6-1997

A Robust Estimate for an Autoregressive Time Series

Jeffrey Terpstra
Western Michigan University

A ROBUST ESTIMATE FOR AN AUTOREGRESSIVE TIME SERIES

by

Jeffrey Terpstra

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Mathematics and Statistics

Western Michigan University
Kalamazoo, Michigan
June 1997

A ROBUST ESTIMATE FOR AN AUTOREGRESSIVE TIME SERIES

Jeffrey Terpstra, Ph.D.

Western Michigan University, 1997

A weighted rank-based estimate for estimating the parameter of an autoregressive time series is considered. When the weights are constant, the estimate is equivalent to using Jaeckel's estimate and Wilcoxon scores. The estimate can be shown to be asymptotically normal at rate $\sqrt{n}$. In a linear regression setting this estimate has the desired properties of a continuous totally bounded influence function and a positive breakdown point. It is shown via examples and Monte Carlo that these properties are preserved in an autoregressive time series setting.

UMI Number: 9732888

ACKNOWLEDGMENTS

I would like to acknowledge everyone who contributed to the completion of this dissertation and hence my degree. Specifically, special thanks go out to my advisor Joseph McKean and Joshua Naranjo whose periodical meetings, suggestions, and insights were greatly appreciated. Furthermore, I would also like to thank my family and friends for their support and understanding throughout my entire education.

Jeffrey Terpstra

TABLE OF CONTENTS

Table of Contents—Continued

CHAPTER

CHAPTER

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 The Linear Model

One of the most popular and widely used models in statistics is the linear regression model,

$$Y_i \;=\; \beta_0 + \boldsymbol{X}_i'\boldsymbol{\beta} + \varepsilon_i, \quad \text{i} = 1,2,\ldots,\text{n}.$$

In the above model $\boldsymbol{X}_i' = (X_{i1}, X_{i2}, \ldots, X_{ip})$ represents a vector of perfectly observed "design" information and Y_i denotes an observable random variable of interest. In the classical setting the random errors, ε_i, are assumed to be independent and identically distributed (iid) according to some distribution function, F. Lastly, $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$ denotes an unknown parameter vector and β_0 an unknown intercept term. One of the major goals in linear regression is estimation of these unknown quantities. If the error terms are Gaussian, it is well-known that the least squares estimate of $\boldsymbol{\beta}$ is optimal (Hampel et al., 1986, pg. 309).

However, when the error terms are not normal the least squares estimate can be altered (Huber, 1972). To account for such situations many robust estimation procedures have been developed. For example, re-weighted least squares and least median squares are discussed in Rousseeuw and Leroy (1987). For a

discussion of maximum likelihood type estimates, or M-estimates, one is referred to Huber (1981). As a final example, rank-based estimates, or R-estimates, are treated in Hettmansperger (1984). All of these estimates, with the exception of least median squares, depend on some sort of weighting scheme. The purpose of these weighting schemes is to downweight or discount potential outliers. One criteria for selecting a weighting scheme is to find one that performs well under a variety of situations for F.

In an observational study the $\boldsymbol{X}_i$'s contain realizations of random variables associated with the particular sampling unit. Thus, it is possible that outliers can be introduced into the design points. Outliers appearing in design points can have drastic effects on estimates , even for some of the robust estimates mentioned above. For example, estimates that do not have a totally bounded influence function are sensitive to design points with outliers, or leverage points. To account for these types of situations many robust estimation procedures have been generalized. For instance, generalized M-estimates, or GM-estimates, are discussed by (Simpson et al., 1992; Coakley & Hettmansperger, 1993) while modified rank-based estimates are treated by Sievers (1983), Naranjo and Hettmansperger (1994), and Chang (1996).

In time series analysis one observes realizations of a given random variable as a function of time. A widely used model in time series analysis is the stationary autoregressive model of order p, denoted by AR(p). If we let $X_1, X_2, \ldots, X_n$ denote

random variables pertaining to the realizations the model can be written as,

$$X_i = \alpha + \boldsymbol{Y}'_{i-1}\boldsymbol{\rho} + \varepsilon_i, \quad \text{i} = 1,2, \ldots, \text{n}, \tag{1.1}$$

where $\boldsymbol{Y}'_{i-1} = (X_{i-1}, X_{i-2}, \ldots, X_{i-p})$ is a design point containing the previous p time series values; the random errors, ε_i, are assumed independent and identically distributed according to some distribution function, F; and $\boldsymbol{Y}'_0 = (X_0, X_{-1}, \ldots, X_{1-p})$ is an observable random vector independent of $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$. Furthermore, the solutions to the following equation,

$$X^p - \rho_1 X^{p-1} - \rho_2 X^{p-2} - \ldots - \rho_p = 0, \tag{1.2}$$

lie in the interval $(-1, 1)$. This is the condition that assures the process is stationary. Since the AR(p) is basically a linear regression model with the design points containing the previous p lags, it is customary to expect that both classical and robust estimation procedures for linear regression parameters can be used to obtain estimates for the autoregressive parameter vector. In fact, many properties and results obtained for $\hat{\boldsymbol{\beta}}_n$ can be shown to hold for $\hat{\boldsymbol{\rho}}_n$ as well. However, the tools used to obtain these results can be quite different since independence among observations is no longer present in the data. For an account of M-estimates and GM-estimates one is referred to Denby and Martin (1979), Martin and Yohai (1991), and Bustos (1982). For estimates of $\boldsymbol{\rho}$ based on ranking procedures one can see Koul (1993) and Ferretti, Kelmansky, and Yohai (1991). Finally, Rousseeuw and Leroy (1987) discuss an example where they use re-weighted least squares and least median squares to obtain estimates of $\boldsymbol{\rho}$.

1.2 A Generalized Rank-Based Estimate

The proposed estimate of $\boldsymbol{\rho}$ will be a value of $\boldsymbol{\rho}$ that minimizes the following dispersion function,

$$\begin{aligned} D(\boldsymbol{\rho}) &= \sum_{1 \le i < j \le n} b_{ij} \left| \varepsilon_i - \varepsilon_j \right| \\ &= \sum_{1 \le i < j \le n} b_{ij} \left| (X_i - X_j) - (\boldsymbol{Y}_{i-1} - \boldsymbol{Y}_{j-1})' \boldsymbol{\rho} \right|, \end{aligned}$$

with b_{ij} denoting weights used for the $(i,j)^{th}$ comparison. For the case of the linear regression model, this estimate was first considered by Sievers (1983). Sievers showed this estimate to be a modified (or generalized) rank-based method. His paper also discusses the general theory for the estimate and eludes to the fact that the use of weights can be used to achieve an estimate with a totally bounded influence function. Naranjo and Hettmansperger (1994) expand on this idea of using the weights to achieve a bounded influence function. Specifically, they show that, for a special case of the weights, the estimate possesses a bounded influence function and has a positive breakdown point. Lastly, Chang (1996) shows that, for another special case of the weights, one obtains an estimate with a bounded influence function and a 50% breakdown point!

Upon examination of these papers one can divide the weights, b_{ij}, into three basic classes; constant, non-random, and random. Constant weights are simply obtained by letting $b_{ij} = K$, a constant. For the linear regression model and constant weights Hettmansperger and McKean (1978a) have shown that the dispersion function is equivalent to Jaeckel's (1972) dispersion function using Wilcoxon

scores. Since Koul (1993) has applied Jaeckel's dispersion function to the AR(p), one is referred to his paper for an account of using constant weights for the AR(p).

In the case of the linear regression model the non-random weight class includes weights that are functions of the design points. However, in the case of the AR(p) this would make the weights random. Thus, in the case of the AR(p) non-random weights are essentially weights determined by the researcher and not the data; although one might consider weights that are functions of time.

For the AR(p) we can sub-divide the random weight class into two categories. The first category is obtained when the weights are functions of the design points, $b_{ij} = b(\boldsymbol{Y}_{i-1}, \boldsymbol{Y}_{j-1})$. In this paper the weights used by Naranjo and Hettmansperger (1994) are considered. These weights are defined as $b_{ij} = h_i h_j$ where $h_i = h(\boldsymbol{Y}_{i-1})$. Although not immediate from the notation, these weights may depend on a location and scatter estimate for the set $\{\boldsymbol{Y}_{i-1}\}$. The effect of using these type of weights is essentially to downweight all leverage points. Since not all leverage points are "bad" leverage points, one may be interested in using weights that distinguish between "good" and "bad" leverage points. This leads to the second category of weights under the class of random weights. Following Chang (1996) one may consider weights that depend on the design points as well as the corresponding residuals from some initial estimate. That is, $b_{ij} = b(X_i, \boldsymbol{Y}_{i-1}; X_j, \boldsymbol{Y}_{j-1})$. Again, the notation may be misleading since these weights actually depend on an initial estimate of $\boldsymbol{\rho}$ and a location and scatter esti-

mate for $\{\boldsymbol{Y}_{i-1}\}$. In this dissertation we will develop the theory for the first type of random weights and leave the development of the theory for the second type for future work. We will, however, consider the second type of random weights in examples and simulations for the sake of comparisons.

Regardless of the class of weights, if one assumes only that the weights are positive and bounded it is easily shown that $D(\boldsymbol{\rho})$ is non-negative and piecewise linear. Furthermore, $D(\boldsymbol{\rho})$ is convex since the absolute value of a linear function is convex and the linear combination of convex functions is convex (Rockafellar, 1970, pg. 32–33). Hence, a minimum of $D(\boldsymbol{\rho})$ is guaranteed. Although the minimum is not necessarily unique, the diameter of the set of solutions tends to zero asymptotically. This basically follows from the fact (see Section 2.7 and or Section 3.6.2) that for any $\hat{\boldsymbol{\rho}}_n$ such that $D(\hat{\boldsymbol{\rho}}_n)$ is equal to the minimum, $\hat{\boldsymbol{\rho}}_n$ is a consistent estimator of $\boldsymbol{\rho}$ when $\boldsymbol{\rho}$ is the true parameter.

The gradient of $D(\boldsymbol{\rho})$ exists except at a finite number of points and is given by $\boldsymbol{\nabla} D(\boldsymbol{\rho}) = -\boldsymbol{S}(\boldsymbol{\rho})$ where, for $k = 1, 2, \ldots, p$,

$$S_k(\boldsymbol{\rho}) \;=\; 2 \sum_{1 \le i < j \le n} b_{ij}(X_{j-k} - X_{i-k}) \left(\varphi(\varepsilon_i(\boldsymbol{\rho}), \varepsilon_j(\boldsymbol{\rho})) - \frac{1}{2}\right).$$

Here, $\varepsilon_i(\boldsymbol{\rho}) = X_i - \boldsymbol{Y}'_{i-1}\boldsymbol{\rho}$ and $\varphi(u,v) = \frac{\operatorname{sgn}(v-u)+1}{2} \doteq I(u \le v)$ except at those points where the gradient does not exist. If we assume that the distribution of the $X's$ is continuous then the points where the derivative does not exist will have probability zero, and hence can be ignored. This assumption is essentially a continuity assumption on F. Thus, we may think of the estimate of $\boldsymbol{\rho}$ as

any $\hat{\boldsymbol{\rho}}_n$ such that $\boldsymbol{S}(\hat{\boldsymbol{\rho}}_n) \doteq 0$. The equation is approximate because $\boldsymbol{S}(\boldsymbol{\rho})$ is not continuous. This is readily seen from the inclusion of the indicator function in $\boldsymbol{S}(\boldsymbol{\rho})$. In fact, $\boldsymbol{S}(\boldsymbol{\rho})$ can take on only a finite range of values. The change of values occurs whenever $\boldsymbol{\rho}$ crosses one of the $\frac{n(n-1)}{2}$ hyperplanes, $H_{ij} = \{\boldsymbol{\theta} \in \Re^p : X_i - X_j = (\boldsymbol{Y}_{i-1} - \boldsymbol{Y}_{j-1})'\boldsymbol{\theta}\}$ (Naranjo, 1989, pg. 13).

1.3 Outliers in the AR(p)

In a designed linear regression experiment outliers can appear only in the response. For instance, this may occur if the error distribution, F, is heavy tailed or there is a recording error in the response. However, in an observational study the design points are random. Thus, it is possible for outliers to be present in the design points as well. An outlier appearing in the design points is usually referred to as a leverage point. Rousseeuw and Leroy (1987) classify a leverage point as a "good" leverage point if the design point is outlying but the fit results in a small residual. Conversely, if the design point is outlying and the fit results in a large residual then the leverage point is called a "bad" leverage point. Some robust linear regression estimates are not effected by "good" leverage points and even tend to have smaller variances. However, a number of these estimates may still be sensitive to "bad" leverage points. Thus, robust linear regression estimates that distinguish between "good" and "bad" leverage points have emerged (Coakley & Hettmansperger, 1993; Chang et al., 1996).

Recall that in time series analysis one observes a realization of the time series. Thus, outliers appear only in the response variable. To contrast outlier terminology in the linear regression model with that of the AR(p) it is important to consider how these outliers occur. If an outlier is introduced through the error distribution, F, it not only alters the current observation, but will also effect future observations as well. Thus, resulting in design points that contain outliers, or leverage points. However, due to the nature of the AR(p) these leverage points tend to be "good" leverage points. Secondly, recording errors and other outside influences can introduce an outlier or several outliers into the series. Since the outliers are not a result of the error distribution, F, the current observation is the only observation effected. However, in an AR(p) this will potentially result in p leverage points. When this situation occurs these leverage points usually turn out to be "bad" leverage points. Lastly, it is possible that F combined with outside influences can produce outliers or even "patches" of outliers. The results being both "good" and "bad" leverage points.

To model such occurrences, consider the stationary AR(p) defined by (1.1) and (1.2). If X_i denotes the "core" process the "observed" process can be defined as,

$$X_i^* = (1 - V_i)X_i + V_i Z_i. \tag{1.3}$$

In this model $V_1, V_2, \ldots$ denotes a sequence of binary zero-one random variables that may be correlated to account for patches of outliers and Z_i is a random

variable with a contaminating distribution. Bustos (1982) refers to this model as the Substitutive Outlier (SO) Model while Martin and Yohai (1991) refer to it as a General Replacement Outlier (RO) Model.

There are two special cases worth noting. If $V_i \equiv 0$ then $X_i^* = X_i$ and the observed process is actually the autoregressive time series. This is the situation when outliers enter into the observations through the error distribution, F. Fox (1972) referred to these types of outliers as Type II or "innovation" outliers. Secondly, when $V_i \equiv \frac{1}{2}$ and $Z_i = 2\nu_i + X_i$ one obtains Fox's (1972) Type I outlier model. Denby and Martin (1979) refer to this model as the "Additive" Outlier (AO) Model. The ν_i (not necessarily independent) are usually distributed as $(1-\gamma)\delta_0(\cdot) + \gamma G(\cdot)$ where γ denotes the proportion of contamination, δ_0 is a point mass at zero, and G is some contaminating distribution. Here, outliers are introduced by "adding" contamination to the core process. However, since the contamination at time i only effects the core process at time i future observations of the core process are not effected. This is the situation that gives rise to the "bad" leverage points. Lastly, it should be pointed out that the SO model is similar to the AO model except that instead of adding contamination to the core process, it "replaces" the core process with a contamination process.

To better understand model (1.3) consider the following AR(1) model for the core process, $X_i = 0.8X_{i-1} + \varepsilon_i$. Figure 1 displays time series plots, with associated lag one scatter plots given in Figure 2, for various situations under

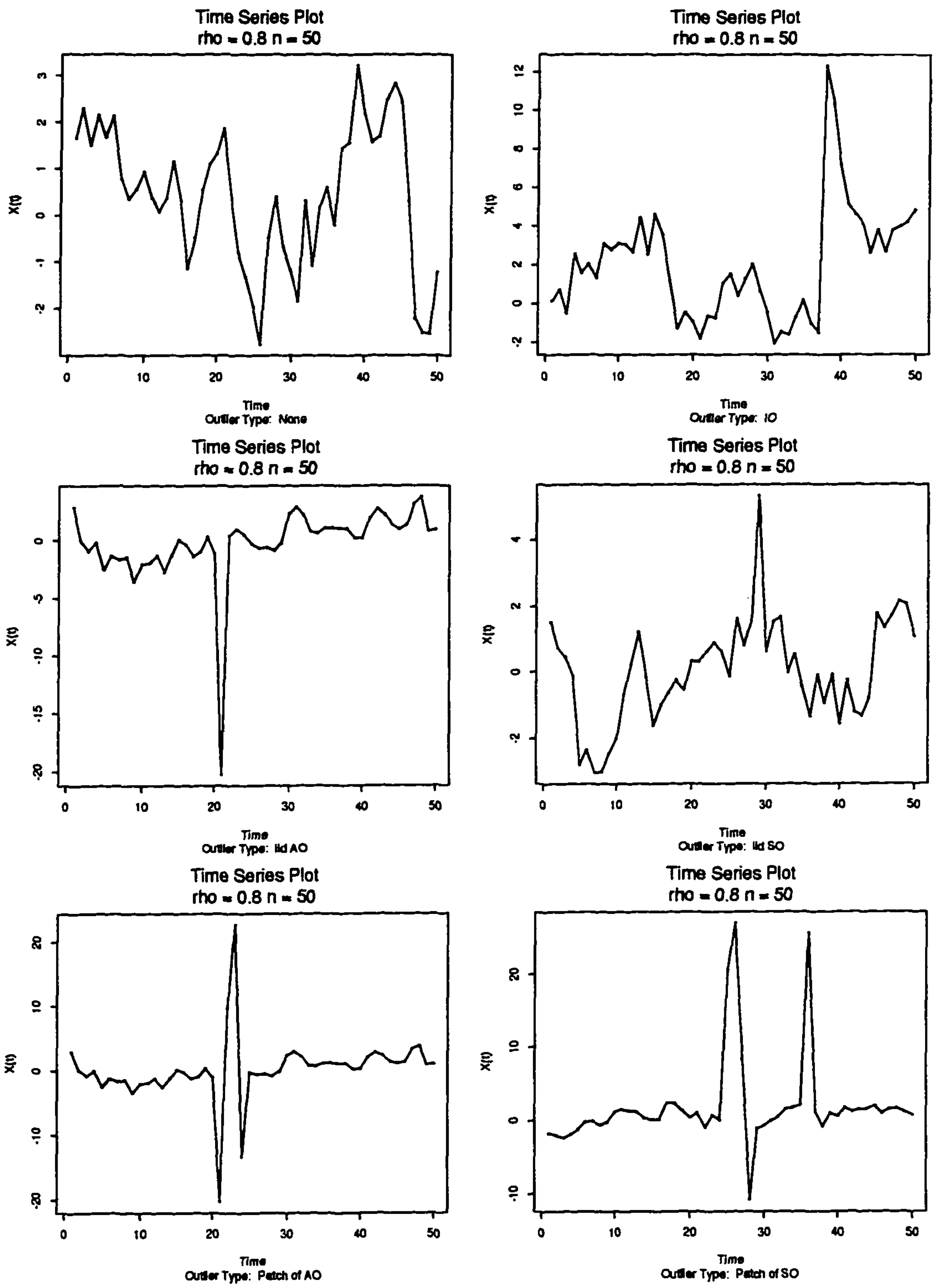

Figure 1. Time Series Plots for Outlier Situations.

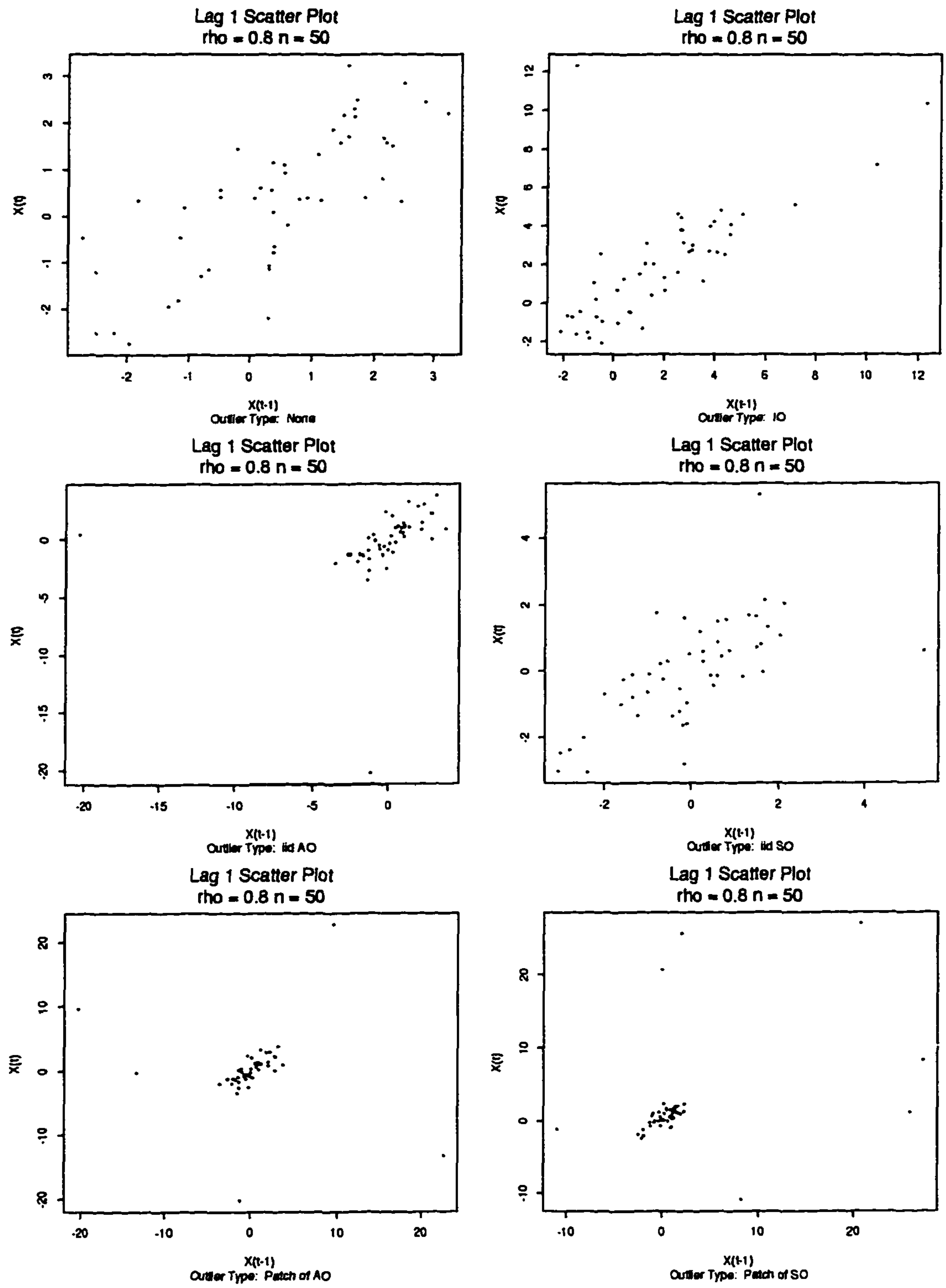

Figure 2. Lag One Scatter Plots for Outlier Situations.

the SO model. The first row in the figures portray IO situations; one when F is not heavy tailed and one when F is heavy tailed. When there are no outliers present one would expect that most estimation procedures would produce similar estimates as depicted in the associated lag one scatter plot. However, the other IO situation portrays on outlier at time index 38. One should note how the process eventually returns to its original state. Also, as pointed out in the lag one scatter plots, the outlier results in three "good" leverage points. The second row of the figures contains examples of an AO and SO model respectively when the ν_i are iid. Both time series plots exhibit an obvious outlier. However, unlike the IO situation these outliers do not effect future observations. Since we are dealing with an AR(1), there are two points of interest, as shown in the lag one scatter plots. When the outlier appears in the response, the design point is still good. This results in an outlier in residual space. Estimates that are bounded in residual space, such as R-estimates and M-estimates, are generally not effected by these types of points. However, since least squares estimates are not bounded in residual space, the least squares estimates will tend to be sensitive to these points. The second point of interest is when the outlier enters into the design. This results in a "bad" leverage point since its residual will not be consistent with the bulk of the data. Since least squares, M-estimates, and R-estimates are not bounded in factor space these estimates will tend to be bias towards zero. Hence, it is precisely these two situations that give rise to generalized R and M estimates, or

any other estimate that may be less sensitive to "bad" leverage points. Finally, the situation when the ν_i are not iid is portrayed in the last row of the figures. The AO and SO patch situations are similar to there iid counter parts in regards to defining outliers in residual space and "bad" leverage points. However, with patches it is also possible to get "good" leverage points.

CHAPTER II

ESTIMATION OF ρ USING NON-RANDOM WEIGHTS

2.1 Defining the Estimate

Consider the AR(1) given by (1.1) and (1.2). That is, assume the observations of the process can be modeled as,

$$X_i = \alpha + \rho X_{i-1} + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

where $\rho \in (-1, 1)$, X_0 is an observable random variable independent of $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$, and the ε_i are iid F. The condition on ρ is equivalent to assuming the process is stationary.

Following Section 1.2 the proposed estimate of ρ will be the value that minimizes the dispersion function,

$$\begin{aligned} D(\rho) &= \sum_{1 \le i < j \le n} b_{ij} \left| \varepsilon_i - \varepsilon_j \right| \\ &= \sum_{1 \le i < j \le n} b_{ij} \left| (X_i - X_j) - (X_{i-1} - X_{j-1})\rho \right| . \end{aligned}$$

The b_{ij} denote weights used for the $(i,j)^{th}$ comparison. Additionally, one can view the estimate of ρ as an approximate solution to the equation $S(\rho) = 0$ where,

$$S(\rho) = 2 \sum_{1 \le i < j \le n} b_{ij} (X_{j-1} - X_{i-1}) \left(\varphi(\varepsilon_i(\rho), \varepsilon_j(\rho)) - \frac{1}{2} \right),$$

and $\varphi(u,v) = \frac{\mathrm{sgn}(v-u)+1}{2} \doteq I(u \le v)$ except at those points where the gradient does not exist. The equation is approximate since $S(\rho)$ is a step function that changes values at the sample slopes, $H_{ij} = \frac{X_i - X_j}{X_{i-1} - X_{j-1}}$.

In this chapter the weights will be considered non-random. For instance, the b_{ij}'s are non-random when $b_{ij} = 1$, $b_{ij} = b(i,j)$, or the b_{ij}'s are predetermined by the investigator. To motivate the practicality of non-random weights consider the following hypothetical example. Suppose one observes a yearly time series which consists of average rainfall amounts in a given region. Typically, there exists a-priori knowledge of those years considered as "drought" years. Thus, one may wish to develop a weighting scheme that discounts the drought years. For instance, one may want to downweight all comparisons involving drought years by a specified amount. Alternatively, one may wish to downweight all comparisons involving drought years, but have the degree of downweighting depend on the elapsed time between the drought year and the comparison year. Furthermore, one may want to disregard the drought years entirely by letting $b_{ij} = 0$ whenever i or j represent a drought year.

As a final note, it is conjectured that the results presented in this chapter generalize to the AR(p). The case when $p > 1$ is not considered for the sake of simplicity. However, it is hoped that consideration of the AR(1) will provide enough insight into the problem so that a generalization of the results to the AR(p) can be made with a minimal amount of effort.

2.2 Assumptions for the Asymptotic Theory

We begin this section by stating a list of assumptions that will be needed for the asymptotic theory of the estimate. Assumptions denoted with a "M" represent conditions pertaining to the model, assumptions denoted with an "E" represent conditions on the error distribution, and assumptions denoted with a "W" represent constraints on the weights.

M1. $X_i = \alpha + \rho X_{i-1} + \varepsilon_i$ where $i = 1, 2, \ldots, n$ and $|\rho| < 1$

M2. X_0 is an observable random variable independent of $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ and is such that $E\left[X_0^4\right] < \infty$

E1. $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are iid F random variables with $E\left[\varepsilon_1\right] = 0$ and $E\left[\varepsilon_1^4\right] < \infty$

E2. The density function, f, of F is such that f is absolutely continuous, $f > 0$ a.e., f has finite Fisher Information, and f' is uniformly bounded

W1. $\limsup_{n\to\infty} |b_{ij}| \leq B_b < \infty$

W2. $\frac{1}{n}\sum_{i=1}^{n} b_{i.}^2 \longrightarrow \eta_b > 0$ where $b_{i.} = \frac{1}{n}\sum_{j=1}^{n} b_{ij}$

We now consider some comments concerning the above assumptions. First, under assumptions M1 and E1 the mean of the process is $\mu_X = E\left[X_1\right] = \frac{\alpha}{1-\rho}$. If we define the "centered" process as $X_i^c = X_i - \mu_X$ and rewrite the model in terms

of the centered process we get the following,

$$\begin{aligned} X_i^c + \mu_X &= \alpha + \rho(X_{i-1}^c + \mu_X) + \varepsilon_i \\ &= \alpha + \rho\left(\frac{\alpha}{1-\rho}\right) + \rho X_{i-1}^c + \varepsilon_i \\ &= \frac{\alpha}{1-\rho} + \rho X_{i-1}^c + \varepsilon_i. \end{aligned}$$

Hence, rewriting M1 in terms of the centered process yields,

$$X_i^c = \rho X_{i-1}^c + \varepsilon_i. \tag{2.1}$$

One should note that the ρ in the centered model is the same ρ that appears in the non-centered model. Thus, since we are primarily interested in the estimation of ρ, one can assume without loss of generality that the process has a zero mean. That is, without loss of generality, one can assume $E[X_1] = 0$. For convenience in subsequent discussions we will drop the X_i^c notation and just write X_i (keeping in mind that $E[X_1] = 0$).

Although the estimation of ρ is our primary concern, a couple of comments concerning the estimation of α are in order. In practice, one should first center the data with an unbiased robust estimate of location, $\hat{\mu}_X$. Then, using the proposed estimate, one can fit (2.1) to obtain a $\hat{\rho}_n$. Once an estimate of ρ has been determined define the residuals as $\hat{\varepsilon}_i = X_i - \hat{\rho}_n X_{i-1}$. Since M1 implies $X_i - \rho X_{i-1} = \alpha + \varepsilon_i$, one can fit the model $\hat{\varepsilon}_i = \alpha + \varepsilon_i$ using a robust estimate of location to obtain an $\hat{\alpha}_n$. Since $\mu_X = \frac{\alpha}{1-\rho}$, one may also consider $\hat{\alpha}_n = \hat{\mu}_X(1 - \hat{\rho}_n)$ as an alternative estimate.

Next, consider the conditions given in assumption E2. The following lemma summarizes a few useful properties concerning the density function. The proofs of these results are well-known (e. g. Koul (1992)), but are presented for the sake of completeness.

Lemma 2.2.1 *Under assumption E2 we have the following:*

1. *f has a uniform bound, B_f*

2. *f is uniformly continuous*

3. *$\tau = \int_{-\infty}^{\infty} f^2(t)dt$ is finite*

proof(1). Without loss of generality, assume f is defined at 0 and let x denote any point such that $f(x)$ is well-defined. Then, using the absolute continuity of f, the fact that $f > 0$ a.e., and the Cauchy Schwarz Inequality one obtains the following,

$$\begin{aligned}
|f(x) - f(0)| &= \left|\int_{-\infty}^{x} f'(t)dt - \int_{-\infty}^{0} f'(t)dt\right| \\
&= \left|\int_{0}^{x} f'(t)dt\right| \\
&= \left|\int_{0}^{x} \frac{f'(t)\sqrt{f(t)}}{\sqrt{f(t)}}dt\right| \\
&\leq \sqrt{\int_{0}^{x}\left(\frac{f'(t)}{\sqrt{f(t)}}\right)^2 dt}\sqrt{\int_{0}^{x} f(t)dt} \\
&\leq \sqrt{\int_{0}^{x}\left(\frac{f'(t)}{\sqrt{f(t)}}\right)^2 dt}\sqrt{|F(x) - F(0)|} \\
&\leq \sqrt{I(f)},
\end{aligned}$$

where $I(f)$ denotes the finite Fisher Information. Thus, for any x such that $f(x)$ is defined we have,

$$f(0)-\sqrt{I(f)} \;\leq\; f(x) \;\leq\; f(0)+\sqrt{I(f)}.$$

This implies that f is uniformly bounded. □

proof(2). Replacing the 0 with a y in the above derivations and applying the Mean Value Theorem yields the following,

$$\begin{aligned} |f(x)-f(y)| &\leq \sqrt{I(f)}\sqrt{|F(x)-F(y)|} \\ &= \sqrt{I(f)}\sqrt{f(\xi)\,|x-y|} \\ &\leq \sqrt{I(f)B_f}\sqrt{|x-y|}, \end{aligned}$$

where ξ is a point between x and y and B_f represents the uniform bound on f. Thus, for any $\varepsilon > 0$, $|x-y| < \frac{\varepsilon^2}{I(f)B_f}$ implies $|f(x)-f(y)| < \varepsilon$. Hence, f is uniformly continuous. □

proof(3). Similar to the proof of 1, one can use the absolute continuity of f, the fact that $f > 0$ a.e., and the Cauchy Schwarz Inequality to show the following,

$$\begin{aligned} \tau &= \int_{-\infty}^{\infty} f^2(t)dt \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{t} f'(s)ds\right) f(t)dt \\ &\leq \int_{-\infty}^{\infty}\left|\int_{-\infty}^{t} f'(s)ds\right| f(t)dt \\ &\leq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f'(s)|\, ds f(t)dt \\ &= \int_{-\infty}^{\infty} |f'(s)|\, ds \end{aligned}$$

$$
\begin{aligned}
&= \left| \int_{-\infty}^{\infty} \frac{|f'(s)|\sqrt{f(s)}}{\sqrt{f(s)}} ds \right| \\
&\leq \sqrt{\int_{-\infty}^{\infty} \left(\frac{f'(s)}{\sqrt{f(s)}} \right)^2 ds} \sqrt{\int_{-\infty}^{\infty} f(s) ds} \\
&= \sqrt{I(f)}.
\end{aligned}
$$

Since $I(f)$ is finite, the proof is complete. □

Finally, consider the conditions on the weights. Although W1 implies that the weights are bounded by B_b, one can assume that $B_b = 1$. The idea behind weighting schemes is to downweight outlying observations. Thus, for our situation it does not seem reasonable to consider weights that are larger than one. Additionally, note that upon dividing the numerator and denominator by n, W1 and W2 together imply the following Noether type condition on the $b_{i.}^2$'s,

$$
\frac{\max_{1 \leq i \leq n} b_{i.}^2}{\sum_{i=1}^{n} b_{i.}^2} \longrightarrow 0.
$$

However, the Noether condition does not imply W1 and W2 simultaneously as is easily seen by letting $b_{ij} = \sqrt{n}$. Thus, W1 and W2 together comprise a stronger condition than the classic Noether type condition. However, since the weights are considered non-random, and are left to the discretion of the experimenter, this is irrelevant.

2.3 Notation Used in the Asymptotic Theory

In this section we define some notation and random variables that will be critical to the theoretical development of the estimate. In many of the proofs

involving expectations it will be necessary to condition on "past" information. For convenience, let $\mathcal{F}_i = \sigma\text{-field}\{X_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i\}$; $i = 1, 2, \ldots, n$ denote the smallest σ-field generated by $\{X_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i\}$ and let $\mathcal{F}_0 = \sigma\text{-field}\{X_0\}$. Conditioning on a σ-field such as $\mathcal{F}_i$ will essentially allow us to consider any random variables depending on the set $\{X_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i\}$ as fixed.

Now, for $u \leq v$, define the following random variable,

$$\Gamma_u^v = \sum_{r=0}^{v-u} \rho^r \varepsilon_{v-r}.$$

Typically, this random variable will denote a finite linear combination of present and future error terms. In what follows we will let $\sigma^2 = E[\varepsilon_1^2]$. The following lemma establishes bounds on the second and fourth moments of Γ_u^v that are independent of u and v.

