_k^{'}$.

For a random variable X and $p > 0$, we write $X \in \mathcal{L}^p$ if $\|X\|_p := (\mathrm{E}(\|X\|^p))^{1/p} < \infty$ and in particular, use $\|X\|$ for the $\mathcal{L}^2$-norm $\|X\|_2$.

Assume $X \in \mathcal{L}^p$, $p > 1$. Define the physical dependence measure order of p as

$$\delta_p(i) = \|X_i - \Omega_0(X_i)\|,$$

which quantifies the dependence of X_i on the innovation of ϵ_0. Let $p^{'} = \min(2, p)$ and define

$$\Theta_p(n) = \sum_{i=n}^{\infty} \delta_p(i),$$
$$\Psi_p(n) = \left(\sum_{i=n}^{\infty} \delta_p(i)^{p^{'}}\right)^{1/p^{'}},$$
$$\Delta_p(n) = \sum_{i=0}^{\infty} \min\{\mathcal{C}_p \Psi_p(n), \delta_p(i)\}$$

where

$$\mathcal{C}_p = \begin{cases} (p-1)^{-1}, \text{ if } 1 < p < 2; \\ \sqrt{p-1}, \text{ if } p \geq 2. \end{cases}$$

Besides, for $i \leq j$, define $\mathcal{F}_i^j =< \epsilon_i, \epsilon_{i+1}, \dots, \epsilon_j >$ be the $\sigma -$ field generated by the innovation $\epsilon_i, \epsilon_{i+1}, \dots, \epsilon_j$, and the projection operator $\mathcal{H}_i^j = \mathbb{E}(\cdot|\mathcal{F}_i^j)$.

Set $\mathcal{F}_i := \mathcal{F}_i^{\infty}, \mathcal{F}^j := \mathcal{F}_{-\infty}^j$, and define $\mathcal{H}_i$ and $\mathcal{H}^j$ similarly. Define projection operator $\mathcal{P}^j(\cdot) = \mathcal{H}^j - \mathcal{H}^{j-1}$, and $\mathcal{P}_i(\cdot) = \mathcal{H}_i - \mathcal{H}_{i+1}$, then $(\mathcal{P}^j(\cdot)_{j\in\mathbf{Z}})$ and $(\mathcal{P}_{-i}(\cdot)_{i\in\mathbf{Z}})$ becomes martingale difference sequence with respect to filtration $\mathcal{F}^j$ and $(\mathcal{F}_{-i})$ respectively.

3.3 Maximum Covariance for Bi-Variate Stationary Process

Consider the bi-variate stationary time series $\mathbf{x}(t) = (x_{1t}, x_{2t})'$ of the form

$$\mathbf{x}(t) = g(\epsilon_t, \epsilon_{t-1}, \dots)$$

where $\epsilon_t, t \in \mathbb{Z}$ are i.i.d two dimensional random vector.

Let $\gamma_{12}(k) = Cov(x_{1t}, x_{2,t+k})$. Assume without loss of generality that

$$\gamma_{12}(k^\star) = \max_{k\in\mathbb{Z}}\{|\gamma_{12}(k)|\} =: \gamma_{12}^\star$$

We also assume that $|\gamma_{12}(k)| < \gamma_{12}^{\star}$ whenever $k \neq 0$. Suppose we have observed x_{1t} for $1 \leq t \leq T$. At each $1 \leq t \leq T$, we also have an observation on the second process, but with a time lead $k^{\star}$, i.e. we observe $x_{2,t+k^{\star}}$. Therefore, the data are

$$x_{1t}, \ldots, x_{1T}, x_{2,k^{\star}+1}, \ldots, x_{2,k^{\star}+T}$$

The time lead $k^{\star}$ is unknown, which we would love to identify from the data. We allow $k^{\star}$ to depend implicitly on T, and always assume $|k^{\star}| \leq cT$ for some constant $0 < c < 1$. Theorem 5 shows that asymptotically, $k^{\star}$ can be identify by

$$\hat{k} = \max_{-T<k<T} |\hat{\gamma}_{12}(k)|$$

where $\hat{\gamma}_{12}(k)$ is the sample cross covariance and defined as:

$$\hat{\gamma}_{12}(k) = \frac{1}{T} \sum_{t=|k|+1}^{T} (x_{1t} - \bar{x}_1)(x_{2,t-k} - \bar{x}_2)$$

3.3.1 Theoretical Results

Theorem 5. *Assume $EX_i = 0$, $X_i \in L^p$ for some $p \geq 4$, and $\Theta_p(m) = O(m^{-\alpha})$, $\Delta_p(m) = O(m^{-\alpha'})$ for some $\alpha \geq \alpha' \geq 0$. If $\alpha > 1/2$ or $\alpha' p > 2$ then there exists some constants c_p such that*

$$\lim_{T\to\infty} P\left(\max_{-T<k<T} |\hat{\gamma_{12}}(k) - E\hat{\gamma_{12}}(k)| \leq c_p \sqrt{\frac{logT}{T}} \right) = 1$$

Corollary 1. *Assume the condition of Theorem 5, Then,*

$$\lim_{T\to\infty} P(\hat{k} = k^{\star}) = 1.$$

3.3.2 Simulation

AR(1) Example:

(x_{1t}, x_{2t}) are generated from the Model (3.5), where x_{1t} follows AR(1), and x_{2t} is a simple shift of x_{1t} by $k^\star$

$$\begin{cases} X_{1t} &= \alpha X_{1,t-1} + \epsilon_t \text{ where } \epsilon_t \sim \mathcal{N}(0,1) \\ X_{2t} &= X_{1,t+k^\star} \end{cases} \tag{3.5}$$

Figure 3.1 shows the moving trend of $(X_1(t), X_2(t))$ in a specific case when $k^\star = 10$.

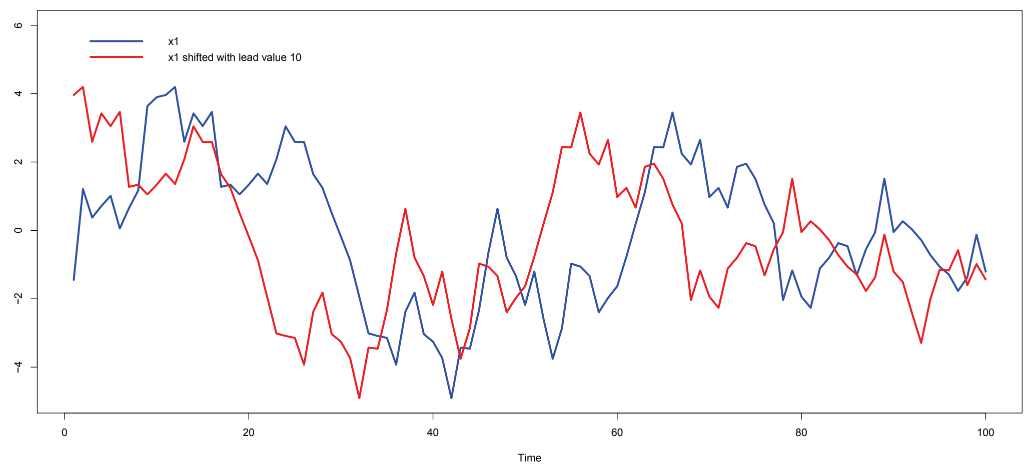


Figure 3.1: moving trend of $(X_1(t), X_2(t))$ with leading shift $k^\star = 10$

Different AR coefficients α are considered in the simulation. We consider the sample size from $T = 50$ up to $T = 500$, and allow the leading values of $k^\star$ to be dependent on the sample size. Every time, for fixed leading value and sample size, we calculate the maximum cross covariance, and compare the corresponding lead (or lag) to the true value $k^\star$. If they are identical, we count it as 1, otherwise, set it as 0. We repeat the procedure 1000 times, based on which, the average probability of identifying $k^\star$ is calculated. Table 3.1 shows the final results. As a characteristic of AR(1) model, larger α indicates stronger dependence between observations. We notice that the smaller the α and true leading value, the larger the identification rate. And as the sample size gets bigger, the probability of identifying the true lead values becomes higher.

		$\log(T)$	$T^{0.5}$	$T^{0.8}$	$T^{0.9}$	$T/3$
$\alpha = 0.3$	$T = 50$	1.000	1.000	0.920	0.060	0.850
	$T = 100$	1.000	1.000	1.000	0.680	0.980
	$T = 250$	1.000	1.000	1.000	1.000	1.000
	$T = 500$	1.000	1.000	1.000	1.000	1.000
$\alpha = 0.5$	$T = 50$	1.000	1.000	0.843	0.046	0.739
	$T = 100$	1.000	1.000	0.999	0.517	0.938
	$T = 250$	1.000	1.000	1.000	1.000	1.000
	$T = 500$	1.000	1.000	1.000	1.000	1.000
$\alpha = 0.8$	$T = 50$	0.998	0.982	0.559	0.020	0.450
	$T = 100$	1.000	1.000	0.938	0.223	0.626
	$T = 250$	1.000	1.000	1.000	0.938	0.928
	$T = 500$	1.000	1.000	1.000	0.999	0.992
$\alpha = 0.9$	$T = 50$	0.982	0.929	0.375	0.015	0.285
	$T = 100$	1.000	0.999	0.778	0.121	0.399
	$T = 250$	1.000	1.000	0.996	0.715	0.695
	$T = 500$	1.000	1.000	1.000	0.992	0.908

Table 3.1: The first column shows the value of α in the generating model 3.5. The second column shows the sample size, and the first row shows the true leading value $k^\star$. The number in each cell is the proportion of identifying the true lead values.

VAR(1) Example:

In this example, $\boldsymbol{x}_t = (x_{1t}, x_{2t})$ are generated from the following VAR(1) Model:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \tag{3.6}$$

where $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}) \sim N(\mathbf{0}, I_2)$

Then We make a shift of the second series $x_2(t)$, i.e. $x_{2t}^{\mathrm{S}} = x_{2,t+k^\star}$.

Figure 3.2 show the moving trend of $(x_{1t}, x_{2t}^{\mathrm{S}}, x_{2t})$. Similar as the AR(1) example, different $k^\star$ depending on sample size are tried. Table 3.2 summarize the final identification rate. Same conclusion can be drawn as that of AR(1) example.

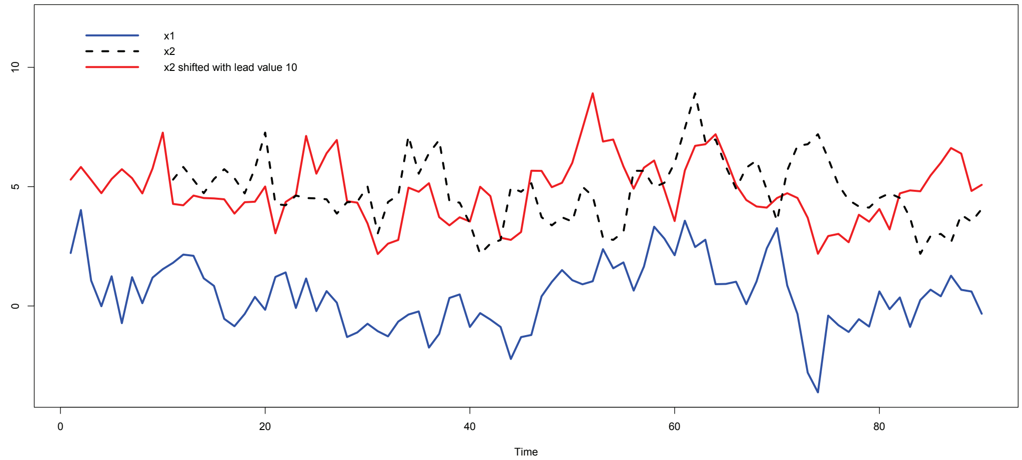


Figure 3.2: Example: series from VAR(1) with second series shifted by lead 10

	$\log(T)$	$T^{0.5}$	$T^{0.7}$	$T/10$
$T = 100$	0.64	0.47	0.15	0.47
$T = 500$	0.90	0.85	0.69	0.76
$T = 1000$	0.94	0.88	0.78	0.81
$T = 10000$	0.99	0.95	0.88	0.86

Table 3.2: Estimated proportion of identifying $k^\star$ out of 1000 simulations

3.4 Window Sum Approach

Section 3.3 describes the maximum cross covariance approach to identify the lead-lag value between two series. The approach is time-consuming if both sample size T and time lead $k^\star$ is large. A direct approach is window sum approach. The approach splits the samples into blocks with size $B = T^\eta$, compute the sum in each block, and then calculate the cross covariance using the sum. Let B be the window size, which implicitly depends on T. It's naturally to assume as $B \to \infty$ and $1/N = B/T \to 0$. Define the window sum:

$$\bar{x}_{1i} = \sum_{(i-1)*B+1}^{(i*B)\vee T} x_{1t}, \text{ and } \bar{x}_{2i} = \sum_{(i-1)*B+1}^{(i*B)\vee T} x_{2,t+k^\star} \tag{3.7}$$

for $1 \leq i \leq N$. Then we compute

$$\bar{\gamma}_{12}(j) = \frac{1}{T} \sum_{1+j^-}^{N-j^+} \bar{x}_{1,t+j} \bar{x}_{2,t+k^\star}$$

and find out

$$\bar{j} = \max_{-N<j<N} |\bar{\gamma}_{12}(j)|$$

Let $\bar{k}^\star = [k^\star/B]$. Theorem 6 shows that with high probability, $\bar{j} \in \{\bar{k}^\star - 1, \bar{k}^\star, \bar{k}^\star + 1\}$

3.4.1 Theoretical Results

Theorem 6. *Assume the condition of Theorem 5.*

$$\lim_{T\to\infty} P\left(\max_{-N<j<N} |\bar{\gamma}_{12}(j) - E\bar{\gamma}_{12}(j)| \leq c_p \sqrt{\frac{\log N}{N}} \right) = 1$$

Corollary 2. *Assume the condition of Theorem 6, assume the cross spectral density $f_{12}(\theta)$ of $\{x_j(t)\}$ is nonzero at $\theta = 0$. Then*

$$\lim_{T\to\infty} P\left(\bar{j} \in \{\bar{k}^\star - 1, \bar{k}^\star, \bar{k}^\star + 1\}\right) = 1$$

3.4.2 Simulation

AR(1) Example

$(x_{1t}, x_{1,t+k^\star})$ is generated from the same AR(1) Model define in Eq. (3.5) with $\alpha = 0.8$. In this exercise, we adopt the window sum approach described in section 3.4, with different window size $B = T^{0.3}, T^{0.4}, T^{0.5}$ considered, The identification rate are summarized in Table 3.3. The results show that for fixed leading value $k^\star$, the smaller the window size B, the higher the identification rate. And also we notice that for fixed sample size and block size, the larger the true lead value is, the smaller the identification rate is.

		$\log(T)$	$T^{0.5}$	$T^{0.7}$	$T/4$
$B = T^{0.3}$	$T = 100$	1.000	0.999	0.778	0.778
	$T = 500$	1.000	1.000	1.000	0.999
	$T = 1000$	1.000	1.000	1.000	0.999
	$T = 10000$	1.000	1.000	1.000	0.999
$B = T^{0.4}$	$T = 100$	0.999	0.969	0.711	0.711
	$T = 500$	1.000	1.000	0.991	0.836
	$T = 1000$	1.000	1.000	1.000	0.980
	$T = 10000$	1.000	1.000	1.000	1.000
$B = T^{0.5}$	$T = 100$	0.959	0.961	0.498	0.498
	$T = 500$	1.000	1.000	0.835	0.653
	$T = 1000$	1.000	1.000	1.000	0.960
	$T = 10000$	1.000	1.000	1.000	0.999

Table 3.3: The first column shows block size. The second column shows the sample size, and the first row in shows the true leading value $k^{\star}$. The number in each cell is the proportion of identifying the true lead values.

VAR(1) Example

$(x_{1t}, x_{1,t+k^{\star}})$ is generated from the same VAR(1) Model in define in Eq. (3.6). In this exercise, we adopt the WSA described in section 3.4, with different window size $B = T^{0.3}, T^{0.4}, T^{0.5}$ considered. The results are summarized in Table 3.4. Same conclusion as that in the AR(1) example can be reached .

		$\log(T)$	$T^{0.5}$	$T^{0.7}$	$T/4$
$B = T^{0.3}$	$T = 100$	0.883	0.773	0.346	0.338
	$T = 500$	1.000	1.000	0.989	0.858
	$T = 1000$	1.000	1.000	1.000	0.990
	$T = 10000$	1.000	1.000	1.000	0.999
$B = T^{0.4}$	$T = 100$	0.905	0.768	0.313	0.311
	$T = 500$	0.999	1.000	0.938	0.666
	$T = 1000$	1.000	1.000	0.999	0.929
	$T = 10000$	1.000	1.000	1.000	1.000
$B = T^{0.5}$	$T = 100$	0.857	0.810	0.289	0.271
	$T = 500$	0.995	0.999	0.575	0.267
	$T = 1000$	0.968	1.000	1.000	0.664
	$T = 10000$	1.000	1.000	1.000	1.000

Table 3.4: The first column shows block size. The second column shows the sample size, and the first row shows the true leading value $k^{\star}$. The number in each cell is the proportion of identifying the true lead values.

3.5 Maximum cross correlation among multiple series

We consider a collection of p time series of length T, denoted by (x_{jt}), $1 \le j \le p$, $1 \le t \le T$. Assume the p series are independent. We allow each individual series to have its own temporal dependence. To quantify it, we assume each series $\{X_{i\cdot}\}$ has the causal representation. Let $\delta_q^{(i)}(k)$ be its physical dependence measures, and $\Delta_q^{(i)}(k)$ the corresponding tail sums. Set $\Delta_q(k) := \sup_i \Delta_q^{(i)}(k)$. To begin with, we will introduce Lemma 1:

Lemma 1. *For each pair of series $x_{j\cdot}$ and $x_{k\cdot}$, Let $\tau_{jk}^2 = \sum_{s\in\mathbb{Z}} \gamma_{jj}(s)\gamma_{kk}(s)$. If $\Delta_2^{(j)} < \infty$ and $\Delta_2^{(k)} < \infty$, then*

$$\sqrt{T}\hat{\gamma}_{jk}(s) \to N(0, \tau_{jk}^2)$$

According to Lemma 1, the sample cross covariances from different pairs of series can have different asymptotic variances. Due to this reason, we need to standardized them to have the same scale asymptotically. The asymptotic variance τ_{jk}^2 can be estimated as

$$\hat{\tau}_{ij}^2 = \sum_{k=-\nu_T}^{\nu_T} \hat{\gamma}_{ii}(k)\hat{\gamma}_{jj}(k),$$

where ν_T satisfy the condition $\nu_T \to \infty$ and $\nu_T/T \to 0$. The first statistic we consider is

$$M_1 = \max_{|k|\le s_T, 1\le i<j\le p} |\hat{\gamma}_{ij}(k)|/\hat{\tau}_{ij}. \tag{3.8}$$

On the other hand, as it is often assumed that the error terms in time series models (for example VAR model) are white noise, as a diagnostic procedure, it is necessary to perform white noise test. Therefore, for the second statistic, we assume each $\{X_{i\cdot}\}$ is a white noise, and the test statistic is

$$M_{2m} = \max_{|k|\le s_T, 1\le i<j\le p} Q_{ij}(k, m), \tag{3.9}$$

where $Q_{ij}(m, k) = \sum_{l=k+1}^{k+m} \hat{\rho}_{ij}^2(l)$, and

$$\hat{\rho}_{ij}(k) := \hat{\gamma}_{ij}(k)/\sqrt{\hat{\gamma}_{ii}(0)\hat{\gamma}_{jj}(0)}.$$

Our main result is summarized in Theorem 7 and Theorem 8, establishing the Gumbel convergence of maximum deviations (M_1 and M_{2m}) across all possible pair and over a wide range of lags.