Lemma 2.3.1 *Under M1, E1, and $u \leq v$ we have the following:*

1. $E[\Gamma_u^v] = 0$

2. $E\left[(\Gamma_u^v)^2\right] = \frac{\sigma^2}{1-\rho^2}(1 - \rho^{2(v-u+1)}) \leq \frac{\sigma^2}{1-\rho^2} < \infty$

3. $E\left[(\Gamma_u^v)^4\right] \leq E[\varepsilon_1^4]\left(\frac{1-\rho^{4(v-u+1)}}{1-\rho^4}\right) + 2\sigma^4\left(\frac{1-\rho^{2(v-u+1)}}{1-\rho^2}\right)^2 \leq$
   $E[\varepsilon_1^4]\left(\frac{2}{1-\rho^4}\right) + 2\sigma^4\left(\frac{2}{1-\rho^2}\right)^2 < \infty$

proof(1). The proof follows directly from the fact that $E[\varepsilon_1] = 0$ and that an expectation of a sum is the sum of expectations. □

proof(2). The proof follows from the independence of the errors, the fact that

$E[\varepsilon_1] = 0$, and the formula for a finite sum of a geometric series.

$$\begin{aligned} E\left[(\Gamma_u^v)^2\right] &= E\left[\sum_{r=0}^{v-u} \rho^{2r}\varepsilon_{v-r}^2 + \sum_{(r\neq q)} \rho^r\rho^q\varepsilon_{v-r}\varepsilon_{v-q}\right] \\ &= \sum_{r=0}^{v-u} \rho^{2r} E\left[\varepsilon_{v-r}^2\right] + \sum_{(r\neq q)} E\left[\rho^r\rho^q\varepsilon_{v-r}\varepsilon_{v-q}\right] \\ &= \sigma^2 \sum_{r=0}^{v-u} \rho^{2r} \\ &= \sigma^2 \left(\frac{1-\rho^{2(v-u+1)}}{1-\rho^2}\right) \quad \square \end{aligned}$$

proof(3). The proof follows much like that of (2). The idea is to expand the fourth order sum into the different terms and take expectations of these terms. Some of the terms will drop out do to the independence of the ε's and the fact that $E[\varepsilon_1] = 0$. The remaining terms comprise what is left in (3). $\square$

The basic idea in many of the proofs involving expectations is to rewrite a present value of the process as a sum of a past value of the process and a linear combination of past and present error terms. This idea along with some results pertaining to the moments of the process are established in the following lemma.

Lemma 2.3.2 *Under assumptions M1, M2, and E1 we have the following:*

1. *For* $u \leq v$, $X_v = \rho^{v-u}X_u + \Gamma_{u+1}^v$
2. *For* $v = 1, 2, \ldots, n$, $E[X_v^2] = E[X_1^2] = \frac{\sigma^2}{1-\rho^2} < \infty$
3. *For* $v = 1, 2, \ldots, n$, $E[X_v^4] = E[X_1^4] < \infty$
4. *For* $u \leq v$, $E\left[(X_{v-1} - X_{u-1})^4\right] \leq 32E[X_1^4] < \infty$

proof(1). The proof follows by exploiting the recursive nature of the AR(1) process. □

proof(2). The proof can be found in most standard textbooks on time series. For example, see Abraham and Ledolter (1983, pg. 200). □

proof(3). The proof follows from the stationarity of the X's and the fact that,

$$\begin{aligned} E\left[X_1^4\right] &= E\left[(\rho X_0 + \varepsilon_1)^4\right] \\ &\leq E\left[2^4(|\rho X_0|^4 + |\varepsilon_1|^4)\right] \\ &\leq 16E\left[(|X_0|^4 + |\varepsilon_1|^4)\right] \\ &< \infty. \ \square \end{aligned}$$

proof(4). The proof follows from an application of part 3 on the following inequality,

$$E\left[(X_{v-1} - X_{u-1})^4\right] \leq E\left[2^4(|X_{v-1}|^4 + |X_{u-1}|^4)\right]. \ \square$$

Another important property pertaining to the values of the process is given in the next lemma.

Lemma 2.3.3 *Assuming only stationarity of the X's and $E[X_1^2] < \infty$ we have the following:*

$$\frac{1}{\sqrt{n}} \max_{1 \leq t \leq n} |X_t| = o_p(1)$$

proof. The proof of this lemma can be found in Koul (1992, pg. 227). However, it is presented here for the sake of completeness. Choose $\epsilon > 0$. Then we can show

the following,

$$\begin{aligned}
P\left[\frac{1}{\sqrt{n}}\max_{1\le t\le n}|X_t|>\epsilon\right] &= P\left[\max_{1\le t\le n}|X_t|>\sqrt{n}\epsilon\right]\\
&= P\left[\max_{1\le t\le n}|X_t|^2>n\epsilon^2\right]\\
&\le \sum_{t=1}^{n}P\left[|X_t|^2>n\epsilon^2\right]\\
&= \sum_{t=1}^{n}E\left[I(|X_t|^2>n\epsilon^2)\right]\\
&\le \sum_{t=1}^{n}E\left[\frac{|X_t|^2}{n\epsilon^2}I(|X_t|^2>n\epsilon^2)\right]\\
&= \frac{1}{n\epsilon^2}\sum_{t=1}^{n}E\left[|X_t|^2 I(|X_t|^2>n\epsilon^2)\right]\\
&= \frac{1}{\epsilon^2}E\left[|X_1|^2 I(|X_1|^2>n\epsilon^2)\right]\\
&= \frac{1}{\epsilon^2}E\left[|X_1|^2 I(|X_1|>\sqrt{n}\epsilon)\right]\\
&= \frac{1}{\epsilon^2}\int_{|X_1|>\sqrt{n}\epsilon}|X_1|^2 dP.
\end{aligned}$$

The lemma now follows by taking the limit as $n \longrightarrow \infty$ through the above inequality. □

Now, for $\Delta \in \Re$ define the following random variable,

$$t_{uv}(\Delta) = \frac{\Delta}{\sqrt{n}}(X_{v-1} - X_{u-1}).$$

This random variable will play a key role in proving the linearity result given in Theorem 2.5.1. The following result is a direct consequence of Lemma 2.3.3.

Corollary 2.3.1 *Under the assumptions of Lemma 2.3.3 we have for all $\Delta \in \Re$:*

$$\max_{1\le u,v\le n}|t_{uv}(\Delta)| = o_p(1)$$

proof.

$$\begin{aligned}
\max_{1\le u,v\le n} |t_{uv}(\Delta)| &= \max_{1\le u,v\le n} |\frac{\Delta}{\sqrt{n}}(X_{v-1} - X_{u-1})| \\
&\le \frac{|\Delta|}{\sqrt{n}} \max_{1\le u,v\le n} (|X_{v-1}| + |X_{u-1}|) \\
&\le \frac{2|\Delta|}{\sqrt{n}} \max_{1\le t\le n} |X_t| \\
&= 2|\Delta| \frac{1}{\sqrt{n}} \max_{1\le t\le n} |X_t| \\
&= o_p(1) \ \square
\end{aligned}$$

The last random variable that will be needed for the asymptotic theory is defined below,

$$W_{uv} = \begin{cases} -1 & \text{if } \varepsilon_u < \varepsilon_v < \varepsilon_u + t_{uv}(\Delta) \\ 0 & \text{otherwise} \\ 1 & \text{if } \varepsilon_u + t_{uv}(\Delta) < \varepsilon_v < \varepsilon_u. \end{cases}$$

Some important results pertaining to W_{uv} are presented next.

Lemma 2.3.4 *Under assumption E2 and $u < v$ we have the following:*

$$E\left[W_{uv} \mid \mathcal{F}_{v-1}\right] = -f(\xi_{uv}(\Delta))t_{uv}(\Delta)$$

where $|\xi_{uv}(\Delta) - \varepsilon_u| \le |t_{uv}(\Delta)|$.

proof. The proof follows from the definition of W_{uv}, the independence of ε_v and $(\varepsilon_u, t_{uv}(\Delta))$, and an application of the Mean Value Theorem applied to F.

$$\begin{aligned}
&E\left[W_{uv} \mid \mathcal{F}_{v-1}\right] \\
&= (-1)P\left[\varepsilon_u < \varepsilon_v < \varepsilon_u + t_{uv}(\Delta) \mid \mathcal{F}_{v-1}\right] + (1)P\left[\varepsilon_u + t_{uv}(\Delta) < \varepsilon_v < \varepsilon_u \mid \mathcal{F}_{v-1}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \begin{cases} -[F(\varepsilon_u + t_{uv}(\Delta)) - F(\varepsilon_u)] & \text{if } t_{uv}(\Delta) > 0 \\ F(\varepsilon_u) - F(\varepsilon_u + t_{uv}(\Delta)) & \text{if } t_{uv}(\Delta) < 0 \end{cases} \\
&= F(\varepsilon_u) - F(\varepsilon_u + t_{uv}(\Delta)) \\
&= -f(\xi_{uv}(\Delta)) t_{uv}(\Delta)
\end{aligned}
$$

where $|\xi_{uv}(\Delta) - \varepsilon_u| \le |t_{uv}(\Delta)|$. Note that Corollary 2.3.1 implies that $|\xi_{uv}(\Delta) - \varepsilon_u| = o_p(1)$. □

Corollary 2.3.2 *Under assumptions M1, M2, E1, E2, and $u < v$ we have the following:*

$$
E[W_{uv}] = \frac{-\Delta}{\sqrt{n}} E[f(\xi_{uv}(\Delta))(X_{v-1} - X_{u-1})] = o(1)
$$

proof.

$$
\begin{aligned}
E[W_{uv}] &= E[E[W_{uv} \mid \mathcal{F}_{v-1}]] \\
&= E[-f(\xi_{uv}(\Delta)) t_{uv}(\Delta)] \\
&= \frac{-\Delta}{\sqrt{n}} E[f(\xi_{uv}(\Delta))(X_{v-1} - X_{u-1})] \\
&\le \frac{|\Delta|}{\sqrt{n}} E[|f(\xi_{uv}(\Delta))|\,|X_{v-1} + X_{u-1}|] \\
&\le \frac{2|\Delta| B_f}{\sqrt{n}} E[|X_1|] \\
&= o(1) \quad \square
\end{aligned}
$$

It should be noted that $|W_{uv}| = {W_{uv}}^2 = {W_{uv}}^4$ where,

$$
{W_{uv}}^2 = \begin{cases} 1 & \text{if } \varepsilon_u < \varepsilon_v < \varepsilon_u + t_{uv}(\Delta) \text{ or } \varepsilon_u + t_{uv}(\Delta) < \varepsilon_v < \varepsilon_u \\ 0 & \text{otherwise.} \end{cases}
$$

This random variable will also need to be considered. Thus, we present some results pertaining to ${W_{uv}}^2$.

Lemma 2.3.5 *Under assumption E2 and $u < v$ we have the following:*

$$E\left[{W_{uv}}^2 \mid \mathcal{F}_{v-1}\right] = f(\xi_{uv}(\Delta))|t_{uv}(\Delta)|$$

where $|\xi_{uv}(\Delta) - \varepsilon_u| \leq |t_{uv}(\Delta)|$.

proof. The proof follows from the definition of ${W_{uv}}^2$, the independence of ε_v and $(\varepsilon_u, t_{uv}(\Delta))$, and an application of the Mean Value Theorem applied to F.

$$\begin{aligned}
&E\left[{W_{uv}}^2 \mid \mathcal{F}_{v-1}\right] \\
&= (1)P\left[\{\varepsilon_u + t_{uv}(\Delta) < \varepsilon_v < \varepsilon_u\} \bigcup \{\varepsilon_u < \varepsilon_v < \varepsilon_u + t_{uv}(\Delta)\} \mid \mathcal{F}_{v-1}\right] \\
&= P\left[\varepsilon_u + t_{uv}(\Delta) < \varepsilon_v < \varepsilon_u \mid \mathcal{F}_{v-1}\right] + P\left[\varepsilon_u < \varepsilon_v < \varepsilon_u + t_{uv}(\Delta) \mid \mathcal{F}_{v-1}\right] \\
&= \begin{cases} F(\varepsilon_u + t_{uv}(\Delta)) - F(\varepsilon_u)] & \text{if } t_{uv}(\Delta) > 0 \\ F(\varepsilon_u) - F(\varepsilon_u + t_{uv}(\Delta)) & \text{if } t_{uv}(\Delta) < 0 \end{cases} \\
&= \begin{cases} f(\xi_{uv}(\Delta))t_{uv}(\Delta) & \text{if } t_{uv}(\Delta) > 0 \\ -f(\xi_{uv}(\Delta))t_{uv}(\Delta) & \text{if } t_{uv}(\Delta) < 0 \end{cases} \\
&= f(\xi_{uv}(\Delta))|t_{uv}(\Delta)|
\end{aligned}$$

where $|\xi_{uv}(\Delta) - \varepsilon_u| \leq |t_{uv}(\Delta)|$. Again, it should be noted that Corollary 2.3.1 implies that $|\xi_{uv}(\Delta) - \varepsilon_u| = o_p(1)$. □

Corollary 2.3.3 *Under assumptions M1, M2, E1, E2, and $u < v$ we have the following:*

$$E\left[{W_{uv}}^2\right] = \frac{|\Delta|}{\sqrt{n}} E\left[f(\xi_{uv}(\Delta))|X_{v-1} - X_{u-1}|\right] = o(1)$$

proof.

$$\begin{aligned}
E\left[W_{uv}{}^2\right] &= E\left[E\left[W_{uv}{}^2 \mid \mathcal{F}_{v-1}\right]\right] \\
&= E\left[f(\xi_{uv}(\Delta))|t_{uv}(\Delta)|\right] \\
&= \frac{|\Delta|}{\sqrt{n}} E\left[f(\xi_{uv}(\Delta))|X_{v-1} - X_{u-1}|\right] \\
&\leq \frac{2|\Delta| B_f}{\sqrt{n}} E\left[|X_1|\right] \\
&= o(1) \quad \square
\end{aligned}$$

Lemma 2.3.6 *Let $\psi_{ij} = (X_{j-1} - X_{i-1})(W_{ij} + \tau t_{ij}(\Delta))$. Then, under assumptions M1, M2, E1, and E2, we have the following:*

1. *$E\left[\psi_{ij}^2\right]$ is uniformly bounded*

2. *$E\left[\psi_{ij}^2\right] = o(1)$ for all (i, j)*

proof(1). The Cauchy Schwarz Inequality and part 4 of Lemma 2.3.2 imply the following,

$$\begin{aligned}
E\left[\psi_{ij}^2\right] &= E\left[(X_{j-1} - X_{i-1})^2 (W_{ij} + \tau t_{ij}(\Delta))^2\right] \\
&\leq \sqrt{E\left[(X_{j-1} - X_{i-1})^4\right] E\left[(W_{ij} + \tau t_{ij}(\Delta))^4\right]} \\
&\leq \sqrt{K E\left[(W_{ij} + \tau t_{ij}(\Delta))^4\right]},
\end{aligned}$$

where $K = 32E[X_1^4]$. Now,

$$\begin{aligned}
E\left[(W_{ij} + \tau t_{ij}(\Delta))^4\right] &\leq E\left[16 W_{ij}{}^4 + 16\tau^4 t_{ij}^4(\Delta)\right] \\
&= 16E\left[W_{ij}{}^4\right] + \frac{16\tau^4\Delta^4}{n^2} E\left[(X_{j-1} - X_{i-1})^4\right].
\end{aligned}$$

Thus, using the bound on W_{ij}, the finiteness of τ, and part 4 of Lemma 2.3.2 we can bound the right-hand side of the above inequality by a constant that depends on τ, Δ, and $E[X_1^4]$. □

proof(2). The proof of 2 follows from the same inequality, part 4 of Lemma 2.3.2, and the fact that $E\left[W_{ij}{}^4\right] = o(1)$ for all (i,j) by Corollary 2.3.3. □

Finally, consider a random variable of the form,

$$T_n = \sum_{i<j}\sum_{k<l} t_{ijkl},$$

where $\{t_{ijkl}\}$ is a set of random variables defined for $1 \leq i < j \leq n$ and $1 \leq k < l \leq n$. In what follows it will be neccessary to take the expectation of such a random variable. Because of the recursive nature of the AR(1), the order of the subscripts will be critical when evaluating a term such as $E[t_{ijkl}]$. Thus, for the sake of reference, Table 1 yields the possible cases in which the subscripts can be arranged. The first thing to note is that under cases 1, 3, 5, 9, 11, and 13 the sum is $O(n^4)$. Under cases 2, 4, 6, 8, 10, and 12 the sum is $O(n^3)$; and under case 7 the sum is $O(n^2)$. Secondly, if $E[t_{ijkl}]$ is symmetric in the indices then there exists symmetry in the table. That is, if $E[t_{ijkl}]$ is symmetric in the indices then one needs only consider cases 1 through 7.

2.4 The Asymptotic Distribution of the Gradient

The purpose of this section is to derive the asymptotic distribution of the gradient, $S_n(\rho_0)$, where ρ_0 denotes the true autoregressive parameter. Our proof

Table 1

Subscript Arrangements in a Summation With Four Indices

Case	Subscript Order	$j-i<l-k$	$j-i=l-k$	$j-i>l-k$
1	$i<j<k<l$	X	X	X
2	$i<j=k<l$	X	X	X
3	$i<k<j<l$	X	X	X
4	$i<k<j=l$			X
5	$i<k<l<j$			X
6	$i=k<j<l$	X		
7	$i=k<l=j$		X	
8	$i=k<l<j$			X
9	$k<i<j<l$	X		
10	$k<i<j=l$	X		
11	$k<i<l<j$	X	X	X
12	$k<l=i<j$	X	X	X
13	$k<l<i<j$	X	X	X

will involve martingale results; therefore, it will be convenient to first state a definition pertaining to martingales and then a central limit theorem for martingale differences.

2.4.1 Martingale Results

The following definition can be found on page 257 of Koul (1992).

Definition 2.4.1 *Let* $\{\mathcal{F}_{n,t} : 1 \le t \le n; n \ge 1\}$ *be an array of sub* σ*-fields such that* $\mathcal{F}_{n,t} \subset \mathcal{F}_{n,t+1}$, $1 \le t \le n$. *Furthermore, let* $S_{n,j} = \sum_{t=1}^{j} Z_{n,t}$, $1 \le j \le n$. *Then,* $\{S_{n,j}, \mathcal{F}_{n,j} : 1 \le j \le n, n \ge 1\}$ *is a zero-mean square-integrable martingale array with differences* $\{Z_{n,j} : 1 \le j \le n, n \ge 1\}$ *if the following three conditions*

are satisfied:

1. $Z_{n,j}$ *is* $\mathcal{F}_{n,j}$ *measurable*

2. $E\left[Z_{n,j}^2\right] < \infty$

3. $E\left[Z_{n,j} \,|\mathcal{F}_{n,j-1}\right] = 0.$

We are now ready to state the martingale central limit theorem that will be used to prove the asymptotic normality of the estimate. The following theorem, along with a proof, can be found on page 58 of Hall and Heyde (1980).

Theorem 2.4.1 *Let* $\{S_{n,j}, \mathcal{F}_{n,j} : 1 \leq j \leq n, n \geq 1\}$ *be a zero-mean square-integrable martingale array with differences* $Z_{n,j}$. *Furthermore, let* η^2 *be an a.s. finite random variable, and assume the following four conditions are satisfied:*

1. $\max_{1\leq j\leq n} |Z_{n,j}| = o_p(1)$

2. $\sum_{j=1}^n Z_{n,j}^2 = \eta^2 + o_p(1)$

3. $E\left[\max_{1\leq j\leq n} Z_{n,j}^2\right] = O(1)$

4. $\mathcal{F}_{n,j} \subseteq \mathcal{F}_{n+1,j}$ *for* $1 \leq j \leq n$ *and* $n \geq 1$.

Then S_{nn} *converges in distribution to a random variable* Z *whose characteristic function at* t *is* $E\left[exp\left(-\frac{1}{2}\eta^2 t^2\right)\right]$. *In particular, when* η^2 *is a constant, the following is true,*

$$S_{nn} = \sum_{j=1}^{n} Z_{n,j} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \eta^2\right).$$

2.4.2 Preliminary Results

In this section we present two results that will be used in the proof of the asymptotic distribution of the gradient, Theorem 2.4.2. Although the following is an abuse of notation, define $g_1 \equiv g_1(\mathcal{F}_{u-1})$ to be an arbitrary random variable that depends on a subset of $\{X_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{u-1}\}$. The two lemmas pertain to expectations that we will encounter.

Lemma 2.4.1 *Assume assumptions M1, M2, and E1 hold. If $u < v$ then we have the following expectation.*

$$E\left[g_1(\mathcal{F}_u)X_{v-1}\right] = \rho^{v-u-1}\left(\rho E\left[X_{u-1}g_1(\mathcal{F}_u)\right] + E\left[g_1(\mathcal{F}_u)\varepsilon_u\right]\right)$$

proof.

$$\begin{aligned}
E\left[g_1(\mathcal{F}_u)X_{v-1}\right] &= E\left[g_1(\mathcal{F}_u)\left(\rho^{v-u}X_{u-1} + \Gamma_u^{v-1}\right)\right] \\
&= \rho^{v-u}E\left[g_1(\mathcal{F}_u)X_{u-1}\right] + E\left[g_1(\mathcal{F}_u)\Gamma_u^{v-1}\right] \\
&= \rho^{v-u}E\left[g_1(\mathcal{F}_u)X_{u-1}\right] + \\
&\quad\; E\left[g_1(\mathcal{F}_u)\left(\varepsilon_{v-1} + \rho\varepsilon_{v-2} + \ldots + \rho^{v-u-1}\varepsilon_u\right)\right] \\
&= \rho^{v-u}E\left[g_1(\mathcal{F}_u)X_{u-1}\right] + \rho^{v-u-1}E\left[g_1(\mathcal{F}_u)\varepsilon_u\right] \\
&= \rho^{v-u-1}\left(\rho E\left[X_{u-1}g_1(\mathcal{F}_u)\right] + E\left[g_1(\mathcal{F}_u)\varepsilon_u\right]\right). \;\square
\end{aligned}$$

Lemma 2.4.2 *Assume assumptions M1, M2, and E1 hold. If $u < v$ then we have the following expectation.*

$$E\left[g_1(\mathcal{F}_u)X_{v-1}^2\right] =$$

$$\rho^{2(v-u)}E\left[X_{u-1}^2 g_1(\mathcal{F}_u)\right] + 2\rho^{v-u}\rho^{v-u-1}E\left[X_{u-1}g_1(\mathcal{F}_u)\varepsilon_u\right]$$
$$+\rho^{2(v-u-1)}E\left[g_1(\mathcal{F}_u)\varepsilon_u^2\right] + E\left[g_1(\mathcal{F}_u)\right]E\left[\left(\Gamma_{u+1}^{v-1}\right)^2\right]$$

proof.

$$\begin{aligned}
&E\left[g_1(\mathcal{F}_u)X_{v-1}^2\right] \\
&= E\left[g_1(\mathcal{F}_u)\left(\rho^{v-u}X_{u-1} + \Gamma_u^{v-1}\right)^2\right] \\
&= E\left[g_1(\mathcal{F}_u)\left(\rho^{2(v-u)}X_{u-1}^2 + 2\rho^{v-u}X_{u-1}\Gamma_u^{v-1} + \left(\Gamma_u^{v-1}\right)^2\right)\right] \\
&= \rho^{2(v-u)}E\left[g_1(\mathcal{F}_u)X_{u-1}^2\right] + 2\rho^{v-u}E\left[g_1(\mathcal{F}_u)X_{u-1}\Gamma_u^{v-1}\right] + E\left[g_1(\mathcal{F}_u)\left(\Gamma_u^{v-1}\right)^2\right] \\
&= \rho^{2(v-u)}E\left[X_{u-1}^2 g_1(\mathcal{F}_u)\right] + \\
&\quad 2\rho^{v-u}E\left[X_{u-1}g_1(\mathcal{F}_u)\left(\varepsilon_{v-1} + \rho\varepsilon_{v-2} + \ldots + \rho^{v-u-1}\varepsilon_u\right)\right] \\
&\quad +E\left[g_1(\mathcal{F}_u)\left(\rho^{v-u-1}\varepsilon_u + \Gamma_{u+1}^{v-1}\right)^2\right] \\
&= \rho^{2(v-u)}E\left[X_{u-1}^2 g_1(\mathcal{F}_u)\right] + 2\rho^{v-u}\rho^{v-u-1}E\left[X_{u-1}g_1(\mathcal{F}_u)\varepsilon_u\right] \\
&\quad +E\left[g_1(\mathcal{F}_u)\left(\rho^{2(v-u-1)}\varepsilon_u^2 + 2\rho^{v-u-1}\varepsilon_u\Gamma_{u+1}^{v-1} + \left(\Gamma_{u+1}^{v-1}\right)^2\right)\right] \\
&= \rho^{2(v-u)}E\left[X_{u-1}^2 g_1(\mathcal{F}_u)\right] + 2\rho^{v-u}\rho^{v-u-1}E\left[X_{u-1}g_1(\mathcal{F}_u)\varepsilon_u\right] \\
&\quad +\rho^{2(v-u-1)}E\left[g_1(\mathcal{F}_u)\varepsilon_u^2\right] + E\left[g_1(\mathcal{F}_u)\right]E\left[\left(\Gamma_{u+1}^{v-1}\right)^2\right]. \quad \square
\end{aligned}$$

2.4.3 Asymptotic Normality of the Gradient

Now we are ready to prove that the gradient is asymptotically normal. This result is presented in the following theorem.

Theorem 2.4.2 *Under assumptions M1, M2, E1, E2, W1, and W2 we have the following:*

$$S_n(\rho_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\eta_b^2 \sigma_X^2}{3}\right).$$

proof. First, rewrite the gradient as a sum of two components as follows.

$$\begin{aligned}
S_n(\rho_0) &= \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} \operatorname{sgn}\left((X_i - \rho_0 X_{i-1}) - (X_j - \rho_0 X_{j-1})\right) X_{i-1} \\
&= \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} \operatorname{sgn}(\varepsilon_i - \varepsilon_j) X_{i-1} \\
&= \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} \left(2I(\varepsilon_j \le \varepsilon_i) - 1\right) X_{i-1} \quad a.e. \\
&= \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} X_{i-1} \left(2I(\varepsilon_j \le \varepsilon_i) - 2F(\varepsilon_i) + 2F(\varepsilon_i) - 1\right) \\
&= \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} X_{i-1} \left(2F(\varepsilon_i) - 1 + 2I(\varepsilon_j \le \varepsilon_i) - 2F(\varepsilon_i)\right) \\
&= \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} X_{i-1} \left(2F(\varepsilon_i) - 1\right) + \\
&\quad \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} X_{i-1} \left(2I(\varepsilon_j \le \varepsilon_i) - 2F(\varepsilon_i)\right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{i-1} \left(2F(\varepsilon_i) - 1\right) \left(\frac{1}{n} \sum_{j=1}^{n} b_{ij}\right) + \\
&\quad \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} X_{i-1} \left(I(\varepsilon_j \le \varepsilon_i) - F(\varepsilon_i)\right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_{i.} X_{i-1} \left(2F(\varepsilon_i) - 1\right) + \\
&\quad \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} X_{i-1} \left(I(\varepsilon_j \le \varepsilon_i) - F(\varepsilon_i)\right) \\
&= S_{n1}(\rho_0) + S_{n2}(\rho_0),
\end{aligned}$$

where $b_{i.} = \frac{1}{n} \sum_{j=1}^{n} b_{ij}$. In what follows $S_{n1}(\rho_0)$ will be shown to be asymptotically normal using Theorem 2.4.1 and $S_{n2}(\rho_0)$ will be shown to be $o_p(1)$. Thus,

the asymptotic normality of $S_n(\rho_0)$ will follow from an application of Slutsky's Theorem (Serfling, 1980, pg. 19). We will start by showing that $S_{n1}(\rho_0)$ is asymptotically normal. For convenience, let $u_i = 2F(\varepsilon_i) - 1$. It should be noted that u_i is uniformly distributed over $(-1, 1)$ so that $E[u_i] = 0$ and $V[u_i] = \frac{1}{3}$. Now let

$$Z_{n,t} = \frac{1}{\sqrt{n}} b_{t.} X_{t-1} u_t.$$

Then, $Z_{n,t}$ is $\mathcal{F}_{n,t}$ measurable since $\mathcal{F}_{n,t} = \sigma\text{-field}\{X_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_t\}$ and $b_{t.}$ is non-random. It should be noted that assumption W1 implies that $|b_{t.}| \le B_b$ and assumption M1 implies the process is stationary. These two properties and the fact that X_{t-1} is independent of u_t yield the following inequality,

$$\begin{aligned} E\left[Z_{n,t}^2\right] &= E\left[\frac{1}{n} b_{t.}^2 X_{t-1}^2 u_t^2\right] \\ &= \frac{1}{n} b_{t.}^2 E\left[X_{t-1}^2\right] E\left[u_t^2\right] \\ &= \frac{b_{t.}^2 \sigma_X^2}{3n} \\ &\le \frac{B_b^2 \sigma_X^2}{3} \\ &< \infty. \end{aligned}$$

Next, it follows from the independence of u_t and $\mathcal{F}_{n,t-1}$ and the fact that $E[u_t] = 0$ that,

$$\begin{aligned} E\left[Z_{n,t} \mid \mathcal{F}_{n,t-1}\right] &= E\left[\frac{1}{\sqrt{n}} b_{t.} X_{t-1} u_t \mid \mathcal{F}_{n,t-1}\right] \\ &= \frac{1}{\sqrt{n}} b_{t.} X_{t-1} E\left[u_t \mid \mathcal{F}_{n,t-1}\right] \\ &= 0. \end{aligned}$$

Now let,

$$
\begin{aligned}
S_{n,j} &= \frac{1}{\sqrt{n}}\sum_{t=1}^{j} b_{t.}X_{t-1}u_t \\
&= \sum_{t=1}^{j} Z_{n,t}.
\end{aligned}
$$

Since the three conditions in Definition 2.4.1 are satisfied, $\{S_{n,j}, \mathcal{F}_{n,j}\}$ is a zero-mean square-integrable martingale array with differences $Z_{n,j}$. Thus, we only need to show that the four conditions of Theorem 2.4.1 are satisfied to prove that $S_{n1}(\rho_0)$ is asymptotically normal. Consider the first condition of the theorem. Using W1 to bound $b_{i.}$ and Lemma 2.3.3 we get the following,

$$
\begin{aligned}
\max_{1\le i\le n}|Z_{n,i}| &= \max_{1\le i\le n}\left|\frac{1}{\sqrt{n}}b_{i.}X_{i-1}u_i\right| \\
&\le B_b\left(\frac{1}{\sqrt{n}}\max_{1\le i\le n}|X_{i-1}|\right) \\
&= o_p(1).
\end{aligned}
$$

Thus, the first condition of the theorem is satisfied. Since the second condition is the most difficult one to show, we will save it for last and proceed with the third condition. Again, using W1 to bound $b_{i.}$ and the independence of X_{i-1} and u_i we get the following,

$$
\begin{aligned}
E\left[\max_{1\le i\le n} Z_{n,i}^2\right] &\le E\left[\sum_{i=1}^{n} Z_{n,i}^2\right] \\
&= \sum_{i=1}^{n} E\left[Z_{n,i}^2\right] \\
&= \sum_{i=1}^{n} E\left[\frac{1}{n}b_{i.}^2 X_{i-1}^2 u_i^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} b_{i.}^2 E\left[X_{i-1}^2\right]E\left[u_i^2\right]
\end{aligned}
$$

$$\leq \frac{B_b^2 \sigma_X^2}{3}.$$

Thus, taking the limit as $n \longrightarrow \infty$ in the above inequality implies that,

$$E\left[\max_{1 \leq i \leq n} Z_{n,i}^2\right] = O(1).$$

Hence, the third condition is satisfied. For verification of condition four one is referred to the "Remarks" paragraph on page 59 of Hall and Heyde (1980). To complete the proof we need to verify that the second condition is satisfied. Specifically, we need to show the following,

$$\frac{1}{n}\sum_{i=1}^{n} b_{i.}^2 X_{i-1}^2 u_i^2 \xrightarrow{p} \frac{\eta_b^2 \sigma_X^2}{3}.$$

To simplify notation, let $w_i = b_{i.}^2$. Now,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} w_i X_{i-1}^2 u_i^2 - \frac{\eta_b^2 \sigma_X^2}{3} \\
&= \frac{1}{n}\sum_{i=1}^{n} w_i X_{i-1}^2 u_i^2 - \frac{1}{n}\sum_{i=1}^{n} w_i \frac{\sigma_X^2}{3} + \frac{1}{n}\sum_{i=1}^{n} w_i \frac{\sigma_X^2}{3} - \frac{\eta_b^2 \sigma_X^2}{3} \\
&= \frac{1}{n}\sum_{i=1}^{n} w_i \left(X_{i-1}^2 u_i^2 - \frac{\sigma_X^2}{3}\right) + \left(\frac{1}{n}\sum_{i=1}^{n} w_i - \eta_b^2\right)\frac{\sigma_X^2}{3} \\
&= T_{n1} + T_{n2} \text{ say.}
\end{aligned}$$

Assumption W2 implies that $T_{n2} = o(1)$. Hence, we only need to show $T_{n1} = o_p(1)$ in order to verify the condition. Using Chebyshev's Inequality, it suffices to show $E[T_{n1}] = o(1)$ and $V[T_{n1}] = o(1)$. Consider the expectation first. The independence of X_{i-1} and u_i imply the following,

$$E[T_{n1}] = E\left[\frac{1}{n}\sum_{i=1}^{n} w_i \left(X_{i-1}^2 u_i^2 - \frac{\sigma_X^2}{3}\right)\right]$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n} w_i E\left[X_{i-1}^2 u_i^2 - \frac{\sigma_X^2}{3}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} w_i \left(E\left[X_{i-1}^2\right] E\left[u_i^2\right] - \frac{\sigma_X^2}{3}\right) \\
&= 0.
\end{aligned}
$$

Thus, the fact that $E[T_{n1}] = o(1)$ is trivial. Next, consider the variance of T_{n1},

$$
\begin{aligned}
&V[T_{n1}] \\
&= V\left[\frac{1}{n}\sum_{i=1}^{n} w_i \left(X_{i-1}^2 u_i^2 - \frac{\sigma_X^2}{3}\right)\right] \\
&= \frac{1}{n^2} V\left[\sum_{i=1}^{n} w_i \left(X_{i-1}^2 u_i^2 - \frac{\sigma_X^2}{3}\right)\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} w_i^2 V\left[X_{i-1}^2 u_i^2\right] + \frac{2}{n^2}\sum_{i<j} w_i w_j COV\left[X_{i-1}^2 u_i^2, X_{j-1}^2 u_j^2\right] \\
&= T_{n11} + T_{n12} \text{ say.}
\end{aligned}
$$

Consider T_{n11} first. Since the sum in T_{n11} is $O(n)$, $w_i^2 \leq B_b^2$, and we are dividing by n^2 it suffices to show that $V\left[X_{i-1}^2 u_i^2\right] \leq K < \infty$ for some K and all i. This follows from the independence of X_{i-1} and u_i, stationarity, and part 3 of Lemma 2.3.2,

$$
\begin{aligned}
V\left[X_{i-1}^2 u_i^2\right] &\leq E\left[X_{i-1}^4 u_i^4\right] \\
&= E\left[X_{i-1}^4\right] E\left[u_i^4\right] \\
&= E\left[X_1^4\right] E\left[u_1^4\right] \\
&\leq E\left[X_1^4\right] \\
&< \infty.
\end{aligned}
$$