Theorem 7. *Assume the p series are independent. Their physical dependence measures satisfy the uniform rate $\Delta_q(k) = O(k^{-\alpha})$ for some $\alpha > 0$ and $q > 2$. Assume the p spectral densities of the p series are uniformly bounded below from zero. Assume $p = O(T^{\gamma})$ for some $\gamma > 0$, and $s_T = O(T^{\eta})$ for some $0 \le \eta < 1$. Let $n = n_T = (2s_T + 1)p(p-1)/2$. Then if*

$$\begin{cases} \eta + 2\gamma < \alpha q & \text{if } \alpha \le 1/2 - 2/q; \\ \eta + 2\gamma < q/2 - 1 & \text{if } \alpha > 1/2 - 1/q; \end{cases}$$

we have

$$\lim_{T\to\infty} P(TM_1^2 - 2\log n + \log(\log n) + \log \pi \le z) = \exp\left(-e^{-z/2}\right).$$

Theorem 8. *Assume the same conditions of Theorem 7. Furthermore, assume each series is a white noise. Then*

$$\lim_{T\to\infty} P\left[TM_{2m} - 2\log n - (m-2)\log(\log n) + 2\log\Gamma(m/2) \le z\right] = \exp\left(-e^{-z/2}\right).$$

3.5.1 Useful Intermediate result

In order to accomplish the main theoretical results in Theorem 7 and Theorem 8, in this section, we provide some useful intermediate results. Lemma 2 presents the limiting distribution of the maximum of independent observations from chi-squared distribution after appropriate centering and scaling. Lemma 3 extends the result by replacing the independent chi-squared random variables with sum of consecutive squared standard normal. We notice that some of the sums are correlated with each other due to the overlapping terms. For example,see $Q_{k,m} = \sum_{i=k}^{i=k+m-1} Z_i^2$ and $Q_{k+1,m} = \sum_{i=k+1}^{i=k+m} Z_i^2$. Later, we find that these with overlapping terms contribute little in probability to the limiting distribution as sample size gets larger enough. As a consequence, we could conclude the same asymptomatic distribution. Lemma 3 could be regarded as an approximation of Lemma 4. In Lemma 4, we generalize the results to the maximum of sum of consecutive observation with unknown distribution. By making use of Gaussian approximation Zaitsev (1987), we show the similar convergence.

Lemma 2. *Suppose* $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} \chi^2_m$, *and* $M_{m,0} = \max\{X_1, X_2, ..., X_n\}$, *then*

$$\lim_{n\to\infty} P\left(\frac{M_{m,0} - d_n}{c_n} \leq z\right) = e^{e^{-z}},$$

where $d_n = 2\log n + (m-2)\log(\log n) - 2\log\Gamma(m/2)$ *and* $c_n = 2$.

This is directly from Embrechts et al. (1997) (in table 3.4.4 on page 156).

Lemma 3. *Let* $Z_i, i = 1, 2, 3, ..., n$ *be* n *independent random variables from* $N(0,1)$, *and define*

$$M^2_{m,1} = \max_{1\leq k\leq n-m+1} Q_{k,m},$$

where $Q_{k,m} = \sum_{i=k}^{i=k+m-1} Z_i^2$. *Then*

$$\lim_{T\to\infty} P\left[M^2_{m,1} - 2\log n - (m-2)\log(\log n) + 2\log\Gamma(m/2) \leq z\right] = \exp\left(\exp(-z/2)\right).$$

Let $Y = (y_{ij})_{1\leq i\leq n, 1\leq j\leq p}$ be a data matrix whose n rows are independent and identically distributed (i.i.d.) as some population distribution with mean 0_p, and covariance matrix I_p.

The sample mean of j-th column is: $\bar{y}_j = \frac{1}{n}\sum_{i=1}^{i=n} y_{i,j}, (j = 1, 2, ..., p)$.

Define:

$$Q'_{1,m} = n\left(\bar{y}_1^2 + \bar{y}_2^2 + \cdots + \bar{y}_m^2\right)$$
$$Q'_{2,m} = n\left(\bar{y}_2^2 + \bar{y}_3^2 + \cdots + \bar{y}_{m+1}^2\right)$$
$$\cdots$$
$$Q'_{p-m+1,m} = n\left(\bar{y}_{p-m+1}^2 + \bar{y}_{p-m+2}^2 + \cdots + \bar{y}_p^2\right)$$

Lemma 4. *Suppose* m *is constant, and* $E(|Y_j|^q) < \infty$. *Let* $M_{m,2} = \max\left\{Q'_{1,m}, \ldots, Q'_{p-m+1,m}\right\}$.

Suppose the dimension $p = n^r$*, where* $1 + r < \frac{q}{2}$,

$$\lim_{p\to\infty} P\left[M_{m,2}^2 - 2\log p - (m-2)\log(\log p) + 2\log\Gamma(m/2) \le z\right] = \exp\left(\exp\left(-z/2\right)\right).$$

3.5.2 Simulation

Wild bootstrap

As it is well known that Gumbel convergence is slow (See examples in Hall (1979)). The empirical distributions of $TM_{2m} - 2\log n - (m-2)\log(\log n) + 2\log\Gamma(m/2)$ is not close to the limiting distribution if the sample size is not large enough. Therefore, it is not reasonable to use the limiting distribution to approximate the finite sample distribution. As the limiting distribution does not depend on the data generating scheme, we draw the samples from i.i.d standard normal distribution, and then use its corresponding quantile as approximated critical values, based on which, we are able to compute the p-values. We will consider the following models. Please note that $\epsilon_i(t)$ are i.i.d standard normal distribution in the whole section 3.5.2.

$$\text{I.I.D. } X_i(t) = \epsilon_i(t) \tag{3.10}$$

$$\text{ARMA } X_i(t) = \alpha_1 X_i(t-1) + \theta_1\epsilon_i(t-1) + \epsilon_i(t) \tag{3.11}$$

$$\text{Bilinear } X_i(t) = (a + b\epsilon_i(t))X_i(t-1) + \epsilon_i(t) \tag{3.12}$$

$$\text{ARCH } X_i(t) = \sigma_t\epsilon_i(t), \text{ where } \sigma_t^2 = a + b_1 X_i(t-1) \tag{3.13}$$

$$\text{GARCH } X_i(t) = \sigma_t\epsilon_i(t), \text{ where } \sigma_t^2 = a + b_1 X_i(t-1) + b_2\sigma_{t-1}^2 \tag{3.14}$$

$$\text{Linear Process } X_i(t) = \sum_{j=1}^{50} \frac{\epsilon_i(t)}{j^{\alpha_j}}, \text{ where } 2 < \alpha_j < 4 \tag{3.15}$$

$$\text{AR and MA } X_i(t) = \begin{cases} \phi_j X_i(t-1) + \epsilon_{i,t} & \text{if } j = 1, \ldots, [p/2] \\ \theta_j \epsilon_{i,t-1} + \epsilon_{i,t} & \text{if } j = [p/2]+1, \ldots, p \end{cases}. \tag{3.16}$$

$$\text{TAR } X_i(t) \quad = \begin{cases} \phi_1 X_i(t-1) + \epsilon_i(t) & \text{if } X_i(t-1) < 0 \\ \phi_1 X_i(t-1) + \epsilon_i(t) & \text{Otherwise} \end{cases}. \tag{3.17}$$

- Step 1: generate samples x_{ij} from some pre-specified model (e.g. model 3.10) where $i = 1, \ldots, T$, $j = 1, 2, \ldots, p$;

- Step 2: for each pair $1 \leq j_1 < j_2 \leq p$, compute correlation at lag k

$$\rho_{j_1 j_2}(k) = \frac{\hat{\gamma}_{j_1 j_2}(k)}{\hat{\tau}_{j_1 j_2}},$$

  where $\hat{\gamma}_{j_1 j_2}(k) = \frac{1}{T} \sum_{1 \leq t < t+k \leq T} = (x_{j_1,t} - \bar{x}_{j_1})(x_{j_2,t+k} - \bar{x}_{j_2})$, and $\hat{\tau}_{j_1 j_2} = \sqrt{\sum_{s=-v_T}^{v_T} \hat{\gamma}_{j_1 j_1}(s) \hat{\gamma}_{j_2 j_2}(s)}$.

- Step3: Find $M_1^2 = \max_{1 \leq j_1 < j_2 \leq p, -S_T \leq k \leq S_T} \{\rho^2_{j_1 j_2}(k)\}$;

- Step4: rescale and center M_1^2 with the formula below:

$$M_n^* = TM_1^2 - 2\log(n) + \log\log(n) + 2\log\Gamma(1/2)$$

- Repeat Step 1-4 for 5000 times, and record all the values from step 4.

Using step 1-4, we tried all the models listed in Eq. (3.10) through Eq. (3.17). Let dimension $p = 50$, sample size $T = 600$, lag range $S_T = 10$, and the lag number used to compute variance $v_T = 10$. For Model (3.11) and Model (3.16), the AR coefficients are generated from Beta(1,5), and the MA coefficients are generated from Beta(3,2). For Model (3.12), a, b are generated from $\mathrm{U}(0, 0.4)$. For Model (3.13), a is generated from $\mathrm{U}(0, 0.3)$, and b_1 is generated from $\mathrm{U}(0, 0.5)$. For Model (3.14), a, b_1 are generated in the same way as that of Model (3.13), and b_2 is generated from $\mathrm{U}(0, 0.3)$. For Model (3.15), $\alpha_j, j = 1, \ldots, 50$ are generated from $-\mathrm{U}(2, 4)$ For Model (3.17), ϕ_1 is generated from $-\mathrm{U}(1, 1.5)$, and ϕ_2 is generated from $\mathrm{U}(0.3, 0.8)$. All the coefficients are only generated once and then fixed across all the repetitions.

First, we calculate the 90% and 95% quantile of the results from Model (3.10), i.e., i.i.d standard normal distribution. The critical values are $q_{0.9} = 6.133861$ and $q_{0.95} = 7.450773$ respectively. Then for each other Model, we average the number that exceeds the critical values, i.e. $\#(M_n^* > q_\alpha)/5000$. The results are summarized in table 3.5. We see that for linear process (ARMA, mixed AR and MA, and Linear Process), the empirical rejection rate (ERR) is around the pre specified level α. However, for the non-linear process, the ERR is larger than nominal size, especially for the BiProduct Process. Fig. (3.3) and (3.4) shows the CDF from each model against that from the i.i.d standard normal:

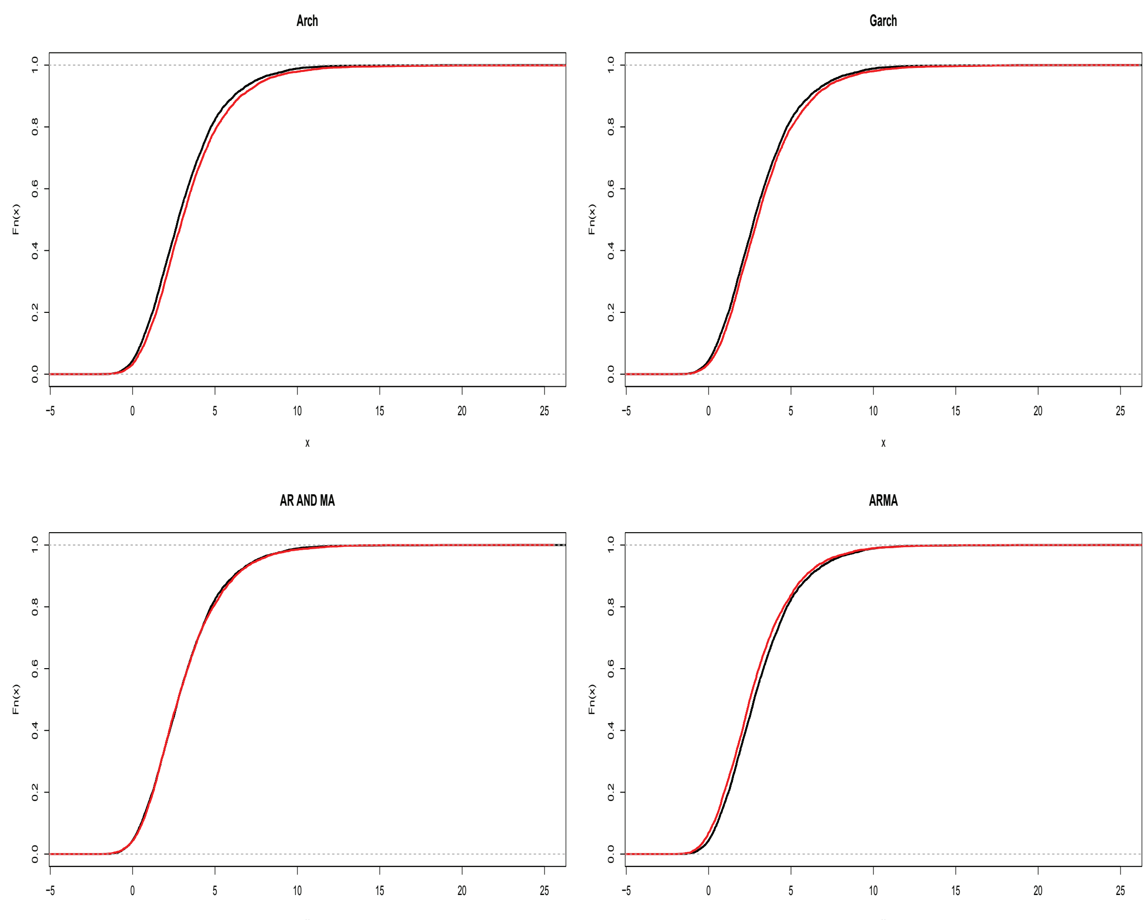


Figure 3.3: CDF from each model against that from i.i.d Standard Normal. Black line in each subplot represents the CDF from the i.i.d standard normal model, and red line represents the model in comparison with

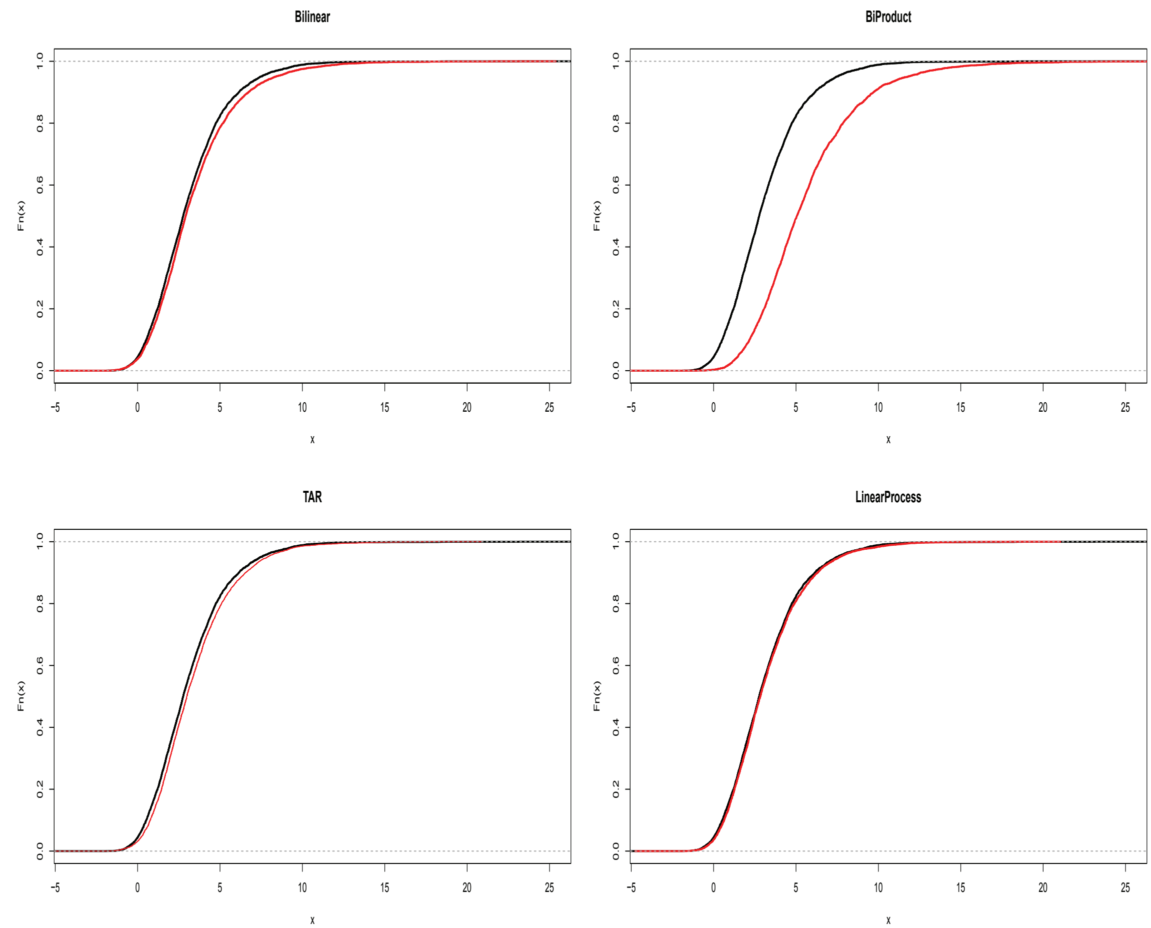


Figure 3.4: CDF from each model against that from i.i.d Standard Normal. Black line in each subplot represents the CDF from the i.i.d standard normal model, and red line represents the model in comparison with

The rejection rates at different significance level α are shown below:

	$\alpha = 10\%$	$\alpha = 5\%$
AR and MA	0.1056	0.0538
ARMA	0.0872	0.0422
ARCH	0.124	0.0682
Garch	0.1202	0.0598
Bilinear	0.1292	0.0718
TAR	0.1224	0.0616
Linear Process	0.1086	0.0552
BiProduct	0.3524	0.2336

Table 3.5: Empirical Rejection rate with i.i.d

Moving window bootstrap

We notice that for the nonlinear process, the rejection probability is larger than the nominal size. As a remediation, we also conduct the moving-window Bootstrap. For some block size b_T, the ith block is denoted by $\mathcal{B}_i = c(X_j(i), X_j(i{+}1), \ldots, X_j(i{+}b_n{-}1))$, where $i = 1, 2, \ldots, T - b_T + 1$. For simplicity, assume $h_T = T/b_T$ is an integer. The procedure is as following:

1. Step 1. generate samples x_{ij} from some pre-specified model (e.g. model 3.10) where $i = 1, \ldots, T$, $j = 1, 2, \ldots, p$. Calculate M_1^2 using formula (7).

2. Step 2. For the j-th series, sample h_T times with replacement from $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{T-b_T+1}\}$ to obtain blocks $\{\mathcal{B}^\star_{i_1}, \ldots, \mathcal{B}^\star_{i_{h_T}}\}$, and then they are laid end-to-end to form $\mathbf{X}^\star_j = \Big(X^\star_j(i_1), X^\star_j(i_1+1), \ldots, X^\star_j(i_{h_T}+b_n-1)\Big)$. Perform the re-sampling procedure for all the series $1 \leq j \leq p$

3. Step 3. pretend that $\mathbf{X} = \left(\mathbf{X}^\star_1, \mathbf{X}^\star_2, \ldots, \mathbf{X}^\star_p\right)$ is random samples of size T, calculate the $M^*_B = TM^\star_1 - 2\log(n) + \log\log(n) + 2\log\Gamma(1/2)$, where $M^\star_1$ is defined as Eq. (3.8) using the new samples.