Now consider the covariance term in T_{n12}. Since $i < j$ implies u_j is independent of (X_{i-1}, u_i, X_{j-1}) we have the following,

$$\begin{aligned}
&COV\left[X_{i-1}^2u_i^2, X_{j-1}^2u_j^2\right] \\
&= E\left[X_{i-1}^2u_i^2X_{j-1}^2u_j^2\right] - E\left[X_{i-1}^2u_i^2\right]E\left[X_{j-1}^2u_j^2\right] \\
&= E\left[X_{i-1}^2u_i^2X_{j-1}^2\right]E\left[u_j^2\right] - E\left[X_{i-1}^2u_i^2\right]E\left[X_{j-1}^2\right]E\left[u_j^2\right] \\
&= \frac{1}{3}E\left[X_{i-1}^2u_i^2X_{j-1}^2 - E\left[X_{i-1}^2u_i^2\right]X_{j-1}^2\right] \\
&= \frac{1}{3}E\left[\left(X_{i-1}^2u_i^2 - E\left[X_{i-1}^2u_i^2\right]\right)X_{j-1}^2\right] \\
&= \frac{1}{3}E\left[A_iX_{j-1}^2\right] \text{ say.}
\end{aligned}$$

Now, using part 1 of Lemma 2.3.2 to write X_{j-1} in terms of X_{i-1} and Γ_i^{j-1} we get,

$$\begin{aligned}
&COV\left[X_{i-1}^2u_i^2, X_{j-1}^2u_j^2\right] \\
&= \frac{1}{3}E\left[A_i\left({\rho_0}^{j-i}X_{i-1} + \Gamma_i^{j-1}\right)^2\right] \\
&= \frac{1}{3}E\left[A_i\left({\rho_0}^{2(j-i)}X_{i-1}^2 + {\rho_0}^{j-i}X_{i-1}\Gamma_i^{j-1} + \left(\Gamma_i^{j-1}\right)^2\right)\right] \\
&= \frac{1}{3}{\rho_0}^{j-i}E\left[{\rho_0}^{j-i}A_iX_{i-1}^2 + 2A_iX_{i-1}\Gamma_i^{j-1}\right] + \frac{1}{3}E\left[A_i\left(\Gamma_i^{j-1}\right)^2\right] \\
&= C_1 + C_2 \text{ say.}
\end{aligned}$$

Using the fact that $|\rho_0| < 1$, Lemma 2.3.1, Lemma 2.3.2, and the Cauchy Schwarz Inequality it can be shown that the expectation in C_1 can be bounded by a constant K_1 that does not depend on i or j. Thus, $|C_1| \leq |\rho_0|^{j-i}K_1$. Now consider

C_2,

$$
\begin{aligned}
C_2 &= \frac{1}{3}E\left[A_i\left(\Gamma_i^{j-1}\right)^2\right] \\
&= \frac{1}{3}E\left[A_i\left({\rho_0}^{j-i-1}\varepsilon_i + \Gamma_{i+1}^{j-1}\right)^2\right] \\
&= \frac{1}{3}E\left[A_i\left({\rho_0}^{2(j-i-1)}\varepsilon_i^2 + 2{\rho_0}^{j-i-1}\varepsilon_i\Gamma_{i+1}^{j-1} + \left(\Gamma_{i+1}^{j-1}\right)^2\right)\right] \\
&= \frac{1}{3}{\rho_0}^{2(j-i-1)}E\left[A_i\varepsilon_i^2\right] + \frac{2}{3}{\rho_0}^{j-i-1}E\left[A_i\varepsilon_i\Gamma_{i+1}^{j-1}\right] + \frac{1}{3}E\left[A_i\left(\Gamma_{i+1}^{j-1}\right)^2\right] \\
&= C_{21} + C_{22} + C_{23} \text{ say.}
\end{aligned}
$$

Since (A_i, ε_i) is independent of Γ_{i+1}^{j-1}, $E[A_i] = 0$, and $E\left[\Gamma_{i+1}^{j-1}\right] = 0$, it follows that $C_{22} = C_{23} = 0$. Thus, $C_2 = C_{21} = \frac{1}{3}{\rho_0}^{2(j-i-1)}E\left[A_i\varepsilon_i^2\right]$. Again, the Cauchy Schwarz Inequality, Lemma 2.3.2, M1, and E1 imply that there exists a constant K_2 such that $|C_{21}| \leq |\rho_0|^{j-i-1}K_2$. Putting the pieces together we have,

$$
\begin{aligned}
\left|COV\left[X_{i-1}^2u_i^2, X_{j-1}^2u_j^2\right]\right| &= |C_1 + C_2| \\
&\leq |C_1| + |C_2| \\
&\leq |\rho_0|^{j-i}K_1 + |\rho_0|^{j-i-1}K_2 \\
&= |\rho_0|^{j-i-1}|\rho_0|K_1 + |\rho_0|^{j-i-1}K_2 \\
&\leq |\rho_0|^{j-i-1}K \text{ say.}
\end{aligned}
$$

Consider T_{n12} once again. Using W1 to bound w_i, the above inequality, and the formula for the sum of a convergent geometric series we have the following,

$$
\begin{aligned}
|T_{n12}| &= \left|\frac{2}{n^2}\sum_{i<j} w_iw_jCOV\left[X_{i-1}^2u_i^2, X_{j-1}^2u_j^2\right]\right| \\
&\leq \frac{2}{n^2}\sum_{i<j}|w_iw_j|\left|COV\left[X_{i-1}^2u_i^2, X_{j-1}^2u_j^2\right]\right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{2B_b^2}{n^2}\sum_{i<j}|\rho_0|^{j-i-1}K \\
&= \frac{2B_b^2K}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}|\rho_0|^{j-i-1} \\
&\leq \frac{2B_b^2K}{n^2}\sum_{i=1}^{n-1}\sum_{j=0}^{\infty}|\rho_0|^{j} \\
&= \frac{2B_b^2K}{1-|\rho_0|}\left(\frac{n-1}{n^2}\right) \\
&= o(1).
\end{aligned}
$$

Therefore, since T_{n11} and T_{n12} are $o(1)$, T_{n1} is $o_p(1)$. That is,

$$
\frac{1}{n}\sum_{i=1}^{n} b_{i.}^2 X_{i-1}^2 u_i^2 \xrightarrow{p} \frac{\eta_b^2\sigma_X^2}{3}.
$$

Hence, the second condition of Theorem 2.4.1 is satisfied. Since all four conditions of Theorem 2.4.1 have been verified, we have proven that,

$$
S_{n1}(\rho_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\eta_b^2\sigma_X^2}{3}\right).
$$

To complete the proof we must show that $S_{n2}(\rho_0) = o_p(1)$. To proceed, we will write $S_{n2}(\rho_0)$ into three components and then show that each component is $o_p(1)$,

$$
\begin{aligned}
S_{n2}(\rho_0) &= \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}X_{i-1}\left(I(\varepsilon_j \leq \varepsilon_i) - F(\varepsilon_i)\right) \\
&= \frac{2}{n^{\frac{3}{2}}}\sum_{i<j} b_{ij}X_{i-1}\left(I(\varepsilon_j \leq \varepsilon_i) - F(\varepsilon_i)\right) + \\
&\quad \frac{2}{n^{\frac{3}{2}}}\sum_{i=j} b_{ij}X_{i-1}\left(I(\varepsilon_j \leq \varepsilon_i) - F(\varepsilon_i)\right) + \\
&\quad \frac{2}{n^{\frac{3}{2}}}\sum_{i>j} b_{ij}X_{i-1}\left(I(\varepsilon_j \leq \varepsilon_i) - F(\varepsilon_i)\right) \\
&= S_{n21} + S_{n22} + S_{n23} \text{ say.}
\end{aligned}
$$

Consider S_{n21} first. To show that $S_{n21} = o_p(1)$ using Chebyshev's Inequality it suffices to show that $E[S_{n21}] = o(1)$ and $V[S_{n21}] = o(1)$. Consider the expectation first. The following calculations show that upon conditioning on $\mathcal{F}_{j-1}$ the expectation is zero, and thus $o(1)$,

$$\begin{aligned}
E[S_{n21}] &= \frac{2}{n^{\frac{3}{2}}}\sum_{i<j} b_{ij} E\left[X_{i-1}\left(I\left(\varepsilon_j \le \varepsilon_i\right) - F(\varepsilon_i)\right)\right] \\
&= \frac{2}{n^{\frac{3}{2}}}\sum_{i<j} b_{ij} E\left[X_{i-1} E\left[I\left(\varepsilon_j \le \varepsilon_i\right) - F(\varepsilon_i) \mid \mathcal{F}_{j-1}\right]\right] \\
&= \frac{2}{n^{\frac{3}{2}}}\sum_{i<j} b_{ij} E\left[X_{i-1}\left(F(\varepsilon_i) - F(\varepsilon_i)\right)\right] \\
&= 0.
\end{aligned}$$

Next, consider the variance term. If we let $\theta_{ij} = b_{ij}X_{i-1}\left(I\left(\varepsilon_j \le \varepsilon_i\right) - F(\varepsilon_i)\right)$ then we have the following,

$$\begin{aligned}
V[S_{n21}] &= E\left[S_{n21}^2\right] \\
&= E\left[\frac{2}{n^{\frac{3}{2}}}\sum_{i<j}\theta_{ij}\frac{2}{n^{\frac{3}{2}}}\sum_{k<l}\theta_{kl}\right] \\
&= \frac{4}{n^3}E\left[\sum_{i<j}\theta_{ij}\sum_{k<l}\theta_{kl}\right] \\
&= \frac{4}{n^3}E\left[\sum_{i<j}\theta_{ij}^2 + \sum_{i<j}\sum_{k<l}\theta_{ij}\theta_{kl}\right] \\
&= \frac{4}{n^3}\sum_{i<j}E\left[\theta_{ij}^2\right] + \frac{4}{n^3}\sum_{i<j}\sum_{k<l}E\left[\theta_{ij}\theta_{kl}\right] \\
&= S_{n211} + S_{n212} \text{ say.}
\end{aligned}$$

Since the sum in $S_{n211} = O(n^2)$ and we are dividing by n^3 we only need to show that $E\left[\theta_{ij}^2\right]$ is uniformly bounded in order to show that $S_{n211} = o(1)$. Using W1 to bound the weights, the bound on F, and the stationarity assumption it is easy

to verify that $E\left[\theta_{ij}^2\right] \leq 4B_b^2\sigma_X^2 < \infty$. Thus, $S_{n211} = o(1)$. Now consider S_{n212}. Because the sum in this term is $O(n^4)$ it will be convenient to break the sum into the possible cases given in Table 1. We note that under case 7, S_{n212} reduces to S_{n211} which has already been shown to be $o(1)$. Consider cases 1, 2, 3, 5, 6, 8, 9, 11, 12, and 13. In each of these cases, either j or l is the largest subscript and $j \neq l$. Thus, upon conditioning on $\mathcal{F}_{j-1}$ or $\mathcal{F}_{l-1}$ it is immediate that $E\left[\theta_{ij}\theta_{kl}\right] = 0$. Because all of the above cases are treated the same, only the details for case 1 $(i < j < k < l)$ are shown,

$$\begin{aligned} E\left[\theta_{ij}\theta_{kl}\right] &= E\left[\theta_{ij}E\left[\theta_{kl} \mid \mathcal{F}_{l-1}\right]\right] \\ &= b_{kl}E\left[\theta_{ij}X_{k-1}E\left[I\left(\varepsilon_l \leq \varepsilon_k\right) - F(\varepsilon_k) \mid \mathcal{F}_{l-1}\right]\right] \\ &= b_{kl}E\left[\theta_{ij}X_{k-1}\left(F(\varepsilon_k) - F(\varepsilon_k)\right)\right] \\ &= 0. \end{aligned}$$

Hence, all cases, with the exception of case 4 and case 10 are $o(1)$. Under cases 4 and 10, $j = l$ so the sum in S_{n212} is $O(n^3)$. Since we are dividing by n^3 we will exploit the iterative nature of the process to reduce the sum in S_{n212} to $O(n^2)$. The details will be presented for case 4 only since the details for case 10 are exactly the same except that the roles of k and i are interchanged. Consider the expectation in the sum of S_{n212} under case 4 $(i < k < j = l)$,

$$\begin{aligned} &E\left[\theta_{ij}\theta_{kj}\right] \\ &\quad = b_{ij}b_{kj}E\left[X_{i-1}X_{k-1}(I\left(\varepsilon_j \leq \varepsilon_i\right) - F(\varepsilon_i))(I\left(\varepsilon_j \leq \varepsilon_k\right) - F(\varepsilon_k))\right] \end{aligned}$$

$$
\begin{aligned}
&= b_{ij}b_{kj}E\left[X_{i-1}X_{k-1}E\left[(I(\varepsilon_j \le \varepsilon_i) - F(\varepsilon_i))(I(\varepsilon_j \le \varepsilon_k) - F(\varepsilon_k)) \mid \mathcal{F}_{j-1}\right]\right] \\
&= b_{ij}b_{kj}E\left[X_{i-1}X_{k-1}\left(F(\min(\varepsilon_i, \varepsilon_k)) - F(\varepsilon_k)F(\varepsilon_i)\right)\right] \\
&= b_{ij}b_{kj}E\left[X_{i-1}X_{k-1}\left(\min(F(\varepsilon_i), F(\varepsilon_k)) - F(\varepsilon_i)F(\varepsilon_k)\right)\right] \\
&= b_{ij}b_{kj}E\left[X_{i-1}X_{k-1}E\left[\min(F(\varepsilon_i), F(\varepsilon_k)) - F(\varepsilon_i)F(\varepsilon_k) \mid \mathcal{F}_{k-1}\right]\right] \\
&= b_{ij}b_{kj}E\left[X_{i-1}X_{k-1}\left(F(\varepsilon_i) - \frac{1}{2}F^2(\varepsilon_i) - \frac{1}{2}F(\varepsilon_i)\right)\right] \\
&= \frac{1}{2}b_{ij}b_{kj}E\left[X_{i-1}\left(F(\varepsilon_i) - F^2(\varepsilon_i)\right)X_{k-1}\right].
\end{aligned}
$$

Now, applying Lemma 2.4.1 with $g(\mathcal{F}_i) = X_{i-1}(F(\varepsilon_i) - F^2(\varepsilon_i))$ and using the independence of X_{i-1} and ε_i in the last expectation of that Lemma yields the following,

$$
E\left[\theta_{ij}\theta_{kj}\right] = \frac{1}{2}b_{ij}b_{kj}\rho_0{}^{k-i-1}\left(\rho_0 E\left[X_{i-1}^2\left(F(\varepsilon_i) - F^2(\varepsilon_i)\right)\right]\right).
$$

Now use the bound on the weights, the fact that $|\rho_0| < 1$, the bound on F, and stationarity to show the following,

$$
\begin{aligned}
\left|E\left[\theta_{ij}\theta_{kj}\right]\right| &\le |\rho_0|^{k-i-1}\frac{1}{2}B_b^2\sigma_x^2 \\
&= |\rho_0|^{k-i-1}K \text{ say.}
\end{aligned}
$$

Now use the above inequality and the formula for the sum of a convergent geometric series to show the following,

$$
\begin{aligned}
\left|\frac{4}{n^3}\sum_{i<k<j}E\left[\theta_{ij}\theta_{kj}\right]\right| &\le \frac{4}{n^3}\sum_{i<k<j}|\rho_0|^{k-i-1}K \\
&= \frac{4K}{n^3}\sum_{i=1}^{n-2}\sum_{k=i+1}^{n-1}\left(|\rho_0|^{k-i-1}\sum_{k+1}^{n}1\right)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{4Kn}{n^3}\sum_{i=1}^{n-2}\sum_{k=0}^{\infty}|\rho_0|^k \\
&\leq \frac{4K}{1-|\rho_0|}\frac{n^2}{n^3} \\
&= o(1).
\end{aligned}
$$

Thus, the proof of $S_{n212} = o(1)$ is complete. Since S_{n211} and S_{212} are $o(1)$ we have shown that $S_{n21} = o_p(1)$.

Next, we need to show $S_{n22} = o_p(1)$. This follows from an application of the Ergodic Theorem as follows,

$$
\begin{aligned}
|S_{n22}| &= \left|\frac{2}{n^{\frac{3}{2}}}\sum_{i=j} b_{ij}X_{i-1}(I(\varepsilon_j \leq \varepsilon_i) - F(\varepsilon_i))\right| \\
&\leq \frac{2B_b}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}|X_{i-1}|\,|1-F(\varepsilon_i)| \\
&= o(1)O_p(1) \\
&= o_p(1).
\end{aligned}
$$

In order to complete the proof of $S_{n2}(\rho_0) = o_p(1)$ we need to show $S_{n23} = o_p(1)$. We will use Chebyshev's Inequality and show that $E\left[S_{n23}^2\right] = o(1)$. Let θ_{ij} denote the random variable defined in S_{n21}. Then,

$$
\begin{aligned}
E\left[S_{n23}^2\right] &= E\left[\frac{2}{n^{\frac{3}{2}}}\sum_{i>j}\theta_{ij}\frac{2}{n^{\frac{3}{2}}}\sum_{k>l}\theta_{kl}\right] \\
&= \frac{4}{n^3}E\left[\sum_{i>j}\theta_{ij}\sum_{k>l}\theta_{kl}\right] \\
&= \frac{4}{n^3}E\left[\sum_{i>j}\theta_{ij}^2 + \sum_{i>j}\sum_{k>l}\theta_{ij}\theta_{kl}\right] \\
&= \frac{4}{n^3}\sum_{i>j}E\left[\theta_{ij}^2\right] + \frac{4}{n^3}\sum_{i>j}\sum_{k>l}E\left[\theta_{ij}\theta_{kl}\right]
\end{aligned}
$$

$$= \quad S_{n231} + S_{n232} \text{ say.}$$

The proof of $S_{n231} = o(1)$ is identical to the proof of $S_{n211} = o(1)$. Thus, we only need to show that $S_{n232} = o(1)$. Re-labeling the subscripts yields,

$$S_{n232} = \frac{4}{n^3} \sum_{i<j} \sum_{k<l} E\left[\theta_{ji}\theta_{lk}\right].$$

The techniques used to handle this term are similar to those of S_{n212}. We will break the sum in S_{n232} into the possible cases given in Table 1 and show that each case is $o(1)$. Again, case 7 reduces to S_{n231} which has already been shown to be $o(1)$. Now, it is easy to see that each of the cases 1 through 6 have a symmetric counter part in cases 8 through 13. Thus, we only need to consider cases 1 through 6. Table 1 implies that cases 2, 4, and 6 are $O(n^3)$ while cases 1, 3, and 5 are $O(n^4)$. Since the evaluation of each of the six cases would be too time consuming and tedious only two cases will be considered. However, the two cases considered seem to be the most detailed. The remaining cases, although somewhat less detailed, are treated using similar techniques. We will begin with case 4 ($i < j < k = l$) which is $O(n^3)$. The proof is similar to that used in case 4 of S_{n212}. Consider the expectation in the sum under case 4,

$$\begin{aligned}
E\left[\theta_{ji}\theta_{jk}\right] &= b_{ji}b_{jk}E\left[X_{j-1}X_{j-1}(I\left(\varepsilon_i \le \varepsilon_j\right) - F(\varepsilon_j))(I\left(\varepsilon_k \le \varepsilon_j\right) - F(\varepsilon_j))\right] \\
&= b_{ji}b_{jk}E\left[X_{j-1}^2 E\left[(I\left(\varepsilon_i \le \varepsilon_j\right) - F(\varepsilon_j))(I\left(\varepsilon_k \le \varepsilon_j\right) - F(\varepsilon_j)) \mid \mathcal{F}_{j-1}\right]\right] \\
&= b_{ji}b_{jk}E\left[X_{j-1}^2\left(\frac{1}{3} - F(\max(\varepsilon_i, \varepsilon_k)) + \frac{1}{2}F^2(\varepsilon_i) + \frac{1}{2}F^2(\varepsilon_k)\right)\right] \\
&= b_{ji}b_{jk}E\left[\left(\frac{1}{3} - \max(F(\varepsilon_i), F(\varepsilon_k)) + \frac{1}{2}F^2(\varepsilon_i) + \frac{1}{2}F^2(\varepsilon_k)\right) X_{j-1}^2\right]
\end{aligned}$$

$$= b_{ji}b_{jk}E\left[g\left(\mathcal{F}_k\right)X_{j-1}^2\right] \text{ say.}$$

Applying Lemma 2.4.2 with

$$g(\mathcal{F}_k) = \left(\frac{1}{3} - \max(F(\varepsilon_i), F(\varepsilon_k)) + \frac{1}{2}F^2(\varepsilon_i) + \frac{1}{2}F^2(\varepsilon_k)\right),$$

and using the fact that $E\left[g(\mathcal{F}_k)\right] = 0$ in the last expectation of that Lemma yields the following,

$$\begin{aligned} E\left[\theta_{ji}\theta_{jk}\right] &= b_{ji}b_{jk}{\rho_0}^{j-k}\left({\rho_0}^{j-k}E\left[X_{k-1}^2 g(\mathcal{F}_k)\right] + \right. \\ &\quad \left. 2{\rho_0}^{j-k-1}E\left[X_{k-1}g(\mathcal{F}_k)\varepsilon_k\right]{\rho_0}^{j-k-2}E\left[g(\mathcal{F}_k)\varepsilon_k^2\right]\right). \end{aligned}$$

Now use the bound on the weights, the fact that $|\rho_0| < 1$, the bound on $g(\mathcal{F}_k)$, and stationarity to show the following,

$$\left|E\left[\theta_{ji}\theta_{jk}\right]\right| \leq |\rho_0|^{j-k}K,$$

where K is a constant that depends on B_b, σ_X^2, and $E\left[\varepsilon_1^2\right]$. Now use the $|\rho_0|^{j-k}$ piece to reduce the sum to $O(n^2)$ as was done in the proof of case 4 of S_{n212}. Since we are dividing by n^3 we have that $S_{n232} = o(1)$ under case 4.

Now consider the expectation under case 5 ($i < k < l < j$),

$$\begin{aligned} E\left[\theta_{ji}\theta_{lk}\right] &= b_{ji}b_{lk}E\left[X_{j-1}X_{l-1}(I\left(\varepsilon_i \leq \varepsilon_j\right) - F(\varepsilon_j))(I\left(\varepsilon_k \leq \varepsilon_l\right) - F(\varepsilon_l))\right] \\ &= b_{ji}b_{lk}E\left[X_{j-1}X_{l-1}(I\left(\varepsilon_k \leq \varepsilon_l\right) - F(\varepsilon_l))E\left[(I\left(\varepsilon_i \leq \varepsilon_j\right) - F(\varepsilon_j)) \mid \mathcal{F}_{j-1}\right]\right] \\ &= b_{ji}b_{lk}E\left[X_{j-1}X_{l-1}(I\left(\varepsilon_k \leq \varepsilon_l\right) - F(\varepsilon_l))\left(\frac{1}{2} - F(\varepsilon_i)\right)\right] \\ &= b_{ji}b_{lk}E\left[\left(\frac{1}{2} - F(\varepsilon_i)\right)X_{l-1}(I\left(\varepsilon_k \leq \varepsilon_l\right) - F(\varepsilon_l))X_{j-1}\right] \\ &= b_{ji}b_{lk}E\left[g\left(\mathcal{F}_l\right)X_{j-1}\right] \text{ say.} \end{aligned}$$

Applying Lemma 2.4.1 to the last expectation above immediately yields the following,

$$E\left[\theta_{ji}\theta_{lk}\right] = b_{ji}b_{lk}{\rho_0}^{j-l-1}\left(\rho_0 E\left[X_{l-1}g\left(\mathcal{F}_l\right)\right] + E\left[g\left(\mathcal{F}_l\right)\varepsilon_l\right]\right).$$

The ${\rho_0}^{j-l-1}$ piece will reduce the sum from $O(n^4)$ to $O(n^3)$. However, this is not enough since we are only dividing by n^3. Thus, we need to extract a factor of ${\rho_0}^{l-k-1}$ from the remaining two expectations above. Consider the first of these two expectations,

$$\begin{aligned} E\left[X_{l-1}g\left(\mathcal{F}_l\right)\right] &= E\left[\left(\frac{1}{2} - F(\varepsilon_i)\right) X_{l-1}^2 (I\left(\varepsilon_k \leq \varepsilon_l\right) - F(\varepsilon_l))\right] \\ &= E\left[\left(\frac{1}{2} - F(\varepsilon_i)\right) X_{l-1}^2 E\left[\left(I\left(\varepsilon_k \leq \varepsilon_l\right) - F(\varepsilon_l)\right) \mid \mathcal{F}_{l-1}\right]\right] \\ &= E\left[\left(\frac{1}{2} - F(\varepsilon_i)\right)\left(\frac{1}{2} - F(\varepsilon_k)\right) X_{l-1}^2\right] \\ &= E\left[g\left(\mathcal{F}_k\right) X_{l-1}^2\right] \text{ say.} \end{aligned}$$

Now apply Lemma 2.4.2 with $g(\mathcal{F}_k) = \left(\frac{1}{2} - F(\varepsilon_i)\right)\left(\frac{1}{2} - F(\varepsilon_k)\right)$. Using the independence of (ε_i, X_{k-1}) and ε_k, and the fact that $E\left[g(\mathcal{F}_k)\right] = 0$ the third and fourth expectations in that Lemma are 0. Thus,

$$\left|E\left[X_{l-1}g\left(\mathcal{F}_l\right)\right]\right| \leq \left|\rho_0\right|^{l-k} K,$$

where K is a constant that depends on σ_X^2, $E\left[\varepsilon_1^2\right]$, and a bound on F. Now let

$$g^*(\varepsilon_k) = E\left[\left(I\left(\varepsilon_k \leq \varepsilon_l\right) - F(\varepsilon_l)\right)\varepsilon_l \mid \mathcal{F}_{l-1}\right],$$

and note that $g^*(\varepsilon_k)$ is bounded by $2E[|\varepsilon_1|]$. Then the second expectation is as follows,

$$\begin{aligned} E[g(\mathcal{F}_l)\varepsilon_l] &= E\left[\left(\frac{1}{2} - F(\varepsilon_i)\right) X_{l-1}(I(\varepsilon_k \le \varepsilon_l) - F(\varepsilon_l))\varepsilon_l\right] \\ &= E\left[\left(\frac{1}{2} - F(\varepsilon_i)\right) X_{l-1}E\left[(I(\varepsilon_k \le \varepsilon_l) - F(\varepsilon_l))\varepsilon_l \mid \mathcal{F}_{l-1}\right]\right] \\ &= E\left[\left(\frac{1}{2} - F(\varepsilon_i)\right) g^*(\varepsilon_k)X_{l-1}\right] \\ &= E[g(\mathcal{F}_k)X_{l-1}] \text{ say.} \end{aligned}$$

Now apply Lemma 2.4.1 with $g(\mathcal{F}_k) = \left(\frac{1}{2} - F(\varepsilon_i)\right) g^*(\varepsilon_k)$ to the last expectation and use the fact that $E\left[\left(\frac{1}{2} - F(\varepsilon_i)\right)\right] = 0$ along with independence to show,

$$E[g(\mathcal{F}_l)\varepsilon_l] = {\rho_0}^{l-k-1}\left(\rho_0 E\left[\left(\frac{1}{2} - F(\varepsilon_i)\right) X_{k-1}g^*(\varepsilon_k)\right]\right).$$

Using the bound on F, the bound on $g^*(\varepsilon_k)$, and the fact that $|\rho_0| < 1$ yields the following,

$$|E[g(\mathcal{F}_l)\varepsilon_l]| \le |\rho_0|^{l-k-1} K,$$

where K is a constant that depends on σ_X^2, $E[\varepsilon_1^2]$, and the bound on F. Thus, combining the two parts we get,

$$\begin{aligned} |E[\theta_{ji}\theta_{lk}]| &= \left|b_{ji}b_{lk}{\rho_0}^{j-l-1}(\rho_0 E[X_{l-1}g(\mathcal{F}_l)] + E[g(\mathcal{F}_l)\varepsilon_l])\right| \\ &\le B_b^2 |\rho_0|^{j-l-1}(|E[X_{l-1}g(\mathcal{F}_l)]| + |E[g(\mathcal{F}_l)\varepsilon_l]|) \\ &\le B_b^2 |\rho_0|^{j-l-1}\left(|\rho_0|^{l-k} K + |\rho_0|^{l-k-1} K\right) \\ &\le |\rho_0|^{j-l-1} |\rho_0|^{l-k-1} K \text{ say.} \end{aligned}$$

Now, putting everything together for S_{n232} (under case 5) we have the following,

$$\begin{aligned}
|S_{n232}| &\leq \frac{4}{n^3}\sum_{i<j}\sum_{k<l}|E[\theta_{ji}\theta_{lk}]| \\
&\leq \frac{4}{n^3}\sum_{i<j}\sum_{k<l}|\rho_0|^{j-l-1}|\rho_0|^{l-k-1}K \\
&= \frac{4K}{n^3}O(n^2) \\
&= o(1).
\end{aligned}$$

Hence, we have that $S_{n23} = o_p(1)$ under case 5. As mentioned earlier, the remaining cases can be treated in a similar fashion. Since we have shown that $S_{n21} = S_{n22} = S_{n23} = o_p(1)$, it is true that $S_{n2} = o_p(1)$. One will recall that this was the last piece we needed for the asymptotic normality of $S_n(\rho_0)$. Thus, we have shown that,

$$S_n(\rho_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\eta_b^2\sigma_X^2}{3}\right),$$

which completes the proof. □

2.5 Asymptotic Linearity of the Gradient

2.5.1 Preliminary Results

In this section we present results that will be used in the proof of the asymptotic linearity result, Theorem 2.5.1. Although the following is an abuse of notation, define $g_1 \equiv g_1(\mathcal{F}_{u-1})$ to be an arbitrary random variable that depends on a subset of $\{X_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{u-1}\}$. Furthermore, let $g_2 \equiv g_2(\varepsilon_u)$ represent a

random variable that depends only on ε_u. It follows from assumption E1 that g_1 and g_2 are independent. We will begin by presenting some lemmas pertaining to some expectations that we will encounter.

Lemma 2.5.1 *Assume assumptions M1, M2, and E1 hold. If $u < v$ then we have the following expectation.*

$$E\left[g_1(\mathcal{F}_{u-1})g_2(\varepsilon_u)\Gamma_u^v\right] = \rho^{v-u}E\left[g_1(\mathcal{F}_{u-1})\right]E\left[g_2(\varepsilon_u)\varepsilon_u\right]$$

proof. Condition on $\mathcal{F}_{u-1}$, apply the definition of Γ_u^v, and then part 1 of Lemma 2.3.1 to show the following,

$$\begin{aligned}
E\left[g_1(\mathcal{F}_{u-1})g_2(\varepsilon_u)\Gamma_u^v\right] &= E\left[g_1(\mathcal{F}_{u-1})E\left[g_2(\varepsilon_u)\Gamma_u^v \mid \mathcal{F}_{u-1}\right]\right] \\
&= E\left[g_1(\mathcal{F}_{u-1})\right]E\left[g_2(\varepsilon_u)\Gamma_u^v\right] \\
&= E\left[g_1(\mathcal{F}_{u-1})\right]E\left[g_2(\varepsilon_u)(\rho^{v-u}\varepsilon_u + \Gamma_{u+1}^v)\right] \\
&= E\left[g_1(\mathcal{F}_{u-1})\right]\left(\rho^{v-u}E\left[g_2(\varepsilon_u)\varepsilon_u\right] + E\left[g_2(\varepsilon_u)\right]E\left[\Gamma_{u+1}^v\right]\right) \\
&= \rho^{v-u}E\left[g_1(\mathcal{F}_{u-1})\right]E\left[g_2(\varepsilon_u)\varepsilon_u\right]. \ \square
\end{aligned}$$

Lemma 2.5.2 *Assume assumptions M1, M2, and E1 hold. If $u < v$ then we have the following expectation.*

$$\begin{aligned}
E\left[g_1(\mathcal{F}_{u-1})g_2(\varepsilon_u)(\Gamma_u^v)^2\right] &= \sigma_X^2 E\left[g_1(\mathcal{F}_{u-1})\right]E\left[g_2(\varepsilon_u)\right] + \\
&\rho^{2(v-u)}\left(E\left[g_1(\mathcal{F}_{u-1})\right]E\left[g_2(\varepsilon_u)\varepsilon_u^2\right] - \sigma_X^2 E\left[g_1(\mathcal{F}_{u-1})\right]E\left[g_2(\varepsilon_u)\right]\right)
\end{aligned}$$

proof.

$$E\left[g_1(\mathcal{F}_{u-1})g_2(\varepsilon_u)(\Gamma_u^v)^2\right]$$

$$
\begin{aligned}
&= E\left[g_1(\mathcal{F}_{u-1})E\left[g_2(\varepsilon_u)(\Gamma_u^v)^2 \mid \mathcal{F}_{u-1}\right]\right] \\
&= E\left[g_1(\mathcal{F}_{u-1})\right]E\left[g_2(\varepsilon_u)\left(\sum_{r=0}^{v-u}\rho^{2r}\varepsilon_{v-r}^2 + \sum_{r\neq q}\rho^r\rho^q\varepsilon_{v-r}\varepsilon_{v-q}\right)\right] \\
&= E\left[g_1(\mathcal{F}_{u-1})\right]E\left[\sum_{r=0}^{v-u}\rho^{2r}g_2(\varepsilon_u)\varepsilon_{v-r}^2 + \sum_{r\neq q}\rho^r\rho^q\varepsilon_{v-r}g_2(\varepsilon_u)\varepsilon_{v-q}\right] \\
&= E\left[g_1(\mathcal{F}_{u-1})\right]\times \\
&\quad \left(E\left[\rho^{2(v-u)}g_2(\varepsilon_u)\varepsilon_u^2 + \sum_{r=0}^{v-u-1}\rho^{2r}g_2(\varepsilon_u)\varepsilon_{v-r}^2\right] + E\left[\sum_{r\neq q}\rho^r\rho^q\varepsilon_{v-r}g_2(\varepsilon_u)\varepsilon_{v-q}\right]\right) \\
&= E\left[g_1(\mathcal{F}_{u-1})\right]\left(\rho^{2(v-u)}E\left[g_2(\varepsilon_u)\varepsilon_u^2\right] + E\left[g_2(\varepsilon_u)\right]E\left[\sum_{r=0}^{v-u-1}\rho^{2r}\varepsilon_{v-r}^2\right]\right) \\
&= E\left[g_1(\mathcal{F}_{u-1})\right]\left(\rho^{2(v-u)}E\left[g_2(\varepsilon_u)\varepsilon_u^2\right] + E\left[g_2(\varepsilon_u)\right]\sigma^2\left(\frac{1-\rho^{2(v-u)}}{1-\rho^2}\right)\right) \\
&= E\left[g_1(\mathcal{F}_{u-1})\right]\left(\rho^{2(v-u)}E\left[g_2(\varepsilon_u)\varepsilon_u^2\right] + \sigma_X^2\left(1-\rho^{2(v-u)}\right)E\left[g_2(\varepsilon_u)\right]\right) \\
&= E\left[g_1(\mathcal{F}_{u-1})\right]\left(\rho^{2(v-u)}E\left[g_2(\varepsilon_u)\varepsilon_u^2\right] + \sigma_X^2E\left[g_2(\varepsilon_u)\right] - \sigma_X^2\rho^{2(v-u)}E\left[g_2(\varepsilon_u)\right]\right) \\
&= \sigma_X^2E\left[g_1(\mathcal{F}_{u-1})\right]E\left[g_2(\varepsilon_u)\right] + \\
&\quad \rho^{2(v-u)}\left(E\left[g_1(\mathcal{F}_{u-1})\right]E\left[g_2(\varepsilon_u)\varepsilon_u^2\right] - \sigma_X^2E\left[g_1(\mathcal{F}_{u-1})\right]E\left[g_2(\varepsilon_u)\right]\right). \quad \square
\end{aligned}
$$

Lemma 2.5.3 *Suppose $\{\xi_{ni}\}$ is a sequence of random variables such that,*

$$\max_{1\leq i\leq n}|\xi_{ni}-\xi_i| \xrightarrow{p} 0.$$

Furthermore, let f denote an arbitrary function that is uniformly continuous. Then the following is true,

$$\max_{1\leq i\leq n}|f(\xi_{ni})-f(\xi_i)| \xrightarrow{p} 0.$$

proof. Arbitrarily choose $\varepsilon > 0$ and $\delta > 0$. Since f is uniformly continuous we can find a $K(\varepsilon)$ such that for $|x-y| < K(\varepsilon)$ we have $|f(x)-f(y)| < \varepsilon$. Now let

$\Omega_n = \{\omega : \max_{1\leq i\leq n} |\xi_{ni} - \xi_i| < K(\varepsilon)\}$. Then we have the following,

$$\begin{aligned}
&P\left[\left\{\omega : \max_{1\leq i\leq n} |f(\xi_{ni}) - f(\xi_i)| > \varepsilon\right\}\right] \\
\leq\ & P\left[\left\{\omega : \max_{1\leq i\leq n} |f(\xi_{ni}) - f(\xi_i)| > \varepsilon\right\} \bigcap \Omega_n\right] \\
+\ & P\left[\left\{\omega : \max_{1\leq i\leq n} |f(\xi_{ni}) - f(\xi_i)| > \varepsilon\right\} \bigcap \Omega_n^c\right] \\
=\ & P_1 + P_2 \text{ say.}
\end{aligned}$$

Consider P_1 first. If $\omega \in \Omega_n$ then $|f(\xi_{ni}) - f(\xi_i)| < \varepsilon$ for all i by the uniform continuity of f. Thus, the intersection is empty which implies $P_1 = 0$. Now consider P_2,

$$\begin{aligned}
P_2 &= P\left[\left\{\omega : \max_{1\leq i\leq n} |f(\xi_{ni}) - f(\xi_i)| > \varepsilon\right\} \bigcap \Omega_n^c\right] \\
&\leq P[\Omega_n^c].
\end{aligned}$$

Since $\max_{1\leq i\leq n} |\xi_{ni} - \xi_i| \xrightarrow{p} 0$, we can find a $N(\varepsilon, \delta)$ such that for all $n \geq N(\varepsilon, \delta)$ we have $P[\Omega_n^c] < \delta$. Thus, for all $n \geq N(\varepsilon, \delta)$ we have the following,

$$P\left[\left\{\omega : \max_{1\leq i\leq n} |f(\xi_{ni}) - f(\xi_i)| > \varepsilon\right\}\right] < \delta,$$

which proves the result. □

Next, we present a standard probability result along with a lemma pertaining to the expectation of a bounded random variable that converges in probability to zero.

Lemma 2.5.4 *Let Y be a nonnegative random variable which is integrable over a set E. Then, given $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that for every set $A \subset E$*

with $P[A] < \delta(\epsilon)$, we have the following inequality,

$$\int_A Y dP < \epsilon.$$

proof. This lemma is just a restatement of Proposition 14 on page 88 of Royden (1988) in terms of probability measures. □

Lemma 2.5.5 *Suppose X_n is a random variable such that $X_n = o_p(1)$ and $|X_n| \leq B$. Furthermore, suppose Y is another random variable such that $E[|Y|] < \infty$. Then, $E[YX_n] = o(1)$.*

proof. Choose $\epsilon > 0$ and define $\Omega_n = \{\omega : |X_n(\omega)| < \frac{\epsilon}{2E[|Y|]}\}$. Then we have the following,

$$\begin{aligned}
|E[YX_n]| &= |\int Y(\omega)X_n(\omega)dP| \\
&\leq \int |Y(\omega)X_n(\omega)|dP \\
&= \int_{\Omega_n} |Y(\omega)X_n(\omega)|dP + \int_{\Omega_n^c} |Y(\omega)X_n(\omega)|dP \\
&\leq \frac{\epsilon}{2E[|Y|]} \int_{\Omega_n} |Y(\omega)|dP + B \int_{\Omega_n^c} |Y(\omega)|dP \\
&\leq \frac{\epsilon}{2E[|Y|]} E[|Y|] + B \int_{\Omega_n^c} |Y(\omega)|dP
\end{aligned}$$

Now, since $X_n = o_p(1)$ it follows that $\lim_{n\to\infty} P[\Omega_n^c] = 0$. Thus, one can apply Lemma 2.5.4 to get the following inequality,

$$\int_{\Omega_n^c} |Y(\omega)|dP < \frac{\epsilon}{2B}.$$

Now, taking the limit as $n \to \infty$ first and then taking the limit as $\epsilon \to 0$ completes the proof. That is,

$$\lim_{n\to\infty} E[YX_n] = 0. \quad \square$$

2.5.2 Asymptotic Linearity Result

Before we state and prove the linearity result, consider the following lemma.

Lemma 2.5.6 *Let* $C_n = \sum_{i<j} b_{ij} (X_{j-1} - X_{i-1})^2$ *and* $C = b_{..}\sigma_X^2$ *where* $b_{..} = \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n b_{ij}$. *Then, under assumptions M1, M2, E1, and W1, we have the following,*

$$\frac{1}{n^2}C_n \xrightarrow{p} C.$$

proof.

$$\begin{aligned}
\frac{1}{n^2}C_n &= \frac{1}{n^2}\sum_{i<j} b_{ij} (X_{j-1} - X_{i-1})^2 \\
&= \frac{1}{2n^2}\sum_{i=1}^n \sum_{j=1}^n b_{ij} (X_{j-1} - X_{i-1})^2 \\
&= \frac{1}{2n^2}\sum_{i=1}^n \sum_{j=1}^n b_{ij} \left(X_{j-1}^2 - 2X_{j-1}X_{i-1} + X_{i-1}^2\right) \\
&= \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n b_{ij}X_{i-1}^2 - \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n b_{ij}X_{i-1}X_{j-1} \\
&\quad \text{since } b_{ij} = b_{ji} \\
&= C_{n1} - C_{n2} \text{ say.}
\end{aligned}$$

Note that if $b_{ij} \equiv 1$ then,

$$\begin{aligned}
C_{n2} &= \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n X_{i-1}X_{j-1} \\
&= \frac{1}{n}\sum_{i=1}^n X_{i-1}\frac{1}{n}\sum_{j=1}^n X_{j-1} \\
&= o_p(1),
\end{aligned}$$

and $C_{n1} \xrightarrow{p} \sigma_X^2$ by the Ergodic Theorem. Thus, it seems reasonable to conjecture that $C_{n1} \xrightarrow{p} C$ and $C_{n2} \xrightarrow{p} 0$. Consider C_{n1} first. Since $C_{n1} \xrightarrow{p} C$ if and only if $C_{n1} - C \xrightarrow{p} 0$ it suffices to prove the latter. Thus,

$$\begin{aligned} C_{n1} - C &= \frac{1}{n}\sum_{i=1}^{n} b_{i.}X_{i-1}^2 - b_{..}\sigma_X^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} b_{i.}\left(X_{i-1}^2 - \sigma_X^2\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} c_i \text{ say.} \end{aligned}$$

To show our result using Chebyshev's Inequality it suffices to prove,

$$E\left[\left(\frac{1}{n}\sum_{i=1}^{n} c_i\right)^2\right] = o(1).$$

Now,

$$\begin{aligned} E\left[\left(\frac{1}{n}\sum_{i=1}^{n} c_i\right)^2\right] &= \frac{1}{n^2}E\left[\sum_{i=1}^{n} c_i^2 + 2\sum_{i<j} c_ic_j\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n} E\left[c_i^2\right] + \frac{2}{n^2}\sum_{i<j} E\left[c_ic_j\right] \\ &= C_{n11} + C_{n12} \text{ say.} \end{aligned}$$

Using W1 to bound $b_{i.}$, the finite fourth moment assumptions, and stationarity it is easily show that,

$$\begin{aligned} E\left[c_i^2\right] &= b_{i.}^2 E\left[X_{i-1}^4 - 2\sigma_X^2 X_{i-1}^2 + \sigma_X^4\right] \\ &\leq K < \infty, \end{aligned}$$

where K is a constant that depends on B_b and $E\left[X_1^4\right]$. Hence, the sum in C_{n11} is $O(n)$. Since we are dividing by n^2 it follows that $C_{n11} = o(1)$. Next, consider the

expectation in C_{n12},

$$\begin{aligned}
E[c_i c_j] &= b_{i.}b_{j.}E\left[\left(X_{i-1}^2 - \sigma_X^2\right)\left(X_{j-1}^2 - \sigma_X^2\right)\right] \\
&= b_{i.}b_{j.}E\left[\left(X_{i-1}^2 - \sigma_X^2\right)\left(\left({\rho_0}^{j-i}X_{i-1} + \Gamma_i^{j-1}\right)^2 - \sigma_X^2\right)\right] \\
&= b_{i.}b_{j.}{\rho_0}^{j-i}E\left[\left(X_{i-1}^2 - \sigma_X^2\right)\left({\rho_0}^{j-i}X_{i-1}^2 + 2X_{i-1}\Gamma_i^{j-1}\right)\right] + \\
&\quad b_{i.}b_{j.}E\left[\left(\Gamma_i^{j-1}\right)^2 - \sigma_X^2\right] \\
&= b_{i.}b_{j.}{\rho_0}^{j-i}E\left[\left(X_{i-1}^2 - \sigma_X^2\right)\left({\rho_0}^{j-i}X_{i-1}^2 + 2X_{i-1}\Gamma_i^{j-1}\right)\right] + \\
&\quad b_{i.}b_{j.}\left(\sigma_X^2 - {\rho_0}^{2(j-i)} - \sigma_X^2\right) \\
&\quad \text{using part 2 of Lemma 2.3.1} \\
&= b_{i.}b_{j.}{\rho_0}^{2(j-i)}E\left[\left(X_{i-1}^2 - \sigma_X^2\right)^2\right] \\
&\quad \text{since } X_{i-1} \text{ is independent of } \Gamma_i^{j-1} \text{ and } E\left[\Gamma_i^{j-1}\right] = 0.
\end{aligned}$$

Thus, it follows that,

$$\begin{aligned}
|E[c_i c_j]| &\le |b_{i.}b_{j.}|\left|\rho_0^2\right|^{j-i}\left|E\left[\left(X_1^2 - \sigma_X^2\right)^2\right]\right| \\
&\le \left|\rho_0^2\right|^{j-i}K,
\end{aligned}$$

where K is a constant that depends on B_b and $E[X_1^4]$. Hence,

$$\begin{aligned}
|C_{n12}| &\le \frac{2}{n^2}\sum_{i<j}|E[c_i c_j]| \\
&\le \frac{2}{n^2}\sum_{i<j}\left|\rho_0^2\right|^{j-i}K \\
&= \frac{2K}{n^2}O(n) \\
&= o(1).
\end{aligned}$$

Since we have shown that $C_{n11} = o(1)$ and $C_{n12} = o(1)$ it follows that $C_{n1} \xrightarrow{p} C$. To complete the proof we must show that $C_{n2} = o_p(1)$. Again, using Chebyshev's Inequality it suffices to show that $E\left[C_{n2}^2\right] = o(1)$. Now,

$$\begin{aligned}
E\left[C_{n2}^2\right] &= E\left[\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}X_{i-1}X_{j-1}\right)^2\right] \\
&= E\left[\left(\frac{2}{n^2}\sum_{i<j} b_{ij}X_{i-1}X_{j-1}\right)^2\right] \\
&= \frac{4}{n^4}\sum_{i<j} b_{ij}^2 E\left[X_{i-1}^2X_{j-1}^2\right] \\
&+ \frac{4}{n^4}\sum_{i<j}\sum_{k<l} b_{ij}b_{kl}E\left[X_{i-1}X_{j-1}X_{k-1}X_{l-1}\right] \\
&= C_{n21} + C_{n22} \text{ say.}
\end{aligned}$$

Consider C_{n21} first. Using the Cauchy Schwarz Inequality, the finite fourth moment assumption, stationarity, and W1 to bound b_{ij} it is easily shown that the term in the sum can be bounded by a constant K that is free of the subscripts. Thus, the sum is $O(n^2)$. Since we are dividing by n^4 it follows that $C_{n21} = o(1)$. Now consider C_{n22}. Since this sum contains four subscripts it will be convenient to break the sum into the different cases given in Table 1. The table indicates that the sum will be either $O(n^3)$ or $O(n^4)$. Consider those cases where the sum is $O(n^3)$. Using the Cauchy Schwarz Inequality, the finite fourth moment assumptions, stationarity, and W1 to bound the weights it can be shown that the term in the sum can be bounded by a constant that is free of the subscripts. Since we are dividing by n^4 we have that $C_{n22} = o(1)$ for those cases in which the sum is $O(n^3)$. Finally, consider the cases where the sum is $O(n^4)$. Since all of these

cases are treated the same way, only the details for case 1 ($i < j < k < l$) will be shown. Consider the expectation under case 1,

$$\begin{aligned} E\left[X_{i-1}X_{j-1}X_{k-1}X_{l-1}\right] &= E\left[X_{i-1}X_{j-1}X_{k-1}\left({\rho_0}^{l-k}X_{k-1} + \Gamma_k^{l-1}\right)\right] \\ &= {\rho_0}^{l-k}E\left[X_{i-1}X_{j-1}X_{k-1}^2\right] + E\left[X_{i-1}X_{j-1}X_{k-1}\right]E\left[\Gamma_k^{l-1}\right] \\ &= {\rho_0}^{l-k}E\left[X_{i-1}X_{j-1}X_{k-1}^2\right]. \end{aligned}$$

The last expression follows since Γ_k^{l-1} is independent of $(X_{i-1}, X_{j-1}, X_{k-1})$ and $E\left[\Gamma_k^{l-1}\right] = 0$. Again, using the Cauchy Schwarz Inequality, the finite fourth moment assumptions, stationarity, and W1 to bound the weights it follows that the term in the fourth order sum is bounded by $|\rho_0|^{l-k} K$ where K is free of any subscripts. Hence, under case 1 we have,

$$\begin{aligned} &\left|\frac{4}{n^4}\sum_{i<j<k<l} b_{ij}b_{kl}E\left[X_{i-1}X_{j-1}X_{k-1}X_{l-1}\right]\right| \\ &\leq \frac{4}{n^4}\sum_{i<j<k<l} |b_{ij}b_{kl}|\left|E\left[X_{i-1}X_{j-1}X_{k-1}X_{l-1}\right]\right| \\ &\leq \frac{4}{n^4}\sum_{i<j<k<l} |\rho_0|^{l-k} K \\ &= \frac{4}{n^4}O(n^3) \\ &= o(1). \end{aligned}$$

Thus, we have shown that $C_{n22} = o(1)$ for all cases. Since we have shown that $C_{n21} = o(1)$ and $C_{n22} = o(1)$ it follows that $C_{n2} = o_p(1)$. That is, we have proved that $\frac{1}{n^2}C_n \xrightarrow{p} C$. □

Now we are ready to state and prove the main result of this chapter; the asymptotic linearity result. As was done in Section 2.4.3 let $S_n(\rho) = \frac{1}{n^{\frac{3}{2}}}S(\rho)$.

Furthermore, let C_n be defined as in Lemma 2.5.6. Then we have the following theorem.

Theorem 2.5.1 *Under assumptions M1, M2, E1, E2, and W1 the following holds for all* $\Delta \in \Re$,

$$S_n\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) - S_n(\rho_0) + 2\tau\left(\frac{1}{n^2}C_n\right)\Delta \;=\; o_p(1).$$

proof. Choose a $\Delta \in \Re$ and let this Δ be fixed throughout the proof. Next, define the following:

$$\begin{aligned}
\varphi(\varepsilon_i(\rho), \varepsilon_j(\rho)) &= \frac{\operatorname{sgn}(\varepsilon_j(\rho) - \varepsilon_i(\rho)) + 1}{2} \\
T_n(\rho) &= \frac{1}{n^{\frac{3}{2}}}\sum_{i<j} b_{ij}(X_{j-1} - X_{i-1})\,\varphi(\varepsilon_i(\rho), \varepsilon_j(\rho)) \\
a_{ij} &= b_{ij}(X_{j-1} - X_{i-1}).
\end{aligned}$$

Then, it is straight forward to show the following,

$$\begin{aligned}
&S_n\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) - S_n(\rho_0) + 2\tau\left(\frac{1}{n^2}C_n\right)\Delta \\
&= 2T_n\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) - \sum_{i<j} a_{ij} - 2T_n(\rho_0) + \sum_{i<j} a_{ij} + 2\tau\left(\frac{1}{n^2}C_n\right)\Delta \\
&= 2\left[T_n\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) - T_n(\rho_0) + \tau\left(\frac{1}{n^2}C_n\right)\Delta\right].
\end{aligned}$$

Therefore, in order to prove the linearity result it suffices to show the following,

$$\begin{aligned}
R_n(\Delta) &= T_n\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) - T_n(\rho_0) + \tau\left(\frac{1}{n^2}C_n\right)\Delta \\
&= o_p(1).
\end{aligned}$$

Using Chebyshev's Inequality, it suffices to show that $E\left[R_n^2(\Delta)\right] = o(1)$. If we let

$$\begin{aligned}
W_{ij} &= \varphi\left(\varepsilon_i\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right), \varepsilon_j\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right)\right) - \varphi\left(\varepsilon_i(\rho_0), \varepsilon_j(\rho_0)\right) \\
&= \begin{cases} -1 & \text{if } \varepsilon_i < \varepsilon_j < \varepsilon_i + t_{ij}(\Delta) \\ 0 & \text{otherwise} \\ 1 & \text{if } \varepsilon_i + t_{ij}(\Delta) < \varepsilon_j < \varepsilon_i, \end{cases}
\end{aligned}$$

then it follows that,

$$\begin{aligned}
R_n(\Delta) &= T_n\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) - T_n(\rho_0) + \tau\left(\frac{1}{n^2}C_n\right)\Delta \\
&= \frac{1}{n^{\frac{3}{2}}}\sum_{i<j} a_{ij}\varphi\left(\varepsilon_i\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right), \varepsilon_j\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right)\right) \\
&- \frac{1}{n^{\frac{3}{2}}}\sum_{i<j} a_{ij}\varphi\left(\varepsilon_i(\rho_0), \varepsilon_j(\rho_0)\right) + \frac{1}{n^{\frac{3}{2}}}\tau C_n \frac{\Delta}{\sqrt{n}} \\
&= \frac{1}{n^{\frac{3}{2}}}\sum_{i<j} a_{ij}W_{ij} + \frac{1}{n^{\frac{3}{2}}}\tau C_n \frac{\Delta}{\sqrt{n}} \\
&= \frac{1}{n^{\frac{3}{2}}}\sum_{i<j} b_{ij}\left(X_{j-1} - X_{i-1}\right)\left(W_{ij} + \tau\left(X_{j-1} - X_{i-1}\right)\frac{\Delta}{\sqrt{n}}\right) \\
&= \frac{1}{n^{\frac{3}{2}}}\sum_{i<j} b_{ij}\left(X_{j-1} - X_{i-1}\right)\left(W_{ij} + \tau t_{ij}(\Delta)\right) \\
&= \frac{1}{n^{\frac{3}{2}}}\sum_{i<j} b_{ij}\psi_{ij} \text{ say.}
\end{aligned}$$

Thus, it follows that,

$$\begin{aligned}
E\left[R_n^2(\Delta)\right] &= E\left[\left(\frac{1}{n^{\frac{3}{2}}}\sum_{i<j} b_{ij}\psi_{ij}\right)^2\right] \\
&= \frac{1}{n^3}E\left[\sum_{i<j} b_{ij}^2\psi_{ij}^2 + \sum_{i<j}\sum_{k<l} b_{ij}\psi_{ij}b_{kl}\psi_{kl}\right] \\
&= \frac{1}{n^3}\sum_{i<j} b_{ij}^2 E\left[\psi_{ij}^2\right] + \frac{1}{n^3}\sum_{i<j}\sum_{k<l} b_{ij}b_{kl}E\left[\psi_{ij}\psi_{kl}\right] \\
&= R_{n1} + R_{n2} \text{ say.}
\end{aligned}$$

Consider R_{n1} first. By part 1 of Lemma 2.3.6, $E\left[\psi_{ij}^2\right]$ can be uniformly bounded by a constant K that is free of the subscripts. Also, W1 implies that the weights are bounded. Thus, we see that the sum in R_{n1} is $O(n^2)$. Since we are dividing by n^3 it follows that $R_{n1} = o(1)$. Next, consider R_{n2}. R_{n2} contains a sum with four indices. Thus, it will be convenient to break the sum into the possible cases given in Table 1. Due to symmetry, only cases one through six need to be considered. We note that case 7 reduces to R_{n1} which has already been show to be $o(1)$. To verify the symmetry, consider case 12. Under case 12 we have the following,

$$\begin{aligned}
\frac{1}{n^3}\sum_{k<l=i<j} b_{ij}b_{kl}E\left[\psi_{ij}\psi_{kl}\right] &= \frac{1}{n^3}\sum_{k<l=i<j} b_{ij}b_{ki}E\left[\psi_{ij}\psi_{ki}\right] \\
&= \frac{1}{n^3}\sum_{a<b=c<d} b_{cd}b_{ac}E\left[\psi_{cd}\psi_{ac}\right] \\
&\quad \text{where } a = k,\ b = l,\ c = i, \text{ and } d = j \\
&= \frac{1}{n^3}\sum_{a<b=c<d} b_{ac}b_{cd}E\left[\psi_{ac}\psi_{cd}\right] \\
&= \frac{1}{n^3}\sum_{a<b=c<d} b_{ab}b_{cd}E\left[\psi_{ab}\psi_{cd}\right].
\end{aligned}$$

The sum in the last expression is the sum under case 2 in Table 1. Thus, case 2 $\equiv$ case 12. Similar calculations show that case 1 $\equiv$ case 13, case 3 $\equiv$ case 11, case 4 $\equiv$ case 10, case 5 $\equiv$ case 9, and case 6 $\equiv$ case 8. Now, cases 1, 3, and 5 are $O(n^4)$ and cases 2, 4, and 6 are $O(n^3)$. Let us start with the cases that are $O(n^3)$. Since all of the cases are treated the same way consider only case 4. Under case 4 we wish to show that,

$$\frac{1}{n^3}\sum_{i<k<j} b_{ij}b_{kj}E\left[\psi_{ij}\psi_{kj}\right] = o(1).$$

The Cauchy Schwarz Inequality, W1, and part 2 of Lemma 2.3.6 imply the following,

$$\begin{aligned} |b_{ij}b_{kj}E[\psi_{ij}\psi_{kj}]| &\le B_b^2\sqrt{E\left[\psi_{ij}^2\right]E\left[\psi_{kj}^2\right]} \\ &= o(1), \end{aligned}$$

uniformly in (i,k,j). Because we have an average of terms and the above result is uniform in (i,k,j) it follows that the average is $o(1)$. Hence, under case 4, $R_{n2} = o(1)$. Since cases 2 and 6 can be handled in a similar fashion, all cases that are $O(n^3)$ are also $o(1)$. Now consider those terms that are $O(n^4)$. The strategy will be to factor out a $\frac{1}{\sqrt{n}}$ and a ρ_0 raised to some power from the expectation. The ρ_0 term will reduce the sum from $O(n^4)$ to $O(n^3)$ and the $\frac{1}{\sqrt{n}}$ will be used to show that the term is $o(1)$. Because the details under cases 1 and 3 are similar, only the details for case 5 $(i < k < l < j)$ will be given. However, this seems to be the most difficult case since the distance between i and j is maximized. In what follows it will be convenient to let $d_{ij} = (X_{j-1} - X_{i-1})$. Now, Lemma 2.3.4 implies the following,

$$\begin{aligned} &E[\psi_{ij}\psi_{kl}] \\ &= E[d_{ij}(W_{ij} + \tau t_{ij}(\Delta))\psi_{kl}] \\ &= E[d_{ij}\psi_{kl}E[W_{ij} + \tau t_{ij}(\Delta) \mid \mathcal{F}_{j-1}]] \\ &= E[d_{ij}\psi_{kl}(-f(\xi_{ij}(\Delta))t_{ij}(\Delta) + \tau t_{ij}(\Delta))] \\ &= E[d_{ij}\psi_{kl}(\tau - f(\xi_{ij}(\Delta)))t_{ij}(\Delta)] \end{aligned}$$

$$
\begin{aligned}
&= \frac{\Delta}{\sqrt{n}} E\left[d_{ij}^2 \psi_{kl} \left(\tau - f\left(\xi_{ij}(\Delta)\right)\right)\right] \\
&= \frac{\Delta}{\sqrt{n}} E\left[d_{ij}^2 \psi_{kl} \left(\tau - f(\varepsilon_i) + f(\varepsilon_i) - f\left(\xi_{ij}(\Delta)\right)\right)\right] \\
&= \frac{\Delta}{\sqrt{n}} E\left[\left(\tau - f(\varepsilon_i)\right) \psi_{kl} d_{ij}^2\right] + \frac{\Delta}{\sqrt{n}} E\left[\psi_{kl} d_{ij}^2 \left(f(\varepsilon_i) - f\left(\xi_{ij}(\Delta)\right)\right)\right] \\
&= E_{n1} + E_{n2} \text{ say.}
\end{aligned}
$$

Consider E_{n1} first. Since the next largest subscript is l, we will use part 1 of Lemma 2.3.2 to write d_{ij}^2 in terms of X_{i-1}, X_{l-1}, and Γ_l^{j-1} and then condition on $\mathcal{F}_{l-1}$. Now, part 1 of Lemma 2.3.2 implies the following,

$$
\begin{aligned}
d_{ij}^2 &= \left(X_{j-1} - X_{i-1}\right)^2 \\
&= \left({\rho_0}^{j-l} X_{l-1} + \Gamma_l^{j-1} - X_{i-1}\right)^2 \\
&= {\rho_0}^{j-l}\left({\rho_0}^{j-l} X_{l-1}^2 + 2X_{l-1}\Gamma_l^{j-1} - 2X_{i-1}X_{l-1}\right) + \left(X_{i-1} - \Gamma_l^{j-1}\right)^2 \\
&= {\rho_0}^{j-l} D_1 + D_2^2 \text{ say.}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
E_{n1} &= \frac{\Delta}{\sqrt{n}} E\left[\left(\tau - f(\varepsilon_i)\right) \psi_{kl} d_{ij}^2\right] \\
&= \frac{\Delta}{\sqrt{n}} E\left[\left(\tau - f(\varepsilon_i)\right) \psi_{kl} \left({\rho_0}^{j-l} D_1 + D_2^2\right)\right] \\
&= \frac{\Delta}{\sqrt{n}} {\rho_0}^{j-l} E\left[\left(\tau - f(\varepsilon_i)\right) \psi_{kl} D_1\right] + \frac{\Delta}{\sqrt{n}} E\left[\left(\tau - f(\varepsilon_i)\right) \psi_{kl} D_2^2\right] \\
&= E_{n11} + E_{n12} \text{ say.}
\end{aligned}
$$

Now consider E_{n11}. The Cauchy Schwarz Inequality and the bound on f imply the following,

$$
\left|E\left[\left(\tau - f(\varepsilon_i)\right) \psi_{kl} D_1\right]\right| \le B_f \sqrt{E\left[\psi_{kl}^2\right] E\left[D_1^2\right]}.
$$

Part 1 of Lemma 2.3.6, the finite fourth moment assumptions, and stationarity imply that the above expectations can be bounded by a constant that is free of any subscripts. Thus,

$$\begin{aligned} |E_{n11}| &\leq \frac{|\Delta|}{\sqrt{n}} |\rho_0|^{j-l} K \\ &= O\left(\frac{1}{\sqrt{n}}\right) |\rho_0|^{j-l}. \end{aligned}$$

Now consider the expectation given in E_{n12},

$$\begin{aligned} & E\left[(\tau - f(\varepsilon_i))\, \psi_{kl} D_2^2\right] \\ &= E\left[(\tau - f(\varepsilon_i))\, \psi_{kl} \left(X_{i-1} - \Gamma_l^{j-1}\right)^2\right] \\ &= E\left[(\tau - f(\varepsilon_i))\, \psi_{kl} \left(X_{i-1}^2 - 2X_{i-1}\Gamma_l^{j-1} + \left(\Gamma_l^{j-1}\right)^2\right)\right] \\ &= E\left[X_{i-1}^2 (\tau - f(\varepsilon_i))\, \psi_{kl}\right] - 2E\left[X_{i-1} (\tau - f(\varepsilon_i))\, \psi_{kl} \Gamma_l^{j-1}\right] \\ &+ E\left[(\tau - f(\varepsilon_i))\, \psi_{kl} \left(\Gamma_l^{j-1}\right)^2\right] \\ &= E_{n121} + E_{n122} + E_{n123} \text{ say.} \end{aligned}$$

Consider E_{n121} first. Lemma 2.3.4 can be used to show the following,

$$\begin{aligned} E_{n121} &= E\left[X_{i-1}^2 (\tau - f(\varepsilon_i))\, \psi_{kl}\right] \\ &= E\left[X_{i-1}^2 (\tau - f(\varepsilon_i))\, d_{kl} \left(W_{kl} + \tau t_{kl}(\Delta)\right)\right] \\ &= E\left[X_{i-1}^2 (\tau - f(\varepsilon_i))\, d_{kl} E\left[W_{kl} + \tau t_{kl}(\Delta) \mid \mathcal{F}_{l-1}\right]\right] \\ &= E\left[X_{i-1}^2 (\tau - f(\varepsilon_i))\, d_{kl} \left(-f\left(\xi_{kl}(\Delta)\right) t_{kl}(\Delta) + \tau t_{kl}(\Delta)\right)\right] \\ &= E\left[X_{i-1}^2 (\tau - f(\varepsilon_i))\, d_{kl} \left(\tau - f\left(\xi_{kl}(\Delta)\right)\right) t_{kl}(\Delta)\right] \\ &= \frac{\Delta}{\sqrt{n}} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i))\, d_{kl}^2 \left(\tau - f(\varepsilon_k) + f(\varepsilon_k) - f\left(\xi_{kl}(\Delta)\right)\right)\right] \end{aligned}$$

$$
\begin{aligned}
&= \frac{\Delta}{\sqrt{n}} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) (\tau - f(\varepsilon_k)) d_{kl}^2\right] + \\
&\quad \frac{\Delta}{\sqrt{n}} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) d_{kl}^2 (f(\varepsilon_k) - f(\xi_{kl}(\Delta)))\right] \\
&= E_{n1211} + E_{n1212} \text{ say.}
\end{aligned}
$$

Now use part 1 of Lemma 2.3.2 to show the following,

$$
\begin{aligned}
E_{n1211} &= \frac{\Delta}{\sqrt{n}} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) (\tau - f(\varepsilon_k)) d_{kl}^2\right] \\
&= \frac{\Delta}{\sqrt{n}} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) (\tau - f(\varepsilon_k)) (X_{l-1} - X_{k-1})^2\right] \\
&= \frac{\Delta}{\sqrt{n}} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) (\tau - f(\varepsilon_k)) \left({\rho_0}^{l-k} X_{k-1} + \Gamma_k^{l-1} - X_{k-1}\right)^2\right] \\
&= \frac{\Delta}{\sqrt{n}} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) (\tau - f(\varepsilon_k)) \left(\left({\rho_0}^{l-k} - 1\right) X_{k-1} + \Gamma_k^{l-1}\right)^2\right] \\
&= \frac{\Delta}{\sqrt{n}} \left({\rho_0}^{l-k} - 1\right)^2 E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) X_{k-1}^2 (\tau - f(\varepsilon_k))\right] + \\
&\quad 2\frac{\Delta}{\sqrt{n}} \left({\rho_0}^{l-k} - 1\right) E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) X_{k-1} (\tau - f(\varepsilon_k)) \Gamma_k^{l-1}\right] + \\
&\quad \frac{\Delta}{\sqrt{n}} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) (\tau - f(\varepsilon_k)) \left(\Gamma_k^{l-1}\right)^2\right].
\end{aligned}
$$

The first expectation in the last expression is zero since ε_k is independent of $\mathcal{F}_{k-1}$ and $E[\tau - f(\varepsilon_k)] = 0$. For the second expectation, let $g_1(\mathcal{F}_{k-1}) = X_{i-1}^2(\tau - f(\varepsilon_i))X_{k-1}$ and $g_2(\varepsilon_k) = \tau - f(\varepsilon_k)$ and apply Lemma 2.5.1 to show the following,

$$
\begin{aligned}
&2\frac{\Delta}{\sqrt{n}} \left({\rho_0}^{l-k} - 1\right) E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) X_{k-1} (\tau - f(\varepsilon_k)) \Gamma_k^{l-1}\right] = \\
&\quad 2\frac{\Delta}{\sqrt{n}} \left({\rho_0}^{l-k} - 1\right) {\rho_0}^{l-k-1} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) X_{k-1}\right] E\left[(\tau - f(\varepsilon_k)) \varepsilon_k\right].
\end{aligned}
$$

For the third term, note that $X_{i-1}^2 (\tau - f(\varepsilon_i))$ and $(\tau - f(\varepsilon_k)) \left(\Gamma_k^{l-1}\right)^2$ are independent. Since ε_i and X_{i-1} are independent and $E[(\tau - f(\varepsilon_i))] = 0$ it follows

that,

$$\frac{\Delta}{\sqrt{n}} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) (\tau - f(\varepsilon_k)) \left(\Gamma_k^{l-1}\right)^2\right] = 0.$$

Thus, the above comments imply,

$$E_{n1211} = \frac{2\Delta}{\sqrt{n}} \left({\rho_0}^{l-k} - 1\right) {\rho_0}^{l-k-1} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) X_{k-1}\right] E\left[(\tau - f(\varepsilon_k)) \varepsilon_k\right].$$

Using the fact that $|\rho_0| < 1$, f is bounded, and τ is finite the Cauchy Schwarz Inequality implies that the above expectations can be bounded by a constant that only depends on a subset of Δ, τ, B_f, $E[X_1^4]$, and $E[\varepsilon_1^4]$. Hence,

$$|E_{n1211}| \leq \frac{1}{\sqrt{n}} |\rho_0|^{l-k-1} K.$$

Now consider the expectation in E_{n1212}. Part 1 of Lemma 2.2.1, Corollary 2.3.1, and Lemma 2.5.3 imply that $(f(\varepsilon_k) - f(\xi_{kl}(\Delta)))$ is a bounded random variable that is $o_p(1)$ uniformly in (k, l). Hence, the finite fourth moment assumptions, stationarity, the bound on f, and Lemma 2.5.5 imply that this expectation is $o(1)$ uniformly in (i, k, l). Thus,

$$\begin{aligned} |E_{n1212}| &= \frac{\Delta}{\sqrt{n}} E\left[X_{i-1}^2 (\tau - f(\varepsilon_i)) d_{kl}^2 (f(\varepsilon_k) - f(\xi_{kl}(\Delta)))\right] \\ &= O\left(\frac{1}{\sqrt{n}}\right) o(1), \end{aligned}$$

uniformly in (i, k, l). Combining the last two results yields,

$$\begin{aligned} |E_{n121}| &= |E_{n1211} + E_{n1212}| \\ &\leq |E_{n1211}| + |E_{n1212}| \end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{\sqrt{n}}|\rho_0|^{l-k-1}K + O\left(\frac{1}{\sqrt{n}}\right)o(1) \\
&= O\left(\frac{1}{\sqrt{n}}\right)|\rho_0|^{l-k-1}O\left(\frac{1}{\sqrt{n}}\right)o(1),
\end{aligned}$$

where the convergence is uniform over all subscripts. Now consider E_{n122}.