4. Repeat Step 1, 2 and 3 1000 times, and calculate the empirical rejection rate $\#(M^*_B > M_T)/1000$.

Different block size are tried. Table 3.6 shows the empirical rejection rates when dimension $p = 10$. All the results are in percentage, for example, 10.2% is written as 10.2 in the table. The Column named 'Test' described the model under which the series are generated. We see that when $p = 10$ the empirical rejection rates are close to the nominal ones. Furthermore, we increase the variable dimension p to 20, and run the bootstrap again. The results are summarize in Table 3.7.

Test	$b_T = 10$		$b_T = 15$		$b_T = 20$		$b_T = 25$		$b_T = 30$	
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$
IID	10.2	5.5	9.7	4.3	10	5.8	10.4	3.8	10.4	5.6
ARMA	10.2	4.7	9.5	5.7	9.8	5.8	8.1	3.4	10.4	5.2
ARCH	9.2	5.3	9.1	5.0	10.8	5.9	10.1	5.1	8.7	4.9
GARCH	11.5	5.4	9.5	4.8	10.9	5.5	10.1	4.9	10.6	6.2
BiLINEAR	8.1	3.8	10.3	5.7	10.1	5.0	9.8	4.6	11.5	6.1
BiProduct	11.1	5.9	10.6	4.8	11.1	5.5	11.2	5.2	10.2	4.8
TAR	9.7	4.9	11.1	5.6	9.1	4.2	10.3	5.3	10.1	4.9

Table 3.6: Rejection Probabilities in percentage, where $p = 10$, $s_T = 10$, Different block size b_T is considered

Test	$b_T = 10$		$b_T = 15$		$b_T = 20$		$b_T = 25$		$b_T = 30$	
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$
IID	12.0	6.2	9.2	4.8	9.4	5.2	10.2	6.4	8.4	3.6
ARMA	9.2	4.8	7.8	5.4	8.0	4.6	9.6	4.4	9.8	5.0
ARCH	11.4	4.8	9.6	5.6	12.8	7.4	9.8	5.2	8.0	4.0
GARCH	9.2	5.4	8.2	3.6	10.4	5.4	10.2	5.4	10.4	6.8
BiLINEAR	11.0	5.4	8.6	4.6	11.4	5.8	10.6	5.0	10.2	4.6
BiProduct	11.8	7.2	12.2	5.6	11.4	4.4	12.0	6.6	11.0	6.0
TAR	11.0	6.0	10.4	5.0	10.2	5.0	8.6	5.4	10.8	6.4

Table 3.7: Rejection Probabilities in percentage, where $p = 20$, $s_T = 10$, Different block size b_T is considered

BOB bootstrap

In addition to window-moving alike bootstrap, we also conduct BOB procedure as described in Horowitz et al. (2006). From the jth series $X_{1j}, X_{2j}, \ldots, X_{nj}$, for the specified number of lag s_T and block size b_T, form $\mathbf{Y}_{ij} = (X_{ij}, X_{i+1j}, \ldots, X_{i+s_T j})^T$, $1 \leq i \leq T - s_T$, and blocks $\mathbf{B}_{k,j} = (\mathbf{Y}_{kj}, \ldots, \mathbf{Y}_{k+b_T-1,j})$ where $1 \leq k \leq T - s_T - b_T + 1$. For simplicity, assume $h_T = T/b_T$ is an integer. And we form such block for each series $j = 1, 2, \ldots, p$.

1. Step 1. generate samples x_{ij} from some pre-specified model (e.g. model 3.10) where $i = 1, \ldots, T$, $j = 1, 2, \ldots, p$. Calculate M_1 using formula (3.8). Note that the coefficients are generated in the same way as described in the wild bootstrap approach. And they are set constant over all repetition once it's generated.

2. Step 2. For jth series, we sample h_T blocks with replacement from $\{\mathbf{B}_{1,j}, \mathbf{B}_{2,j}, \ldots, \mathbf{B}_{T-s_T-b_T+1,j}\}$, to obtain blocks $\{\mathbf{B}^\star_{1j}, \mathbf{B}^\star_{2,j}, \ldots, \mathbf{B}^\star_{T-s_T-b_T+1,j}\}$, which are laid end to end to form a series of vector $Y^\star_{1,j}, Y^\star_{2,j}, \ldots, Y^\star_{T-s_T,j}$. We do such sampling over all series $j = 1, 2, \ldots, p$

3. Step 3. Pretend $Y^\star_{1j}, Y^\star_{2j}, \ldots, Y^\star_{T-s_Tj}$, for $j = 1, 2, \ldots, p$ are random samples from s_T dimensions with sample size T. Denote the k th row element in $Y^\star_{1,j}$ by $Y^\star_{1,j}(k)$, for $k = 1, \ldots, s_T$. Thus, the cross covariance of $\gamma_{j,l}$ at lag k could be calculated by: $\gamma_{jl}(k) = \dfrac{1}{T} \sum_{t=1}^T Y^\star_{tj}(1) Y^\star_{tj}(k+1)$. Furthermore, we can calculate the cross correlation of $\hat{\rho}_{jl}(k) = \dfrac{\hat{\gamma}_{jl}(k)}{\hat{\tau}_{jl}}$ where $\hat{\tau}_{jl} = \sqrt{\sum_{s=-v_T}^{v_T} \hat{\gamma}_{jj}(s)\hat{\gamma}_{ll}(s)}$. Afterwards, we can calculate $M^\star_B = \max_{1 \le j_1 < j_2 \le p, -s_T \le k \le s_T} \rho^2_{j,l}(k)$

4. Step 4. Repeat Step 2 and Step 3 1000 times, and calculate the and calculate the empirical rejection rate $\#(M^\star_B > M_T)/1000$.

The simulation results are summarized in Table 3.8.

Test	$b_T = 10$		$b_T = 15$		$b_T = 20$		$b_T = 25$		$b_T = 30$	
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$
IID	10.7	4.7	10.4	5.5	9.7	4.9	8.9	4.8	10.3	5.4
ARMA	12.5	6.4	10.5	6.2	10.6	4.6	10.1	5.6	10.4	4.5
ARCH	7.9	3.2	8.7	4.5	7.9	4.5	7.5	2.9	10.8	5.6
GARCH	8.4	4.2	8.8	4.5	8.9	4.9	9.7	5.0	9.6	3.3
BiLINEAR	8.7	5.4	8.6	4.6	10.1	5.6	8.9	4.2	8.5	4.1
BiProduct	7.0	2.7	8.6	3.5	8.4	4.4	10.4	5.0	10.0	5.2
TAR	8.7	4.8	9.0	5.0	10.1	4.9	9.1	4.7	11.3	5.7

Table 3.8: Rejection Probabilities in percentage, where $p = 10$, $s_T = 10$, $v_T = 3$. Different block size b_T is considered

3.5.3 Real Data Analysis

Consider the simple returns of monthly indexes of U.S. government bonds with maturities in 30 years, 20 years, 10 years, 5 years, and 1 year. The data is obtained from Wharton Research Data Services (WRDS), and have 600 observations starting January 1942 to December 1991. Figure 3.5 shows their historical trend: the volatility of the 1-year bond returns is much smaller than that of returns with longer maturities.

Table 3.9 gives the lag 1 and lag 2 cross-correlation matrices of $\boldsymbol{r}_t$. Most of the significant cross correlations are at lag 1. And there exhibits stronger linearity between longer term bonds than those between shorter bonds. Here we consider the cross correlations excluding the concurrent ones. More specifically, we want to test $\rho_{ij}(l) = 0$ $\forall l \neq 0$ and $i \neq j$. We perform the bootsrap test and the pvalue is 0,

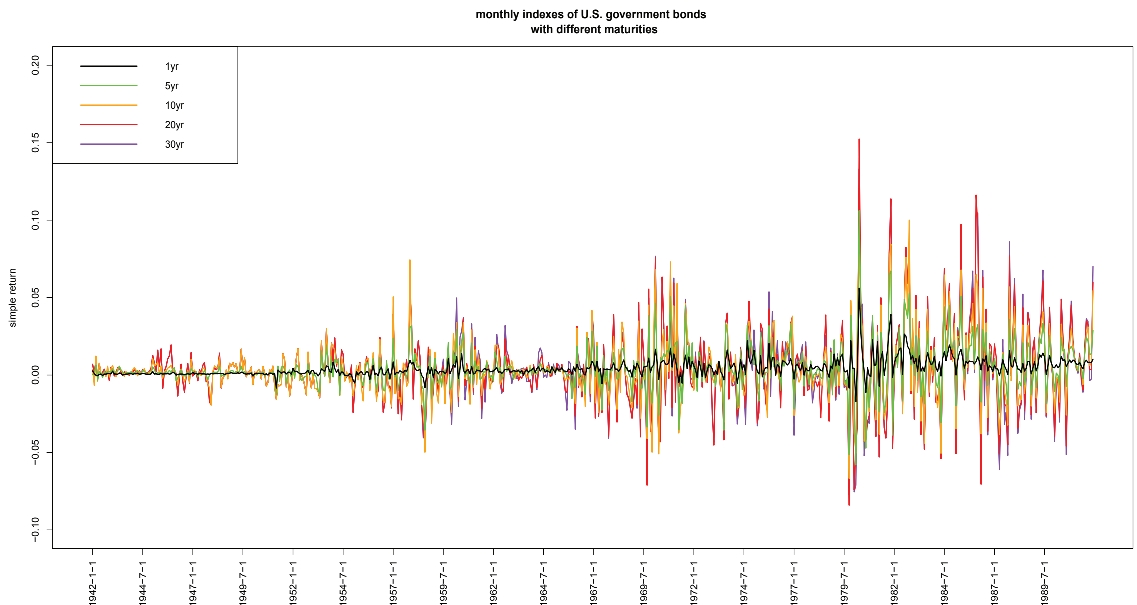


Figure 3.5: Monthly indexes of U.S. government bonds with maturities in 30 years, 20 years, 10 years, 5 years, and 1 year

Lag 1					Lag 2				
0.08	0.06	0.10	0.12	0.16	-0.02	-0.01	-0.00	-0.03	0.03
0.09	0.07	0.12	0.13	0.18	-0.02	-0.01	-0.01	-0.04	0.02
0.07	0.06	0.08	0.12	0.18	-0.01	0.00	0.01	-0.02	0.08
0.13	0.10	0.14	0.13	0.22	-0.04	-0.03	-0.01	-0.04	0.08
0.17	0.15	0.22	0.22	0.40	-0.03	-0.01	0.02	0.02	0.23

Table 3.9: Sample Cross-Correlation Matrices of Monthly Simple Returns of Five Indexes of U.S. Government Bonds: January 1942 to December 1991

Latent factors factor model Lam et al. (2011b) is considered to capture the linear dynamic panel dependence of the bond indexes $\boldsymbol{r}_t$:

$$\begin{aligned} \boldsymbol{r}_t &= \boldsymbol{f}_t + \boldsymbol{\epsilon}_t \\ &= A\boldsymbol{x}_t + \boldsymbol{\epsilon}_t \end{aligned} \tag{3.18}$$

Where $\boldsymbol{x}_t$ is $r \times 1$ latent process with $r < p$, A is $p \times r$ unknown constant matrix, and $\boldsymbol{\epsilon}_t \sim \text{WN}(\mathbf{0}_p, \Sigma_\epsilon)$. More detailed assumptions are described in Lam et al. (2011a).

Let $L = \sum_{k=1}^{k=k_0} \Sigma_r(k)\Sigma_r(k)'$, where $\Sigma_r(k) = cov(\boldsymbol{r}_{t+k}, \boldsymbol{r}_{t+k})$.

We choose the number of factor using the method propose by Lam et al. (2011b). Denote the j-th eigen value of $\hat{L}$ as $\hat{\lambda}_j$, then if we have strong r factors, the following holds:

$$\frac{\lambda_{j+1}}{\lambda_j} \asymp 1, j = 1, 2, \ldots, r-1, \text{ and } \frac{\lambda_{j+1}}{\lambda_j} = O_p(\frac{1}{n}) \tag{3.19}$$

Suggested by the eigen value of $\hat{L}$ (See Figure 3.6), we select the factor number $r = 1$.

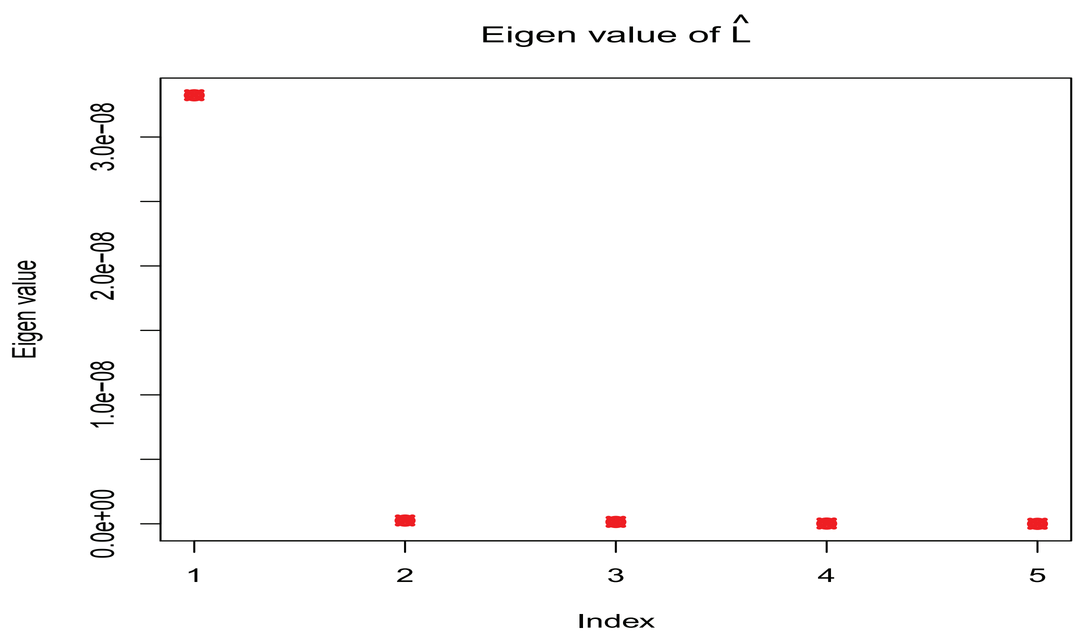


Figure 3.6: Eigen value of $\hat{L}$

We tried different k_0, and the corresponding loading matrix $A_{p\times 1}$ is shown in Table 3.10. The first factor under different k_0 is also shown in Figure 3.7. After we calculate

$\boldsymbol{f}_t$ in Model 3.18, we can also compute the residual $\hat{\boldsymbol{\epsilon}}_t$. Now we want to test whether there is any non-concurrent cross correlation among $\{\hat{\boldsymbol{\epsilon}}_t\}$. We perform the bootstrap test, and report the p values in table 3.11. Note that s_t and b_t are defined as lag range and block size. It seems that the results from $k_0 = 1, 2$ are very similar, and the assumption of uncorrelated residuals is reasonable.

$k_0 = 1$	$k_0 = 2$
-0.53	-0.53
-0.57	-0.57
-0.38	-0.38
-0.43	-0.44
-0.25	-0.25

Table 3.10: Constant Matrix A, each column represents the solution with different k_0

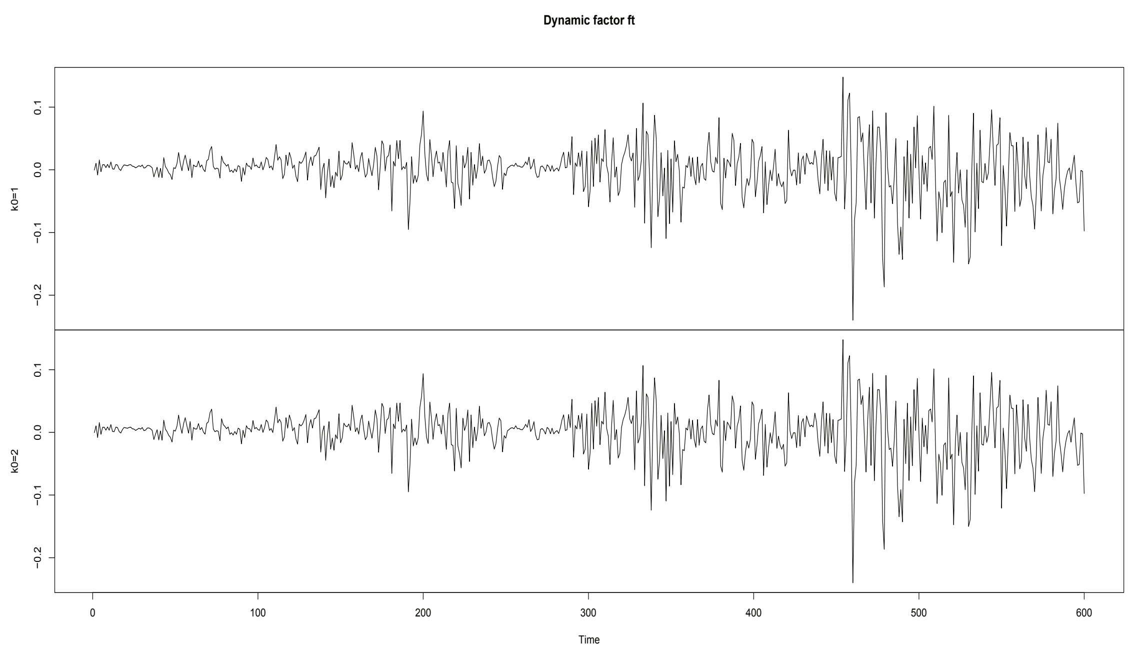


Figure 3.7: First factor from different k_0, the upper one is from when $k_0 = 1$ the bottom one is from when $k_0 = 2$

		$b_t = 10$	$b_t = 15$	$b_t = 20$	$b_t = 25$	$b_t = 30$
$k_0 = 1$	$s_t = 5$	5.41%	5.11%	5.71%	7.41%	8.51%
	$s_t = 7$	5.71%	6.31%	9.31%	8.81%	12.31%
	$s_t = 10$	6.71%	7.61%	10.01%	9.31%	12.41%
		$b_t = 10$	$b_t = 15$	$b_t = 20$	$b_t = 25$	$b_t = 30$
$k_0 = 2$	$s_t = 5$	5.21%	5.81%	6.01%	5.41%	9.31%
	$s_t = 7$	7.41%	6.71%	8.71%	8.91%	11.71%
	$s_t = 10$	8.21%	7.71%	10.61%	11.01%	13.61%

Table 3.11: H0: The residual is uncorrelated with each other. P value of maximum CCF over lags $-S_T, -S_T+1, \ldots, -1, 1, \ldots, S_T-1, S_T$ from bootstrap

3.6 Proof

3.6.1 Proof of Theorem 5

Proof. Without loss of generality, assume the true lead of $\boldsymbol{X}_2$ over $\boldsymbol{X}_1$ is $k^\star > 0$. Let $m_T = T^\beta$, where $\beta \in (0,1)$, define the m-dependent approximation series: $\tilde{X}_j(i) = \mathbb{E}(X_j(i) \mid F_{i-m_T})$, where $j = 1, 2$. Define:

$$R_{T,k} = \sum_{k+1}^{T} \left(X_1(i)X_2(i-k+k^\star) - \gamma_{12}(k)\right),$$

$$\tilde{R}_{T,k} = \sum_{k+1}^{T} \left(\tilde{X}_1(i)\tilde{X}_2(i-k+k^\star) - \tilde{\gamma}_{12}(k)\right),$$

where $\gamma_{12}(k) = \mathbb{E}x_1(i)x_2(i-k+k^\star)$ and $\tilde{\gamma}_{12}(k) = \mathbb{E}\tilde{x}_1(i)\tilde{x}_2(i-k+k^\star)$.