$$\begin{aligned}
E_{n122} &= -2E\left[X_{i-1}(\tau - f(\varepsilon_i))\psi_{kl}\Gamma_l^{j-1}\right] \\
&= -2E\left[X_{i-1}(\tau - f(\varepsilon_i))\psi_{kl}\left({\rho_0}^{j-l-1}\varepsilon_l + \Gamma_{l+1}^{j-1}\right)\right] \\
&= -2{\rho_0}^{j-l-1}E\left[X_{i-1}(\tau - f(\varepsilon_i))\psi_{kl}\varepsilon_l\right],
\end{aligned}$$

since Γ_{l+1}^{j-1} is independent of $\mathcal{F}_l$ and $E\left[\Gamma_{l+1}^{j-1}\right] = 0$. Now apply the Cauchy Schwarz Inequality to bound the remaining expectation by a constant that is free of any subscripts. Thus,

$$|E_{n122}| \leq |\rho_0|^{j-l-1}K.$$

Finally, consider E_{n123}. The definition of Γ_l^{j-1} and the independence of Γ_{l+1}^{j-1} with $\mathcal{F}_l$ imply the following,

$$\begin{aligned}
E_{n123} &= E\left[(\tau - f(\varepsilon_i))\psi_{kl}\left(\Gamma_l^{j-1}\right)^2\right] \\
&= E\left[(\tau - f(\varepsilon_i))\psi_{kl}\left({\rho_0}^{j-l-1}\varepsilon_l + \Gamma_{l+1}^{j-1}\right)^2\right] \\
&= {\rho_0}^{2(j-l-1)}E\left[(\tau - f(\varepsilon_i))\psi_{kl}\varepsilon_l^2\right] + 2{\rho_0}^{j-l-1}E\left[(\tau - f(\varepsilon_i))\psi_{kl}\varepsilon_l\right]E\left[\Gamma_{l+1}^{j-1}\right] + \\
&\quad E\left[(\tau - f(\varepsilon_i))\psi_{kl}\right]E\left[\left(\Gamma_{l+1}^{j-1}\right)^2\right] \\
&= {\rho_0}^{2(j-l-1)}E\left[(\tau - f(\varepsilon_i))\psi_{kl}\varepsilon_l^2\right] + E\left[(\tau - f(\varepsilon_i))\psi_{kl}\right]\left(\sigma_X^2 - \frac{{\rho_0}^{2(j-l-1)}}{1-{\rho_0}^2}\right) \\
&= {\rho_0}^{2(j-l-1)}\left(E\left[(\tau - f(\varepsilon_i))\psi_{kl}\varepsilon_l^2\right] - \frac{1}{1-{\rho_0}^2}E\left[(\tau - f(\varepsilon_i))\psi_{kl}\right]\right) + \\
&\quad E\left[\sigma_X^2(\tau - f(\varepsilon_i))\psi_{kl}\right].
\end{aligned}$$

The two expectations in the first term can be bounded by a constant, K, that depends on a subset of Δ, τ, B_f, $E[X_1^4]$, and $E[\varepsilon_1^4]$ after an application of the Cauchy Schwarz Inequality. Note that the second term is exactly the same as E_{n121} except that there is a σ_x^2 where there was a X_{i-1}^2 in E_{n121}. Since the X_{i-1}^2 didn't contribute anything to the final result on E_{n121}, the same result holds for this term as well. Hence,

$$\begin{aligned}
|E_{123}| &\leq \left|\rho_0^2\right|^{(j-l-1)} K + \frac{1}{\sqrt{n}} |\rho_0|^{l-k-1} K + O\left(\frac{1}{\sqrt{n}}\right) o(1) \\
&= \left|\rho_0^2\right|^{(j-l-1)} K + O\left(\frac{1}{\sqrt{n}}\right) |\rho_0|^{l-k-1} + O\left(\frac{1}{\sqrt{n}}\right) o(1),
\end{aligned}$$

where the convergence is uniform in all subscripts. Combining these three parts we get the following,

$$\begin{aligned}
|E_{n12}| &= \frac{|\Delta|}{\sqrt{n}} |E_{n121} + E_{n122} + E_{n123}| \\
&\leq \frac{|\Delta|}{\sqrt{n}} (|E_{n121}| + |E_{n122}| + |E_{n123}|) \\
&\leq \frac{2|\Delta|}{\sqrt{n}} \left(|\rho_0|^{j-l-1} K + \frac{1}{\sqrt{n}} |\rho_0|^{l-k-1} K + O\left(\frac{1}{\sqrt{n}}\right) o(1) \right) \\
&= O\left(\frac{1}{\sqrt{n}}\right) |\rho_0|^{j-l-1} + O\left(\frac{1}{n}\right) |\rho_0|^{l-k-1} + O\left(\frac{1}{n}\right) o(1),
\end{aligned}$$

where the convergence is uniform in all subscripts. Therefore,

$$\begin{aligned}
|E_{n1}| &= |E_{n11} + E_{n12}| \\
&\leq |E_{n11}| + |E_{n12}| \\
&\leq \frac{|\Delta|}{\sqrt{n}} |\rho_0|^{j-l} K + O\left(\frac{1}{\sqrt{n}}\right) |\rho_0|^{j-l-1} + O\left(\frac{1}{n}\right) |\rho_0|^{l-k-1} + O\left(\frac{1}{n}\right) o(1) \\
&= O\left(\frac{1}{\sqrt{n}}\right) |\rho_0|^{j-l} + O\left(\frac{1}{\sqrt{n}}\right) |\rho_0|^{j-l-1} + O\left(\frac{1}{n}\right) |\rho_0|^{l-k-1} + O\left(\frac{1}{n}\right) o(1)
\end{aligned}$$

$$= O\left(\frac{1}{\sqrt{n}}\right)|\rho_0|^{j-l-1} + O\left(\frac{1}{n}\right)|\rho_0|^{l-k-1} + O\left(\frac{1}{n}\right)o(1),$$

where the convergence is uniform in the subscripts. Recall that under case 5 we want to show,

$$\begin{aligned}
\frac{1}{n^3}\sum_{i<k<l<j} b_{ij}b_{kl}E\left[\psi_{ij}\psi_{kl}\right] &= \frac{1}{n^3}\sum_{i<k<l<j} b_{ij}b_{kl}\left(E_{n1}+E_{n2}\right) \\
&= \frac{1}{n^3}\sum_{i<k<l<j} b_{ij}b_{kl}E_{n1} + \frac{1}{n^3}\sum_{i<k<l<j} b_{ij}b_{kl}E_{n2} \\
&= o(1).
\end{aligned}$$

Consider just the sum with E_{n1},

$$\begin{aligned}
&\left|\frac{1}{n^3}\sum_{i<k<l<j} b_{ij}b_{kl}E_{n1}\right| \\
&\leq \frac{1}{n^3}\sum_{i<k<l<j} \left|b_{ij}b_{kl}E_{n1}\right| \\
&\leq \frac{B_b^2}{n^3}\sum_{i<k<l<j} \left|E_{n1}\right| \\
&\leq \frac{B_b^2}{n^3}\sum_{i<k<l<j}\left(O\left(\frac{1}{\sqrt{n}}\right)|\rho_0|^{j-l-1} + O\left(\frac{1}{n}\right)|\rho_0|^{l-k-1} + O\left(\frac{1}{n}\right)o(1)\right).
\end{aligned}$$

Now, the terms with ρ_0 can be used to reduce the sum to $O(n^3)$. Since we still have a factor of $\frac{1}{\sqrt{n}}$ or $\frac{1}{n}$ left over, this implies that the first two sums are $o(1)$. For the last sum, note that the factor of $\frac{1}{n}$ makes the last term $O(1)$. Since we have a factor that is $o(1)$ uniformly in the subscripts left over, the last sum is also $o(1)$. Hence, the sum that includes E_{n1} is $o(1)$. To complete the proof we need to show that the sum containing E_{n2} is also $o(1)$. Appling the definition of ψ_{kl} yields,

$$E_{n2} = \frac{\Delta}{\sqrt{n}}E\left[\psi_{kl}d_{ij}^2\left(f(\varepsilon_i) - f\left(\xi_{ij}(\Delta)\right)\right)\right]$$

$$
\begin{aligned}
&= \frac{\Delta}{\sqrt{n}} E\left[d_{kl}\left(W_{kl} + \tau t_{kl}(\Delta)\right) d_{ij}^2 \left(f(\varepsilon_i) - f\left(\xi_{ij}(\Delta)\right)\right)\right] \\
&= \frac{\Delta}{\sqrt{n}} E\left[d_{kl} W_{kl} d_{ij}^2 \left(f(\varepsilon_i) - f\left(\xi_{ij}(\Delta)\right)\right)\right] + \\
&\quad \frac{\Delta}{\sqrt{n}} E\left[d_{kl} \tau t_{kl}(\Delta) d_{ij}^2 \left(f(\varepsilon_i) - f\left(\xi_{ij}(\Delta)\right)\right)\right] \\
&= E_{n21} + E_{n22} \text{ say.}
\end{aligned}
$$

Consider E_{n21} first. An application of the Mean Value Theorem on f and the bound on f' yield the following,

$$
\begin{aligned}
|E_{n21}| &= \left|\frac{\Delta}{\sqrt{n}} E\left[d_{kl} W_{kl} d_{ij}^2 \left(f(\varepsilon_i) - f\left(\xi_{ij}(\Delta)\right)\right)\right]\right| \\
&\le \frac{|\Delta|}{\sqrt{n}} E\left[\left|d_{kl} W_{kl} d_{ij}^2\right| \left|f(\varepsilon_i) - f\left(\xi_{ij}(\Delta)\right)\right|\right] \\
&\le \frac{|\Delta|}{\sqrt{n}} E\left[\left|d_{kl} W_{kl} d_{ij}^2\right| \left|f'\left(c_{ij}(\Delta)\right)\left(\varepsilon_i - \xi_{ij}(\Delta)\right)\right|\right] \\
&\le \frac{|\Delta|}{\sqrt{n}} E\left[\left|d_{kl} W_{kl} d_{ij}^2\right| \left|B_{f'} t_{ij}(\Delta)\right|\right] \\
&\le \frac{\Delta^2 B_{f'}}{n} E\left[\left|d_{kl} W_{kl} d_{ij}^3\right|\right] \\
&= \frac{\Delta^2 B_{f'}}{n} E\left[\left|d_{kl} d_{ij}^3\right| |W_{kl}|\right].
\end{aligned}
$$

Now, W_{kl} is bounded by definition and Corollary 2.3.3 implies that $W_{kl} = o_p(1)$ for all (k,l). Furthermore, the finite fourth moment assumptions and stationarity imply that $E\left[d_{kl} d_{ij}^3\right] \le K$ for all (i,j,k,l). Thus, Lemma 2.5.5 implies,

$$
E\left[\left|d_{kl} d_{ij}^3\right| |W_{kl}|\right] = o(1),
$$

for all (i,j,k,l). Hence, it follows that,

$$
|E_{n21}| = O\left(\frac{1}{n}\right) o(1),
$$

uniformly in (i,j,k,l). Now consider E_{n22},

$$\begin{aligned}|E_{n22}| &= \left|\frac{\Delta}{\sqrt{n}}E\left[d_{kl}\tau t_{kl}(\Delta)d_{ij}^2\left(f(\varepsilon_i)-f\left(\xi_{ij}(\Delta)\right)\right)\right]\right| \\ &\leq \frac{\Delta^2\tau}{n}E\left[\left|d_{kl}^2d_{ij}^2\right|\left|f(\varepsilon_i)-f\left(\xi_{ij}(\Delta)\right)\right|\right]\end{aligned}$$

Part 1 of Lemma 2.2.1, Corollary 2.3.1, and Lemma 2.5.3 imply that $|f(\varepsilon_i)-f\left(\xi_{ij}(\Delta)\right)|$ is a bounded random variable that is $o_p(1)$ uniformly in (i,j). Hence, the finite fourth moment assumptions, stationarity, and Lemma 2.5.5 imply that the expectation in E_{n22} is $o(1)$ uniformly in (i,j,k,l). That is,

$$E\left[\left|d_{kl}^2d_{ij}^2\right|\left|f(\varepsilon_i)-f\left(\xi_{ij}(\Delta)\right)\right|\right] = o(1),$$

for all (i,j,k,l). Hence, it follows that,

$$|E_{n22}| = O\left(\frac{1}{n}\right)o(1),$$

uniformly in the subscripts. Putting the results on E_{n21} and E_{n22} together we have that,

$$\begin{aligned}|E_{n2}| &= |E_{n21}+E_{n22}| \\ &\leq |E_{n21}|+|E_{n22}| \\ &= O\left(\frac{1}{n}\right)o(1)+O\left(\frac{1}{n}\right)o(1) \\ &= O\left(\frac{1}{n}\right)o(1),\end{aligned}$$

uniformly in the subscripts. Now consider the sum that contains E_{n2}. Since the above result is uniform in (i,j,k,l) we have the following,

$$\left|\frac{1}{n^3}\sum_{i<k<l<j}b_{ij}b_{kl}E_{n2}\right| \leq \frac{1}{n^3}\sum_{i<k<l<j}|b_{ij}b_{kl}E_{n2}|$$

$$
\begin{aligned}
&\leq \frac{B_b^2}{n^3} \sum_{i<k<l<j} |E_{n2}| \\
&\leq \frac{B_b^2}{n^3} \sum_{i<k<l<j} O\left(\frac{1}{n}\right) o(1) \\
&= o(1).
\end{aligned}
$$

Hence, under case 5, we have shown the following,

$$
\frac{1}{n^3} \sum_{i<k<l<j} b_{ij} b_{kl} E\left[\psi_{ij}\psi_{kl}\right] = o(1).
$$

The proofs for cases 1 and 3 are similar but less detailed since the (i, j) subscripts are "closer" together. Therefore, the proof of $R_{n2} = o(1)$ is complete. Since we have already shown that $R_{n1} = o(1)$, it follows that $E\left[R_n^2(\Delta)\right] = o(1)$. Hence, by Chebyshev's Inequality, it follows that $R_n(\Delta) = o_p(1)$. That is,

$$
S_n\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) - S_n(\rho_0) + 2\tau\left(\frac{1}{n^2}C_n\right)\Delta = o_p(1). \ \square
$$

As a final remark, Lemma 2.5.6 implies the following,

$$
S_n\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) - S_n(\rho_0) + 2\tau C\Delta = o_p(1).
$$

2.6 Asymptotic Uniform Linearity and Uniform Quadraticity

2.6.1 Convex Function Results

In this section we collect some definitions and results pertaining to convex functions. The results presented will play a critical role in establishing the uniformity part of the linearity and quadraticity results. Hence, this section serves as a reference for some important results that will be needed later on.

In what follows H will represent a real valued function defined on $\mathcal{R}^p$. Let us begin by stating a couple of definitions pertaining to convex functions and their gradients. The following definition is given as Theorem 4.1 on page 25 of Rockafellar (1970).

Definition 2.6.1 *A function H is called convex if*

$$H(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda H(\boldsymbol{x}) + (1-\lambda)H(\boldsymbol{y}),$$

for $0 < \lambda < 1$. Furthermore, a convex function H is called proper if H is defined on an open set $C \subseteq \mathcal{R}^p$ and $H(\boldsymbol{x}) < \infty$ for all $\boldsymbol{x} \in C$.

Proper convex functions are defined on page 24 of Rockafellar (1970). The next definition is related to the gradient of a convex function and is given on page 214 of Rockafellar (1970).

Definition 2.6.2 *A vector $\boldsymbol{D}_H(\boldsymbol{x}_0)$ is called a subgradient of H at $\boldsymbol{x}_0$ if*

$$H(\boldsymbol{x}) - H(\boldsymbol{x}_0) \geq \boldsymbol{D}_H(\boldsymbol{x}_0)'(\boldsymbol{x} - \boldsymbol{x}_0),$$

for all $\boldsymbol{x} \in C \subseteq \Re^p$.

We now state a sequence of theorems that will be used later on. For a more intensive discussion of these theorems one is referred to Rockafellar (1970) and Heiler and Willers (1988). Since the proofs of these theorems are given in either Rockafellar and or Heiler and Willers, we will not present them here. Instead, we will just give the references to where the proofs can be found.

Theorem 2.6.1 *Let H be a proper convex function defined on an open set C. Then H has a subgradient at each point in C.*

proof. The proof can be found on page 217 of Rockafellar (1970). □

Theorem 2.6.2 *Let H be a proper convex function defined on an open set C. Then H is continuous on C and differentiable almost everywhere on C.*

proof. The proofs can be found on pages 82 and 246 respectively in Rockafellar (1970). □

Theorem 2.6.3 *Let H be a proper convex function defined on an open set C. If H is differentiable at $\boldsymbol{x}_0$ then the gradient of H at $\boldsymbol{x}_0$, $\boldsymbol{\nabla} H(\boldsymbol{x}_0)$, is the unique subgradient of H at $\boldsymbol{x}_0$, $\boldsymbol{D}_H(\boldsymbol{x}_0)$.*

proof. The proof can be found on page 242 of Rockafellar (1970). □

Theorem 2.6.4 *Let $\{H_n\}$ be a sequence of proper convex functions defined on C. In addition, suppose the sequence converges for all $\boldsymbol{x} \in C'$ where C' is a dense subset of C and the limit function H is differentiable. Then*

$$\lim_{n\to\infty} \boldsymbol{\nabla} H_n(\boldsymbol{x}) = \boldsymbol{\nabla} H(\boldsymbol{x}),$$

where $\boldsymbol{x} \in C$. Furthermore,

$$\lim_{n\to\infty} \sup_{\boldsymbol{x}\in C''} |\boldsymbol{\nabla} H_n(\boldsymbol{x}) - \boldsymbol{\nabla} H(\boldsymbol{x})| = 0,$$

where C'' is any compact subset of C.

proof. The proof is given as Lemma 4.3 in Heiler and Willers (1988) and is a modification of a theorem found on page 248 of Rockafellar (1970). □

Theorem 2.6.5 *Let $\{H_n\}$ be a sequence of proper convex functions defined on C and let C' be a dense subset of C. In addition, suppose the limit function H is differentiable. Then*

$$\lim_{n\to\infty} \nabla H_n(\boldsymbol{x}) = \nabla H(\boldsymbol{x}), \quad \textit{for all } \boldsymbol{x} \in C'$$

and

$$\lim_{n\to\infty} H_n(\boldsymbol{x}_0) = H(\boldsymbol{x}_0), \quad \textit{for at least one } \boldsymbol{x}_0 \in C'$$

imply that

$$\lim_{n\to\infty} H_n(\boldsymbol{x}) = H(\boldsymbol{x}), \quad \textit{for all } \boldsymbol{x} \in C.$$

Furthermore,

$$\lim_{n\to\infty} \sup_{\boldsymbol{x}\in C''} |H_n(\boldsymbol{x}) - H(\boldsymbol{x})| = 0,$$

where C'' is any compact subset of C.

proof. The proof is given as Lemma 4.4 of Heiler and Willers (1988). □

2.6.2 Probability Results

In this section we present two probability results that will be used in conjunction with the above convex function results to prove the uniformity part of our linearity result.

Theorem 2.6.6 *Let $X, X_1, X_2, \ldots$ be a sequence of p-dimensional vector random variables. Then $X_n \xrightarrow{p} X$ if and only if every subsequence, $\{X_{\phi(n)}\}$, contains a further subsequence, $\{X_{\Phi(\phi(n))}\}$, such that $X_{\Phi(\phi(n))} \overset{wp1}{\to} X$.*

proof. The proof can be found on page 103 of Tucker (1967). □

Theorem 2.6.7 *Let $X, X_1, X_2, \ldots$ be a sequence of p-dimensional vector random variables. Then $X_n \overset{wp1}{\to} X$ implies that $X_{\phi(n)} \overset{wp1}{\to} X$ where $\{X_{\phi(n)}\}$ represents an arbitrary subsequence of $\{X_n\}$.*

proof. The proof is trivial. □

2.6.3 Uniform Linearity and Quadraticity

In what follows ρ_0 will denote the true parameter in the AR(1) model and Δ will represent an arbitrary real number. Recall from Section 2.1 that $D(\rho)$ represents the dispersion function and $S(\rho)$ is the negative of its gradient. Now define the following functions of Δ,

$$\begin{aligned}
D_n(\Delta) &= \frac{1}{n} D\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) \\
S_n(\Delta) &= -\frac{d}{d\Delta} D_n(\Delta) = \frac{1}{n^{\frac{3}{2}}} S\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) \\
Q_n(\Delta) &= D_n(0) - S_n(0)\Delta + \tau C \Delta^2.
\end{aligned}$$

In an effort to motivate the results we need consider the following heuristic discussion. Assuming a first order Taylors Expansion of $S_n(\Delta)$ about 0 is possible,

the above relationships imply the following,

$$
\begin{aligned}
S_n(\Delta) &\doteq S_n(0) + S_n'(0)\,\Delta \\
&= \frac{1}{n^{\frac{3}{2}}} S(\rho_0) + \frac{1}{n^2} S_n'(\rho_0)\,\Delta.
\end{aligned}
$$

However, Theorem 2.5.1 implies that,

$$
S_n(\Delta) \doteq S_n(0) - 2\tau C\Delta.
$$

Thus, the above two statements suggest that $\frac{1}{n^2} S_n'(\rho_0)$ behaves like $-2\tau C$. Furthermore, assuming a second order Taylors Expansion of $D_n(\Delta)$ about 0 is possible, we would have the following,

$$
\begin{aligned}
D_n(\Delta) &\doteq D_n(0) + D_n'(0)\,\Delta + \frac{1}{2} D_n''(0)\,\Delta^2 \\
&= D_n(0) - S_n(0)\Delta - \frac{1}{2}\left(\frac{1}{n^2} S_n'(\rho_0)\right)\Delta^2.
\end{aligned}
$$

Since $\frac{1}{n^2} S_n'(\rho_0)$ behaves like $-2\tau C$ it seems reasonable to replace this term in the above expression and expect the following to hold,

$$
\begin{aligned}
D_n(\Delta) &\doteq D_n(0) - S_n(0)\Delta + \tau C\Delta^2 \\
&= Q_n(\Delta).
\end{aligned}
$$

In keeping with the tradition of rank-based estimates, it is this closeness between $D_n(\Delta)$ and $Q_n(\Delta)$ that needs to be characterized.

To begin with we will state the results that we want to obtain. We will say that we have asymptotic linearity (AL) if for all $\Delta \in \Re$ the following result holds,

$$
S_n(\Delta) - S_n(0) + 2\tau C\Delta \xrightarrow{p} 0. \tag{2.2}
$$

Secondly, we will say we have asymptotic uniform linearity (AUL) if for all $c > 0$ the following result holds,

$$\sup_{|\Delta| \leq c} |S_n(\Delta) - S_n(0) + 2\tau C\Delta| \xrightarrow{p} 0. \tag{2.3}$$

Lastly, we will say that we have asymptotic uniform quadraticity (AUQ) if for all $c > 0$ the following result holds,

$$\sup_{|\Delta| \leq c} |D_n(\Delta) - Q_n(\Delta)| \xrightarrow{p} 0. \tag{2.4}$$

In their 1988 paper Heiler and Willers (1988) show that (2.2), (2.3), and (2.4) are equivalent in the context of linear regression. Due to the linear nature of the AR(1) process, this result can be expected to hold. In fact, upon examination of their proof of the result it is apparent that their is nothing specific to the linear regression model that is needed. The proof exploits the convexity of the functions under consideration. However, for the sake of completeness, we will present the theorem and proof given in Heiler and Willers (1988) in the context of the AR(1).

Theorem 2.6.8 *Under M1, M2, E1, E2, and W1; AL, AUL, and AUQ are equivalent.*

proof. Following Heiler and Willers, we first show that (2.2) implies (2.4). To begin with, let $\lambda \in (0,1)$, $\Delta_1 \in \Re$, and $\Delta_2 \in \Re$ and note that the following derivations show that $D_n(\Delta)$ is a convex function of Δ,

$$D_n(\lambda\Delta_1 + (1-\lambda)\Delta_2) = \frac{1}{n} D\left(\rho_0 + \frac{\lambda\Delta_1 + (1-\lambda)\Delta_2}{\sqrt{n}}\right)$$

$$
\begin{aligned}
&= \frac{1}{n} D\left(\rho_0 + \lambda\rho_0 - \lambda\rho_0 + \frac{\lambda\Delta_1}{\sqrt{n}} + \frac{(1-\lambda)\Delta_2}{\sqrt{n}}\right) \\
&= \frac{1}{n} D\left(\lambda\left(\rho_0 + \frac{\Delta_1}{\sqrt{n}}\right) + (1-\lambda)\left(\rho_0 + \frac{\Delta_2}{\sqrt{n}}\right)\right) \\
&\leq \lambda\frac{1}{n} D\left(\rho_0 + \frac{\Delta_1}{\sqrt{n}}\right) + (1-\lambda)\frac{1}{n} D\left(\rho_0 + \frac{\Delta_2}{\sqrt{n}}\right) \\
&\qquad \text{since } D_n(\rho) \text{ is a convex function} \\
&= \lambda D_n(\Delta_1) + (1-\lambda) D_n(\Delta_2).
\end{aligned}
$$

Additionally, since $Q_n(\Delta)$ is a quadratic function it is also convex. Thus, both $D_n(\Delta)$ and $Q_n(\Delta)$ are convex functions of Δ. Furthermore, both $D_n(\Delta)$ and $Q_n(\Delta)$ are proper convex functions since each is defined on the open set $C = \Re$ and each is everywhere finite on C. Next, consider the gradients of both $D_n(\Delta)$ and $Q_n(\Delta)$. The gradient of $D_n(\Delta)$ can be shown to be $-S_n(\Delta)$ almost everywhere as follows,

$$
\begin{aligned}
\nabla D_n(\Delta) &= \frac{d}{d\Delta}\frac{1}{n} D\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) \\
&= \frac{1}{n}\frac{d}{d\Delta} D\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) \\
&= \frac{1}{n}\left[-S\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right)\right]\frac{1}{\sqrt{n}} \quad \text{a.e.} \\
&= -\frac{1}{n^{\frac{3}{2}}} S\left(\rho_0 + \frac{\Delta}{\sqrt{n}}\right) \\
&= -S_n(\Delta).
\end{aligned}
$$

Since $Q_n(\Delta)$ is a quadratic function the gradient follows immediately,

$$
\begin{aligned}
\nabla Q_n(\Delta) &= \frac{d}{d\Delta}\left(D_n(0) - S_n(0)\Delta + \tau C\Delta^2\right) \\
&= -S_n(0) + 2\tau C\Delta.
\end{aligned}
$$

Combining the last two results yields the following,

$$\nabla D_n(\Delta) - \nabla Q_n(\Delta) \;=\; -\left[S_n(\Delta) - S_n(0) + 2\tau C\Delta\right]. \qquad (2.5)$$

Since Theorem 2.5.1 and Lemma 2.5.6 imply that the right hand side of (2.5) is $o_p(1)$, we have for all $\Delta \in \Re$,

$$\nabla D_n(\Delta) \;=\; \nabla Q_n(\Delta) + o_p(1). \qquad (2.6)$$

Now, if we were to integrate both sides of (2.6) it would seem reasonable to expect that for a specific Δ the following,

$$D_n(\Delta) \;=\; Q_n(\Delta) + o_p(1).$$

However, this is not enough since we need the result to be uniform for all Δ satisfying $|\Delta| \leq c$.

To prove uniformity proceed as follows. Let $N = \{1, 2, \ldots\}$ and let $N_0 \subseteq N$ be an arbitrary infinant index set. Also, let C' denote the set of rational numbers. Since C' is countable we can define $C' = \{\Delta_n\}$ to be the sequence of rationals. Now, for $n \in N$, Theorem 2.5.1 implies the following,

$$S_n(\Delta_1) - S_n(0) + 2\tau C\Delta_1 \;\xrightarrow{p}\; 0.$$

By Theorem 2.6.6 there exists a $N_1 \subseteq N_0$ such that for $n \in N_1$ we have,

$$S_n(\Delta_1) - S_n(0) + 2\tau C\Delta_1 \;\xrightarrow{wp1}\; 0.$$

For $n \in N$, a second application of Theorem 2.5.1 yields,

$$S_n(\Delta_2) - S_n(0) + 2\tau C\Delta_2 \;\xrightarrow{p}\; 0.$$

Again, Theorem 2.6.6 implies there exists a $N_2 \subseteq N_1 \subseteq N_0$ such that for $n \in N_2$ we have the following,

$$S_n(\Delta_2) - S_n(0) + 2\tau C\Delta_2 \overset{wp1}{\longrightarrow} 0.$$

By Theorem 2.6.7 and the fact that $N_2 \subseteq N_1$ we also have for $n \in N_2$,

$$S_n(\Delta_1) - S_n(0) + 2\tau C\Delta_1 \overset{wp1}{\longrightarrow} 0.$$

Continue in this fashion so that on the i^{th} application of Theorem 2.5.1 we have, for $n \in N$,

$$S_n(\Delta_i) - S_n(0) + 2\tau C\Delta_i \overset{p}{\longrightarrow} 0.$$

Then, Theorem 2.6.6 implies that there exists a $N_i \subseteq \ldots \subseteq N_2 \subseteq N_1 \subseteq N_0$ such that for $n \in N_i$,

$$S_n(\Delta_i) - S_n(0) + 2\tau C\Delta_i \overset{wp1}{\longrightarrow} 0.$$

Furthermore, by Theorem 2.6.7 and the fact that $N_i \subseteq \ldots \subseteq N_2 \subseteq N_1 \subseteq N_0$ we have for $n \in N_i$ and $j = 1, 2, \ldots, i$,

$$S_n(\Delta_j) - S_n(0) + 2\tau C\Delta_j \overset{wp1}{\longrightarrow} 0.$$

Now let $\tilde{n}_i$ be the i^{th} element of N_i. Since $\ldots \subseteq N_i \subseteq N_{i-1} \subseteq \cdots \subseteq N_1 \subseteq N_0 \subseteq N$ and each N_i is an infinant index set we have that $\tilde{n}_1 < \tilde{n}_2 < \ldots < \tilde{n}_i < \ldots$. Now, define the following infinant index set, $\check{N} = \{\tilde{n}_1, \tilde{n}_2, \ldots, \tilde{n}_i, \ldots\}$. It should

be noted that by definition of $\tilde{N}$, $\tilde{N} \subseteq N_0 \subseteq N$ where N_0 was an arbitrary infinant index set. By the construction of $\tilde{N}$ we now have for $n \in \tilde{N}$ and for $\Delta \in C'$,

$$S_n(\Delta) - S_n(0) + 2\tau C\Delta \xrightarrow{wp1} 0. \tag{2.7}$$

Now define the following functions of Δ,

$$\begin{aligned} H_n(\Delta) &= D_n(\Delta) - D_n(0) + S_n(0)\Delta \\ H(\Delta) &= \tau C\Delta^2. \end{aligned}$$

It follows from the convexity of $D_n(\Delta)$ that $H_n(\Delta)$ is a proper convex function of Δ. Since $H(\Delta)$ is a quadratic function it too is a proper convex function. Using the definitions of $H_n(\Delta)$, $H(\Delta)$, and $Q_n(\Delta)$ it follows that,

$$\begin{aligned} & D_n(\Delta) - Q_n(\Delta) \\ &= (H_n(\Delta) + D_n(0) - S_n(0)\Delta) - \left(D_n(0) - S_n(0)\Delta + \tau C\Delta^2\right) \\ &= H_n(\Delta) - \tau C\Delta^2 \\ &= H_n(\Delta) - H(\Delta). \end{aligned}$$

Thus, it follows that,

$$\begin{aligned} \nabla D_n(\Delta) - \nabla Q_n(\Delta) &= \nabla H_n(\Delta) - \nabla H(\Delta) \\ &= -\left[S_n(\Delta) - S_n(0) + 2\tau C\Delta\right]. \end{aligned}$$

It now follows from (2.7) that for $n \in \tilde{N}$ and for $\Delta \in C'$ we have,

$$\nabla H_n(\Delta) \xrightarrow{wp1} \nabla H(\Delta).$$

Hence, outside of a set with probability zero we have for $n \in \tilde{N}$ and for $\Delta \in C'$,

$$\lim_{n\to\infty} \nabla H_n(\Delta) = \nabla H(\Delta).$$

Since $H_n(0) = 0$ it follows that for $n \in \tilde{N}$,

$$\lim_{n\to\infty} H_n(0) = H(0) = 0.$$

Thus, off the set with probability zero, Theorem 2.6.5 implies that for $n \in \tilde{N}$ and for $\Delta \in C = \Re$,

$$\lim_{n\to\infty} H_n(\Delta) = H(\Delta).$$

Furthermore, off the set with probability zero, Theorem 2.6.5 implies that for $n \in \tilde{N}$,

$$\lim_{n\to\infty} \sup_{|\Delta|\le c} |H_n(\Delta) - H(\Delta)| = 0,$$

where $c > 0$. Thus, for $n \in \tilde{N}$,

$$\sup_{|\Delta|\le c} |H_n(\Delta) - H(\Delta)| \xrightarrow{wp1} 0.$$

Since $\tilde{N} \subseteq N_0 \subseteq N$ and N_0 was arbitrary, Theorem 2.6.6 implies that for $n \in N$,

$$\sup_{|\Delta|\le c} |H_n(\Delta) - H(\Delta)| \xrightarrow{p} 0.$$

Since $H_n(\Delta) - H(\Delta) = D_n(\Delta) - Q_n(\Delta)$ we have shown that,

$$\sup_{|\Delta|\le c} |D_n(\Delta) - Q_n(\Delta)| \xrightarrow{p} 0.$$

This completes the proof of AL $\Rightarrow$ AUQ.

Next, we would like to show that (2.4) implies (2.3). Define N and N_0 as in the proof above. Now, (2.4) implies that for $n \in N$ and for $\Delta \in \Re$,

$$D_n(\Delta) \xrightarrow{p} Q_n(\Delta).$$

Now repeat the same "diagonal" argument as above except replace the linearity piece with this quadratic piece. In doing so, for an arbitrary infinant index set $N_0 \subseteq N$ we can construct another infinant index set $\tilde{N} \subseteq N_0 \subseteq N$ such that for $n \in \tilde{N}$ and $\Delta \in C'$,

$$D_n(\Delta) - Q_n(\Delta) \xrightarrow{wp1} 0.$$

Letting $H_n(\Delta)$ and $H(\Delta)$ be the same functions defined above we have for $n \in \tilde{N}$ and $\Delta \in C'$,

$$H_n(\Delta) \xrightarrow{wp1} H(\Delta).$$

That is, outside of a set with probability zero we have for $n \in \tilde{N}$ and $\Delta \in C'$,

$$\lim_{n\to\infty} H_n(\Delta) = H(\Delta).$$

Now, off this set of probability zero, Theorem 2.6.4 implies that for $n \in \tilde{N}$ and $\Delta \in C = \Re$,

$$\lim_{n\to\infty} \nabla H_n(\Delta) \quad = \quad \nabla H(\Delta).$$

Hence, for $n \in \tilde{N}$ and $\Delta \in C$ we have,

$$\nabla H_n(\Delta) \xrightarrow{wp1} \nabla H(\Delta).$$

Furthermore, Theorem 2.6.4 implies that for $c > 0$ and $n \in \tilde{N}$,

$$\sup_{|\Delta|\le c} |\nabla H_n(\Delta) - \nabla H(\Delta)| \xrightarrow{wp1} 0.$$

Since $\tilde{N} \subseteq N_0 \subseteq N$ and N_0 was arbitrary, Theorem 2.6.6 implies that for $n \in N$,

$$\sup_{|\Delta|\le c} |\nabla H_n(\Delta) - \nabla H(\Delta)| \xrightarrow{p} 0.$$

By definition of $H_n(\Delta)$ and $H(\Delta)$ we now have,

$$\begin{aligned}\nabla H_n(\Delta) - \nabla H(\Delta) &= \nabla D_n(\Delta) - \nabla Q_n(\Delta) \\ &= -\left[S_n(\Delta) - S_n(0) + 2\tau C\Delta\right].\end{aligned}$$

Thus, we have shown that,

$$\sup_{|\Delta|\le c} |S_n(\Delta) - S_n(0) + 2\tau C\Delta| \xrightarrow{p} 0.$$

This completes the proof of AUQ $\Rightarrow$ AUL.

Lastly, we need to show that (2.3) implies (2.2). However, since $c > 0$ in (2.3) is arbitrary, the proof is trivial. Hence, AUL $\Rightarrow$ AL. □

2.7 Asymptotic Normality of the Estimate

We are now ready to derive the asymptotic distribution of the estimate, $\hat{\rho}_n$. In this section it will be convenient to change notation by letting $\Delta = \sqrt{n}(\rho - \rho_0)$ where ρ_0 still represents the true parameter for the AR(1). Substituting this value of Δ into the functions defined in Section 2.6.3 and thinking of these functions as

functions of ρ instead of Δ we obtain the following,

$$D_n(\rho) = \frac{1}{n}D(\rho)$$

$$S_n(\rho) = \frac{1}{n^{\frac{3}{2}}}S(\rho)$$

$$Q_n(\rho) = D_n(\rho_0) - \sqrt{n}S_n(\rho_0)(\rho-\rho_0) + n\tau C(\rho-\rho_0)^2.$$

For the sake of reference the asymptotic uniform quadraticity result can be restated in terms of the above notation as,

$$\sup_{|\sqrt{n}(\rho-\rho_0)|\le c} |D_n(\rho) - Q_n(\rho)| \xrightarrow{P} 0, \tag{2.8}$$

where $c > 0$ is arbitrary.

Now, taking the derivative of $Q_n(\rho)$ with respect to ρ yields,

$$Q_n'(\rho) = -\sqrt{n}S_n(\rho_0) + 2n\tau C(\rho-\rho_0).$$

If $\tilde{\rho}_n$ is such that $Q_n'(\tilde{\rho}_n) = 0$ then $\tilde{\rho}_n$ denotes the value of ρ for which $Q_n(\rho)$ is minimized. It follows that,

$$\tilde{\rho}_n = \rho_0 + \frac{1}{2\sqrt{n}\tau}C^{-1}S_n(\rho_0).$$

Equivalently, $\tilde{\rho}_n$ is such that,

$$\sqrt{n}(\tilde{\rho}_n - \rho_0) = \frac{1}{2\tau}C^{-1}S_n(\rho_0). \tag{2.9}$$

Because $\tilde{\rho}_n$ depends on the true value of the process, $\tilde{\rho}_n$ is not a statistic. However, we can still derive its asymptotic distribution. In Section 2.4 we showed that,

$$S_n(\rho_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\eta_b^2\sigma_x^2}{3}\right). \tag{2.10}$$

It follows (see Serfling (1980, pg. 19)) from (2.9) and (2.10) that,

$$\sqrt{n}\,(\tilde{\rho}_n - \rho_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{12\tau^2} C^{-1}\left(\eta_b^2 \sigma_X^2\right) C^{-1}\right). \tag{2.11}$$

Since convergence in distribution implies bounded in probability (see Serfling (1980, pg. 52)), it follows that $\sqrt{n}\,(\tilde{\rho}_n - \rho_0) = O_p(1)$.

Now consider the estimate, $\hat{\rho}_n$. We want to determine the asymptotic distribution of $\sqrt{n}\,(\hat{\rho}_n - \rho_0)$. Adding in and subtracting out $\tilde{\rho}_n$ within this expression yields,

$$\sqrt{n}\,(\hat{\rho}_n - \rho_0) = \sqrt{n}\,(\hat{\rho}_n - \tilde{\rho}_n) + \sqrt{n}\,(\tilde{\rho}_n - \rho_0).$$

If we can show that $\sqrt{n}\,(\hat{\rho}_n - \tilde{\rho}_n) = o_p(1)$ then it will follow (see Serfling (1980, pg. 19)) from (2.11) that,

$$\sqrt{n}\,(\hat{\rho}_n - \rho_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{12\tau^2} C^{-1}\left(\eta_b^2 \sigma_X^2\right) C^{-1}\right).$$

The fact that $\sqrt{n}\,(\hat{\rho}_n - \tilde{\rho}_n) = o_p(1)$ is given in the following theorem. The proof of this theorem is due to Jaeckel (1972). Jaeckel first proved the result in the context of the linear regression model. However, upon examination of Jaeckel's proof, there is nothing unique about the linear regression model that is required. Thus, because of the linear structure of the AR(1), this result can be expected to hold. For the sake of completeness, we will present Jaeckel's result and proof in the context of the AR(1).

Theorem 2.7.1 *Under the conditions of Theorem 2.4.2 we have,*

$$\sqrt{n}\,(\hat{\rho}_n - \tilde{\rho}_n) = o_p(1).$$

proof. Choose $\varepsilon > 0$ and $\delta > 0$ and let $\tilde{\rho}_n = \min_\rho Q_n(\rho)$. By (2.11) we have that,

$$\sqrt{n}(\tilde{\rho}_n - \rho_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{12\tau^2}C^{-1}\left(\eta_b^2\sigma_X^2\right)C^{-1}\right).$$

Since convergence in distribution implies bounded in probability (Serfling, 1980, pg. 52), we can apply the definition of bounded in probability to the quantity $\frac{\delta}{2}$ and find a $N_1 = N_1(\delta)$ and B_δ such that,

$$P\left[\sqrt{n}\,|\tilde{\rho}_n - \rho_0| \geq B_\delta\right] \leq \frac{\delta}{2},$$

for $n \geq N_1$. If we let $E_1 = \{\omega : \sqrt{n}\,|\tilde{\rho}_n(\omega) - \rho_0| < B_\delta\}$ then for $n \geq N_1$ we have $P[E_1^c] \leq \frac{\delta}{2}$. Now define,

$$T_\varepsilon = \min\left\{Q_n(\rho) : \sqrt{n}\,|\tilde{\rho}_n - \rho| = \varepsilon\right\} - Q_n(\tilde{\rho}_n).$$

At first, it appears that T_ε depends on n. However, it can be shown that $T_\varepsilon = \tau C\varepsilon^2$ so that T_ε does not depend on n. For the multivariate case, $T_\varepsilon = \tau C^*\varepsilon^2$ where C^* is the minimum eigenvalue of the matrix C (Hettmansperger, 1984, pg. 278). Secondly, since $\tilde{\rho}_n$ is the unique minimum of $Q_n(\rho)$ and $\sqrt{n}\,|\tilde{\rho}_n - \rho| = \varepsilon$, we have that $T_\varepsilon > 0$. Since $T_\varepsilon > 0$ and is free of n we can apply the AUQ result given by (2.8) with $c = B_\delta + \varepsilon$ and $\frac{T_\varepsilon}{2}$ and find a $N_2 = N_2(\varepsilon, \delta)$ such that for all $n \geq N_2$ we have the following,

$$P\left[\sup_{\sqrt{n}|\rho-\rho_0|\leq B_\delta+\varepsilon} |D_n(\rho) - Q_n(\rho)| \geq \frac{T_\varepsilon}{2}\right] \leq \frac{\delta}{2}.$$

If we let

$$E_2 = \left\{\omega : \sup_{\sqrt{n}|\rho-\rho_0|\leq B_\delta+\varepsilon} |D_n(\omega;\rho) - Q_n(\omega;\rho)| < \frac{T_\varepsilon}{2}\right\},$$

then we have for $n \geq N_2$, $P[E_2^c] \leq \frac{\delta}{2}$. Now let $N^* = N^*(\varepsilon, \delta) = \max(N_1, N_2)$. For arbitrary sets A and B, standard probability arguments can be used to show $P[A \cap B] \geq 1 - P[A^c] - P[B^c]$. Thus, for $n > N^*$ we have $P[E_1 \cap E_2] \geq 1 - \delta$. That is, for $n > N^*$ we have the following,

$$P\Big[\Big\{\omega : \sqrt{n}\,|\tilde{\rho}_n(\omega) - \rho_0| < B_\delta\Big\} \cap \\ \Big\{\omega : \sup_{\sqrt{n}|\rho-\rho_0| \leq B_\delta + \varepsilon} |D_n(\omega;\rho) - Q_n(\omega;\rho)| < \frac{T_\varepsilon}{2}\Big\}\Big] \geq 1 - \delta. \qquad (2.12)$$

Next, define $P_{1\varepsilon} = \{\rho : \sqrt{n}\,|\tilde{\rho}_n - \rho| = \varepsilon\}$ and let $\rho \in P_{1\varepsilon}$ and $\omega \in E_1 \cap E_2$. Since $\omega \in E_1 \cap E_2$ it must be true that $\sqrt{n}\,|\tilde{\rho}_n(\omega) - \rho_0| < B_\delta < B_\delta + \varepsilon$. Hence,

$$|D_n(\omega; \tilde{\rho}_n(\omega)) - Q_n(\omega; \tilde{\rho}_n(\omega))| \;\leq\; \frac{T_\varepsilon}{2}.$$

Using just the upper bound on the above inequality we get,

$$D_n(\omega; \tilde{\rho}_n(\omega)) \;\leq\; Q_n(\omega; \tilde{\rho}_n(\omega)) + \frac{T_\varepsilon}{2}.$$

Next, since $\rho \in P_{1\varepsilon}$ and $\omega \in E_1 \cap E_2$ the following inequality is obtained,

$$\begin{aligned} \sqrt{n}\,|\rho - \rho_0| &= \sqrt{n}\,|\rho - \tilde{\rho}_n(\omega) + \tilde{\rho}_n(\omega) - \rho_0| \\ &\leq \sqrt{n}\,|\tilde{\rho}_n(\omega) - \rho_0| + \sqrt{n}\,|\tilde{\rho}_n(\omega) - \rho| \\ &\leq B_\delta + \varepsilon. \end{aligned}$$

Thus, from the E_2 part of (2.12) we have for $\omega \in E_1 \cap E_2$ and $\rho \in P_{1\varepsilon}$,

$$|D_n(\omega;\rho) - Q_n(\omega;\rho)| \;\leq\; \frac{T_\varepsilon}{2}.$$

Now use the lower bound of this inequality to get,

$$Q_n(\omega;\rho) - \frac{T_\varepsilon}{2} \le D_n(\omega;\rho).$$

Hence, $\rho \in P_{1\varepsilon}$ and $\omega \in E_1 \cap E_2$ imply $\omega \in F_1 \cap F_2$ where,

$$\begin{aligned} F_1 &= \left\{\omega : D_n(\omega;\tilde{\rho}_n(\omega)) \le Q_n(\omega;\tilde{\rho}_n(\omega)) + \frac{T_\varepsilon}{2}\right\} \text{ and} \\ F_2 &= \left\{\omega : D_n(\omega;\rho) \ge Q_n(\omega;\rho) - \frac{T_\varepsilon}{2}\right\}. \end{aligned}$$

Since it is true for arbitrary sets A and B that $A \subseteq B$ implies that $P[A] \le P[B]$ we must have that for $\rho \in P_{1\varepsilon}$ and $n > N^*$, $P[F_1 \cap F_2] \ge 1 - \delta$. Now, when $\omega \in F_1 \cap F_2$ the defintion of T_ε implies the following,

$$\begin{aligned} D_n(\omega;\rho) &\ge Q_n(\omega;\rho) - \frac{T_\varepsilon}{2} \\ &> \min_{\sqrt{n}|\tilde{\rho}_n - \rho|} Q_n(\omega;\rho) - \frac{T_\varepsilon}{2} \\ &= Q_n(\tilde{\rho}_n(\omega)) + T_\varepsilon - \frac{T_\varepsilon}{2} \\ &= Q_n(\tilde{\rho}_n(\omega)) + \frac{T_\varepsilon}{2} \\ &\ge D_n(\omega;\tilde{\rho}_n(\omega)). \end{aligned}$$

Thus, we see that $\rho \in P_{1\varepsilon}$ and $\omega \in F_1 \cap F_2$ imply $\omega \in \{\omega : D_n(\omega;\rho) > D_n(\omega;\tilde{\rho}_n(\omega))\}$. Thus, for $n > N^*$ and $\rho \in P_{1\varepsilon}$ we have,

$$P[\{\omega : D_n(\omega;\rho) > D_n(\omega;\tilde{\rho}_n(\omega))\}] \ge 1 - \delta.$$

Finally, let $P_{2\varepsilon} = \{\rho : \sqrt{n}\,|\tilde{\rho}_n - \rho| \ge \varepsilon\}$. If $\omega \in \{\omega : D_n(\omega;\rho) > D_n(\omega;\tilde{\rho}_n(\omega))\}$ and $\rho \in P_{2\varepsilon}$ then the convexity of $D_n(\rho)$ implies that for $n > N^*$,

$$P[\{\omega : D_n(\omega;\rho) > D_n(\omega;\tilde{\rho}_n(\omega))\}] \ge 1 - \delta.$$

That is, if $\rho \in P_{2\varepsilon}$ then $D_n(\rho) > D_n(\tilde{\rho}_n)$ with high probability. However, since $D_n(\tilde{\rho}_n) \geq \min_\rho D_n(\rho) = D_n(\hat{\rho}_n)$, $\hat{\rho}_n$ must be in $P_{2\varepsilon}^c$ with high probability. That is, for $n > N^*$,

$$P\left[\left\{\omega : \sqrt{n}\,|\tilde{\rho}_n(\omega) - \hat{\rho}_n(\omega)| < \varepsilon\right\}\right] \geq 1 - \delta.$$

This proves that $\sqrt{n}\,(\hat{\rho}_n - \tilde{\rho}_n) = o_p(1)$. □

CHAPTER III

ESTIMATION OF $\boldsymbol{\rho}$ USING RANDOM WEIGHTS

3.1 Defining the Estimate

Consider the AR(p) given by (1.1) and (1.2). That is, assume the observations of the process can be modeled as,

$$
\begin{aligned}
X_i &= \alpha + \rho_1 X_{i-1} + \rho_2 X_{i-2} + \ldots + \rho_p X_{i-p} + \varepsilon_i \\
&= \alpha + \boldsymbol{Y}'_{i-1}\boldsymbol{\rho} + \varepsilon_i; \quad i = 1, 2, \ldots n,
\end{aligned}
$$

where $p \geq 1$, $\boldsymbol{Y}'_{i-1} = (X_{i-1}, X_{i-2}, \ldots, X_{i-p})$, $\boldsymbol{\rho}' = (\rho_1, \rho_2, \ldots, \rho_p)$, and $\boldsymbol{Y}_0$ is an observable random vector independent of $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$. Additionally, assume the ε_i are iid F and the solutions to the following equation,

$$
X^p - \rho_1 X^{p-1} - \rho_2 X^{p-2} - \ldots - \rho_p = 0,
$$

lie in the interval $(-1, 1)$. This condition is equivalent to assuming the process is stationary.

Following section 1.2 the proposed estimate of $\boldsymbol{\rho}$ will be the value that minimizes the dispersion function,

$$
\begin{aligned}
D(\boldsymbol{\rho}) &= \sum_{1 \leq i < j \leq n} b_{ij} \left| \varepsilon_i - \varepsilon_j \right| \\
&= \sum_{1 \leq i < j \leq n} b_{ij} \left| (X_i - X_j) - (\boldsymbol{Y}_{i-1} - \boldsymbol{Y}_{j-1})'\boldsymbol{\rho} \right|.
\end{aligned}
$$

Again, the b_{ij} denote weights used for the $(i,j)^{th}$ comparison. Alternatively, one can view the estimate of $\boldsymbol{\rho}$ as an approximate solution of the equation $\boldsymbol{S}(\boldsymbol{\rho}) = \mathbf{0}$ where,

$$\begin{aligned} \boldsymbol{S}(\boldsymbol{\rho}) &= -\boldsymbol{\nabla} D(\boldsymbol{\rho}) \\ &= 2 \sum_{1 \le i < j \le n} b_{ij} \left(\boldsymbol{Y}_{j-1} - \boldsymbol{Y}_{i-1}\right) \left(\varphi(\varepsilon_i(\boldsymbol{\rho}), \varepsilon_j(\boldsymbol{\rho})) - \frac{1}{2}\right), \end{aligned}$$

and $\varphi(u,v) = \frac{\operatorname{sgn}(v-u)+1}{2} \doteq I(u \le v)$ except at those points where the gradient does not exist. The equation is approximate since $\boldsymbol{S}(\boldsymbol{\rho})$ is a simple function that changes values whenever $\boldsymbol{\rho}$ crosses one of the $\frac{n(n-1)}{2}$ hyperplanes (Naranjo, 1989),

$$H_{ij} = \{\boldsymbol{\theta} \in \Re^p : X_i - X_j = (\boldsymbol{Y}_{i-1} - \boldsymbol{Y}_{j-1})'\boldsymbol{\theta}\}.$$

In this chapter the weights will be considered random. For instance, the b_{ij}'s are random when $b_{ij} = b\left(\boldsymbol{Y}_{i-1}, \boldsymbol{Y}_{j-1}\right)$. Random weights are useful in situations when one desires the data to determine the weights. This may occur in problems where the observed time series is very large and or there is no a-priori knowledge of outlying observations. Typically, weighting schemes are chosen such that observations that deviate from the "center" of the data are downweighted while those observations that are consistent with the bulk of the data are left alone. Thus, the weighting function usually depends implicitly on some robust measure of location and scatter.

3.2 Assumptions for the Asymptotic Theory

We begin this section by stating a list of assumptions that will be needed for the asymptotic theory of the estimate. Assumptions denoted with a "M" represent conditions pertaining to the model, assumptions denoted with an "E" represent conditions on the error distribution, and assumptions denoted with a "W" represent constraints on the weights.

M1. $X_i = \alpha + \rho_1 X_{i-1} + \rho_2 X_{i-2} + \ldots + \rho_p X_{i-p} + \varepsilon_i$ where $i = 1, 2, \ldots, n$ and $p \geq 1$

M2. The roots of $X^p - \rho_1 X^{p-1} - \rho_2 X^{p-2} - \ldots - \rho_p = 0$ are in $(-1, 1)$

M3. $\boldsymbol{Y}_0$ is an observable random vector independent of $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ and is such that $E\left[||\boldsymbol{Y}_0||^2\right] < \infty$.

E1. $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are iid F random variables with $E[\varepsilon_1] = 0$ and $E[\varepsilon_1^2] < \infty$

E2. The density function, f, of F is such that f is absolutely continuous, $f > 0$ a.e., and f has finite Fisher Information

W1. $b_{ij} = h\left(\boldsymbol{Y}_{i-1}; \hat{\boldsymbol{\theta}}_n\right) h\left(\boldsymbol{Y}_{j-1}; \hat{\boldsymbol{\theta}}_n\right)$ where $\hat{\boldsymbol{\theta}}_n$ represents an estimated parameter vector that contains the elements of $\hat{\boldsymbol{\mu}}_n$ and the upper triangular portion of $\hat{\boldsymbol{\Sigma}}_n$.

W2. $\hat{\boldsymbol{\mu}}_n$ and $\hat{\boldsymbol{\Sigma}}_n$ are robust measures of location and scatter for $\{\boldsymbol{Y}_{i-1}\}$ such that $\sqrt{n}\,(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}) = \boldsymbol{O}_p(1)$ and $\sqrt{n}\left(\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}\right) = \boldsymbol{O}_p(1)$ for some $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

W3. h is twice differentiable

W4. $\left\|\boldsymbol{D}^{(k)}h\,(\boldsymbol{Y};\cdot)\right\| \leq B_1; \quad k = 0, 1$

W5. $\left\|\boldsymbol{D}^{(k)}h\,(\boldsymbol{Y};\cdot)\,Y_l\right\| \leq B_2; \quad k = 0, 1, 2; \;\; l = 1, 2, \ldots, p$

W6. $E\left[h\,(\boldsymbol{Y};\cdot)\,\boldsymbol{Y}\right] = \boldsymbol{0}$

W7. The number of elements in $\{h\,(\boldsymbol{Y}_0), h\,(\boldsymbol{Y}_1), \ldots, h\,(\boldsymbol{Y}_{n-1})\}$ such that $h\,(\boldsymbol{Y}_i) \neq K$ is at most $[n^r]$ where $r < \frac{1}{2}$

We now consider some comments concerning the above assumptions. First, under assumptions M1 and E1 it is well-known (e.g. Abraham and Ledolter (1983, pg. 207)) that the mean of the process is $\mu_X = E\left[X_1\right] = \frac{\alpha}{1-\rho_1-\rho_2-\ldots-\rho_p}$ and the variance of the process is $\sigma_X^2 = V\left[X_1\right] = \frac{\sigma_\varepsilon^2}{1-c_1\rho_1-c_2\rho_2-\cdots-c_p\rho_p} < \infty$ where c_i denotes the lag i correlation coefficient between X_j and X_{j-i}. If we define the "centered" process as $X_i^c = X_i - \mu_X$ and rewrite the model in terms of the centered process we get the following,

$$\begin{aligned}
X_i^c + \mu_X &= \alpha + \rho_1(X_{i-1}^c + \mu_X) + \rho_2(X_{i-2}^c + \mu_X) + \ldots + \rho_p(X_{i-p}^c + \mu_X) + \varepsilon_i \\
&= \alpha + (\rho_1 + \rho_2 + \ldots + \rho_p)\left(\frac{\alpha}{1 - \rho_1 - \rho_2 - \ldots - \rho_p}\right) + \\
&\quad\; \rho_1 X_{i-1}^c + \rho_2 X_{i-2}^c + \ldots + \rho_p X_{i-p}^c + \varepsilon_i
\end{aligned}$$

$$= \frac{\alpha}{1-\rho_1-\rho_2-\ldots-\rho_p} + \rho_1 X_{i-1}^c + \rho_2 X_{i-2}^c + \ldots + \rho_p X_{i-p}^c + \varepsilon_i.$$

Hence, rewriting M1 in terms of the centered process yields,

$$\begin{aligned} X_i^c &= \rho_1 X_{i-1}^c + \rho_2 X_{i-2}^c + \ldots + \rho_p X_{i-p}^c + \varepsilon_i \\ &= \boldsymbol{\rho}' \boldsymbol{Y}_{i-1}^c + \varepsilon_i. \end{aligned} \tag{3.1}$$

Note that the $\boldsymbol{\rho}$ in the centered model is the same $\boldsymbol{\rho}$ that appears in the non-centered model. Thus, since we are primarily interested in the estimation of $\boldsymbol{\rho}$, one can assume without loss of generality that the process has a zero mean. That is, without loss of generality, one can assume $E[X_1] = 0$. For convenience in subsequent discussions we will drop the X_i^c and $\boldsymbol{Y}_{i-1}^c$ notation and just write X_i and $\boldsymbol{Y}_{i-1}$ (keeping in mind that $E[X_1] = 0$).

Although the estimation of $\boldsymbol{\rho}$ is our primary concern, a couple of comments concerning the estimation of α are in order. In practice, one should first center the data with an unbiased robust estimate of location, $\widehat{\mu}_X$. Then, using the proposed estimate, one can fit (3.1) to obtain a $\widehat{\boldsymbol{\rho}}_n$. Once an estimate of $\boldsymbol{\rho}$ has been determined define the residuals as $\widehat{\varepsilon}_i = X_i - \widehat{\boldsymbol{\rho}}_n' \boldsymbol{Y}_{i-1}$. Since M1 implies $X_i - \boldsymbol{\rho}' \boldsymbol{Y}_{i-1} = \alpha + \varepsilon_i$, one can fit the model $\widehat{\varepsilon}_i = \alpha + \varepsilon_i$ using a robust estimate of location to obtain an $\widehat{\alpha}_n$. Since $\mu_X = \frac{\alpha}{1-\rho_1-\rho_2-\ldots-\rho_p}$, one may also consider $\widehat{\alpha}_n = \widehat{\mu}_X (1 - \widehat{\rho}_{n1} - \widehat{\rho}_{n2} - \ldots - \widehat{\rho}_{np})$ as an alternative estimate.

Now consider the conditions on the weights. Let $\boldsymbol{\theta}$ denote a vector analogous to $\widehat{\boldsymbol{\theta}}_n$ that contains the corresponding elements of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Then, assump-

tions W1, W2, W3, and W4 can be used to show the following,

$$\begin{aligned}\left|b_{ij}\left(\boldsymbol{Y}_{i-1},\boldsymbol{Y}_{j-1};\widehat{\boldsymbol{\theta}}_n\right)-b_{ij}\left(\boldsymbol{Y}_{i-1},\boldsymbol{Y}_{j-1};\boldsymbol{\theta}\right)\right| &\leq K\left\|\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}\right\| \\ &= O_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

where the convergence is uniform in (i, j) and $K = 2B_1^2$. This will essentially allow us to assume that the weights are only functions of $\boldsymbol{Y}_{i-1}$ and $\boldsymbol{Y}_{j-1}$, a condition that is needed in order to apply a theorem given by Koul (1991). Secondly, since h is twice differentiable, it follows that both h and $\boldsymbol{D}h$ are continuous and differentiable. That is, h and $\boldsymbol{D}h$ are both "smooth". Now consider the following function, $\varphi(\boldsymbol{Y}) = h(\boldsymbol{Y})\boldsymbol{\lambda}'\boldsymbol{Y}$. This function will play a critical role in proving the main result of this section. It follows from W3 that $\boldsymbol{D}^2\varphi$ exists so that φ and $\boldsymbol{D}\varphi$ are both "smooth". Next, the bounds given in W4 and W5 imply that φ, $\boldsymbol{D}\varphi$, and $\boldsymbol{D}^2\varphi$ are all bounded. As we shall also see, some of these bounds are required in order to apply a result by Koul (1991). In addition, Naranjo and Hettmansperger (1994) use W5 to obtain a bounded influence function in the classic linear regression model. Although the influence function will not be considered in this paper, it is conjectured that the influence function for the proposed estimate and model will be bounded under assumption W5. Now consider the expectation given in W6. This assumption was also used by Naranjo and Hettmansperger (1994) in their proof of the influence function. It is implied if $\boldsymbol{Y}$ has a symmetric distribution and h is an even function. Alternatively, it is implied if h is constant and $\boldsymbol{Y}$ has mean zero. In what follows $\boldsymbol{Y}$ will typically

denote a value of $\boldsymbol{Y}_{i-1}$ and or a $p \times 1$ vector of ε_i's. Hence, a symmetric distribution for ε_1 and an even h implies W6. Lastly, W7 implies that only a certain number of $h(\boldsymbol{Y}_{i-1};\cdot)$'s are allowed to deviate from K. Although this assumption may be argued from a practical point of view, it is somewhat "stringent" since it depends on the observed data. For instance, if one designs h in such a way that $h(\boldsymbol{Y}_{i-1};\cdot) \neq K$ when $\boldsymbol{Y}_{i-1}$ is say six standard deviations from the "center" then this assumption is probably realistic. It should be noted that this assumption is only used in an alternative, much simpler, proof of the asymptotic normality of the estimate. The smoothness assumptions and bounds given in W3–W6 are preferred over W7.

Now consider the functional form of h. Typically h is defined as,

$$h(\boldsymbol{Y}) = h^*(\|\boldsymbol{Y}\|_s),$$

where $\|\boldsymbol{Y}\|_s = (\boldsymbol{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})$ denotes a "statistical" distance. Examples of h^*'s that satisfy the above assumptions are given by,

$$h^*(y) = \frac{a}{b + d\left(\frac{y}{c}\right)^m} \quad \text{and}$$

$$h^*(y) = \frac{a}{be^{d\left(\frac{y}{c}\right)^m} + be^{-d\left(\frac{y}{c}\right)^m}},$$

where a, b, c, d, and m are suitably chosen constants. It should be noted that these choices for h^* yield "smooth" analogs to the h used by Naranjo and Hettmansperger (1994).

Finally, E2 implies that f is uniformly bounded, uniformly continuous, and $\tau = \int_{-\infty}^{\infty} f^2(t)dt$ is finite (see Lemma 2.2.1). Additionally, M2, M3, and E1 imply that any finite linear combination of the X's has a finite second moment. As a special case, we have $E\left[X_i^2\right] = E\left[X_1^2\right] < \infty$.

3.3 Some Preliminaries

3.3.1 Ergodic Theory

In order to prove the asymptotic normality of the estimate it will be necessary to invoke a "law of large number" type theorem. Since the sums we will be dealing with do not contain independent terms the classical "law of large number" theorems are not applicable. However, a more general theorem called the Ergodic Theorem, can be used to handle such situations. Therefore, in this section we briefly state some definitions and theorems (without proof) pertaining to Ergodic Theory. The following overview is a selected summary from Sections 24 and 36 of Billingsley (1995).

Let us begin by stating some definitions. The first definition can be found on page 311 of Billingsley (1995).

Definition 3.3.1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A mapping $T : \Omega \longrightarrow \Omega$ is a measure-preserving transformation if for each $F \in \mathcal{F}$, $T^{-1}(F) \in \mathcal{F}$ and $P\left[T^{-1}(F)\right] = P\left[F\right]$.*

It should be noted that if T is measure-preserving then it follows from the defini-

tion that $P[T^{-n}(F)] = P[F]$ for $n \geq 0$. Furthermore, if T is invertible then T is measure-preserving. The next two definitions can be found on pages 313 and 314 of Billingsley (1995).

Definition 3.3.2 *The $\mathcal{F}$-set F is invariant under T if $T^{-1}(F) = F$. Furthermore, it is a non-trivial invariant set if $0 < P[F] < 1$.*

Definition 3.3.3 *T is ergodic if there are no non-trivial invariant sets in $\mathcal{F}$.*

The last definition can be found on page 314 of Billingsley (1995).

Definition 3.3.4 *A measurable function g is invariant if $g(T(\omega)) = g(\omega)$ for all $\omega \in \Omega$.*

We now have the terminology to state the relevant theorems. In what follows let $\xi_t = (\ldots, \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \ldots)$ denote a sequence whose origin is at the position occupied by ε_t. Furthermore, let $\varphi : \Re^\infty \longrightarrow \Re$ be a measurable function. Then we have the following theorem, which is given as Theorem 36.4 on page 495 of Billingsley (1995).

Theorem 3.3.1 *If $\xi_0 = (\ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots)$ is stationary and ergodic, in particular if the ε_t are iid F, and $W_t = \varphi(\xi_t)$ then $\chi_0 = (\ldots, W_{-1}, W_0, W_1, \ldots)$ is stationary and ergodic.*

proof. A discussion of the proof can be found on page 495 of Billingsley (1995).

□

As an application of this theorem consider the AR(p) given by M1, M2, and M3. Since the ε_i are assumed iid F, $\xi_t = (\ldots, \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \ldots)$ is stationary and ergodic. Now, it is a well-known fact (Fuller, 1996, pg. 59–61) that the observations of an AR(p) can be written as,

$$X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k},$$

where $\{\psi_k\}$ are real numbers that satisfy $\psi_0 = 1$ and $\sum_{k=0}^{\infty} |\psi_k| < \infty$. If we let B_{-k} denote the backwards shift operator (i.e. $B_{-k}(\xi_0) = \varepsilon_{-k}$) and define $\varphi(\xi) = \sum_{k=0}^{\infty} \psi_k B_{-k}(\xi)$ then it follows that,

$$\begin{aligned} X_t &= \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \\ &= \sum_{k=0}^{\infty} \psi_k B_{-k}(\xi_t) \\ &= \varphi(\xi_t). \end{aligned}$$

Thus, Theorem 3.3.1 implies that an AR(p) is stationary and ergodic.

Now consider another measurable function, $\phi : \Re^{\infty} \longrightarrow \Re$. If Z_t is such that,

$$\begin{aligned} Z_t &= \phi(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots) \\ &= \phi(\chi_t), \end{aligned}$$

then Theorem 3.3.1 implies that $\{Z_t\}$ is stationary and ergodic. Hence, we see that a sequence of random variables defined in terms of the AR(p) process is again stationary and ergodic. This is the result that essentially allows us to apply the Ergodic Theorem to various sums whose terms are functions of the AR(p) process.

The Ergodic Theorem is stated as Theorem 24.1 on page 314 of Billingsley (1995).

Theorem 3.3.2 *Suppose T is a measure-preserving transformation on $(\Omega, \mathcal{F}, P)$ and g is measurable and integrable. Then,*

$$\frac{1}{n}\sum_{k=1}^{n} g\left(T^{k-1}\omega\right) \xrightarrow{wp1} \widehat{g}(\omega)$$

where $\widehat{g}$ is invariant and integrable and $E[\widehat{g}] = E[g]$. If T is ergodic, then $\widehat{g} = E[g]$ with probability one.

proof. The proof is given on pages 317-319 of Billingsley (1995). □

3.3.2 Useful Lemmas

We now present some lemmas that will be critical to the proof of the asymptotic linearity result. In what follows define $\varepsilon_i(\boldsymbol{\rho}) = X_i - \boldsymbol{\rho}'\boldsymbol{Y}_{i-1}$. If $\boldsymbol{\rho}_0$ denotes the true parameter vector then $\varepsilon_i(\boldsymbol{\rho}_0) = \varepsilon_i$. Now define the following function,

$$W_\gamma(x, \boldsymbol{\rho}) = \frac{1}{n}\sum_{i=1}^{n} \gamma(\boldsymbol{Y}_{i-1}) I(\varepsilon_i(\boldsymbol{\rho}) \leq x),$$

where $\gamma : \Re^p \longrightarrow \Re$. W_γ is essentially a randomly weighted empirical distribution function. The following theorem is given in Koul (1991).

Theorem 3.3.3 *Assume the following conditions hold,*

1. $X_i = \boldsymbol{\rho}'\boldsymbol{Y}_{i-1} + \varepsilon_i, \quad 1 \leq i \leq n$

2. *γ is a bounded function*

3. $\frac{1}{\sqrt{n}} \max_{1 \le i \le n} \|\boldsymbol{Y}_{i-1}\| = o_p(1)$

4. $\frac{1}{n} \sum_{i=1}^{n} \|\gamma(\boldsymbol{Y}_{i-1}) \boldsymbol{Y}_{i-1}\| = O_p(1)$

5. *F has a uniformly continuous density f with $f > 0$ a.e.*

Then, for any $0 < b < \infty$,

$$\begin{aligned} \sup_{x \in \Re, \|\boldsymbol{\Delta}\| \le b} \Big| \sqrt{n} W_\gamma \left(x, \boldsymbol{\rho}_0 + \tfrac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - \sqrt{n} W_\gamma (x, \boldsymbol{\rho}_0) & \\ - \boldsymbol{\Delta}' \left(\tfrac{1}{n} \textstyle\sum_{i=1}^{n} \gamma(\boldsymbol{Y}_{i-1}) \boldsymbol{Y}_{i-1}\right) f(x) \Big| &= o_p(1). \end{aligned}$$

proof. The proof is given in Section 2.2 of Koul (1991). □

A few comments concerning the assumptions of this theorem are in order. Under assumptions M1-M3 the model assumption (assumption 1) is trivial. In our applications of this theorem, γ will typically be defined as $\gamma(\boldsymbol{Y}_{i-1}) = h(\boldsymbol{Y}_{i-1})$ and or $\gamma(\boldsymbol{Y}_{i-1}) = h(\boldsymbol{Y}_{i-1}) X_{i-k}$, $k = 1, 2, \ldots, p$. Thus, assumptions W4 and W5 imply that γ will always be a bounded function. Now, using Lemma 2.3.3 one can show the following,

$$\begin{aligned} \frac{1}{\sqrt{n}} \max_{1 \le i \le n} \|\boldsymbol{Y}_{i-1}\| &\le \sqrt{p} \frac{1}{\sqrt{n}} \max_{1-p \le i \le n} |X_i| \\ &= o_p(1). \end{aligned}$$

Hence, the third condition is satisfied. When $\gamma(\boldsymbol{Y}_{i-1}) = h(\boldsymbol{Y}_{i-1})$ W4 implies the fourth condition. Furthermore, when $\gamma(\boldsymbol{Y}_{i-1}) = h(\boldsymbol{Y}_{i-1}) X_{i-k}$ W5, the finite

second moment assumptions, and the Ergodic Theorem imply the following,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\|\gamma(\boldsymbol{Y}_{i-1})\boldsymbol{Y}_{i-1}\| &= \frac{1}{n}\sum_{i=1}^{n}|h(\boldsymbol{Y}_{i-1})X_{i-k}|\,\|\boldsymbol{Y}_{i-1}\| \\
&\leq \frac{B_2}{n}\sum_{i=1}^{n}\|\boldsymbol{Y}_{i-1}\| \\
&= B_2 E\left[\|\boldsymbol{Y}_0\|\right] + o_p(1) \\
&= O_p(1).
\end{aligned}
$$

Therefore, in both cases the fourth condition of the theorem is satisfied. Lastly, E2 and Lemma 2.2.1 imply that f is uniformly continuous. Thus, all of the conditions of the theorem are implied by the assumptions given in Section 3.2. Therefore, we will be able to apply this theorem when applicable.