In order to prove $\max_{-T+1<k<T-1} |R_{T,k}| = o_p(\sqrt{T\log T})$, it's sufficient to show (a) and (b) listed below:

(a) $\max_{-T+1<k<T-1} \left|R_{T,k} - \tilde{R}_{T,k}\right| = o_p(\sqrt{T\log T})$

(b) $\lim_{T\to\infty} P\left(\max_{-T<k<T} \left|\tilde{R}_{T,k}\right| \le c_p\sqrt{T\log T}\right) = 1$

<u>(a): Effect of m dependent approximation:</u>

By Proposition 8 in Xiao and Wu (2014), we have:

$$\left\| R_{T,k} - \tilde{R}_{T,k} \right\|_{p/2} \leq C_p \sqrt{T} k_p \Delta_p(m_n + 1), \text{ where } k_p = \|X_0\|_p$$

Therefore, $\forall \delta_T > 0$ which goes to 0 slowly,

$$\begin{aligned}
&\sum_{k=1-T}^{T-1} P(|R_{T,k} - \tilde{R}_{T,k}| > \delta_T \sqrt{TlogT}) \\
&\leq \sum_{k=1-T}^{T-1} \frac{E|R_{T,k} - \tilde{R}_{T,k}|^{p/2}}{(\delta \sqrt{TlogT})^{p/2}} \\
&= \sum_{k=1-T}^{T-1} \frac{C_p m_n^{-\alpha' p/2} T^{p/4}}{T^{p/4} \log(T)^{p/4} \delta_T^{p/2}} \\
&= \frac{2C_p T^{1-\alpha' p\beta/2}}{\log(T)^{p/4} \delta_T^{p/2}}
\end{aligned}$$

With the assumption that $\alpha' p/2 > 1$, we can always find $\beta \in (0,1)$, such that $1 - \dfrac{\alpha' \beta p}{2} < 0$. Therefore:

$$\begin{aligned}
&P\left(\max_{-T+1<k<T-1} \left| R_{T,k} - \tilde{R}_{T,k} \right| > \delta_T \sqrt{TlogT} \right) \\
&< \sum_{k=1-T}^{T-1} P\left(\left| R_{T,k} - \tilde{R}_{T,k} \right| > \delta_T \sqrt{TlogT} \right) \\
&= \frac{2C_p T^{1-\alpha' p\beta/2}}{\delta_T^{p/2} \log(T)^{p/4}} \xrightarrow{T\to\infty} 0
\end{aligned}$$

<u>(b): Upper bound of the m dependent series:</u>

We prove (b) by splitting k into two cases: (1) $k \geq 3m_T$; (2) $k < 3m_T$.

When $k \geq 3m_T$:

(i) We split [k+1,T] into following blocks, with size $k - m_T$:

$$H_j : [k + (j-1)(k - m_T) + 1, k + j(k - m_T)], (1 \leq j < w_T)$$

$$H_{w_T} : [k + (w_T - 1)(k - m_T) + 1, T],$$

where w_T is the smallest integer that $k + w_T(k - m_T) \geq T$.

(ii) Furthermore we split H_j into smaller blocks with size $2m_T$

$$K_{j,l} : [k + (j-1)(k - m_T) + (l-1)2m_T + 1, k + (j-1)(k - m_T) + 2lm_T],$$
$$(1 \leq l \leq v_j - 1)$$
$$K_{j,v_j} : [k + (j-1)(k - m_T) + (v_j - 1)2m_T + 1, k + (j-1)(k - m_T) + |H_j|],$$

where v_j is the smallest integer that $2v_j m_T \geq |H_j|$.

Define $U_{k,j,l} = \sum_{i \in K_{j,l}} \tilde{X}_1(i)\tilde{X}_2(i - k + k^\star) - \tilde{\gamma}(k)$

$$\tilde{R}^{u,o}_{T,k} = \sum_{u \equiv j (mod\ 3)} \sum_{o \equiv l (mod\ 2)} U_{k,j,l}$$

Note that $\tilde{R}^{u,o}_{n,k}(u = 1, 2, 3, o = 1, 2)$ is sum of independent random variables. Besides, we observe that $\tilde{X}_1(i)$ is independent of $\tilde{X}_2(i - k + k^\star)$ because $k \geq 3m_n$. Therefore, the upper bound of $\|U_{k,j,l}\|_p$ is:

$$\begin{aligned}
&\|U_{k,j,l}\|_p \\
\leq &\sqrt{2m_T} \left\| \tilde{X}_1(i)\tilde{X}_2(i - k + k^\star) \right\|_p \\
\leq &\left\| \tilde{X}_1(i) \right\|_p \left\| \tilde{X}_2(i - k + k^\star) \right\|_p \\
= &\sqrt{2m_T} k_p^2
\end{aligned}$$

Furthermore, let $\lambda > 0$, we have:

$$\begin{aligned}
&P\left(\left| \tilde{R}_{T,k} \right| > 6\lambda\sqrt{TlogT} \right) \\
< &\sum_{u=0}^{2}\sum_{o=0}^{1} P\left(\tilde{R}^{u,o}_{T,k} \geq \lambda\sqrt{TlogT} \right) \\
\leq &\sum_{u=0}^{2}\sum_{o=0}^{1} \left\{ P(U_{k,j,l} \geq \lambda\sqrt{T \log T}) + \exp\left\{ -\frac{c_p \log T}{m_T} \right\} + \left(\frac{c_p T m_T^{p/2}}{\sqrt{TlogT}(\sqrt{T})^{p-1}} \right)^{c_p\sqrt{\log T}} \right\} \\
= &I + II + III
\end{aligned} \tag{3.20}$$

Eq. (3.20) is held by Corollary 1.7 in Nagaev (1979), and we resolve all constants into c_p.

For II, $II = C_P \frac{T}{2m_T} \left(\frac{1}{T}\right)^{T^\beta} = C_p T^{1-T^\beta-\beta}$, which implies that

$$\sum_{k=3m_T}^{T} II \leq C_p T^{2-\beta-T^\beta} \xrightarrow{T\to\infty} 0 \tag{3.21}$$

On the other hand, for I and III:

$$I \leq \frac{T}{2m_T} \frac{(\sqrt{2m_T} k_p^2)^p}{(\sqrt{TlogT})^p} = \frac{T^{(1-p/2)(1-\beta)}}{\log T},$$

$$III \leq C_p T^{1-\beta} \left(\frac{T^{1+(\beta-1)p/2}}{\sqrt{\log T}}\right)^{c_p\sqrt{\log T}}.$$

Since $p > 4$ we could find $\beta \in (0, 1)$, such that:

$$1 + (1 - p/2)(1 - \beta) < 0 \text{ and } 1 + (\beta - 1)p/2 < 0.$$

Therefore,

$$\sum_{k=3m_T}^{T} I \leq \frac{T^{1+(1-p/2)(1-\beta)}}{\log T} \xrightarrow{T\to\infty} 0, \tag{3.22}$$

$$\sum_{k=3m_T}^{T} III = C_p T^{2-\beta} \left(\frac{T^{1+(\beta-1)p/2}}{\sqrt{\log T}}\right)^{c_p\sqrt{\log T}} \xrightarrow{T\to\infty} 0. \tag{3.23}$$

Eq.3.21, Eq. 3.22 and Eq. 3.23 implies that:

$$\sum_{k=3m_T}^{T} P\left(\left|\tilde{R}_{T,k}\right| > 6\lambda\sqrt{TlogT}\right)$$

When $k < 3m_T$:

Split $[k+1,T]$ into blocks with size $4m_T$.

$$H_j : [k+(j-1)(k-m_T)+1, k+j(k-m_T)], (1 \leq j \leq w_T)$$
$$H_{w_n} : [k+(w_T-1)(k-m_T)+1, T],$$

where w_T is the smallest integer that $k+w_T(k-m_T) \geq T$.

$$\tilde{R}^o_{T,k} = \sum_{i \in H_{2N-1}} \tilde{X}_1(i)\tilde{X}_2(i-k+k^\star) - \tilde{\gamma}_{12}(k)$$
$$\tilde{R}^e_{T,k} = \sum_{i \in H_{2N}} \tilde{X}_1(i)\tilde{X}_2(i-k+k^\star) - \tilde{\gamma}_{12}(k).$$

Obviously, $\tilde{R}^o_{T,k}$ and $\tilde{R}^e_{T,k}$ is sum of independent random variables.

Similarly as the case when $k \leq 3m_T$, we show that

$$\sum_{k=1}^{3m_n-1} P\left(\left|\tilde{R}_{T,k} - \tilde{\gamma}_{12}\right| \geq 2\lambda\sqrt{T\log T}\right) \xrightarrow{T\to\infty} 0.$$

Combine both $k > 3m_T$ and $k \leq 3m_T$ cases, we have:

$$\begin{aligned}
&P\left(\max_{-T<k<T} |\hat{\gamma_{12}}(k) - E\hat{\gamma_{12}}(k)| > c_p\sqrt{\frac{logT}{T}}\right) \\
\leq &\sum_{k=-T+1}^{T-1} P\left(|\hat{\gamma_{12}}(k) - E\hat{\gamma_{12}}(k)| > c_p\sqrt{\frac{logT}{T}}\right) \\
= &\sum_{k>3m_T} P\left(|\hat{\gamma_{12}}(k) - E\hat{\gamma_{12}}(k)| > c_p\sqrt{\frac{logT}{T}}\right) + \sum_{k\leq 3m_T} P\left(|\hat{\gamma_{12}}(k) - E\hat{\gamma_{12}}(k)| > c_p\sqrt{\frac{logT}{T}}\right) \\
&\xrightarrow{T\to\infty} 0
\end{aligned}$$

i.e.

$$\lim_{T\to\infty} P\left(\max_{-T<k<T} |\hat{\gamma_{12}}(k) - E\hat{\gamma_{12}}(k)| \leq c_p\sqrt{\frac{logT}{T}}\right) \xrightarrow{T\to\infty} 1$$

$\square$

Proof of Corollary 1

Proof. First we observe that $|\gamma_{12}(k)| \overset{k\to\infty}{\to} 0$.

Here is a short proof of the above observation:

$$\begin{aligned}
|\mathbb{E}\left(x_1(k)x_2(0)\right)| &= \left|\mathbb{E}(\sum_{j=-\infty}^{k} P^j x_1(k) \sum_{w=-\infty}^{0} P^w x_2(0))\right| \\
&= \left|\mathbb{E}(\sum_{j=0}^{-\infty} P^j x_1(k) P^j x_2(0))\right| \\
&\le \sum_{j=0}^{\infty} \mathbb{E}\left|P^j x_1(k)||P^j x_2(0)\right| \\
&\le \sum_{j=0}^{\infty} \delta_2(k+j)\delta_2(k) \\
&\le \Theta(0)\Theta(k) \\
&= O(k^{-\alpha})
\end{aligned}$$

Therefore, there exists such L that when $\mid k \mid > L, \gamma_{12}(k) < \dfrac{1-c}{2}$.

Without losing generality, we assume $\gamma_{12}^\star = \max_{k\in Z}\{|\gamma_{12}(k)|\} = 1$ and $k^\star = cT$ for some $c \in (0,1)$.

On the event

$$\left\{ \max_{-T<k<T} \|\hat{\gamma}_{12}(k) - E(\hat{\gamma}_{12}(k))\| \le c_p\sqrt{\frac{\log T}{T}} \right\}.$$

1. $\hat{\gamma}_{12}(k^\star) \ge \dfrac{T-k^\star}{T}\gamma_{12}(k^\star) - c_p\sqrt{\dfrac{\log T}{T}} = (1-c) - c_p\sqrt{\dfrac{\log T}{T}}$
   so $\lim_{T\to\infty} \hat{\gamma}_{12}(k^\star) \ge (1-c)$.

2. if $|k-k^\star| > L$, $|\hat{\gamma}_{12}(k)| < (1-c)/2 + c_p\sqrt{\dfrac{\log T}{T}}$
   $\lim_{T\to\infty} \hat{\gamma}_{12}(k) \le (1-c)/2$.

3. if $|k-k^\star| \le L, |\hat{\gamma}_{12}(k)| < \dfrac{T-k}{T}|\gamma_{12}(k)| + c_p\sqrt{\dfrac{\log T}{T}} < \dfrac{T-L+k^\star}{T} + c_p\sqrt{\dfrac{\log T}{T}}$
   $\lim_{T\to\infty} \hat{\gamma}_{12}(k) \le (1-c)$.

Therefore, for $k \neq k^\star$

$$\lim_{T\to\infty} \hat{\gamma}_{12}(k^\star) > \lim_{T\to\infty} \hat{\gamma}_{12}(k)$$

This implies that as if the sample size T goes to infinity, maximum cross correlation method can asymptotically identify the true lead $k^\star$.

For $k^\star < 0$, with the same argument, the theorem still holds true. $\square$

3.6.2 Proof of Theorem 6

Proof. Define $\boldsymbol{\epsilon}_i^B = (\epsilon_{(i-1)B+1}, \epsilon_{(i-1)B+2}, \ldots, \epsilon_{iB\vee T})$, thus $\boldsymbol{x}_j$ is:

$$\boldsymbol{x}_j(i) = g_2(\infty, ..., \boldsymbol{\epsilon}_{i-1}^B, \boldsymbol{\epsilon}_i^B), (j = 1, 2)$$

With similar m_T-dependence approach as the proof in Theorem 5, define

$$\tilde{\boldsymbol{x}}_j(i) = E\left(\boldsymbol{x}_j(i)|(\boldsymbol{\epsilon}_{i-m_n}^B, ..., \boldsymbol{\epsilon}_{i-1}^B)\right)$$

It's not difficult to see that for $\{\tilde{\boldsymbol{x}}_j(i)\}$,

$$\begin{cases} \bar{\Delta}_p(m_n + 1) = O((m_n B)^{-\alpha'}) \\ \bar{\Theta}(m_n) = O((m_n B)^{-\alpha}) \end{cases}$$

Without losing of generality, assume $k^\star > 0$. Define:

$$\bar{R}_{N,k} = \sum_{k+1}^{N} = \bar{x}_1(i)\bar{x}_2(i - k + k^\star) - \bar{\gamma}_{12}(k),$$
$$\tilde{\bar{R}}_{N,k} = \sum_{k+1}^{N} = \tilde{\bar{x}}_1(i)\tilde{\bar{x}}_2(i - k + k^\star) - \tilde{\bar{\gamma}}_{12}(k),$$

where $\bar{\gamma}_{12}(k) = E(\bar{x}_1(i)\bar{x}_2(i - k + k^\star)), \tilde{\bar{\gamma}}_{12}(k) = E(\tilde{\bar{x}}_1(i)\tilde{\bar{x}}_2(i - k + k^\star))$

Similarly, by Proposition 8 Xiao and Wu (2014), together with the fact $\|\bar{x}_j(i)\|_p \leq BK_p$, we have:

$$\left\|\bar{R}_{N,k} - \tilde{\bar{R}}_{N,k}\right\|_{p/2} \leq c_p\sqrt{N}(mB)^{-\alpha'} BK_p,$$

For any $\delta > 0$, which converges to 0 slowly,

$$\begin{aligned}
& P\left(\mid \bar{R}_{N,k} - \tilde{\bar{R}}_{N,k} \mid > \delta B \sqrt{N \log N} \right) \\
\leq & \left(c_p K_p \frac{\sqrt{N} (mB)^{-\alpha'} B}{\delta B \sqrt{N \log N}} \right)^{p/2} \\
= & \left(\frac{c_p K_p m_n^{-\alpha'} B^{-\alpha'}}{\delta \sqrt{\log(N)}} \right)^{p/2}
\end{aligned}$$

Assume $B = T^{\eta}$, thus $N = T^{1-\eta}$

$$\begin{aligned}
& \sum_{k=0}^{N} P(\left| \bar{R}_{N,k} - \tilde{\bar{R}}_{N,k} \right| > \delta B \sqrt{N \log N}) \\
\leq & \frac{c_{2,p} N^{1-\alpha' p\beta/2} B^{-\alpha' p/2}}{(\delta \sqrt{\log N})^{p/2}}
\end{aligned}$$

Under the condition of Theorem 6, we can always find $\beta \in (0, 1)$, s.t.

$$\frac{N^{1-\alpha p\beta/2} B^{-\alpha p/2}}{(\log N)^{p/4}} \xrightarrow{T\to\infty} 0.$$

Therefore:

$$\begin{aligned}
& P(\max_{0<k<N} \left| \bar{R}_{N,k} - \tilde{\bar{R}}_{N,k} \right| > \delta B \sqrt{N \log N}) \\
\leq & \sum_{k=0}^{N} P(\left| \bar{R}_{N,k} - \tilde{\bar{R}}_{N,k} \right| > \delta B \sqrt{N \log N}) \\
& \xrightarrow{T\to\infty} 0
\end{aligned}$$

i.e., $\left\| \bar{R}_{N,k} - \tilde{\bar{R}}_{N,k} \right\| = o_p \left(B \sqrt{N log N} \right)$.

Similar as the proof of Theorem 5, we can show that

$$\sum_{k=0}^{N} P\left(\left| \tilde{\bar{R}}_{N,k} - \tilde{\bar{\gamma}}_{12}(k) \right| \geq B \sqrt{N log N} \right) = 0.$$

Therefore $\lim_{T\to\infty} P\left(\max_{N<k<N} |\bar{\gamma}_{12}(j) - E\bar{\gamma}_{12}(j)| \leq c_p \sqrt{\frac{log N}{N}} \right)$. $\square$

3.6.3 Proof of Corollary 2

Proof. First, Let's compute $E(\bar{\gamma}_{12}(j))$

$$
\begin{aligned}
E(\bar{\gamma}_{12}(j)) &= \frac{1}{T}E\left(\sum_{i=1+j^-}^{N-j^+}\bar{x}_1(i+j)\bar{x}_2(i+k^\star)\right)\\
&= \frac{1}{T}\sum_{i=1+j^-}^{N-j^+}E\left\{\sum_{t_1=(i+j-1)B+1}^{(i+j)B\vee T}x_1(t_1)\sum_{t_2=(i-1)B+1}^{(i)B\vee T}x_2(t_2+k^\star)\right\}\\
&= \frac{1}{T}\sum_{i=1+j^-}^{N-j^+}\left\{\sum_{k=-jB}^{-(j-1)B}\gamma_{12}(k+k^\star)((-j+1)B-k+1)\right. \qquad (3.24)\\
&\quad\left.+\sum_{k=-(j+1)B}^{-jB}\gamma_{12}(k+k^\star)((j+1)B+k)\right\}\\
&= \begin{cases} 0, & \text{as } B\to\infty \text{ and } k^\star\notin[(j-1)B,(j+1)B+1];\\ \neq 0, & \text{as } B\to\infty \text{ and } k^\star\in[(j-1)B,(j+1)B+1]. \end{cases}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&E(\bar{\gamma}_{12}(j-1))+E(\bar{\gamma}_{12}(j))+E(\bar{\gamma}_{12}(j-1))\\
=&\frac{N-j}{N}\{(B+1)\sum_{k=-(j+1)B}^{-(j-1)B}\gamma_{12}(k+k^\star)\\
&+\underline{\sum_{k=-(j-1)B}^{-(j-2)B}\gamma_{12}(k+k^\star)((-j+2)B-k+1)+\sum_{k=-(j+2)B}^{-(j+1)B}\gamma_{12}(k+k^\star)((j+2)B+k)\}}
\end{aligned}
$$

If $k^\star\in[(j-1)B,(j+1)B+1]$, then the underlined part goes to 0 as B $\to\infty$. This implies that

$$E(\bar{\gamma}_{12}(j-1))+E(\bar{\gamma}_{12}(j))+E(\bar{\gamma}_{12}(j-1))\xrightarrow{B\to\infty}2\pi f_{12}(0).$$

By assumption that $f_{12}(0)\neq 0$, therefore, at least one of the three terms are nonzero.