The second result is a direct consequence of Lemma 2.5.3 and is thus given as a corollary.

Corollary 3.3.1 *Under assumption E2 and the finite second moment assumptions we have for all $\boldsymbol{\Delta} \in \Re^p$,*

$$\max_{1\leq i\leq n}\left|f\left(\varepsilon_i - \frac{1}{\sqrt{n}}\boldsymbol{\Delta}'\boldsymbol{Y}_{i-1}\right) - f(\varepsilon_i)\right| = o_p(1).$$

proof. Let $\xi_{ni} = \frac{1}{\sqrt{n}}\boldsymbol{\Delta}'\boldsymbol{Y}_{i-1}$. Then Lemma 2.3.3 implies $\max_{1\leq i\leq n}|\xi_{ni}| = o_p(1)$. Since E2 and Lemma 2.2.1 imply f is uniformly continuous, Lemma 2.5.3 implies the result. □

3.4 The Asymptotic Distribution of the Gradient

The purpose of this section is to derive the asymptotic distribution of the gradient, $\boldsymbol{S}_n(\boldsymbol{\rho}_0)$, where $\boldsymbol{\rho}_0$ denotes the true autoregressive parameter vector. As was done in Section 2.4, martingales will be used to prove the result. However, before we begin we will state some notation and lemmas that will be used throughout the sequel.

3.4.1 Simplifying Notation

Some notation that will be used throughout the remainder of this section is stated here for the sake of a convenient reference.

$$\begin{aligned}
\boldsymbol{P} &= \begin{pmatrix} \rho_1 & \rho_2 & \cdots & \rho_{p-1} & \rho_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}, \quad \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_i \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\
h_{i-1} &= h(\boldsymbol{Y}_{i-1}) \\
\varphi_{i-1} &= \varphi(\boldsymbol{Y}_{i-1}) = h(\boldsymbol{Y}_{i-1})\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1} \\
\psi_{ij} &= I(\varepsilon_j \le \varepsilon_i) - F(\varepsilon_i) \\
\boldsymbol{\Gamma}_i^j &= \sum_{r=0}^{j-i} \boldsymbol{P}^r \boldsymbol{\varepsilon}_{j-r} \text{ where } i \le j \\
\boldsymbol{c}_i &= c(\boldsymbol{Y}_{i-1}) = \boldsymbol{P}^{j-i}\boldsymbol{Y}_{i-1} + \boldsymbol{P}^{j-i-1}\boldsymbol{\varepsilon}_i \text{ where } i < j
\end{aligned}$$

3.4.2 Preliminary Results

We begin by proving two lemmas pertaining to the matrix $\boldsymbol{P}$. These lemmas will be used (sometimes without mention) throughout the proof of the main result of this section.

Lemma 3.4.1 *Let λ_i, $i = 1, 2, \ldots, p$, denote the eigenvalues of $\boldsymbol{P}$ and define $\lambda_M = \max_i |\lambda_i|$. Then, $\lambda_M < 1$.*

proof. Direct computations can be used to show the following (e.g. Fuller (1996, pg. 60)),

$$\det(\lambda \boldsymbol{I} - \boldsymbol{P}) = \left(\lambda^p - \rho_1 \lambda^{p-1} - \cdots - \rho_{p-1}\lambda - \rho_p\right)(-1)^{2p-1}.$$

Equating the above expression to zero and solving for λ yields the eigenvalues of $\boldsymbol{P}$. However, the roots to the following equation,

$$X^p - \rho_1 X^{p-1} - \cdots - \rho_{p-1} X - \rho_p = 0,$$

lie in $(-1, 1)$ by assumption. Thus, it follows that $\lambda_M < 1$. □

Lemma 3.4.2 *Let $k \geq p$ and define $P_{ij}^{(k)}$ to be the (i, j) entry of $\boldsymbol{P}^k$. Then there exists a K such that $\left|P_{ij}^{(k)}\right| \leq K\lambda_M^k$ for $1 \leq i, j \leq p$.*

proof. Note that $\boldsymbol{P}$ is a $p \times p$ matrix. Thus, by Schur's Theorem (Goldberg (1991, pg. 292)) there exists a unitary matrix $\boldsymbol{U}$ and upper triangular matrix $\boldsymbol{T}$ such that $\boldsymbol{P} = \boldsymbol{U}\boldsymbol{T}\boldsymbol{U}^{H}$. Furthermore, it follows (e.g. Goldberg (1991, pg. 294)) that the diagonal elements of $\boldsymbol{T}$ are the eigenvalues of $\boldsymbol{P}$. It follows immediately from

the above equality that $\boldsymbol{P}^k = \boldsymbol{U}\boldsymbol{T}^k\boldsymbol{U}^{\mathrm{H}}$. Now let u_{ij}, $t_{ij}^{(k)}$, and u_{ij}^{H} denote the (i,j) entries of $\boldsymbol{U}$, $\boldsymbol{T}$, and $\boldsymbol{U}^{\mathrm{H}}$ respectively. Then, the (i,j) entry of $\boldsymbol{P}^k$ can be written as,

$$P_{ij}^{(k)} = \sum_{a=1}^{p}\sum_{b=1}^{p} u_{ia} t_{ab}^{(k)} u_{bj}^{\mathrm{H}}.$$

Now let $C_{ij} = \max_{a,b}\left|u_{ia}u_{bj}^{\mathrm{H}}\right|$ and $t_{\mathrm{M}} = \max_{a,b}\left|t_{ab}^{(k)}\right|$. Thus, it follows that,

$$\left|P_{ij}^{(k)}\right| \leq p^2 C_{ij} t_{\mathrm{M}}.$$

Next, consider the matrix $\boldsymbol{T}$ and a related matrix $\boldsymbol{S}(\delta)$ where $\delta \geq 0$,

$$\boldsymbol{T} = \begin{pmatrix} \lambda_1 & t_{12} & t_{13} & \dots & t_{1p} \\ 0 & \lambda_2 & t_{23} & \dots & t_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & t_{p-1,p} \\ 0 & 0 & 0 & \dots & \lambda_p \end{pmatrix}, \quad \boldsymbol{S}(\delta) = \begin{pmatrix} \delta & |t_{12}| & |t_{13}| & \dots & |t_{1p}| \\ 0 & \delta & |t_{23}| & \dots & |t_{2p}| \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & |t_{p-1,p}| \\ 0 & 0 & 0 & \dots & \delta \end{pmatrix}$$

respectively. It follows (e.g. Goldberg (1991, pg. 340)) that for $k \geq p$,

$$\boldsymbol{S}^k(\delta) = \delta^k \boldsymbol{I} + \sum_{a=1}^{p-1} C_a^k \delta^{k-a} \boldsymbol{S}^a(0),$$

where C_a^k represents the usual binomial coefficient. Furthermore, it can be shown (e.g. Goldberg (1991, pg. 341)) that,

$$\left|t_{ij}^{(k)}\right| \leq s_{ij}^{(k)},$$

where $t_{ij}^{(k)}$ and $s_{ij}^{(k)}$ denote the (i,j) entries of $\boldsymbol{T}^k$ and $\boldsymbol{S}^k(\delta)$ respectively. Let us now consider the (i,j) entry of $\boldsymbol{S}^k(\delta)$,

$$s_{ij}^{(k)} = B_0\delta^k + \sum_{a=1}^{p-1} C_a^k \delta^{k-a} s_{ij}^{(a)}(0)$$

$$= B_0\delta^k + C_1^k s_{ij}^{(1)}(0)\delta^{-1}\delta^k + \cdots + C_{p-1}^k s_{ij}^{(p-1)}(0)\delta^{-(p-1)}\delta^k,$$

where $B_0 = 0, 1$ according to $i \neq j$ or $i = j$. Next, consider the term $\delta^{-a}s_{ij}^{(a)}(0)C_a^k\delta^k$. Assuming $\delta < 1$ L'Hôpitals rule can be used to show that $C_a^k\delta^k \longrightarrow 0$ as $k \longrightarrow \infty$. Thus, there exists a B_{ija} such that $\left|\delta^{-a}s_{ij}^{(a)}(0)C_a^k\delta^k\right| < B_{ija}$. Now let $\delta = \frac{b}{\lambda_M}$ where $0 < b < \lambda_M < 1$ and λ_M represents the maximum (in absolute value) of the eigenvalues of $\boldsymbol{P}$. Recall that Lemma 3.4.1 implies that $\lambda_M < 1$. Hence, it follows that,

$$\left|\delta^{-a}s_{ij}^{(a)}(0)C_a^k\delta^k\right| \leq B_{ija}\lambda_M^k.$$

Let us now return to the (i, j) entry of $\boldsymbol{S}^k(\delta)$,

$$\begin{aligned}
\left|s_{ij}^{(k)}\right| &= \left|B_0\delta^k + C_1^k s_{ij}^{(1)}(0)\delta^{-1}\delta^k + \cdots + C_{p-1}^k s_{ij}^{(p-1)}(0)\delta^{-(p-1)}\delta^k\right| \\
&\leq \left|B_0\delta^k\right| + \left|C_1^k s_{ij}^{(1)}(0)\delta^{-1}\delta^k\right| + \cdots + \left|C_{p-1}^k s_{ij}^{(p-1)}(0)\delta^{-(p-1)}\delta^k\right| \\
&\leq \left(B_0 + B_{ij1} + \cdots + B_{ij,p-1}\right)\lambda_M^k.
\end{aligned}$$

It now follows that,

$$\left|t_{ij}^{(k)}\right| \leq \left|s_{ij}^{(k)}\right| \leq B_{ij}\lambda_M^k \text{ say.}$$

Now define $B = \max_{ij} B_{ij}$ and $C = \max_{ij} C_{ij}$. We then have the following,

$$\begin{aligned}
\left|P_{ij}^{(k)}\right| &\leq p^2 C_{ij} t_M \\
&\leq p^2 C B \lambda_M^k \\
&\leq K\lambda_M^k \text{ say.}
\end{aligned}$$

This completes the proof. □

Now let $k+1 \leq l-1$ and consider the random variable $\varphi(\boldsymbol{Y}_{l-1})$. Exploiting the iterative nature of the AR(p) and applying a multivariate version of the Mean Value Theorem (e.g. Theorem 40.4 of Bartle (1964, pg. 365)) yields the following,

$$\begin{aligned}
\varphi(\boldsymbol{Y}_{l-1}) &= \varphi\left(\boldsymbol{P}^{l-k}\boldsymbol{Y}_{k-1} + \boldsymbol{\Gamma}_k^{l-1}\right) \\
&= \varphi\left(\boldsymbol{P}^{l-k}\boldsymbol{Y}_{k-1} + \boldsymbol{P}^{l-k-1}\boldsymbol{\varepsilon}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right) \\
&= \varphi\left(\mathbf{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right) \quad \text{say} \\
&= \varphi\left(\boldsymbol{\Gamma}_{k+1}^{l-1}\right) + \boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})'\boldsymbol{c}_k,
\end{aligned}$$

where $\boldsymbol{\xi}_{kl}$ lies on the line segment connecting $\boldsymbol{Y}_{l-1} = \boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}$ and $\boldsymbol{\Gamma}_{k+1}^{l-1}$. That is, $\boldsymbol{\xi}_{kl} = \mathbf{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1} - \lambda \mathbf{c}_k$ for some $\lambda \in (0,1)$. It will be helpful to view $\boldsymbol{\xi}_{kl}$ as $\boldsymbol{\xi}_{kl} = \boldsymbol{\xi}_{kl}\left(\boldsymbol{Y}_{k-1}; \boldsymbol{\Gamma}_k^{l-1}\right)$. Hence, we will view $\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})$ as a function of $\boldsymbol{Y}_{k-1}$. In doing so, consider the point point where $\boldsymbol{Y}_{k-1} = -\boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_k$. When $\boldsymbol{Y}_{k-1}$ attains this specific value $\boldsymbol{c}_k = \mathbf{0}$. Recall that $\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})$ is motivated from an application of the Mean Value Theorem on φ and the points $\boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}$ and $\boldsymbol{\Gamma}_{k+1}^{l-1}$. However, when $\boldsymbol{c}_k = \mathbf{0}$ the expansion is not necessary. Therefore, when $\boldsymbol{Y}_{k-1} = -\boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_k$ we essentially have a "removable singularity". With this in mind, it follows from the continuity of φ and $\boldsymbol{c}_k$ that $\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})$ is a continuous function of $\boldsymbol{Y}_{k-1}$. This fact is also implied by the following lemma.

Lemma 3.4.3 *$\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})$ is differentiable with respect to $\boldsymbol{Y}_{k-1}$ and the derivative is uniformly bounded.*

proof. Recall the relationship obtained by an application of the Mean Value Theorem on φ,

$$\begin{aligned}\varphi(\boldsymbol{Y}_{l-1}) &= \varphi\left(\boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right) \\ &= \varphi\left(\boldsymbol{\Gamma}_{k+1}^{l-1}\right) + \boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})' \boldsymbol{c}_k.\end{aligned}$$

Equivalently, we may write,

$$\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})' \boldsymbol{c}_k = \varphi\left(\boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right) - \varphi\left(\boldsymbol{\Gamma}_{k+1}^{l-1}\right).$$

Because W3 implies that both φ and $\boldsymbol{c}_k = \boldsymbol{P}^{l-k}\boldsymbol{Y}_{k-1} + \boldsymbol{P}^{l-k-1}\boldsymbol{\varepsilon}_k$ are differentiable with respect to $\boldsymbol{Y}_{k-1}$ it follows from the "removable singularity" comment that $\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})$ is differentiable with respect to $\boldsymbol{Y}_{k-1}$. Now, taking the derivative with respect to $\boldsymbol{Y}_{k-1}$ of both sides of the above equality yields the following,

$$\begin{aligned}\frac{d}{d\boldsymbol{Y}_{k-1}}\left(\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})' \boldsymbol{c}_k\right) &= \left(\frac{d}{d\boldsymbol{Y}_{k-1}}\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})\right)\boldsymbol{c}_k + \boldsymbol{P}^{l-k}\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl}) \\ &= \boldsymbol{P}^{l-k}\boldsymbol{D}\varphi\left(\boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right).\end{aligned}$$

Hence, it follows that,

$$\begin{aligned}\left(\frac{d}{d\boldsymbol{Y}_{k-1}}\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})\right)\boldsymbol{c}_k &= \boldsymbol{P}^{l-k}\left(\boldsymbol{D}\varphi\left(\boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right) - \boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})\right) \\ &= \boldsymbol{P}^{l-k}\left(\boldsymbol{D}\varphi\left(\boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right) - \boldsymbol{D}\varphi\left(\boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1} - \lambda\boldsymbol{c}_k\right)\right).\end{aligned}$$

Now, applying a mean value type theorem for a vector valued function (e.g. Theorem 40.5 of Bartle (1964, pg. 366)) yields the following inequality,

$$\left\|\frac{d}{d\boldsymbol{Y}_{k-1}}\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})\,\boldsymbol{c}_k\right\| = \left\|\boldsymbol{P}^{l-k}\left(\boldsymbol{D}\varphi\left(\boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right) - \boldsymbol{D}\varphi\left(\boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1} - \lambda\boldsymbol{c}_k\right)\right)\right\|$$

$$
\begin{aligned}
&\leq \left\|\boldsymbol{P}^{l-k}\right\| \left\|\boldsymbol{D}\varphi\left(\boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right) - \boldsymbol{D}\varphi\left(\boldsymbol{c}_k + \boldsymbol{\Gamma}_{k+1}^{l-1} - \lambda \boldsymbol{c}_k\right)\right\| \\
&\leq \left\|\boldsymbol{P}^{l-k}\right\| \left\|\boldsymbol{D}^2\varphi\left(\boldsymbol{\xi}_{kl}^*\right)\lambda \boldsymbol{c}_k\right\| \\
&\leq \left\|\boldsymbol{P}^{l-k}\right\| \left\|\boldsymbol{D}^2\varphi\left(\boldsymbol{\xi}_{kl}^*\right)\right\| \left\|\boldsymbol{c}_k\right\|.
\end{aligned}
$$

Recall that the above result holds for all $\boldsymbol{c}_k \neq \mathbf{0}$. Since Lemma 3.4.2 implies that $\left\|\boldsymbol{P}^{l-k}\right\|$ is bounded and assumptions W4 and W5 imply that $\boldsymbol{D}^2\varphi$ is bounded it follows upon letting $\boldsymbol{c}_k = \mathbf{1}$ that $\left\|\frac{d}{d\boldsymbol{Y}_{k-1}}\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{kl}\right)\right\|$ is bounded. That is, the derivative of $\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{kl}\right)$ with respect to $\boldsymbol{Y}_{k-1}$ is uniformly bounded. □

Next, we state and prove two results that are needed in the proof of the asymptotic distribution of the gradient. Although both of these results can be used, only one is actually needed. The first result utilizes W1–W6 whereas the second result uses W7.

Theorem 3.4.1 *Assumptions M1-M3, E1, and W1-W6 imply the following,*

$$
\begin{aligned}
S_{n2}\left(\boldsymbol{\rho}_0\right) &= \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\left(I\left(\varepsilon_j \leq \varepsilon_i\right) - F\left(\varepsilon_i\right)\right) \\
&= o_p(1).
\end{aligned}
$$

proof. Start by writing $S_{n2}\left(\boldsymbol{\rho}_0\right)$ as a sum of three components,

$$
\begin{aligned}
S_{n2}\left(\boldsymbol{\rho}_0\right) &= \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\left(I\left(\varepsilon_j \leq \varepsilon_i\right) - F\left(\varepsilon_i\right)\right) \\
&= \frac{2}{n^{\frac{3}{2}}}\left(\sum_{i<j} + \sum_{i=j} + \sum_{i>j}\right)\left(b_{ij}\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\left(I\left(\varepsilon_j \leq \varepsilon_i\right) - F\left(\varepsilon_i\right)\right)\right) \\
&= S_{n21} + S_{n22} + S_{n23} \text{ say.}
\end{aligned}
$$

We will show that $S_{n2k} = o_p(1)$ for $k = 1, 2, 3$. Consider S_{n22} first. Since $i = j$ the Ergodic Theorem and the bounds for φ, h, and F imply the following,

$$\begin{aligned} S_{n22} &= \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n} h^2(\boldsymbol{Y}_{i-1})\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}(1 - F(\varepsilon_i)) \\ &= \frac{2}{\sqrt{n}} \times \frac{1}{n}\sum_{i=1}^{n} \varphi(\boldsymbol{Y}_{i-1})\, h(\boldsymbol{Y}_{i-1})(1 - F(\varepsilon_i)) \\ &= o(1)\left(E\left[\varphi(\boldsymbol{Y}_0)\, h(\boldsymbol{Y}_0)(1 - F(\varepsilon_1))\right] + o_p(1)\right) \\ &= o_p(1). \end{aligned}$$

Now consider S_{n21}. Using the notation defined above write S_{n21} as follows,

$$\begin{aligned} S_{n21} &= \frac{2}{n^{\frac{3}{2}}}\sum_{i<j} b_{ij}\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\left(I(\varepsilon_j \le \varepsilon_i) - F(\varepsilon_i)\right) \\ &= \frac{2}{n^{\frac{3}{2}}}\sum_{i<j} h(\boldsymbol{Y}_{i-1})\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1} h(\boldsymbol{Y}_{j-1})\left(I(\varepsilon_j \le \varepsilon_i) - F(\varepsilon_i)\right) \\ &= \frac{2}{n^{\frac{3}{2}}}\sum_{i<j} \varphi_{i-1} h_{j-1}\psi_{ij} \text{ say.} \end{aligned}$$

To show $S_{n21} = o_p(1)$ using Chebyshev's Inequality it suffices to show that $E\left[S_{n21}^2\right] = o(1)$. Squaring S_{n21} and applying the expectation yields the following,

$$\begin{aligned} E\left[S_{n21}^2\right] &= \frac{4}{n^3} E\left[\sum_{i<j}(\varphi_{i-1}h_{j-1}\psi_{ij})^2 + \sum_{i<j}\sum_{k<l}(\varphi_{i-1}h_{j-1}\psi_{ij})(\varphi_{k-1}h_{l-1}\psi_{kl})\right] \\ &= \frac{4}{n^3}\sum_{i<j} E\left[(\varphi_{i-1}h_{j-1}\psi_{ij})^2\right] + \\ &\quad \frac{4}{n^3}\sum_{i<j}\sum_{k<l} E\left[(\varphi_{i-1}h_{j-1}\psi_{ij})(\varphi_{k-1}h_{l-1}\psi_{kl})\right] \\ &= E_{n1} + E_{n2} \text{ say.} \end{aligned}$$

Consider E_{n1} first. Assumptions W4 and W5 imply that φ_{i-1}, h_{j-1}, and ψ_{ij} are uniformly bounded. Thus, the expectation in E_{n1} is uniformly bounded. Since

the sum in E_{n1} is $O(n^2)$, the expectation is bounded, and we are dividing by n^3 it follows that $E_{n1} = o(1)$. Now consider E_{n2}. Since E_{n2} contains a sum with four indices, it suffices to show that $E_{n2} = o(1)$ for each of the thirteen cases given in Table 1. Now note that if one sets $i = k'$, $k = i'$, $l = j'$, and $j = l'$ it follows that E_{n2} (under case 5) is equal to E_{n2} (under case 9),

$$\begin{aligned}
E_{n2}(5) &= \sum_{i<k<l<j} E\left[\varphi_{i-1}h_{j-1}\psi_{ij}\varphi_{k-1}h_{l-1}\psi_{kl}\right] \\
&= \sum_{k'<i'<j'<l'} E\left[\varphi_{k'-1}h_{l'-1}\psi_{k'l'}\varphi_{i'-1}h_{j'-1}\psi_{i'j'}\right] \\
&= \sum_{k'<i'<j'<l'} E\left[\varphi_{i'-1}h_{j'-1}\psi_{i'j'}\varphi_{k'-1}h_{l'-1}\psi_{k'l'}\right] \\
&= E_{n2}(9).
\end{aligned}$$

This type of symmetry can be shown to hold for the other cases as well. Specifically, only cases 1 through 7 need to be considered. However, under case 7, E_{n2} reduces to E_{n1} which has already been show to be $o(1)$. Hence, only cases 1 through 6 need to be considered. Let us begin with cases 1, 3, and 5. Under each of these cases the sum in E_{n2} is $O(n^4)$. Note that in cases 1 and 3, l is the largest subscript and it is distinct from any of the other subscripts. Thus, conditioning on $\mathcal{F}_{l-1}$ yields the following result,

$$\begin{aligned}
&E\left[\varphi_{i-1}h_{j-1}\psi_{ij}\varphi_{k-1}h_{l-1}\psi_{kl}\right] \\
&\quad= E\left[\varphi_{i-1}h_{j-1}\psi_{ij}\varphi_{k-1}h_{l-1}E\left[\psi_{kl} \mid \mathcal{F}_{l-1}\right]\right] \\
&\quad= E\left[\varphi_{i-1}h_{j-1}\psi_{ij}\varphi_{k-1}h_{l-1}E\left[I\left(\varepsilon_l \le \varepsilon_k\right) - F\left(\varepsilon_k\right) \mid \mathcal{F}_{l-1}\right]\right] \\
&\quad= E\left[\varphi_{i-1}h_{j-1}\psi_{ij}\varphi_{k-1}h_{l-1}\left(F\left(\varepsilon_k\right) - F\left(\varepsilon_k\right)\right)\right]
\end{aligned}$$

$$= 0.$$

Hence, $E_{n2}(1) = E_{n2}(3) = 0$. Now consider case 5. Note that j is the largest subscripts and that it is distinct from any of the other subscripts. Thus, conditioning on $\mathcal{F}_{j-1}$ and following the same argument as above yields $E_{n2}(5) = 0$. Therefore, $E_{n2} = 0$ under cases 1, 3, and 5. Next, consider cases 2 and 6. Again, l is the largest subscript and it is distinct from the other subscripts. Thus, conditioning on $\mathcal{F}_{l-1}$ and following the above derivations yields $E_{n2}(2) = E_{n2}(6) = 0$. To complete the proof of $S_{n21} = o_p(1)$ we must show that $E_{n2}(4) = o(1)$. Since there is a tie in the subscripts (under case 4) the sum in E_{n2} reduces to $O(n^3)$ as follows,

$$\begin{aligned}
E_{n2}(4) &= \frac{4}{n^3} \sum_{i<k<j=l} E\left[\varphi_{i-1} h_{j-1} \psi_{ij} \varphi_{k-1} h_{l-1} \psi_{kl}\right] \\
&= \frac{4}{n^3} \sum_{i<k<j} E\left[\varphi_{i-1} h_{j-1} \psi_{ij} \varphi_{k-1} h_{j-1} \psi_{kj}\right] \\
&= \frac{4}{n^3} \sum_{i<k<j} E\left[\varphi_{i-1} \varphi_{k-1} h_{j-1}^2 \psi_{ij} \psi_{kj}\right].
\end{aligned}$$

Consider only the expectation in $E_{n2}(4)$. Conditioning on $\mathcal{F}_{j-1}$ yields the following,

$$\begin{aligned}
& E\left[\varphi_{i-1} \varphi_{k-1} h_{j-1}^2 \psi_{ij} \psi_{kj}\right] \\
&= E\left[\varphi_{i-1} \varphi_{k-1} h_{j-1}^2 E\left[\psi_{ij} \psi_{kj} \mid \mathcal{F}_{j-1}\right]\right] \\
&= E\left[\varphi_{i-1} \varphi_{k-1} h_{j-1}^2 E\left[\left(I\left(\varepsilon_j \leq \varepsilon_i\right) - F\left(\varepsilon_i\right)\right)\left(I\left(\varepsilon_j \leq \varepsilon_k\right) - F\left(\varepsilon_k\right)\right) \mid \mathcal{F}_{j-1}\right]\right] \\
&= E\left[\varphi_{i-1} \varphi_{k-1} h_{j-1}^2 \left(F\left(\min\left(\varepsilon_i, \varepsilon_k\right)\right) - F\left(\varepsilon_i\right) F\left(\varepsilon_k\right)\right)\right] \\
&= E\left[\varphi_{i-1} \varphi_{k-1} h_{j-1}^2 F_{ik}\right] \text{ say} \\
&= E\left[\varphi_{i-1} \varphi_{k-1} F_{ik} h^2\left(\mathbf{Y}_{j-1}\right)\right].
\end{aligned}$$

Now note that if $k = j - 1$ then the sum in $E_{n2}(4)$ is reduced to $O(n^2)$. Since W4 and W5 imply that the above expectation is bounded and we are dividing by n^3 it follows that $E_{n2}(4) = o(1)$ when $k = j - 1$. Thus, consider the case where $k < j - 1$. Since $k < j - 1$ we can write $\boldsymbol{Y}_{j-1} = \boldsymbol{P}^{j-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{j-1}$ and use the multivariate Mean Value Theorem (e.g. Theorem 40.4 of Bartle (1964, pg. 365)) on h^2 to show the following,

$$\begin{aligned}
&E\left[\varphi_{i-1}\varphi_{k-1}F_{ik}h^2\left(\boldsymbol{Y}_{j-1}\right)\right] \\
&= E\left[\varphi_{i-1}\varphi_{k-1}F_{ik}h^2\left(\boldsymbol{P}^{j-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{j-1}\right)\right] \\
&= E\left[\varphi_{i-1}\varphi_{k-1}F_{ik}\left(h^2\left(\boldsymbol{\Gamma}_{k+1}^{j-1}\right) + \boldsymbol{D}h^2\left(\boldsymbol{\xi}_{kj}\right)'\boldsymbol{P}^{j-k-1}\boldsymbol{Y}_k\right)\right] \\
&= E\left[\varphi_{i-1}\varphi_{k-1}F_{ik}h^2\left(\boldsymbol{\Gamma}_{k+1}^{j-1}\right)\right] + E\left[\varphi_{i-1}\varphi_{k-1}F_{ik}\boldsymbol{D}h^2\left(\boldsymbol{\xi}_{kj}\right)'\boldsymbol{P}^{j-k-1}\boldsymbol{Y}_k\right] \\
&= e_1 + e_2 \text{ say.}
\end{aligned}$$

Consider e_2 first. The bounds on φ, h, $\boldsymbol{D}h$ and F; the finite second moment of the process; and Lemma 3.4.2 yield the following,

$$\begin{aligned}
|e_2| &\leq E\left[\left|\varphi_{i-1}\varphi_{k-1}F_{ik}\boldsymbol{D}h^2\left(\boldsymbol{\xi}_{kj}\right)'\boldsymbol{P}^{j-k-1}\boldsymbol{Y}_k\right|\right] \\
&\leq K_1 E\left[\left|\boldsymbol{D}h^2\left(\boldsymbol{\xi}_{kj}\right)'\boldsymbol{P}^{j-k-1}\boldsymbol{Y}_k\right|\right] \\
&= K_1 E\left[\left|\sum_{a=1}^{p}\sum_{b=1}^{p}\frac{dh^2}{dY_a}P_{ab}^{(j-k-1)}X_{k-b}\right|\right] \\
&\leq K_1\sum_{a=1}^{p}\sum_{b=1}^{p}\left|\frac{dh^2}{dY_a}P_{ab}^{(j-k-1)}\right| E\left[\left|X_{k-b}\right|\right] \\
&\leq K_2\sum_{a=1}^{p}\sum_{b=1}^{p}\left|\lambda_{\mathrm{M}}\right|^{j-k-1}E\left[\left|X_{k-b}\right|\right] \\
&\leq \left|\lambda_{\mathrm{M}}\right|^{j-k-1}p^2K_2\sigma_X
\end{aligned}$$

$$\leq \; |\lambda_{\mathrm{M}}|^{j-k-1} K \text{ say.}$$

Now consider e_1. Since $\mathbf{\Gamma}_{k+1}^{j-1}$ is independent of $\varphi_{i-1}\varphi_{k-1}F_{ik}$ it follows that,

$$\begin{aligned} e_1 &= E\left[\varphi_{i-1}\varphi_{k-1}F_{ik}h^2\left(\mathbf{\Gamma}_{k+1}^{j-1}\right)\right] \\ &= E\left[\varphi_{i-1}\varphi_{k-1}F_{ik}\right]E\left[h^2\left(\mathbf{\Gamma}_{k+1}^{j-1}\right)\right]. \end{aligned}$$

The bound on h implies that $E\left[h^2\left(\mathbf{\Gamma}_{k+1}^{j-1}\right)\right]$ is also bounded. Thus, we will absorb this into the notation for e_1 and consider only the first expectation. Conditioning on $\mathcal{F}_{k-1}$ and using the fact that,

$$\begin{aligned} E\left[F\left(\min\left(\varepsilon_i,\varepsilon_k\right)\right) \mid \mathcal{F}_{k-1}\right] &= E\left[\min\left(F\left(\varepsilon_i\right),F\left(\varepsilon_k\right)\right) \mid \mathcal{F}_{k-1}\right] \\ &= F\left(\varepsilon_i\right) - \frac{1}{2}F^2\left(\varepsilon_i\right) \end{aligned}$$

yields the following result,

$$\begin{aligned} E\left[\varphi_{i-1}\varphi_{k-1}F_{ik}\right] &= E\left[\varphi_{i-1}\varphi_{k-1}E\left[F_{ik} \mid \mathcal{F}_{k-1}\right]\right] \\ &= E\left[\varphi_{i-1}\varphi_{k-1}E\left[F\left(\min\left(\varepsilon_i,\varepsilon_k\right)\right) - F\left(\varepsilon_i\right)F\left(\varepsilon_k\right) \mid \mathcal{F}_{k-1}\right]\right] \\ &= E\left[\varphi_{i-1}\varphi_{k-1}\left(F\left(\varepsilon_i\right) - \frac{1}{2}F^2\left(\varepsilon_i\right) - \frac{1}{2}F\left(\varepsilon_i\right)\right)\right] \\ &= \frac{1}{2}E\left[\varphi_{i-1}\left(F\left(\varepsilon_i\right) - F^2\left(\varepsilon_i\right)\right)\varphi_{k-1}\right] \\ &= \frac{1}{2}E\left[\varphi_{i-1}G_i\varphi_{k-1}\right] \text{ say} \\ &= \frac{1}{2}E\left[\varphi_{i-1}G_i\varphi\left(\mathbf{Y}_{k-1}\right)\right]. \end{aligned}$$

Now note that if $i = k-1$ then the sum in $E_{n2}(4)$ is reduced to $O(n^2)$. Since W4, W5, and the bound on F imply that the above expectation is bounded and we

are dividing by n^3 it follows that $E_{n2}(4) = o(1)$ when $i = k-1$. Thus, consider the case where $i < k-1$. Since $i < k-1$ we can write $\boldsymbol{Y}_{k-1} = \boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i + \boldsymbol{\Gamma}_{i+1}^{k-1}$ and use the multivariate Mean Value Theorem on φ to show the following,

$$\begin{aligned}
&\frac{1}{2}E\left[\varphi_{i-1}G_i\varphi\left(\boldsymbol{Y}_{k-1}\right)\right] \\
&= \frac{1}{2}E\left[\varphi_{i-1}G_i\varphi\left(\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i + \boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right] \\
&= \frac{1}{2}E\left[\varphi_{i-1}G_i\left(\varphi\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right) + \boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{ik}\right)'\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right)\right] \\
&= \frac{1}{2}E\left[\varphi_{i-1}G_i\varphi\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right] + \frac{1}{2}E\left[\varphi_{i-1}G_i\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{ik}\right)'\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right] \\
&= e_{11} + e_{12} \text{ say.}
\end{aligned}$$

Consider e_{12} first. The bounds on φ, $\boldsymbol{D}\varphi$, and F; the finite second moment of the process; and Lemma 3.4.2 imply the following,

$$\begin{aligned}
|e_{12}| &\le \frac{1}{2}E\left[\left|\varphi_{i-1}G_i\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{ik}\right)'\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right|\right] \\
&\le K_1E\left[\left|\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{ik}\right)'\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right|\right] \\
&= K_1E\left[\left|\sum_{a=1}^{p}\sum_{b=1}^{p}\frac{d\varphi}{dY_a}P_{ab}^{(k-i-1)}X_{i-b}\right|\right] \\
&\le K_1\sum_{a=1}^{p}\sum_{b=1}^{p}\left|\frac{d\varphi}{dY_a}P_{ab}^{(k-i-1)}\right|E\left[|X_{i-b}|\right] \\
&\le K_2\sum_{a=1}^{p}\sum_{b=1}^{p}|\lambda_{\mathrm{M}}|^{k-i-1}E\left[|X_{i-b}|\right] \\
&\le |\lambda_{\mathrm{M}}|^{k-i-1}p^2K_2\sigma_X \\
&\le |\lambda_{\mathrm{M}}|^{k-i-1}K \text{ say.}
\end{aligned}$$