Thus, on the event

$$\max_{-N<j<N}|\bar{\gamma}_{12}(j)-E\bar{\gamma}_{12}(j)|\leq c_p\sqrt{\frac{logN}{N}}$$

$$\begin{cases} |\bar{\gamma}_{12}(j)| > E\bar{\gamma}_{12}(j) - c_p\sqrt{\dfrac{logN}{N}} > E\bar{\gamma}_{12}(j) > 0, & j \in \{\bar{k}^\star - 1, \bar{k}^\star, \bar{k}^\star + 1\}; \\ |\bar{\gamma}_{12}(j)| \leq c_p\sqrt{\dfrac{logN}{N}} \xrightarrow{T\to\infty} 0, & \text{otherwise.} \end{cases}$$

This implies that:

$$\lim_{T\to\infty} P\left(\bar{j} \in \{\bar{k}^\star - 1, \bar{k}^\star, \bar{k}^\star + 1\}\right) = 1$$

□

3.6.4 Proof of Lemma 2

Proof. The density of χ^2_m distribution is:

$$f(x) = \frac{x^{m/2-1}\exp\{x/2\}}{2^{m/2}\Gamma(m/2)},$$

where $\Gamma(x) = \int_0^\infty s^{x-1}e^{-s}ds$.

First, we will show that $F(\chi^2_m > x) \sim 2f(x)$. (Here the sign $\sim$ means that the quotient of the two functions converges to 1 as $x \to +\infty$.)

$$\begin{aligned} &\lim_{x\to\infty} \frac{F(\chi^2_m > x)}{x^{m/2-1}\exp\{x/2\}\dfrac{1}{2^{m/2}\Gamma(m/2)}} \\ &= \lim_{x\to\infty} \frac{\int_x^\infty t^{m/2-1}\exp\{-t/2\}dt}{x^{m/2-1}\exp\{-x/2\}} \\ &\overset{L'H}{=} \lim_{x\to\infty} \frac{-x^{m/2-1}\exp\{-x/2\}}{(m/2-1)x^{m/2-2}\exp\{-x/2\} - 0.5x^{m/2-1}\exp\{-x/2\}} \\ &= 2 \end{aligned}$$

Define $\overline{F} = F(\chi^2_m > x)$, and $a(x) = \dfrac{\overline{F}}{f(x)} = 2$, it's easy to see $a(x)' = 0$ so $\overline{F}$ has representation

$$\overline{F} = \exp\left\{-\int_{-\infty}^{x} \frac{1}{a(t)}dt\right\}.$$

It implies that F is Von Moses function. If we interpret $2f(x)$ as tail of some *df* G, then by Proposition 3.3.28 in Embrechts et al. (1997), G and F have same norming constant c_n (scale) and d_n (center).

By Proposition 3.3.25 in Embrechts et al. (1997)

$$d_n = \inf\left\{x \in R, \overline{G}(x) \leq \frac{1}{n}\right\} \text{ and } c_n = a(d_n) \equiv 2 \tag{3.25}$$

Hence look at the solution of $\log(\overline{G}(x)) = -\log(n)$, i.e.

$$\left(\frac{m}{2}-1\right)\log\left(\frac{x}{2}\right) - \frac{x}{2} - \log\left(\Gamma(m/2)\right) = -\log(n)$$

$d_n = 2\log(n) + (m-2)\log\log(n) - 2\log\left(\Gamma\left(\frac{m}{2}\right)\right) + \gamma_n$, where $\gamma_n = o(1)$.

Furthermore, we could show that the order of $\gamma_n = O\left(\frac{\log\log(n)}{\log(n)}\right)$.

Plug in d_n in Eq. 3.25,

$$\begin{aligned}
&\log\log(n) - \frac{\gamma_n}{m-2} \\
=&\log\left(\log(n) + \frac{m-2}{2}\log\log(n) - \log\Gamma\left(\frac{m}{2}\right) + \frac{\gamma_n}{2}\right) \\
=&\log\log(n) + \log\left(1 + \frac{(m-2)\log\log(n) - \log(\Gamma(m/2)) + \gamma_n/2}{2\log(n)}\right)
\end{aligned}$$

i.e. $-\frac{\gamma_n}{m-2} \approx \frac{(m-2)\log\log(n) - \log(\Gamma(m/2)) + \gamma_n/2}{2\log(n)}$

which implies that $\gamma_n = O\left(\frac{\log\log(n)}{\log(n)}\right)$. □

Remark 14. *Let* $z_{n,m} = z + 2\log(n) + (m-2)\log\log(n) - 2\log(\Gamma(\frac{m}{2}))$. *By Theorem 2,* $P(M_{m,n} \leq z_{n,m}) = (1-\overline{F}(z_{n,m}))^n = e^{e^{-z/2}}$, *take logarithms*

$$n\log(1-\overline{F}(z_{n,m})) = e^{-z/2}$$

since $\log(1-\overline{F}(z_{n,m})) \sim \overline{F}(z_{n,m})$,*this implies that* $n\overline{F}(z_{n,m}) \sim e^{-z/2}$

3.6.5 Proof of Lemma 3

Proof. Suppose we select d distinct numbers $i_1, i_2, \dots, i_d$ from $1, 2, 3, \dots, n$, where n is an integer. What is the total possibilities that there exist at least two numbers whose difference are less than m, i.e $|i_h - i_j| < m$, for any $h, j \in \{1, 2, 3, \dots, d\}$?

The answer is:

$$\binom{n}{d} - \binom{n+m+d-md-1}{d} \leq mn^{d-1}.$$

Let $N = \sum_k \mathbb{1}\{Q_k > z_{n,m}\}$. Thus, $P(M_m \leq z_{n,m}) = P(N = 0)$. It's equivalent to prove $N \xrightarrow[n\to\infty]{\mathbf{D}} Poisson(\exp(-z/2))$.

Using moment method, it's sufficient to prove

$$\mathrm{E}(N(N-1)...(N-d+1)) \to \exp\left(-\frac{zd}{2}\right) \tag{3.26}$$

i.e. $\sum\limits_{1\leq i_1\leq i_2,\dots i_d\leq n-m+1} P\left(Q_{i_1} \geq z_{n,m}, \dots Q_{i_d} \geq z_{n,m}\right) \xrightarrow{n\to\infty} \dfrac{\exp(-zd/2)}{d\,!}$.

We'll prove the result by induction.

- When $d = 1$, directly from Theorem 2,

$$\mathrm{E}(N) = \binom{n-m+1}{1} P(\chi^2_m \geq z_{n,m}) \xrightarrow{n\to\infty} \lambda = \exp\{-z/2\}$$

- When $d = 2$, W.L.G, $1 < i_1 < i_2 < n - m + 1$,

$$\begin{aligned}\mathrm{E}\frac{N(N-1)}{2\,!} &= \sum_{i_2-i_1\geq m} P\left(Q_{i_1} \geq z_{n,m}, Q_{i_2} \geq z_{n,m}\right) + \sum_{i_2-i_1<m} P\left(Q_{i_1} \geq z_{n,m}, Q_{i_2} \geq z_{n,m}\right)\\ &= \mathrm{I} + \mathrm{II}\end{aligned}$$

For I, by independence

$$\lim_{n\to\infty} \sum_{i_2-i_1\geq m} P(Q_{i_1} \geq z_{n,m}, Q_{i_2} \geq z_{n,m}) = \lim_{n\to\infty} \frac{n\,!}{2} P^2(Q_{i_1} \geq z_{n,m}) \to \frac{\lambda^2}{2\,!}$$

For II, with the help of Figure 3.8, let $\delta_n = \dfrac{\log\log(n)}{2}$

$$\begin{aligned}
&P(Q_{i_1} \geq z_{n,m}, Q_{i_2} \geq z_{n,m}) \\
\leq& P(Z^2_{i_1+1} + ...Z^2_{i_1+m-1} \geq z_{n,m} - \delta_n, Q_{i_1} \geq z_{n,m}, Q_{i_2} \geq z_{n,m}) \\
&+ P(Z^2_{i_1+1} + ...Z^2_{i_1+m-1} < z_{n,m} - \delta_n, Q_{i_1} \geq z_{n,m}, Q_{i_2} \geq z_{n,m}) \\
\leq& P(Z^2_{i_1+1} + ...Z^2_{i_1+m-1} \geq z_{n,m} - \delta_n) + P(Q_{i_1} \geq z_{n,m}, Z^2_{i_1+m} + ...Z^2_{i_2+m} \geq \delta_n) \\
\leq& P(\chi^2_{m-1} \geq z_{n,m} - \delta_n) + P(Q_{i_1} \geq z_{n,m}) P(\chi^2_m \geq \delta_n) \\
\leq& P\left(\chi^2_{m-1} \geq z_{n,m-1} + \frac{\log\log(n)}{2}\right) + P\left(Q_{i_1} \geq z_{n,m}\right) P\left(\chi^2_m \geq \delta_n\right)
\end{aligned}$$

This implies that

$$\begin{aligned}
\mathrm{II} &\leq n\{P(\chi^2_{m-1} \geq z_{n,m-1} + \frac{\log\log(n)}{2}) + P(Q_{i_1} \geq z_{n,m}) P(\chi^2_m \geq \delta_n)\} \\
&\leq nP(\chi^2_{m-1} \geq z_{n,m-1} + \frac{\log\log(n)}{2}) + \exp\{-\lambda\} P(\chi^2_m \geq \delta_n) \\
&\xrightarrow{n\to\infty} 0
\end{aligned}$$

Therefore, we prove that when $d = 2$, $\mathrm{E}\dfrac{N(N-1)}{2!} \to \dfrac{\lambda^2}{2!}$

- Assume $i_1 < \cdots < i_l$, Define

$$\mathcal{L}_l = \{(i_k, i_{k+1}), k = 1, \ldots, l-1 \mid Q_{i_j} \geq z_{n,m}, \forall j = 1, \ldots, l\}$$

Suppose for $d-1, d-2$, Eq. (3.26) holds. which follows that if there exists some adjacent pair $(i_k, i_{k+1}) \in \mathcal{L}_l$ $l = d-2, d-1$ satisfying $i_{k+1} - i_k < m$,

$$\sum_{i_{k+1}-i_k<m, i_1, \ldots, i_l} P(Q_{i_1} \geq z_{n,m}, \ldots, Q_{i_l} \geq z_{n,m}) = 0$$

And if $\forall (i_k, i_{k+1}) \in \mathcal{L}_l$ $l = d-2, d-1$ satisfying $i_{k+1} - i_k \geq m$

$$\sum_{i_{k+1}-i_k \geq m, \forall k=1,\ldots,l} P(Q_{i_1} \geq z_{n,m}, \ldots, Q_{i_l} \geq z_{n,m}) = \frac{\exp\{-lz/2\}}{l!}$$

For $d \geq 3$, Define

$$\mathcal{I}_d = \{1 \leq i_1 < i_2 < \cdots < i_d \leq n\}$$
$$S_{k,d} = \{1 \leq i_1 < i_2 < ... < i_d \mid i_{h+1} - i_h < m, \text{for} 1 \leq h \leq k, i_{k+1} - i_k \geq m\}$$
$$S_d = \{i_1 < i_2 < \cdots < i_d \mid i_{h+1} - i_h \geq m, \forall h = 1, 2, \ldots, d\}$$

$$\begin{aligned}
&\mathrm{E}\frac{N(N-1)...(N-d+1)}{d!} \\
=&\sum_{\mathcal{I}_d} P(Q_{i_1} \geq z_{n,m}, \ldots, Q_{i_d} \geq z_{n,m}) \\
=&\underbrace{\sum_{\mathcal{I}_d \in S_{1,d} \cap S_d^c} P(Q_{i_1} \geq z_{n,m}, \ldots, Q_{i_d} \geq z_{n,m})}_{\text{A}} + \underbrace{\sum_{k=2}^{d-1} \sum_{\mathcal{I}_d \in S_{k,d}} P(Q_{i_1} \geq z_{n,m}, \ldots, Q_{i_d} \geq z_{n,m})}_{\text{B}} \\
&+ \underbrace{\sum_{\mathcal{I}_d \in S_{d,d}} P(Q_{i_1} \geq z_{n,m}, \ldots, Q_{i_d} \geq z_{n,m})}_{\text{C}} + \underbrace{\sum_{\mathcal{I}_d \in S_d} \geq z_{n,m}, \ldots, Q_{i_d} \geq z_{n,m})}_{\text{D}}
\end{aligned}$$

We will calculate each term:

$$\begin{aligned}
A =&\sum_{i_1} \sum_{(i_2 \ldots i_d) \in S_{d-1}^c} P(Q_{i_1} \geq z_{n,m}) P(Q_{i_2} \geq z_{n,m}, \ldots, Q_{i_d} \geq z_{n,m}) \\
<&nP(Q_{i_1} \geq z_{n,m}) \sum_{(i_2 \ldots i_d) \in S_{d-1}^c} P(Q_{i_2} \geq z_{n,m}, \ldots, Q_{i_d} \geq z_{n,m}) \\
<&\lambda o(1) = o(1) \qquad \text{(by assumption (3.26) holds for d-1)} \\
B =&\sum_{k=2}^{d-1} \sum_{(i_1, \ldots i_k) \in S_{k,k}} \sum_{(i_{k+1} \ldots i_d) \in S_{d-k}} P(Q_{i_1}, \ldots, Q_{i_k} \geq z_{n,m}) P(Q_{i_{k+1}} \geq z_{n,m}, \ldots, Q_{i_d} \geq z_{n,m}) \\
=&\sum_{k=2}^{d-1} \sum_{(i_1, \ldots i_k) \in S_{k,k}} P(Q_{i_1}, \ldots, Q_{i_k} \geq z_{n,m}) \sum_{(i_{k+1} \ldots i_d) \in S_{d-k}} P(Q_{i_{k+1}} \geq z_{n,m}, \ldots, Q_{i_d} \geq z_{n,m}) \\
\leq&\sum_{k=1}^{d-1} o(1) \frac{\lambda^{d-k}}{d!} = o(1) \qquad \text{(by assumption (3.26) holds for d-1 and d-2)} \\
C <&nm^{d-1} P(Q_{i_1} \geq z_{n,m}, Q_{i_2} \geq z_{n,m}) = 0 \\
D =&\binom{n+m+d-md-1}{d} P^d(Q_{i_1} \geq z_{n,m}) \xrightarrow{n \to \infty} \frac{\lambda^d}{d!}
\end{aligned}$$

Therefore, for the case $d \geq 3$ (3.26) still holds. Since we already prove $d = 2, d = 1$, so (3.26) holds.

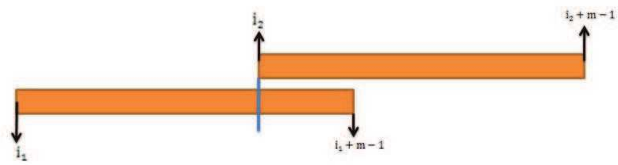


Figure 3.8: overlapped Q_{i_1} and Q_{i_2}

$\square$

3.6.6 Proof of Lemma 4

Proof. Assume $i_1 < i_2 < \cdots < i_d$, and let $\mathbb{K}$ be the collected set of integers such that $h \in \mathbb{K}$ if and only if h can be written as $h = i_j + k$ for some $1 \leq j \leq d$ and $0 \leq k \leq m-1$. Let $\underset{\sim}{x} = \{x_i\}_{j \in \mathbb{K}}$.

Define

$$A = \{\underset{\sim}{x} \mid \underset{\sim}{x} \text{ satisfy } (3.27)\},$$

$$B = \{\underset{\sim}{x} \mid \underset{\sim}{x} \text{ satisfy } (3.28)\},$$

Where (3.27) and (3.28) are defined as:

$$\begin{cases} x_{i_1}^2 + x_{i_1+1}^2 + ... + x_{i_1+m-1}^2 \geq h^2 \\ x_{i_2}^2 + x_{i_2+1}^2 + ... + x_{i_2+m-1}^2 \geq h^2 \\ \cdots \\ x_{i_d}^2 + x_{i_d+1}^2 + ... + x_{i_d+m-1}^2 \geq h^2 \end{cases} \tag{3.27}$$

$$\begin{cases} x_{i_1}^2 + x_{i_1+1}^2 + ... + x_{i_1+m-1}^2 \geq (h-\theta)^2 \\ x_{i_2}^2 + x_{i_2+1}^2 + ... + x_{i_2+m-1}^2 \geq (h-\theta)^2 \\ \cdots \\ x_{i_d}^2 + x_{i_d+1}^2 + ... + x_{i_d+m-1}^2 \geq (h-\theta)^2 \end{cases} \tag{3.28}$$

Define the λ-neighbourhood of set A as

$$A^{\lambda} = \left\{ x \in R^{\kappa}, \inf_{y \in A} |x - y| < \lambda \right\}, \text{ where } \kappa = |\mathbb{K}|.$$

if $\underset{\sim}{y} \in A^{\theta}$, then there $\exists\ \underset{\sim}{x} \in A$,s.t. $|\underset{\sim}{y} - \underset{\sim}{x}| < \theta$. In particular,

$$\left\| \begin{pmatrix} y_{j_1} \\ . \\ . \\ . \\ y_{j_1+m-1} \end{pmatrix} - \begin{pmatrix} x_{l_1} \\ . \\ . \\ . \\ x_{l_1+m-1} \end{pmatrix} \right\| \leq \theta \Longrightarrow \left\| \begin{pmatrix} y_{j_1} \\ . \\ . \\ . \\ y_{j_1+m-1} \end{pmatrix} \right\|^2 \geq (h - \theta)^2$$

$$\Longrightarrow \underset{\sim}{y} \in B$$

Therefore $A^{\theta} \subseteq B$: We divide the proof into two steps. The first step is truncation step, which makes the Gaussian approximation theorem applicable. Next, we will show that the truncation does not change the limiting distribution.

Step 1: Truncation.

Let us define $\mathbb{E}_o(X) = X - \mathbb{E}(X)$ for any random variable X. Set $\delta_n = o(1)$.

Define

$$\tilde{y}_{i,j} = \mathbb{E}_o\, y_{i,j} \mathbb{1}\{|y_{i,j}| \leq K\}, \left(K = \frac{\sqrt{n}\delta_n}{\left(\sqrt{\log p}\right)^3} \right),$$

where $\mathbb{1}\{.\}$ is the indicator function. Define:

$$\tilde{Q}_{1,1} = n(\bar{\tilde{y}}_1^2 + \bar{\tilde{y}}_2^2 + ... + \bar{\tilde{y}}_m^2)$$
$$\tilde{Q}_{2,1} = n(\bar{\tilde{y}}_2^2 + \bar{\tilde{y}}_3^2 + ... + \bar{\tilde{y}}_{m+1}^2)$$
$$\dots$$
$$\tilde{Q}_{p-m+1,1} = n(\bar{\tilde{y}}_{p-m+1}^2 + \bar{\tilde{y}}_{p-m+2}^2 + ... + \bar{\tilde{y}}_p^2)$$

And $M_{p,1} = \max\{\tilde{Q}_{1,1}, \dots, \tilde{Q}_{p-m+1,1}\}$

Step 2: Scale: in order to make the normal approximation theorem applicable, we also normalize $\tilde{y}_{i,j}$

Define

$$\breve{y}_{i,j} = \frac{\tilde{y}_{i,j}}{\sigma_j}, \text{ where } \sigma_j = \sqrt{\mathrm{Var}(\tilde{y}_{i,j})}$$

Similar as in step 1, let $M_{p,2} = \max\{\breve{Q}_1, \ldots, \breve{Q}_{p-m+1}\}$, where $\breve{Q}_j$ could be computed as below :

$$\begin{aligned}
&\breve{Q}_1 = n(\bar{\breve{y}}_1^2 + \bar{\breve{y}}_2^2 + ... + \bar{\breve{y}}_m^2) \\
&\breve{Q}_2 = n(\bar{\breve{y}}_2^2 + \bar{\breve{y}}_3^2 + ... + \bar{\breve{y}}_{m+1}^2) \\
&\ldots \\
&\breve{Q}_{p-m+1} = n(\bar{\breve{y}}_{p-m+1}^2 + \bar{\breve{y}}_{p-m+2}^2 + ... + \bar{\breve{y}}_p^2)
\end{aligned}$$

Let $\breve{N} = \sum_k \mathbb{1}\{\breve{Q}_k > z_{n,m}\}$, thus,$P(\breve{M}_{p,2} \leq z_{n,m}) = P(\breve{N} = 0)$. It's equivalent to prove $\breve{N} \xrightarrow[n\to\infty]{\mathbf{D}} \text{Poisson}(\exp(-z/2))$.

On the hand, Let $z_1, z_2, ...z_p \overset{i.i.d}{\sim} N(0,1)$

$$\begin{aligned}
&Q_1 = (z_1^2 + z_2^2 + ... + z_m^2) \\
&Q_2 = (z_2^2 + z_2^2 + ... + z_{m+1}^2) \\
&\ldots \\
&Q_{p-m+1} = (z_{p-m+1}^2 + z_{p-m+2}^2 + ... + z_p^2)
\end{aligned}$$

And $M_p = \max\{Q_1, \ldots, Q_{p-m+1}\}$,$N = \sum_k \mathbb{1}\{Q_k > z_{p,m}\}$. We already show that

$$N \xrightarrow[n\to\infty]{\mathbf{D}} \text{Poisson}\left(\exp(-z/2)\right)$$

So it's sufficient to prove $\breve{N}$ and N has same limiting distribution since $\mid \breve{y}_{i,j} \mid \leq 2K$, $\underset{\sim}{x} \in \mathbb{B}(\kappa, 2K)$,where $\underset{\sim}{x} \in A$. Let $\theta_p = o\left(\frac{1}{\sqrt{\log{(p)}}}\right)$, by the normal approximation theorem, we could get

$$
\begin{aligned}
&P(\breve{Q}_{i_1}^{1/2} \geq z_{p,m}^{1/2}, ..., \breve{Q}_{i_d}^{1/2} \geq z_{p,m}^{1/2}) \\
\leq &P(Q_{i_1}^{1/2} \geq z_{p,m}^{1/2} - \theta_p, ..., Q_{i_d}^{1/2} \geq z_{p,m}^{1/2} - \theta_p) + C\exp\left\{-\frac{\sqrt{n}\theta_p}{C_d K})\right\} \\
= &P(Q_{i_1} \geq z_{p,m} + \theta_p^2 - 2z_{p,m}^{1/2}\theta_p, ..., Q_{i_d} \geq z_{p,m} + \theta_p^2 - 2z_{p,m}^{1/2}\theta_p) + C\exp\left\{-\frac{\log(p)}{\delta_n}\right\}
\end{aligned}
$$

Therefore, as $n \to \infty$, we could get the following inequality:

$$
\begin{aligned}
&\sum_{1\leq i_1 \leq i_2,...i_d \leq p-m+1} P(\breve{Q}_{i_1} \geq z_{p,m}, ...\breve{Q}_{i_d} \geq z_{p,m}) \\
\leq &\sum_{1\leq i_1 \leq i_2,...i_d \leq p-m+1} \left(P(Q_{i_1} \geq z_{p,m} + o(1), ..., Q_{i_d} \geq z_{p,m} + o(1)) + C\exp\left\{-\frac{\log(p)}{\delta_n}\right\}\right) \\
\leq &\sum_{1\leq i_1 \leq i_2,...i_d \leq p-m+1} P\left(Q_{i_1} \geq z_{p,m} + o(1), ..., Q_{i_d} \geq z_{p,m} + o(1)\right) + C\exp\left\{d\log(p) - \frac{\log(p)}{\delta_n}\right\} \\
&\xrightarrow{n\to\infty} \sum_{1\leq i_1 \leq i_2,...i_d \leq p-m+1} P(Q_{i_1} \geq z_{p,m}, ..., Q_{i_d} \geq z_{p,m}) \\
&\xrightarrow{n\to\infty} \frac{\exp\{-zd/2\}}{d!}
\end{aligned}
$$

Similarly,we could get:

$$
\sum_{1\leq i_1 \leq i_2,...i_d \leq p-m+1} P(\breve{Q}_{i_1} \geq z_{p,m}, ...\breve{Q}_{i_d} \geq z_{p,m}) \geq \frac{\exp\{-zd/2\}}{d!}
$$

Thus we have proved that

$$
\lim_{n\to\infty} \mathbb{E}\left(N(N-1)\dots(N-d+1)\right) = \lim_{n\to\infty} \mathbb{E}\left(\breve{N}(\breve{N}-1)\dots(\breve{N}-d+1)\right).
$$

It implies that $P(N = d) = P(\breve{N} = d)$.

Step 3: Effect of Truncation and normalization.

In this section, we'll show $\left|\tilde{M} - M_{p,2}\right| \xrightarrow{P} 0.$

$$\begin{aligned}
&\left|\tilde{M} - M_{p,2}\right| \\
\leq &|\tilde{M} - M_{p,1}| + |M_{p,1} - M_{p,2}| \\
\leq &\max_{1\leq i\leq p}\left|\tilde{Q}_i - \tilde{Q}_{i,1}\right| + \max_{1\leq i\leq p}\left|\tilde{Q}_{i,1} - \breve{Q}_{i,2}\right|
\end{aligned}$$

So it is sufficient to prove that

$$\begin{cases}
\max_{1\leq i\leq p}\left|\tilde{Q}_i - \tilde{Q}_{i,1}\right| \xrightarrow{P} 0 \\
\max_{1\leq i\leq p}\left|\tilde{Q}_{i,1} - \breve{Q}_{i,2}\right| \xrightarrow{p} 0
\end{cases}$$

$$\begin{aligned}
&\max_{1\leq i\leq p}\left(\tilde{Q}_i - \tilde{Q}_{i,1}\right) \\
=&\max_{1\leq i\leq p} n \sum_{j=i}^{j=i+m-1}\left(\bar{y}_j - \bar{\tilde{y}}_j\right)\left(\bar{y}_j + \bar{\tilde{y}}_j\right) \\
\leq&\max_{1\leq i\leq p}\sqrt{n}\sum_{j=i}^{i+m-1}\left|\bar{y}_j - \bar{\tilde{y}}_j\right| \max_{1\leq i\leq p}\sqrt{n}\sum_{j=i}^{i+m-1}\left(\bar{y}_j + \bar{\tilde{y}}_j\right) \\
\leq&m^2\underbrace{\max_{1\leq i\leq p}\sqrt{n}\left|\bar{y}_i - \bar{\tilde{y}}_i\right|}_{\text{E}}\underbrace{\max_{1\leq i\leq p}\sqrt{n}\left(\bar{y}_i + \bar{\tilde{y}}_i\right)}_{\text{F}}
\end{aligned}$$

For term F, we'll show $F = O\left(\sqrt{log(p)}\right)$, and for E, we'll show $E = o\left(\frac{1}{\sqrt{\log p}}\right)$.

To show $E = o\left(\frac{1}{\sqrt{\log p}}\right)$, it's sufficient to show that, there $\exists\ \{\eta_n\} \xrightarrow{n\to\infty} 0$, such that

$$P\left(E \geq \frac{\eta_n}{\sqrt{\log p}}\right) = 0.$$

$$
\begin{aligned}
&P\left(\max_{1\le i\le p}\sqrt{n}\,|\bar{y}_i-\bar{\tilde{y}}_i|\ge\frac{\eta_n}{\sqrt{\log(p)}}\right)\\
\le&\sum_{i=1}^{i=p}P\left(\sqrt{n}|\bar{y}_i-\bar{\tilde{y}}_i|\ge\frac{\eta_n}{\sqrt{\log(p)}}\right)\\
=&\sum_{i=1}^{i=p}P\left(\left|\sum_{l=1}^{n}(y_{l,i}-\tilde{y}_{l,i})\right|\ge\frac{\sqrt{n}\eta_n}{\sqrt{\log(p)}}\right)\\
=&\sum_{i=1}^{p}P\left(\sum_{l=1}^{n}(y_{l,i}-\tilde{y}_{l,i})\ge\frac{\sqrt{n}\eta_n}{\sqrt{\log(p)}}\right)\\
\le&c_{1,q}p\sum_{l=1}^{n}\int_0 u_i^q dF_l(u_i)\left(\frac{\log(p)}{n\eta_n^2}\right)^{q/2}+\exp\left\{-c_{2,q}\frac{n\eta_n^2}{\log(p)B_n^2}\right\}\\
\le&\underbrace{c_{1,q}n^{1+r}\int_0 u_i^q dF_l(u_i)\left(\frac{\log(p)}{n\eta_n^2}\right)^{q/2}}_{\text{G}}+\underbrace{\exp\left\{-c_{2,q}\frac{n\eta_n^2}{\log(p)B_n^2}\right\}}_{\text{H}}
\end{aligned}
$$

where $u_i = y_{l,i} - \tilde{y}_{l,i}$, and $B_n^2 = \sum_{i=1}^{n}\mathrm{Var}(y_{l,i} - \tilde{y}_{l,i})$.

$$
\begin{aligned}
u_i &= y_{l,i}-\tilde{y}_{l,i}\\
&= y_{l,i}-y_{1,i}\mathbb{1}\left\{|y_{1,i}|\le K\right\}+\mathbb{E}\left(y_{1,i}\mathbb{1}\{|y_{1,i}|\le K\}\right)\\
&= y_{1,i}\mathbb{1}\{|y_{1,i}|>K\}-\mathbb{E}\left(y_{1,i}\mathbb{1}\{\mid y_{1,i}\mid>K\}\right)\\
&= y_{1,i}\mathbb{1}\{|y_{1,i}|>K\}+o(1)
\end{aligned}
$$

$$
\begin{aligned}
B_n^2 &= \sum_{i=1}^{n}\mathbb{E}\left(y_{1,i}\mathbb{1}\left\{|y_{1,i}|>K\right\}-\mathbb{E}\left(y_{1,i}\mathbb{1}\{|y_{1,i}|>K\}\right)\right)^2\\
&\le\sum_{i=1}^{n}\mathbb{E}\left(y_{1,i}^2\mathbb{1}\{|y_{1,i}|>K\}\right)\\
&\le\frac{n\mathbb{E}(Y^q)}{K^{q-2}}
\end{aligned}
$$

Therefore, since $1+r-q/2<0$, we can see that:

$$
\begin{aligned}
G &\le c_{1,q}n^{1+r-q/2}\left(\frac{r\log n}{\eta_n^2}\right)^{q/2}\mathbb{E}(|y_i|^q)\xrightarrow{n\to\infty}0\\
H &\le \exp\left\{-c_{2,q}\frac{\eta_n^2(r\log n)^3}{\delta_n^2}\right\}\xrightarrow{n\to\infty}0
\end{aligned}
$$

Thus, $E = o\left(\frac{1}{\sqrt{\log p}}\right)$

Now we'll show for $F = O(\log(p))$, let $M > 0$,

$$
\begin{aligned}
& P\left(\left|\max_{1\le i\le p}\sqrt{n}(\bar{y}_i + \bar{\breve{y}}_i)\right| > M\sqrt{r\log n}\right) \\
\le & \sum_{i=1}^{i=p} P\left(\sqrt{n}(\bar{y}_i + \bar{\breve{y}}_i) > M\sqrt{r\log n}\right) \\
\le & pP\left(\sum_{l=1}^{l=n}(y_{li} + \breve{y}_{li}) > M\sqrt{nr\log n}\right) \\
\le & n^{1+r}P\left((y_{li} + \breve{y}_{li}) > M\sqrt{nr\log n}\right) + \exp\left\{-c_q M^2 nr\log n/C_n^2\right\} \\
\le & n^{1+r}P\left(y_{li} > M\sqrt{nr\log n}, y_{li} > K\right) + n^{1+r}P\left(2y_{li} > M\sqrt{nr\log n}, y_{li} < K\right) \\
& + \exp\left\{-c_q M^2 nr\log n/C_n^2\right\} \\
= & \underbrace{n^{1+r}P\left(y_{li} > M\sqrt{nr\log n}\right)}_{\text{J}} + \underbrace{\exp\left\{-c_q M^2 nr\log n/C_n^2\right\}}_{\text{K}}
\end{aligned}
$$

where $C_n^2 = \sum_{i=1}^{n}\mathrm{Var}(y_{l,i} + \breve{y}_{l,i}) \le \sum_{i=1}^{n}\left(4\mathbb{E}\left(|y_i|^2\right) + o(1)\right)$.

Thus,

$$
\begin{aligned}
J &\le n^{1+r}\frac{\mathbb{E}(Y^q)}{\left(M\sqrt{nr\log n}\right)^q} \xrightarrow{n\to\infty} 0 \\
K &\le \exp\left\{-c_q M^2 nr\log n/C_n^2\right\} \le \exp\left\{-c_q M^2 r\log n/4\mathbb{E}(|y_i|^2)\right\} \xrightarrow{n\to\infty} 0
\end{aligned}
$$

Thus, $F = O\left(\sqrt{log(p)}\right)$, together with $E = o\left(\frac{1}{\sqrt{log(p)}}\right)$, we've already prove that

$$\left|\tilde{M} - M_{p,1}\right| \xrightarrow{P} 0.$$

Similarly, we could prove $|M_{p,2} - M_{p,1}| \xrightarrow{P} 0$.

Note that $\tilde{M} = \tilde{M} - M_{p,1} + M_{p,1} - M_{p,2} + M_{p,2} \xrightarrow{D} M_{p,2}$ as $n \to \infty$,

$\square$

3.6.7 Proof of Theorem 7

Proof. We will divide the proof into seven steps. The first one is m-approximation step, followed by block sum and truncation steps. These three steps make gaussian approximation result applicable. We consider the effect of using sample means and sample autocovariances in the fourth and fifth steps. Gaussian approximations are give in Step 6 and 7.

Define $\hat{\gamma}_{ij}(k) = \frac{1}{T}\sum_{t=1}^{T-k} X_{t,i}X_{t+k,j}$, and

step 1: m-approximation For an arbitrary integer l, define

$$\tilde{X}_{i,t} = E\left(X_{i,t}|F_{t-l}^{t}\right)$$

For l-dependence approximation, we already have

$$\|\sum_t X_{t,i}X_{t+k,j} - \sum_t \tilde{X}_{t,i}\tilde{X}_{t+k,j}\|_q < c_q\sqrt{T}l^{-\alpha}.$$

In this step we'll show that

$$P\left(\max_{i,j,k}|\frac{\hat{\gamma}_{ij}(k)}{\tau_{i,j}} - \frac{\tilde{\gamma}_{ij}(k)}{\tau_{i,j}}|\right) = o_P\left(\frac{1}{\sqrt{T\log n}}\right) \tag{3.29}$$

Set $m := m_T = O(T^\eta)$ such that $s_T \le m_T$, and define $\tilde{X}_{i,t} = E\left(X_{i,t}|F_{t-m}^{t}\right)$. Applying the nested m-approximation technique as given in Xiao and Wu (2013), we have for any $\delta_T > 0$ which converges to zero slowly enough,

$$P\left(\max_{i,j,k}\left|\frac{\hat{\gamma}_{ij}(k)}{\tau_{ij}} - \frac{\tilde{\gamma}_{ij}(k)}{\tau_{ij}}\right| > \frac{\delta_T}{\sqrt{T\log n}}\right)$$
$$\le \begin{cases} n\delta_T^{-q}\cdot T^{-\alpha q} + o(1) & \text{if } \alpha \le 1/2 - 1/q; \\ n\delta_T^{-q}\cdot T^{1-q/2} + o(1) & \text{if } \alpha > 1/2 - 1/q. \end{cases}$$

Note that $n = O(T^{2\gamma+\eta})$, so (3.29) holds under the conditions of Theorem 7.

Step 2: throw away small blocks

Divide the sample size T into blocks, where the odds block contains small sample size $w = 2m$, while the even ones contains larger ones $S = T^v$, where $S >> w$.

- B_1: $k + w < t < s$,
- B_2: $l + w < t < w + S$,
- B_3: $w + S + 1 < t < 2w + S$,
- B_4: $2w + S + 1 < t < 2w + 2S$,
- $\ldots$
- B_{2u_T-1}: $(w + L)(u_T - 1) < t < T$,

where u_T is the smallest integer that $(w + L)(u_T) > T$.

For any $1 \leq k \leq n$, denote $\tilde{X}_{i,t} = E\left(X_{i,t}|F_{t-l}^t\right)$

$$\begin{aligned}
S_{ij}(k) &= \sum_{t=k+1}^{T} \tilde{X}_{i,t}\tilde{X}_{j,t-k} \\
&= \underbrace{\sum_{j=1}^{u_T} \sum_{t \in B_{2u-1}} \tilde{X}_{i,t}\tilde{X}_{j,t-k}}_{S_{ij}^{(1)}(k)} + \underbrace{\sum_{j=1}^{u_T} \sum_{t \in B_{2u}} \tilde{X}_{i,t}\tilde{X}_{j,t-k}}_{S_{ij}^{(2)}(k)}
\end{aligned}$$

In this step, we will show that

$$\lim_{T\to\infty} \max_{i,j,k} \left\{S_{ij}(k)\right\} \stackrel{D}{=} \lim_{T\to\infty} \max_{i,j,k} \left\{S_{ij}^{(2)}(k)\right\}.$$

In order to prove that, it's equivalent to show that:

$$\lim_{T\to\infty} \max_{i,j,k} \{S_{ij}^{(1)}(k)\} = o_p\left(\frac{\sqrt{T}}{\sqrt{\log n}}\right).$$

For any δ_n,

$$
\begin{aligned}
&P\left(\max_{i,j,k} S_{ij}^{(1)}(k) > \frac{\sqrt{T}\delta_n}{\sqrt{\log n}}\right)\\
<&\sum_{i,j,k} P\left(S_{ij}^{(1)}(k) > \frac{\sqrt{T}\delta_n}{\sqrt{\log n}}\right)\\
=&nP\left(\sum_{j=1}^{u_T}\sum_{t\in B_{2u-1}} \tilde{X}_{i,t}\tilde{X}_{j,t-k} > \frac{\sqrt{T}\delta_n}{\sqrt{\log n}}\right)\\
=&n\sum_{j=1}^{u_T} P\left(\sum_{t\in B_{2u-1}} \tilde{X}_{i,t}\tilde{X}_{j,t-k} > \frac{\sqrt{T}\delta_n}{\sqrt{\log n}}\right) + \exp\left(-\frac{c_q T\delta_n^2}{u_T \mathrm{var}(\sum_{t\in B_{2j-1}} \tilde{X}_{i,t}\tilde{X}_{j,t-k})\log n}\right)\\
=&\underbrace{n\sum_{j=1}^{u_T} P\left(\sum_{t\in B_{2u-1}} \tilde{X}_{i,t}\tilde{X}_{j,t-k} > \frac{\sqrt{T}\delta_n}{\sqrt{\log n}}\right)}_{P_1} + \underbrace{\exp\left(-\frac{c_q T\delta_n^2}{2C_q u_T l\log n}\right)}_{P_2}
\end{aligned}
$$

$$
P_1 < \begin{cases}
nT^{1-v}\left(\frac{c_q(2l)^{q/2-\alpha q}\Delta_q^q}{(\frac{T\delta_n}{\log n})^{q/2}} + C_q\exp(-\frac{(\frac{\sqrt{T}\delta_q}{\sqrt{\log n}})^{q/2}}{2C_q\Delta_q^2 l})\right) & \text{if } \alpha < \frac{1}{2} - \frac{1}{q}\\
nT^{1-v}\left(\frac{c_q(2l)^{q/2-\alpha q}\Delta_q^q}{(\frac{T\delta_n}{\log n})^{q/2}}\log(2l) + C_q\exp(-\frac{(\frac{\sqrt{T}\delta_q}{\sqrt{\log n}})^{q/2}}{2C_q\Delta_q^2 l})\right) & \text{if } \alpha = \frac{1}{2} - \frac{1}{q}\\
nT^{1-v}\left(\frac{c_q(2l)\Delta_q^q}{(\frac{T\delta_n}{\log n})^{q/2}} + C_q\exp(-\frac{(\frac{\sqrt{T}\delta_q}{\sqrt{\log n}})^{q/2}}{2C_q\Delta_q^2 l})\right) & \text{if } \alpha > \frac{1}{2} - \frac{1}{q}
\end{cases}
$$

Note that $0 < \beta < v < 1$,and $l = T^{\beta}, U_T = T^{1-v}$ it's not difficulty to see that

$$P_1 \xrightarrow{T\to\infty} 0, P_2 \xrightarrow{T\to\infty} 0$$

Therefore, we have shown that $P\left(\max_{i,j,k} S_{ij}^{(1)}(k) > \frac{\sqrt{T}\delta_n}{\sqrt{\log n}}\right) \xrightarrow{T\to\infty} 0$

step 3: Truncation

Define:

$$\Xi_{ij}(k,u) = \frac{1}{\tau_{ij}} \sum_{t \in B_{2u}} \tilde{X}_i(t)\tilde{X}_j(t-k)$$

And let

$$\tilde{\Xi}_{ij}(k,u) = \mathrm{E}_0 \Xi_{ij}(k,u) \mathbb{1} \left\{ \Xi_{ij}(k,u) < \frac{\sqrt{T}\delta_T}{(\log n)^{3/2}} \right\}$$

where $\mathbb{E}_0(X) = X - \mathbb{E}(X)$

Besides, we denote the difference between $\Xi_{ij}(k,u)$ and $\tilde{\Xi}_{ij}(k,u)$ as $R_{ij}(k,u)$, i.e.

$$R_{ij}(k,u) = \Xi_{ij}(k,u) - \tilde{\Xi}_{ij}(k,u) = \mathrm{E}_0 \Xi_{ij}(k,u) \mathbb{1} \left\{ \Xi_{ij}(k,u) \geq \frac{\sqrt{T}\delta_T}{(\log n)^{3/2}} \right\}$$

Therefore, in order to prove that the truncation has no effect on limiting distribution,it's sufficient to prove that

$$\max_{i,j,k} \sum_{u=1}^{U_T} R_{ij}(k,u) = o_p\left(\frac{\sqrt{T}}{\sqrt{\log n}} \right)$$

For any $\delta_n > 0$

$$\begin{aligned}
&P\left(\max_{i,j,k} \sum_{u=1}^{U_T} R_{ij}(k,u) > \frac{\delta_T\sqrt{T}}{\sqrt{\log n}} \right) \\
<&\sum_{i,j,k} P\left(\sum_{u=1}^{U_T} R_{ij}(k,u) > \frac{\delta_T\sqrt{T}}{\sqrt{\log n}} \right) \\
<&n \sum_{u=1}^{U_T} P\left(R_{ij}(k,u) > \frac{\delta_T\sqrt{T}}{\sqrt{\log n}} \right) + \exp\left\{ -\frac{c_q \delta^2 T}{\log n \mathrm{var}(\sum_{u=1}^{U_T} R_{ij}(k,u))} \right\} \\
\leq&n \sum_{u=1}^{U_T} P\left(\Xi_{ij}(k,u) > \frac{\delta_T\sqrt{T}}{\sqrt{\log n}} \right) + \exp\left\{ -\frac{c_q \delta^2 T}{\log n \mathrm{var}(\sum_{u=1}^{U_T} R_{ij}(k,u))} \right\} \\
\leq&\underbrace{n \sum_{u=1}^{U_T} P\left(\frac{1}{\sqrt{\sigma_{ij}}} \sum_{t \in B_{2u}} \tilde{X}_i(t)\tilde{X}_j(t-k) > \frac{\delta_T\sqrt{T}}{\sqrt{\log n}} \right)}_{P_3} + \underbrace{\exp\left\{ -\frac{c_q \delta^2 T}{\log n \mathrm{var}(\sum_{u=1}^{U_T} R_{ij}(k,u))} \right\}}_{p_4}
\end{aligned}$$

$$P_3 < \begin{cases} nT^{1-v}\left(\dfrac{c_q(S)^{q/2-\alpha q}\Delta_q^q}{(\frac{T\delta_n}{\log n})^{q/2}} + C_q\exp(-\dfrac{(\frac{\sqrt{T}\delta_q}{\sqrt{\log n}})^{q/2}}{2C_q\Delta_q^2 S})\right) & \text{if } \alpha < \dfrac{1}{2} - \dfrac{1}{q} \\ nT^{1-v}\left(\dfrac{c_q(S)^{q/2-\alpha q}\Delta_q^q}{(\frac{T\delta_n}{\log n})^{q/2}}\log(2l) + C_q\exp(-\dfrac{(\frac{\sqrt{T}\delta_q}{\sqrt{\log n}})^{q/2}}{2C_q\Delta_q^2 S})\right) & \text{if } \alpha = \dfrac{1}{2} - \dfrac{1}{q} \\ nT^{1-v}\left(\dfrac{c_q(S)\Delta_q^q}{(\frac{T\delta_n}{\log n})^{q/2}} + C_q\exp(-\dfrac{(\frac{\sqrt{T}\delta_q}{\sqrt{\log n}})^{q/2}}{2C_q\Delta_q^2 S})\right) & \text{if } \alpha > \dfrac{1}{2} - \dfrac{1}{q} \end{cases}$$

Note that $0 < v < 1$,and $S = T^v, U_T = T^{1-v}$ it's not difficulty to see that

$$P_3 \xrightarrow{T\to\infty} 0, P_4 \xrightarrow{T\to\infty} 0$$

Therefore, we've shown that truncation has no effect on limiting distribution.

step 4: effect of estimated mean $\bar{X}_i$

Set $\bar{X}_i = \dfrac{1}{T}\sum_{t=1}^{T} X_{t,i}$. Define $\hat{\gamma}_{ij}^m(k) = \dfrac{1}{T}\sum_{t=1}^{T-k}(X_{t,i} - \bar{X}_i)(X_{t+k,j} - \bar{X}_j)$

$$\begin{aligned} &\max_{i,j,k}\{\hat{\gamma}_{ij}^m(k) - \hat{\gamma}_{ij}(k)\} \\ =&\max_{i,j,k}\frac{1}{T}\left(\sum_{t=1}^{T-k}(X_{t,i} - \bar{X}_i)(X_{t+k,j} - \bar{X}_j) - X_{t,i}X_{t+k,j}\right) \\ =&\max_{i,j,k}\{-\frac{\bar{X}_j}{T}\sum_{t=k+1}^{T}X_i(t) - \frac{\bar{X}_i}{T}\sum_{t=k+1}^{T}X_j(t-k) + \frac{T-k}{T}\bar{X}_i\bar{X}_j\} \\ \le&\underbrace{\max_{i,j,k}\{|\frac{\bar{X}_j}{T}\sum_{t=k+1}^{T}X_i(t)|\}}_{P_5} + \underbrace{\max_{i,j,k}\{|\frac{\bar{X}_i}{T}\sum_{t=k+1}^{T}X_j(t-k)|\}}_{P_6} + \underbrace{\max_{i,j,k}\{|\frac{T-k}{T}\bar{X}_i\bar{X}_j|\}}_{P_7} \end{aligned}$$

As we already know that

$$\max_i \bar{X}_i = O_P\left(\sqrt{\frac{\log p}{T}}\right)$$

Therefore, $P_7 = O_P\left(\dfrac{\log p}{T}\right) = o_P\left(\sqrt{\dfrac{\log p}{T}}\right)$ Next, we'll show that $P_5 = o_P\left(\sqrt{\dfrac{\log p}{T}}\right)$, in order to do that, it's sufficient to prove

$$\max_{i,j,k}\left\{\left|\frac{1}{T}\sum_{t=k+1}^{T}X_j(t-k)\right|\right\} = o_P\left(\frac{\sqrt{\log p}}{T^{\frac{1}{3}}}\right)$$

For any $\delta_T > 0$, which converges to 0 slowly, we could have:

$$P\left(\max_{i,j,k}\{|\frac{1}{T}\sum_{t=k+1}^{T}X_j(t-k)|\} > \frac{\sqrt{\log p}}{T^{\frac{1}{3}}}\right)$$
$$\leq \begin{cases} n\left(\dfrac{c_q T^{-q/6-\alpha q}\Delta_q^q}{(\log p)^{q/2}} + C_q \exp(-\dfrac{T^{1/3}\log p}{2C_q\Delta_q^2})\right) & \text{if } \alpha < \dfrac{1}{2}-\dfrac{1}{q} \\ n\left(\dfrac{c_q T^{-q/6-\alpha q}\Delta_q^q}{(\log p)^{q/2}}\log T + C_q \exp(-\dfrac{T^{1/3}\log p}{2C_q\Delta_q^2})\right) & \text{if } \alpha = \dfrac{1}{2}-\dfrac{1}{q} \\ n\left(\dfrac{c_q T^{1-2q/3}\Delta_q^q}{(\log p)^{q/2}} + C_q \exp(-\dfrac{T^{1/3}\log p}{2C_q\Delta_q^2})\right) & \text{if } \alpha > \dfrac{1}{2}-\dfrac{1}{q} \end{cases}$$

Therefore,

$$P\left(\max_{i,j,k}\{\hat{\gamma}_{ij}^m(k) - \hat{\gamma}_{ij}(k)\}\right) = o_P\left(\sqrt{\frac{\log p}{T}}\right)$$

Step 5: effect of estimated variance $\hat{\tau}_{ij}$

In this step, we show that the effect of plugging the estimate variance of σ_{ij} is negligible:

$$\max\left\{\frac{\hat{\gamma}_{ij}(k)}{\hat{\tau}_{ij}} - \frac{\hat{\gamma}_{ij}(k)}{\tau_{ij}}\right\} = o_p\left(\sqrt{\frac{1}{T\log n}}\right) \tag{3.30}$$

As we already show that $\max_{i,j,k}|\hat{\gamma}_{i,j}(k)| = O_p\left(\sqrt{\frac{\log n}{T}}\right)$, in order to prove Eq. (3.30), it's equivalent to show that

$$\max_{1\leq i\leq j\leq p}\|\tau_{i,j}^2 - \hat{\tau}_{i,j}^2\| = o_p\left(\frac{1}{\sqrt{\log n}}\right)$$

where $\hat{\tau}_{ij} = \sqrt{\sum_{k\in Z}\hat{\gamma}_{ii}(k)\hat{\gamma}_{jj}(k)}$ and $\tau_{ij} = \sqrt{\sum_{k\in Z}\gamma_{ii}(k)\gamma_{jj}(k)}$

$$
\begin{aligned}
&\hat{\tau}_{ij}^2 - \tau_{ij}^2 \\
&\leq \underbrace{\sum_{k \geq \upsilon_T} \|\hat{\gamma}_{ii}(k)\hat{\gamma}_{jj}(k) - \gamma_{ii}(k)\gamma_{jj}(k)\|}_{R_1} + \underbrace{\sum_{k \leq \upsilon_T} \|\hat{\gamma}_{ii}(k)\hat{\gamma}_{jj}(k) - \gamma_{ii}(k)\gamma_{jj}(k)\|}_{R_2}
\end{aligned}
$$

Let $\upsilon_T = T^{\upsilon}$, It's obvious to see that $R_1 < \upsilon_T^{-2\alpha}$. For R_2,

$$
\begin{aligned}
R_2 &= \sum_{k \leq \upsilon_T} \|\hat{\gamma}_{ii}(k)\hat{\gamma}_{jj}(k) - \gamma_{ii}(k)\gamma_{jj}(k)\| \\
&\leq \sum_{k \leq \upsilon_T} \|\hat{\gamma}_{ii}(k)\left(\hat{\gamma}_{jj}(k) - \gamma_{jj}(k)\right)\| + \|(\gamma_{jj}(k) - \hat{\gamma}_{jj}(k))(\hat{\gamma}_{ii}(k) - \gamma_{ii}(k))\| + \|\hat{\gamma}_{ii}(k)(\hat{\gamma}_{jj}(k) - \gamma_{jj}(k))\|
\end{aligned}
$$

when $2 < q \leq 4$, $1 < q/2 \leq 2$, by Burkholder inequality Burkholder (1973), as long as $\upsilon < 1/2$

$$
\begin{aligned}
&P\left(\|\hat{\gamma}_{jj}(k) - \gamma_{jj}(k)\| \geq \frac{\delta_T}{\sqrt{\log n}\upsilon_T}\right) \\
\leq &\frac{C_q T (\log n)^{q/4} \upsilon_T^{q/2}}{\delta_T^{q/2} T^{q/2}} \\
\leq &T^{1-q/2+\upsilon q/2} \frac{C_q (\log n)^{q/4}}{\delta_T^{q/2}} \\
&\xrightarrow{T\to\infty} 0
\end{aligned}
$$

when $q > 4$, for any $\delta_T > 0$, which converges to 0 slowly, we could have:

$$
\begin{aligned}
&P\left(\hat{\gamma}_{jj}(k) - \gamma_{jj}(k) > \frac{\delta_T}{\sqrt{\log n}\upsilon_T}\right) \\
\leq &\begin{cases}
\left(\dfrac{c_q T^{-q/4-\alpha q/2+\upsilon q/2}\Delta_{q/2}^q(\sqrt{\log n})^{q/2}}{\delta_T^{q/2}} + C_q \exp(-\dfrac{T\delta_T^2}{2C_q\Delta_q^2})\right) & \text{if } \alpha < \dfrac{1}{2} - \dfrac{1}{q} \\
\left(\dfrac{c_q T^{-q/4-\alpha q/2+\upsilon q/2}\Delta_{q/2}^q(\sqrt{\log n})^{q/2}}{\delta_T^{q/2}} \log T + C_q \exp(-\dfrac{T\delta_T^2}{2C_q\Delta_q^2})\right) & \text{if } \alpha = \dfrac{1}{2} - \dfrac{1}{q} \\
\left(\dfrac{c_q T^{1-q/2+\upsilon q/2}\Delta_{q/2}^q(\sqrt{\log n})^{q/2}}{\delta_T^{q/2}} + C_q \exp(-\dfrac{T\delta_T^2}{2C_q\Delta_q^2})\right) & \text{if } \alpha > \dfrac{1}{2} - \dfrac{1}{q}
\end{cases}
\end{aligned}
$$

Therefore, for any $q > 2$, $P\left(\hat{\gamma}_{jj}(k) - \gamma_{jj}(k) > \frac{\delta_T}{\sqrt{\log n}v_T}\right) = 0$

This implies $P\left(R_2 > \frac{\delta_T}{\sqrt{\log n}v_T}\right) = 0$, and thus the proof of step 5 is complete.

step 6: Consider d-tuples

Consider d-tuples $\{\hat{\gamma}_{i_1,j_1}(k_1), \ldots, \hat{\gamma}_{i_d,j_d}(k_d)\}$,where $i_s \neq j_s$, for $s = 1, 2, \ldots d$,i.e. $\hat{\gamma}_{i_s,j_s}$ is cross covariance instead of variance.

First, we'll show that if we draw d such pairs (i_s, j_s) from $1, 2, \ldots, p$, the probability that these 2d numbers are distinct will go to 1 as $p \to \infty$

Instead of calculating it directly, we calculate the probability that at least two of these 2d numbers are same, i.e. $i_{s_1} = i_{s_2}$, or $j_{s_1} = j_{s_2}$ for some s_1, s_2 which is $\frac{p\binom{d}{2}(p-1)^{2d-2}}{p^d(p-1)^d}$.Therefore, the probability that all these 2d numbers are different is $1 - \frac{p\binom{d}{2}(p-1)^{2d-2}}{p^d(p-1)^d} = 1 - \frac{\binom{2d}{2}}{p}(\frac{p-1}{p})^{d-1} \overset{p\to\infty}{\longrightarrow} 1$

Define

$$A = \{\frac{\hat{\gamma}^2_{i_1,j_1}(k_1)}{\sigma^2_{i_1,j_1}}, \ldots, \frac{\hat{\gamma}^2_{i_d,j_d}(k_d)}{\sigma^2_{i_1,j_1}} \mid \text{for some } i_{s_1} = i_{s_2} \text{ or } j_{s_1} = j_{s_2}, s_1 \neq s_2\}$$
$$B = \{\frac{\hat{\gamma}^2_{i_d,j_d}(k_1)}{\sigma^2_{i_d,j_d}}, \ldots, \frac{\hat{\gamma}^2_{i_d,j_d}(k_d)}{\sigma^2_{i_d,j_d}} \mid i_1, j_1, \ldots, i_d, j_d \text{ are all different}\}$$

Next, we'll show that

$$\lim_{p\to\infty} \mathcal{D}(\max\{A, B\}) = \lim_{p\to\infty} \mathcal{D}(\max\{B\})$$

where $\mathcal{D}$ denotes the limiting distribution.

Since $\max\{A, B\} = \max\{B\} + (\max\{A\} - \max\{B\})\mathbb{1}\{\max\{A\} > \max\{B\}\}$, it's equivalent to prove:

$$\lim_{p\to\infty} P(\max\{A\} > \max\{B\}) = 0$$

let $n_1 = |A|, n_2 = |B|$, as we've shown that $n_1 = o(n_2)$. let $\eta_n = \frac{1}{2}\log(n_2/n_1)$

$$P(T\max\{A\} > 2\log(n_1) + \eta_n) < \frac{\sup_{i,j} EX_i^q X_j^q}{(\sigma_{i,j}(2\log(n_1) + \eta_n))^q} \to 0 \text{ as } p \to \infty$$

$$
\begin{aligned}
P(T\max\{B\} < 2\log(n_2) - \eta_n) &= (1 - P(\mathcal{Z} > \sqrt{2\log(n_2) - \eta_n}))^{n_2} \\
&= \exp\left(n_2 \log(1 - P(\mathcal{Z} > \sqrt{2\log(n_2) - \eta_n}))\right) \\
&\sim \exp(-n_2 P(\mathcal{Z} > \sqrt{2\log(n_2) - \eta_n}))
\end{aligned}
$$

Note that

$$
\begin{aligned}
& n_2 P(\mathcal{Z} > \sqrt{2\log(n_2) - \eta_n}) \\
=& n_2 P(\mathcal{Z}^2 > 2\log(n_2) - \eta_n) \\
\sim& n_2 \frac{2}{\sqrt{2\pi}} (2\log(n_2) - \eta_n)^{-1/2} \exp(-\log(n_2) + \eta_n/2) \\
\sim& \frac{2\exp(\eta_n/2)}{\sqrt{2\pi(2\log n_2 - \eta_n)}} \\
=& \frac{\exp(\log(n_2/n_1)/4)}{\sqrt{2\pi(2\log(n_2) - \eta_n)}} \\
=& \frac{(n_2/n_1)^{1/4}}{\sqrt{2\pi(2\log(n_2) - \log(n_2/n_1)/2)}} \\
\to& \infty \text{ as } n_2 \text{ goes to } \infty
\end{aligned}
$$

This implies that

$$
P(T\max\{B\} \geq 2\log(n_2) - \eta_n) \to 1 \text{ as } n_2 \text{ goes to } \infty
$$

Therefore

$$
P(T\max\{B\} \geq 2\log(n_2) - \eta_n > 2\log(n_1) + \eta_n \geq T\max\{A\}) = 1
$$

i.e. $\lim_{p\to\infty} P(\max\{A\} > \max\{B\}) = 0$

Let Υ_d be the set of $\{(i_1, j_1, k_1), \ldots, (i_d, j_d, k_d)\}$ among which all $i_h, j_h, h = 1, \ldots, d$ are distinct. Furthermore, we could write it as:

$$
\Upsilon_d = \{\{(i_1, j_1, k_1), \ldots, (i_d, j_d, k_d)\} \mid 1 \leq i_h < j_h \leq p, -S_T \leq k_h \leq S_T, h = 1, \ldots, d\}
$$

Based the fact that $\lim_{p\to\infty} \mathcal{D}(\max\{A,B\}) = \lim_{p\to\infty} \mathcal{D}(\max\{B\})$, it's sufficient to consider d-tuples $\gamma_d \in \Upsilon_d$.

Since we've already shown that throwing away the small blocks has no effect on the limiting distribution, we only consider the larger blocks.

$$\hat{\gamma}_{i_1,j_1}(k_1) = \frac{1}{T}\sum_{t\in B_2} \tilde{X}_{i_1,t}\tilde{X}_{j_1,t-k_1} + \frac{1}{T}\sum_{t\in B_4} \tilde{X}_{i_1,t}\tilde{X}_{j_1,t-k_1} + \cdots + \frac{1}{T}\sum_{t\in B_{2u_T-2}} \tilde{X}_{i_1,t}\tilde{X}_{j_1,t-k_1}$$
$$\hat{\gamma}_{i_2,j_2}(k_2) = \frac{1}{T}\sum_{t\in B_2} \tilde{X}_{i_2,t}\tilde{X}_{j_2,t-k_2} + \frac{1}{T}\sum_{t\in B_4} \tilde{X}_{i_2,t}\tilde{X}_{j_2,t-k_2} + \cdots + \frac{1}{T}\sum_{t\in B_{2u_T-2}} \tilde{X}_{i_2,t}\tilde{X}_{j_2,t-k_2}$$
$$\vdots$$
$$\hat{\gamma}_{i_d,j_d}(k_d) = \frac{1}{T}\sum_{t\in B_2} \tilde{X}_{i_d,t}\tilde{X}_{j_d,t-k_d} + \frac{1}{T}\sum_{t\in B_4} \tilde{X}_{i_d,t}\tilde{X}_{j_d,t-k_d} + \cdots + \frac{1}{T}\sum_{t\in B_{2u_T-2}} \tilde{X}_{i_d,t}\tilde{X}_{j_d,t-k_d}$$

For any $\gamma_d \in \Upsilon_d$,

let $\varepsilon(\gamma_d, h) = \begin{pmatrix} \sum_{t\in B_{2h}} \tilde{X}_{i_1,t}\tilde{X}_{j_1,t-k_1} \\ \sum_{t\in B_{2h}} \tilde{X}_{i_2,t}\tilde{X}_{j_2,t-k_2} \\ \cdots \\ \sum_{t\in B_{2h}} \tilde{X}_{i_d,t}\tilde{X}_{j_d,t-k_d} \end{pmatrix}$ where $h = 1, 2, \ldots, u_T - 1$.

And it's easy to see that $\varepsilon(\gamma_d, h_1)$ is independent of $\varepsilon(\gamma_d, h_2)$ if $h_1 \neq h_2$ since $\tilde{X}_{i.}$ is l-dependent approximation of $X_{i.}$. Therefore, we could regard $\varepsilon(\gamma_d, h)(h = 2, \ldots, 2u_T - 2)$ as $u_T - 1$ independent observations from some distribution

step 7: Gaussian Approximation

It's obvious to see that $\varepsilon(\gamma_d, h)$ has mean 0_d,and covariance matrix I_d.

Assume $\mathbf{Z}_1, \ldots, \mathbf{Z}_{u_T-1} \in R^d$ are from i.i.d $\mathcal{N}(0, I_d)$,and

$$\frac{\varepsilon^T(\gamma_d)}{T} = \sum_{h=1}^{U_t-1} \frac{\varepsilon(\gamma_d, h)}{T} > \sqrt{z_n}\mathbf{1}_d \tag{3.31}$$
$$\mathbf{Z}^T = \sum_{h=1}^{U_t-1} \mathbf{Z}_h > (\sqrt{z_n} - \theta_n)\mathbf{1}_d \tag{3.32}$$

Where z_n = Denote set

$$C = \{\varepsilon^T \in R^d : \varepsilon^T \text{ satisfy } (3.31)\}$$

$$D = \{\mathbf{Z}^T \in R^d : \mathbf{Z} \text{ satisfy } (3.32)\}$$

Let $D^{\theta_n} = \{x \in R^d : \inf_{\mathbf{Z} \in D} |x - \varepsilon^T| < \theta_n\}$,where $\theta_n = \dfrac{\delta_T}{\sqrt{\log(n)}}$

and it's easy to see that $D^{\theta_n} \subset C$

By the gaussian approximation theorem, denote $(\mathbf{x})$. as the smallest element in $x \in R^d$, we have:

$$\begin{aligned}
&\sum_{\Upsilon_d} P((\frac{\varepsilon^T(\gamma_d)}{T}). > z_n) \\
\leq &\sum_{\Upsilon_d} P((\mathbf{Z}^T). > (z_n - \theta_n)) + \exp(-\frac{\delta_n \theta_n}{c_{2,d} K_2 / T}) \\
< &\frac{-zd/2}{d!} + \sum_{\Upsilon_d} \exp(-\delta_n \log(n) \sqrt{T}) \\
= &\frac{-zd/2}{d!} + \exp(-\delta_n \sqrt{T}) \\
\to &\frac{-zd/2}{d!}
\end{aligned}$$

We have shown the proof of Theorem 7

□

3.6.8 Proof of Theorem 8

Proof. The proof of Theorem 8 is similar with the one for Theorem 7, and we omit details. □

Bibliography

A. R. Abbas, K. Wolslegel, D. Seshasayee, Z. Modrusan, and H. F. Clark. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS ONE*, 4(7):e6098, 2009.

S. K. Ahn. Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization. *Biometrika*, 75(3):590–593, 1988.

Z. Bai, D. Jiang, J.-F. Yao, and S. Zheng. Corrections to lrt on large-dimensional covariance matrix by rmt. *The Annals of Statistics*, 37(6B):3822–3840, 2009.

P. Basappa and V. Lakdawala. Application of digital signal processing techniques to analyze partial discharge data. In *Electrical Insulation and Dielectric Phenomena, 2000 Annual Report Conference on*, volume 1, pages 341–346, 2000.

S. Basu and G. Michailidis. Regularized estimation in sparse high-dimensional time series models. *Annals of Statistics*, 43(4):1535–1567, 2015.

A. Berndt and A. Ostrovnaya. Information flow between credit default swap, option and equity markets. Working paper, Carnegie-Mellon University, 2007.

P. J. Bickel and E. Levina. Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604, 2008a.

P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227, 2008b.

C. R. Bolen, M. Uduman, and S. H. Kleinstein. Cell subset prediction for blood genomic studies. *BMC bioinformatics*, 12(1):1–10, 2011.

G. E. Box and D. A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526, 1970.

P. Bühlmann and S. Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

D. L. Burkholder. Distribution function inequalities for martingales. *Ann. Probab.*, 1 (1):19–42, 1973.

T. Cai, W. Liu, and Y. Xia. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, 108(501):265–277, 2013.

T. T. Cai and T. Jiang. Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *The Annals of Statistics*, 39(3):1496–1525, 2011.

T. T. Cai, C.-H. Zhang, and H. H. Zhou. Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, 38(4):2118–2144, 2010.

S. X. Chen and Y.-L. Qin. A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835, 2010.

X. Chen, M. Xu, and W. B. Wu. Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, 41(6):2994–3021, 2013.

R. V. Chitturi. Distribution of residual autocorrelations in multiple autoregressive schemes. *Journal of the American Statistical Association*, 69(348):928–934, 1974.

J. P. Cobb, M. N. Mindrinos, et al. Application of genome-wide expression analysis to human health and disease. *Proceedings of the National Academy of Sciences of the United States of America*, 102(13):4801–4806, 2005.

J. Cohen. *Statistical Power Analysis for the Behavioral Sciences, 2nc Edition*. Routledge, 1981. ISBN 0805802835.

C. M. Conover and D. R. Peterson. The lead-lag relationship between the option and stock markets prior to substantial earnings surprises and the effect of securities regulation. *Journal of Financial and Strategic Decisions*, 12(1):41–52, 1999.

H. M. Davey and D. B. Kell. Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. *Microbiol Rev*, 60(4):641–696, Dec 1996.

R. A. Davis, P. Zang, and T. Zheng. Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*, (just-accepted):1–53, 2015.

G. P. Duffy and J. E. Hughes-Clarke. Application of spatial cross correlation to detection of migration of submarine sand dunes. *Journal of Geophysical Research: Earth Surface*, 110(F4), 2005.

J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression: I. *Biometrika*, 37(3/4):409–428, 1950.

J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression. ii. *Biometrika*, 38(1/2):159–177, 1951.

S. N. Durlauf. Spectral based testing of the martingale hypothesis. *Journal of Econometrics*, 50(3):355–376, 1991.

P. Embrechts, T. Mikosch, and C. Klüppelberg. *Modelling Extremal Events: For Insurance and Finance*. Springer-Verlag, London, UK, 1997. ISBN 3-540-60931-8.

T. Erkkilä, S. Lehmusvaara, P. Ruusuvuori, T. Visakorpi, I. Shmulevich, and H. Lähdesmäki. Probabilistic analysis of gene expression measurements from heterogeneous tissues. *Bioinformatics*, 26(20):2571–2577, 2010.

J. C. Escanciano, I. N. Lobato, and L. Zhu. Automatic specification testing for vector autoregressions and multivariate nonlinear time series models. *J. Bus. Econom. Statist.*, 31(4):426–437, 2013.

R. Gaujoux and C. Seoighe. Semi-supervised nonnegative matrix factorization for gene expression deconvolution: A case study. *Infection, Genetics and Evolution*, 12(5):913 – 921, 2012.

R. Gaujoux and C. Seoighe. Cellmix: a comprehensive toolbox for gene expression deconvolution. *Bioinformatics*, 29(17):2211–2212, 2013.

T. Gong, N. Hartmann, I. S. Kohane, V. Brinkmann, F. Staedtler, M. Letzkus, S. Bongiovanni, and J. D. Szustakowski. Optimal deconvolution of transcriptional profiling data using quadratic programming with application to complex clinical blood samples. *PLoS One*, 6(11):e27156, Nov 2011.

W. H. Greene. *Econometric analysis*. Prentice Hall, Upper Saddle River, New Jersey, 5 edition, 2002. ISBN 0-13-066189-9.

P. Hall. On the rate of convergence of normal extremes. *Journal of Applied Probability*, 16(2):433–439, 1979.

J. B. Hill and K. Motegi. A max-correlation white noise test for weakly dependent time series. *Available at SSRN 2732144*, 2016.

Y. Hong. Consistent testing for serial correlation of unknown form. *Econometrica: Journal of the Econometric Society*, 64(4):837–864, 1996.

Y. Hong and Y. Lee. Consistent testing for serial correlation of unknown form under general conditional heteroskedasticity. *Preprint, Cornell University, Department of Economic*, 2003.

J. L. Horowitz, I. Lobato, J. C. Nankervis, and N. Savin. Bootstrapping the box–pierce q test: a robust test of uncorrelatedness. *Journal of Econometrics*, 133(2):841–862, 2006.

J. Hosking. Equivalent forms of multivariate portmanteau statistic. *Journal of the Royal Statistical Society, Series B*, 43:261–262, 1981.

J. R. Hosking. The multivariate portmanteau statistic. *Journal of the American Statistical Association*, 75(371):602–608, 1980.

T. Jiang. The asymptotic distributions of the largest entries of sample correlation matrices. *Annals of Applied Probability*, 14(2):865–880, 2004.

I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, 29(2):295–327, 2001.

A. Kuhn, D. Thu, H. J. Waldvogel, R. L. Faull, and R. Luthi-Carter. Population-specific expression analysis (psea) reveals molecular changes in diseased brain. *Nature methods*, 8(11):945–947, 2011.

H. Lähdesmäki, L. Shmulevich, V. Dunmire, O. Yli-Harja, and W. Zhang. In silico microdissection of microarray data from heterogeneous cell populations. *Bmc Bioinformatics*, 6(1):54, 2005.

C. Lam, Q. Yao, and N. Bathia. Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918, 2011a.

C. Lam, Q. Yao, and N. Bathia. Factor modeling for high dimensional time series. In *Recent Advances in Functional Data Analysis and Related Topics*, pages 203–207. Springer, 2011b.

D. Li, W.-D. Liu, and A. Rosalsky. Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probability theory and related fields*, 148(1-2):5–35, 2010.

R. Y. Liu and K. Singh. Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap (East Lansing, MI, 1990)*, Wiley-Intersci. Publ., 1992.

W.-D. Liu, Z. Lin, and Q.-M. Shao. The asymptotic distribution and berry–esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *The Annals of Applied Probability*, 18(6):2337–2366, 2008.

Y. B. Mainassara. Multivariate portmanteau test for structural varma models with uncorrelated but non-independent error terms. *Journal of Statistical Planning and Inference*, 141(8):2961–2975, 2011.

S. V. Nagaev. Large deviations of sums of independent random variables. *Ann. Probab.*, 7(5):745–789, 10 1979.

H. Nagao. On some test criteria for covariance matrix. *The Annals of Statistics*, 1(4): 700–709, 1973.

E. Nelson-Wong, S. Howarth, D. A. Winter, and J. P. Callaghan. Application of autocorrelation and cross-correlation analyses in human movement and rehabilitation research. *Journal of Orthopaedic & Sports Physical Therapy*, 39(4):287–295, 2009.

S. Péché. Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probability Theory and Related Fields*, 143(3-4):481–516, 2009.

G. Raskutti and M. Yuan. *arXiv preprint arXiv:1512.01215*, 2015.

D. Repsilber, S. Kern, A. Telaar, G. Walzl, G. F. Black, J. Selbig, S. K. Parida, S. H. Kaufmann, and M. Jacobsen. Biomarker discovery in heterogeneous tissue samples -taking the in-silico deconfounding approach. *BMC Bioinformatics*, 11(1):1–15, 2010.

P. Robinson. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47(1):67 – 84, 1991.

S. S. Shen-Orr and R. Gaujoux. Computational deconvolution: Extracting cell type-specific information from heterogeneous samples. *Curr Opin Immunol*, 25(5):10.1016, Oct 2013.

S. S. Shen-Orr, R. Tibshirani, P. Khatri, D. L. Bodian, F. Staedtler, N. M. Perry, T. Hastie, M. M. Sarwal, M. M. Davis, and A. J. Butte. Cell type-specific gene expression differences in complex tissues. *Nature Methods*, 7(4):287–289, Apr 2010.

S. Song and P. J. Bickel. Large vector auto regressions. *arXiv preprint arXiv:1106.3915*, 2011.

A. Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *Journal of Statistical Physics*, 108(5-6):1033–1056, 2002.

R. O. Stuart, W. Wachsman, C. C. Berry, J. Wang-Rodriguez, L. Wasserman, I. Klacansky, D. Masys, K. Arden, S. Goodison, M. McClelland, et al. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 101(2):615–620, 2004.

M. Tao, Y. Wang, Q. Yao, and J. Zou. Large volatility matrix inference via combining low-frequency and high-frequency approaches. *Journal of the American Statistical Association*, 106(495):1025–1040, 2012.

T. Tony Cai, W. Liu, and Y. Xia. Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):349–372, 2014.

C. A. Tracy and H. Widom. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159(1):151–174, 1994.

H. Trevor, T. Robert, and F. Jerome. The elements of statistical learning: data mining, inference and prediction. *New York: Springer-Verlag*, 1(8):371–406, 2001.

A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

D. Venet, F. Pecasse, C. Maenhaut, and H. Bersini. Separation of samples into their constituents using gene expression data. *Bioinformatics*, 17(suppl 1):S279–S287, 2001.

M. Wang, S. R. Master, and L. A. Chodosh. Computational expression deconvolution in a complex mammalian organ. *BMC Bioinformatics*, 7(1):1–26, 2006.

A. R. Whitney, M. Diehn, S. J. Popper, A. A. Alizadeh, J. C. Boldrick, D. A. Relman, and P. O. Brown. Individuality and variation in gene expression patterns in human blood. *Proceedings of the National Academy of Sciences*, 100(4):1896–1901, 2003.

C. K. Wikle and M. B. Hooten. A general science-based framework for dynamical spatio-temporal models. *Test*, 19(3):417–451, 2010.

W. B. Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40):14150–14154, 2005.

H. Xiao and W. B. Wu. Asymptotic theory for maximum deviations of sample covariance matrix estimates. *Stochastic Processes and their Applications*, 123(7):2899 – 2920, 2013.

H. Xiao and W. B. Wu. Portmanteau test and simultaneous inference for serial covariances. *Stat. Sin.*, 24(2):577–599, 2014.

I. V. Yang, L. G. Luna, J. Cotter, J. Talbert, S. M. Leach, R. Kidd, J. Turner, N. Kummer, D. Kervitsky, K. K. Brown, et al. The peripheral blood transcriptome identifies the presence and extent of disease in idiopathic pulmonary fibrosis. *PloS one*, 7(6): e37708, 2012.

A. Y. Zaitsev. On the gaussian approximation of convolutions under multidimensional analogues of sn bernstein's inequality conditions. *Probability theory and related fields*, 74(4):535–566, 1987.

X. Zhang and G. Cheng. Bootstrapping high dimensional time series. *arXiv preprint arXiv:1406.1037*, 2014.

W. Zhou. Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Transactions of the American Mathematical Society*, 359(11):5345–5363, 2007.