Now consider e_{11}. Since $\boldsymbol{\Gamma}_{i+1}^{k-1}$ is independent of $\varphi_{i-1}G_i$, W6 implies the following,

$$e_{11} = \frac{1}{2}E\left[\varphi_{i-1}G_i\varphi\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right]$$

$$= \frac{1}{2}E[\varphi_{i-1}]E[G_i]E\left[\varphi\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right]$$
$$= 0.$$

Putting the above results together yields the following,

$$\begin{aligned}
|E_{n2}(4)| &\leq \frac{4}{n^3}\sum_{i<k<j}\left|E\left[\varphi_{i-1}\varphi_{k-1}h_{j-1}^2\psi_{ij}\psi_{kj}\right]\right| \\
&= \frac{4}{n^3}\sum_{i<k<j}|e_{11}+e_{12}+e_2| \\
&\leq \frac{4}{n^3}\sum_{i<k<j}\left(|\lambda_{\mathrm{M}}|^{k-i-1}+|\lambda_{\mathrm{M}}|^{j-k-1}\right)K \\
&= O\left(n^{-3}\right)\left(O\left(n^2\right)+O\left(n^2\right)\right) \\
&= o(1),
\end{aligned}$$

since $|\lambda_{\mathrm{M}}| < 1$. Since all of the other cases have been shown to be $o(1)$ this completes the proof of $E\left[S_{n21}^2\right] = o(1)$ which implies that $S_{n21} = o_p(1)$. Now consider S_{n23}. Note that upon letting $i' = j$ and $j' = i$ we can write S_{n23} as follows,

$$\begin{aligned}
S_{n23} &= \frac{2}{n^{\frac{3}{2}}}\sum_{i>j} b_{ij}\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\left(I\left(\varepsilon_j \leq \varepsilon_i\right)-F\left(\varepsilon_i\right)\right) \\
&= \frac{2}{n^{\frac{3}{2}}}\sum_{i>j} h_{j-1}\varphi_{i-1}\psi_{ij} \\
&= \frac{2}{n^{\frac{3}{2}}}\sum_{j'>i'} h_{i'-1}\varphi_{j'-1}\psi_{j'i'} \\
&= \frac{2}{n^{\frac{3}{2}}}\sum_{i<j} h_{i-1}\varphi_{j-1}\psi_{ji} \text{ say.}
\end{aligned}$$

To show $S_{n23} = o_p(1)$ using Chebyshev's Inequality it suffices to show that $E\left[S_{n23}^2\right] = o(1)$. Squaring S_{n23} and applying the expectation yields the follow-

ing,

$$\begin{aligned}
E\left[S_{n23}^2\right] &= \frac{4}{n^3}E\left[\sum_{i<j}\left(h_{i-1}\varphi_{j-1}\psi_{ji}\right)^2+\sum_{i<j}\sum_{k<l}\left(h_{i-1}\varphi_{j-1}\psi_{ji}\right)\left(h_{k-1}\varphi_{l-1}\psi_{lk}\right)\right] \\
&= \frac{4}{n^3}\sum_{i<j}E\left[\left(h_{i-1}\varphi_{j-1}\psi_{ji}\right)^2\right]+ \\
&\quad \frac{4}{n^3}\sum_{i<j}\sum_{k<l}E\left[\left(h_{i-1}\varphi_{j-1}\psi_{ji}\right)\left(h_{k-1}\varphi_{l-1}\psi_{lk}\right)\right] \\
&= E_{n1}+E_{n2} \text{ say.}
\end{aligned}$$

Consider E_{n1} first. Assumptions W4 and W5 imply that h_{i-1}, φ_{j-1}, and ψ_{ji} are uniformly bounded. Thus, the expectation in E_{n1} is uniformly bounded. Since the sum in E_{n1} is $O(n^2)$, the expectation is bounded, and we are dividing by n^3 it follows that $E_{n1}=o(1)$. Now consider E_{n2}. Since E_{n2} contains a sum with four indices, it suffices to show that $E_{n2}=o(1)$ for each of the thirteen cases given in Table 1. Again, symmetry reduces the work to considering only cases 1 through 7. However, under case 7, E_{n2} reduces to E_{n1} which has already been shown to be $o(1)$. Thus, only cases 1 through 6 need to be considered. To begin, consider cases 2, 4, and 6. Under each of these cases there is exactly one tie among the subscripts. Thus, the sum in E_{n2} reduces to $O(n^3)$. Consider E_{n2} under case 2,

$$\begin{aligned}
E_{n2}(2) &= \frac{4}{n^3}\sum_{i<j=k<l}E\left[\left(h_{i-1}\varphi_{j-1}\psi_{ji}\right)\left(h_{k-1}\varphi_{l-1}\psi_{lk}\right)\right] \\
&= \frac{4}{n^3}\sum_{i<j<l}E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}\varphi_{l-1}\psi_{lj}\right].
\end{aligned}$$

Considering only the expectation given in $E_{n2}(2)$ and conditioning on $\mathcal{F}_{l-1}$ yields,

$$E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}\varphi_{l-1}\psi_{lj}\right]$$

$$= E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}\varphi_{l-1}E\left[\psi_{lj} \mid \mathcal{F}_{l-1}\right]\right]$$

$$= E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}\varphi_{l-1}E\left[I\left(\varepsilon_j \leq \varepsilon_l\right) - F\left(\varepsilon_l\right) \mid \mathcal{F}_{l-1}\right]\right]$$

$$= E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}\left(\frac{1}{2} - F\left(\varepsilon_j\right)\right)\varphi_{l-1}\right]$$

$$= E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}u_j\varphi\left(\boldsymbol{Y}_{l-1}\right)\right] \text{ say.}$$

Now note that in the case of $j = l - 1$ the sum in $E_{n2}(2)$ reduces to $O(n^2)$. Since our assumptions imply that everything inside of the expectation is bounded and we are dividing by n^3 it follows that $E_{n2}(2) = o(1)$ in the case of $j = l - 1$. Hence, consider the above expectation when $j < l-1$. Using the fact that $\boldsymbol{Y}_{l-1} = \boldsymbol{P}^{l-j-1}\boldsymbol{Y}_j + \boldsymbol{\Gamma}_{j+1}^{l-1}$ in conjunction with the multivariate Mean Value Theorem on φ and independence yields the following,

$$\begin{aligned}
&E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}u_j\varphi\left(\boldsymbol{Y}_{l-1}\right)\right] \\
&= E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}u_j\varphi\left(\boldsymbol{P}^{l-j-1}\boldsymbol{Y}_j + \boldsymbol{\Gamma}_{j+1}^{l-1}\right)\right] \\
&= E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}u_j\left(\varphi\left(\boldsymbol{\Gamma}_{j+1}^{l-1}\right) + \boldsymbol{D}\varphi\left(\xi_{jl}\right)'\boldsymbol{P}^{l-j-1}\boldsymbol{Y}_j\right)\right] \\
&= E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}u_j\right]E\left[\varphi\left(\boldsymbol{\Gamma}_{j+1}^{l-1}\right)\right] + \\
&\quad E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}u_j\boldsymbol{D}\varphi\left(\xi_{jl}\right)'\boldsymbol{P}^{l-j-1}\boldsymbol{Y}_j\right] \\
&= e_1 + e_2 \text{ say.}
\end{aligned}$$

Consider e_2 first. Again, the bounds on h, φ, and $\boldsymbol{D}\varphi$; the finite second moment assumption on the process; and Lemma 3.4.2 can be used to show the following,

$$\begin{aligned}
|e_2| &\leq \left|\lambda_{\mathrm{M}}\right|^{l-j-1} K_1\sigma_X \\
&\leq \left|\lambda_{\mathrm{M}}\right|^{l-j-1} K \text{ say.}
\end{aligned}$$

Now consider e_1. Since φ is bounded we will absorb the second expectation into the notation for e_1 and consider the first expectation only. Before we consider this expectation straight forward calculations can be used to show,

$$
\begin{aligned}
E\left[\psi_{ji}u_j \mid \mathcal{F}_{j-1}\right] &= \frac{1}{12}+\frac{1}{2}\left(F^2\left(\varepsilon_i\right)-F\left(\varepsilon_i\right)\right) \\
&= G\left(\varepsilon_i\right) \text{ say.}
\end{aligned}
$$

Furthermore, it follows that $E\left[G_i\right]=0$. Now, conditioning on $\mathcal{F}_{j-1}$ and using the above fact imply the following,

$$
\begin{aligned}
E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}u_j\right] &= E\left[h_{i-1}\varphi_{j-1}h_{j-1}E\left[\psi_{ji}u_j \mid \mathcal{F}_{j-1}\right]\right] \\
&= E\left[h_{i-1}\varphi_{j-1}h_{j-1}G_i\right] \\
&= E\left[h_{i-1}G_i\varphi_{j-1}h_{j-1}\right] \\
&= E\left[h_{i-1}G_ih^2\left(\boldsymbol{Y}_{j-1}\right)\boldsymbol{Y}_{j-1}\right] \\
&= E\left[h_{i-1}G_i\gamma\left(\boldsymbol{Y}_{j-1}\right)\right] \text{ say.}
\end{aligned}
$$

Now note that in the case of $i=j-1$ the sum in $E_{n2}(2)$ reduces to $O\left(n^2\right)$. Since our assumptions imply that everything inside of the expectation is bounded and we are dividing by n^3 it follows that $E_{n2}(2)=o(1)$ in the case of $i=j-1$. Hence, consider the above expectation when $i<j-1$. Using the fact that $\boldsymbol{Y}_{j-1}=\boldsymbol{P}^{j-i-1}\boldsymbol{Y}_i+\boldsymbol{\Gamma}_{i+1}^{j-1}$ in conjunction with the multivariate Mean Value Theorem on γ and independence yields the following,

$$
E\left[h_{i-1}G_i\gamma\left(\boldsymbol{Y}_{j-1}\right)\right]
$$

$$
\begin{aligned}
&= E\left[h_{i-1}G_i\gamma\left(\boldsymbol{P}^{j-i-1}\boldsymbol{Y}_i + \boldsymbol{\Gamma}_{i+1}^{j-1}\right)\right] \\
&= E\left[h_{i-1}G_i\left(\gamma\left(\boldsymbol{\Gamma}_{i+1}^{j-1}\right) + \boldsymbol{D}\gamma\left(\boldsymbol{\xi}_{ij}\right)'\boldsymbol{P}^{j-i-1}\boldsymbol{Y}_i\right)\right] \\
&= E\left[h_{i-1}\right]E\left[G_i\right]E\left[\gamma\left(\boldsymbol{\Gamma}_{i+1}^{j-1}\right)\right] + E\left[h_{i-1}G_i\boldsymbol{D}\gamma\left(\boldsymbol{\xi}_{ij}\right)'\boldsymbol{P}^{j-i-1}\boldsymbol{Y}_i\right] \\
&= e_{11} + e_{12} \text{ say.}
\end{aligned}
$$

Now note that our assumptions imply that both γ and $\boldsymbol{D}\gamma$ are bounded. Since $E\left[G_i\right] = 0$ it follows that $e_{11} = 0$. Now consider e_{12}. The bounds on h and $\boldsymbol{D}\gamma$, the finite second moment on the process, and Lemma 3.4.2 can be used to show the following,

$$
\begin{aligned}
|e_{12}| &\leq |\lambda_{\mathrm{M}}|^{j-i-1} K_1\sigma_X \\
&\leq |\lambda_{\mathrm{M}}|^{j-i-1} K \text{ say.}
\end{aligned}
$$

Putting all of the above pieces together yields the following,

$$
\begin{aligned}
|E_{n2}(2)| &\leq \frac{4}{n^3}\sum_{i<j<l}\left|E\left[h_{i-1}\varphi_{j-1}h_{j-1}\psi_{ji}\varphi_{l-1}\psi_{lj}\right]\right| \\
&= \frac{4}{n^3}\sum_{i<j<l}|e_1 + e_2| \\
&= \frac{4}{n^3}\sum_{i<j<l}|e_{12} + e_2| \\
&\leq \frac{4}{n^3}\sum_{i<j<l}\left(|e_{12}| + |e_2|\right) \\
&\leq \frac{4}{n^3}\sum_{i<j<l}\left(|\lambda_{\mathrm{M}}|^{j-i-1} + |\lambda_{\mathrm{M}}|^{l-j-1}\right)K \\
&= O\left(n^{-3}\right)O\left(n^2\right) \\
&= o(1),
\end{aligned}
$$

since $|\lambda_{\text{M}}| < 1$. This completes the proof of $E_{n2}(2) = o(1)$. Since l is the largest subscript and it is distinct from the other subscripts the proof of $E_{n2}(6) = o(1)$ is treated in a similar fashion. Now consider E_{n2} under case 4. That is,

$$\begin{aligned} E_{n2}(4) &= \frac{4}{n^3} \sum_{i<k<j=l} E\left[(h_{i-1}\varphi_{j-1}\psi_{ji})(h_{k-1}\varphi_{l-1}\psi_{lk})\right] \\ &= \frac{4}{n^3} \sum_{i<k<l} E\left[h_{i-1}h_{k-1}\varphi_{l-1}^2\psi_{li}\psi_{lk}\right]. \end{aligned}$$

Before we consider the expectation given in $E_{n2}(4)$ straight forward calculations can be used to show,

$$\begin{aligned} E\left[\psi_{li}\psi_{lk} \mid \mathcal{F}_{l-1}\right] &= \frac{1}{3} - F(\max(\varepsilon_i, \varepsilon_k)) + \frac{1}{2}F^2(\varepsilon_i) + \frac{1}{2}F^2(\varepsilon_k) \\ &= F_{ik} \text{ say.} \end{aligned}$$

Furthermore, note that F_{ik} is uniformly bounded through F and that $E\left[F_{ik} \mid \mathcal{F}_{k-1}\right] = 0$. Now, conditioning on $\mathcal{F}_{l-1}$ and using the above fact imply the following,

$$\begin{aligned} E\left[h_{i-1}h_{k-1}\varphi_{l-1}^2\psi_{li}\psi_{lk}\right] &= E\left[h_{i-1}h_{k-1}\varphi_{l-1}^2 E\left[\psi_{li}\psi_{lk} \mid \mathcal{F}_{l-1}\right]\right] \\ &= E\left[h_{i-1}h_{k-1}\varphi_{l-1}^2 F_{ik}\right] \\ &= E\left[h_{i-1}h_{k-1}F_{ik}\varphi^2(\boldsymbol{Y}_{l-1})\right]. \end{aligned}$$

Now note that in the case of $k = l-1$ the sum in $E_{n2}(4)$ reduces to $O(n^2)$. Since our assumptions imply that everything inside of the expectation is bounded and we are dividing by n^3 it follows that $E_{n2}(4) = o(1)$ in the case of $k = l-1$. Hence, consider the above expectation when $k < l-1$. Using the fact that $\boldsymbol{Y}_{l-1} =$

$\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}$ in conjunction with the multivariate Mean Value Theorem on φ^2 and independence yields the following,

$$\begin{aligned}
& E\left[h_{i-1}h_{k-1}F_{ik}\varphi^2\left(\boldsymbol{Y}_{l-1}\right)\right] \\
&= E\left[h_{i-1}h_{k-1}F_{ik}\varphi^2\left(\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right] \\
&= E\left[h_{i-1}h_{k-1}F_{ik}\left(\varphi^2\left(\boldsymbol{\Gamma}_{k+1}^{l-1}\right) + \boldsymbol{D}\varphi^2\left(\boldsymbol{\xi}_{kl}\right)'\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right)\right] \\
&= E\left[h_{i-1}h_{k-1}F_{ik}\right]E\left[\varphi^2\left(\boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right] + E\left[h_{i-1}h_{k-1}F_{ik}\boldsymbol{D}\varphi^2\left(\boldsymbol{\xi}_{kl}\right)'\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right] \\
&= e_1 + e_2 \text{ say.}
\end{aligned}$$

Consider e_1 first. Conditioning on $\mathcal{F}_{k-1}$ and using the fact that $E\left[F_{ik} \mid \mathcal{F}_{k-1}\right] = 0$ implies that the first expectation given in e_1 is 0. Hence, $e_1 = 0$. Now consider e_2. The bounds on h, φ, and $\boldsymbol{D}\varphi$; the finite second moment assumption on the process; and Lemma 3.4.2 can be used to show the following,

$$\begin{aligned}
|e_2| &\leq |\lambda_{\mathrm{M}}|^{l-k-1} K_1\sigma_X \\
&\leq |\lambda_{\mathrm{M}}|^{l-k-1} K \text{ say.}
\end{aligned}$$

Putting the above pieces together yields the following,

$$\begin{aligned}
|E_{n2}(4)| &\leq \frac{4}{n^3}\sum_{i<k<l}\left|E\left[h_{i-1}h_{k-1}\varphi_{l-1}^2\psi_{li}\psi_{lk}\right]\right| \\
&= \frac{4}{n^3}\sum_{i<k<l}|e_1 + e_2| \\
&= \frac{4}{n^3}\sum_{i<k<l}|e_2| \\
&\leq \frac{4}{n^3}\sum_{i<k<l}|\lambda_{\mathrm{M}}|^{l-k-1} K \\
&= O\left(n^{-3}\right)O\left(n^2\right)
\end{aligned}$$

$$= o(1),$$

since $|\lambda_{\mathrm{M}}| < 1$. This completes the proof of $E_{n2}(4) = o(1)$. Hence $E_{n2} = o(1)$ for cases 2, 4, and 6. To complete the proof consider cases 1, 3, and 5, which are all $O(n^4)$. We will consider only case 5 ($i < k < l < j$); cases 1 and 3 can be treated using similar techniques,

$$\begin{aligned} E_{n2}(5) &= \frac{4}{n^3} \sum_{i<k<l<j} E\left[h_{i-1}\varphi_{j-1}\psi_{ji}h_{k-1}\varphi_{l-1}\psi_{lk}\right] \\ &= \frac{4}{n^3} \sum_{i<k<l<j} E\left[h_{i-1}h_{k-1}\psi_{lk}\varphi_{l-1}\varphi_{j-1}\psi_{ji}\right]. \end{aligned}$$

Before we begin note the following,

$$\begin{aligned} E\left[\psi_{ji} \mid \mathcal{F}_{j-1}\right] &= E\left[I(\varepsilon_i \leq \varepsilon_j) - F(\varepsilon_j) \mid \mathcal{F}_{j-1}\right] \\ &= \frac{1}{2} - F(\varepsilon_i) \\ &= u_i \text{ say.} \end{aligned}$$

Now, consider only the expectation given in $E_{n2}(5)$. Conditioning on $\mathcal{F}_{j-1}$ and using the above fact yield the following,

$$\begin{aligned} E\left[h_{i-1}h_{k-1}\psi_{lk}\varphi_{l-1}\varphi_{j-1}\psi_{ji}\right] &= E\left[h_{i-1}h_{k-1}\psi_{lk}\varphi_{l-1}\varphi_{j-1}E\left[\psi_{ji} \mid \mathcal{F}_{j-1}\right]\right] \\ &= E\left[h_{i-1}h_{k-1}\psi_{lk}\varphi_{l-1}\varphi_{j-1}u_i\right] \\ &= E\left[h_{i-1}u_i h_{k-1}\psi_{lk}\varphi_{l-1}\varphi\left(\boldsymbol{Y}_{j-1}\right)\right]. \end{aligned}$$

Suppose first that $l = j - 1$ so that the sum in $E_{n2}(5)$ reduces to $O(n^3)$. Now use the fact that $\boldsymbol{Y}_l = \boldsymbol{P}\boldsymbol{Y}_{l-1} + \boldsymbol{\varepsilon}_l$ and condition on $\mathcal{F}_{l-1}$ to write the above

expectation as follows,

$$\begin{aligned}
E\left[h_{i-1}u_i h_{k-1}\psi_{lk}\varphi_{l-1}\varphi_l\right] &= E\left[h_{i-1}u_i h_{k-1}\psi_{lk}\varphi\left(\boldsymbol{Y}_{l-1}\right)\varphi\left(\boldsymbol{P}\boldsymbol{Y}_{l-1}+\boldsymbol{\varepsilon}_l\right)\right] \\
&= E\left[h_{i-1}u_i h_{k-1}E\left[\psi_{lk}\varphi\left(\boldsymbol{Y}_{l-1}\right)\varphi\left(\boldsymbol{P}\boldsymbol{Y}_{l-1}+\boldsymbol{\varepsilon}_l\right)\mid\mathcal{F}_{l-1}\right]\right] \\
&= E\left[h_{i-1}u_i h_{k-1}g\left(\varepsilon_k;\boldsymbol{Y}_{l-1}\right)\right] \text{ say.}
\end{aligned}$$

Now consider the function g,

$$\begin{aligned}
g\left(\varepsilon_k;\boldsymbol{Y}_{l-1}\right) &= E\left[\psi_{lk}\varphi\left(\boldsymbol{Y}_{l-1}\right)\varphi\left(\boldsymbol{P}\boldsymbol{Y}_{l-1}+\boldsymbol{\varepsilon}_l\right)\mid\mathcal{F}_{l-1}\right] \\
&= \int\left(I\left(\varepsilon_k\leq t\right)-F\left(t\right)\right)\varphi\left(\boldsymbol{Y}_{l-1}\right)\varphi\left(\boldsymbol{P}\boldsymbol{Y}_{l-1}+t\right)f\left(t\right)dt.
\end{aligned}$$

It follows from the bound on φ that g is uniformly bounded in all arguments. Now let ε_k be fixed and consider g as a function of $\boldsymbol{Y}_{l-1}$ only. Assuming interchangeability of differentiation and integration it follows from the differentiability of φ that $g\left(\cdot;\boldsymbol{Y}_{l-1}\right)$ is differentiable and hence continuous. Furthermore, it follows from the bounds on φ and $\boldsymbol{D}\varphi$ that $\boldsymbol{D}g$ is uniformly bounded in all arguments. Let us now return to the expectation under consideration. Note that in the case of $k=l-1$ the sum in $E_{n2}\left(5\right)$ reduces to $O\left(n^2\right)$. Since our assumptions imply that everything inside of the expectation is bounded and we are dividing by n^3 it follows that $E_{n2}\left(5\right)=o(1)$ in the case of $k=l-1$. Hence, consider the above expectation when $k<l-1$. Using the fact that $\boldsymbol{Y}_{l-1}=\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k+\boldsymbol{\Gamma}_{k+1}^{l-1}$ in conjunction with the multivariate Mean Value Theorem on $g\left(\cdot;\boldsymbol{Y}\right)$ and independence yields the following,

$$E\left[h_{i-1}u_i h_{k-1}g\left(\varepsilon_k;\boldsymbol{Y}_{l-1}\right)\right]$$

$$
\begin{aligned}
&= E\left[h_{i-1}u_i h_{k-1} g\left(\varepsilon_k; \boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right] \\
&= E\left[h_{i-1}u_i h_{k-1}\left(g\left(\varepsilon_k; \boldsymbol{\Gamma}_{k+1}^{l-1}\right) + \boldsymbol{D}g\left(\varepsilon_k; \boldsymbol{\xi}_{kl}\right)' \boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right)\right] \\
&= E\left[h_{i-1}u_i h_{k-1}\right] E\left[g\left(\varepsilon_k; \boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right] + E\left[h_{i-1}u_i h_{k-1}\boldsymbol{D}g\left(\varepsilon_k; \boldsymbol{\xi}_{kl}\right)' \boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right] \\
&= e_1 + e_2 \text{ say.}
\end{aligned}
$$

Consider e_2 first. The bounds on h and $\boldsymbol{D}g$, the finite second moment assumption on the process, and Lemma 3.4.2 can be used to show the following,

$$
\begin{aligned}
|e_2| &\leq |\lambda_{\mathrm{M}}|^{l-k-1} K_1 \sigma_X \\
&\leq |\lambda_{\mathrm{M}}|^{l-k-1} K \text{ say.}
\end{aligned}
$$

Now consider e_1. Since g is bounded we will absorb the second expectation into the notation for e_1 and consider the first expectation only. That is,

$$
\begin{aligned}
e_1 &= E\left[h_{i-1}u_i h_{k-1}\right] \\
&= E\left[h_{i-1}u_i h\left(\boldsymbol{Y}_{k-1}\right)\right].
\end{aligned}
$$

Now note that in the case of $i = k-1$ the sum in $E_{n2}(5)$ reduces to $O(n^2)$. Since our assumptions imply that everything inside of the expectation is bounded and we are dividing by n^3 it follows that $E_{n2}(5) = o(1)$ in the case of $i = k-1$. Hence, consider the above expectation when $i < k-1$. Using the fact that $\boldsymbol{Y}_{k-1} = \boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i + \boldsymbol{\Gamma}_{i+1}^{k-1}$ in conjunction with the multivariate Mean Value Theorem on h and independence yields the following,

$$
E\left[h_{i-1}u_i h\left(\boldsymbol{Y}_{k-1}\right)\right]
$$

$$
\begin{aligned}
&= E\left[h_{i-1}u_i h\left(\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i + \boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right] \\
&= E\left[h_{i-1}u_i\left(h\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right) + \boldsymbol{D}h\left(\boldsymbol{\xi}_{ik}\right)' \boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right)\right] \\
&= E\left[h_{i-1}\right]E\left[u_i\right]E\left[h\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right] + E\left[h_{i-1}u_i\boldsymbol{D}h\left(\boldsymbol{\xi}_{ik}\right)' \boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right] \\
&= e_{11} + e_{12} \text{ say.}
\end{aligned}
$$

Note that $E\left[u_i\right] = 0$ implies that $e_{11} = 0$. Thus, we only need to consider e_{12}. The bounds on h and $\boldsymbol{D}h$, the finite second moment assumption on the process, and Lemma 3.4.2 can be used to show the following,

$$
\begin{aligned}
\left|e_{12}\right| &\le \left|\lambda_{\mathrm{M}}\right|^{k-i-1} K_1\sigma_X \\
&\le \left|\lambda_{\mathrm{M}}\right|^{k-i-1} K \text{ say.}
\end{aligned}
$$

Recall that we are still working under the assumption that $l = j - 1$ so that the sum in $E_{n2}(5)$ is $O\left(n^3\right)$. Thus, putting the above pieces together yields the following,

$$
\begin{aligned}
\left|E_{n2}(5)\right| &\le \frac{4}{n^3}\sum_{i<k<l<l+1}\left|E\left[h_{i-1}u_i h_{k-1}\psi_{lk}\varphi_{l-1}\varphi_l\right]\right| \\
&= \frac{4}{n^3}\sum_{i<k<l<l+1}\left|e_1 + e_2\right| \\
&= \frac{4}{n^3}\sum_{i<k<l<l+1}\left|e_{12} + e_2\right| \\
&\le \frac{4}{n^3}\sum_{i<k<l<l+1}\left(\left|e_{12}\right| + \left|e_2\right|\right) \\
&\le \frac{4}{n^3}\sum_{i<k<l<l+1}\left(\left|\lambda_{\mathrm{M}}\right|^{k-i-1} + \left|\lambda_{\mathrm{M}}\right|^{l-k-1}\right)K \\
&= O\left(n^{-3}\right)O\left(n^2\right) \\
&= o(1),
\end{aligned}
$$

since $|\lambda_{\mathrm{M}}| < 1$. Now suppose that $l < j-1$ so the sum in $E_{n2}(5)$ is still $O(n^4)$. Consider the expectation given in $E_{n2}(5)$ under the assumption that $l < j-1$. Using the fact that $\boldsymbol{Y}_{j-1} = \boldsymbol{P}^{j-l-1}\boldsymbol{Y}_l + \boldsymbol{\Gamma}_{l+1}^{j-1}$ in conjunction with the multivariate Mean Value Theorem on φ and independence yields the following,

$$
\begin{aligned}
& E\left[h_{i-1}u_ih_{k-1}\psi_{lk}\varphi_{l-1}\varphi\left(\boldsymbol{Y}_{j-1}\right)\right] \\
&= E\left[h_{i-1}u_ih_{k-1}\psi_{lk}\varphi_{l-1}\varphi\left(\boldsymbol{Y}^{j-l-1}\boldsymbol{Y}_l + \boldsymbol{\Gamma}_{l+1}^{j-1}\right)\right] \\
&= E\left[h_{i-1}u_ih_{k-1}\psi_{lk}\varphi_{l-1}\left(\varphi\left(\boldsymbol{\Gamma}_{l+1}^{j-1}\right) + \boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\right)'\boldsymbol{P}^{j-l-1}\boldsymbol{Y}_l\right)\right] \\
&= E\left[h_{i-1}u_ih_{k-1}\psi_{lk}\varphi_{l-1}\right]E\left[\varphi\left(\boldsymbol{\Gamma}_{l+1}^{j-1}\right)\right] + \\
&\quad\; E\left[h_{i-1}u_ih_{k-1}\psi_{lk}\varphi_{l-1}\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\right)'\boldsymbol{P}^{j-l-1}\boldsymbol{Y}_l\right] \\
&= e_1 + e_2 \text{ say.}
\end{aligned}
$$

This implies the following,

$$
\begin{aligned}
E_{n2}(5) &= \frac{4}{n^3}\sum_{i<k<l<j}(e_1 + e_2) \\
&= E_{n21}(5) + E_{n22}(5) \text{ say.}
\end{aligned}
$$

Consider $E_{n21}(5)$ first. Since φ is bounded we will absorb the second expectation into the notation for e_1 and consider the first expectation only. That is,

$$
\begin{aligned}
e_1 &= E\left[h_{i-1}u_ih_{k-1}\psi_{lk}\varphi_{l-1}\right] \\
&= E\left[h_{i-1}u_ih_{k-1}\psi_{lk}\varphi\left(\boldsymbol{Y}_{l-1}\right)\right].
\end{aligned}
$$

Now consider the case when $k = l - 1$ and note that this reduces the sum in $E_{n21}(5)$ to $O(n^3)$. Thus, we have the following,

$$\begin{aligned} e_1 &= E\left[h_{i-1}u_i h_{k-1}\psi_{k+1,k}\varphi_k\right] \\ &= E\left[h_{i-1}u_i h_{k-1}\psi_{k+1,k}\varphi\left(\boldsymbol{PY}_{k-1}+\varepsilon_k\right)\right] \\ &= E\left[h_{i-1}u_i E\left[h_{k-1}\psi_{k+1,k}\varphi\left(\boldsymbol{PY}_{k-1}+\varepsilon_k\right) \mid \mathcal{F}_{k-1}\right]\right] \\ &= E\left[h_{i-1}u_i g\left(\boldsymbol{Y}_{k-1}\right)\right] \end{aligned}$$

where,

$$\begin{aligned} & g\left(\boldsymbol{Y}_{k-1}\right) \\ &= E\left[h_{k-1}\psi_{k+1,k}\varphi\left(\boldsymbol{PY}_{k-1}+\varepsilon_k\right) \mid \mathcal{F}_{k-1}\right] \\ &= \int\int h\left(\boldsymbol{Y}_{k-1}\right)\left(I\left(t_1 \le t_2\right)-F\left(t_2\right)\right)\varphi\left(\boldsymbol{PY}_{k-1}+t_1\right)f\left(t_1\right)f\left(t_2\right)dt_1dt_2. \end{aligned}$$

It follows from the bounds on h and φ that g is a bounded function of $\boldsymbol{Y}_{k-1}$. Furthermore, assuming interchangeability of differentiation and integration it follows from the differentiability of h and φ that g is differentiable and hence continuous. Lastly, it follows from the bounds on h, $\boldsymbol{D}h$, φ, and $\boldsymbol{D}\varphi$ that $\boldsymbol{D}g$ is uniformly bounded in all arguments. Recall that we are considering the case where $k = l-1$. Thus, in the case of $i = k-1$ the sum in $E_{n21}(5)$ reduces to $O(n^2)$. Since our assumptions imply that h and g are bounded and we are dividing by n^3 it follows that $E_{n21}(5) = o(1)$ in the case of $i = k-1 < k = l-1$. Thus, consider the case where $i < k-1$. Using the fact that $\boldsymbol{Y}_{k-1} = \boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i + \boldsymbol{\Gamma}_{i+1}^{k-1}$ in conjunction with

the multivariate Mean Value Theorem on g and independence yields the following,

$$\begin{aligned}
e_1 &= E\left[h_{i-1}u_i g\left(\boldsymbol{Y}_{k-1}\right)\right] \\
&= E\left[h_{i-1}u_i g\left(\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i + \boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right] \\
&= E\left[h_{i-1}u_i\left(g\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right) + \boldsymbol{D}g\left(\boldsymbol{\xi}_{ik}\right)'\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right)\right] \\
&= E\left[h_{i-1}\right]E\left[u_i\right]E\left[g\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right] + E\left[h_{i-1}u_i\boldsymbol{D}g\left(\boldsymbol{\xi}_{ik}\right)'\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right] \\
&= e_{11} + e_{12} \text{ say.}
\end{aligned}$$

It follows from the fact that $E\left[u_i\right] = 0$ that $e_{11} = 0$. Furthermore, the bounds on h and $\boldsymbol{D}g$, the finite second moment assumption on the process, and Lemma 3.4.2 can be used to show the following,

$$\begin{aligned}
\left|e_{12}\right| &\leq \left|\lambda_{\mathrm{M}}\right|^{k-i-1} K_1\sigma_X \\
&\leq \left|\lambda_{\mathrm{M}}\right|^{k-i-1} K \text{ say.}
\end{aligned}$$

Recall that we are dealing with $E_{n21}(5)$ under the assumption that $k = l - 1$. Thus, putting the above pieces together we obtain,

$$\begin{aligned}
\left|E_{n21}(5)\right| &\leq \frac{4}{n^3}\sum_{i<k<k+1<j}\left|e_1\right| \\
&= \frac{4}{n^3}\sum_{i<k<k+1<j}\left|e_{11} + e_{12}\right| \\
&\leq \frac{4}{n^3}\sum_{i<k<k+1<j}\left|\lambda_{\mathrm{M}}\right|^{k-i-1} K \\
&= O\left(n^{-3}\right)O\left(n^2\right) \\
&= o(1),
\end{aligned}$$

since $|\lambda_{\mathrm{M}}| < 1$. Next, consider $E_{n21}(5)$ under the assumption that $k < l-1$. Again, consider only e_1. Conditioning on $\mathcal{F}_{l-1}$ and using the fact that,

$$E[\psi_{lk} \mid \mathcal{F}_{l-1}] = \frac{1}{2} - F(\varepsilon_k) = u_k,$$

we can show the following,

$$\begin{aligned} e_1 &= E[h_{i-1}u_i h_{k-1}\varphi_{l-1}\psi_{lk}] \\ &= E[h_{i-1}u_i h_{k-1}\varphi_{l-1}E[\psi_{lk} \mid \mathcal{F}_{l-1}]] \\ &= E[h_{i-1}u_i h_{k-1}u_k\varphi(\boldsymbol{Y}_{l-1})]. \end{aligned}$$

Using the fact that $\boldsymbol{Y}_{l-1} = \boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}$ in conjunction with the multivariate Mean Value Theorem on φ and independence yields the following,

$$\begin{aligned} e_1 &= E[h_{i-1}u_i h_{k-1}u_k\varphi(\boldsymbol{Y}_{l-1})] \\ &= E\left[h_{i-1}u_i h_{k-1}u_k\varphi\left(\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right] \\ &= E\left[h_{i-1}u_i h_{k-1}u_k\left(\varphi\left(\boldsymbol{\Gamma}_{k+1}^{l-1}\right) + \boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})'\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right)\right] \\ &= E[h_{i-1}u_i h_{k-1}]E[u_k]E\left[\varphi\left(\boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right] + E\left[h_{i-1}u_i h_{k-1}u_k\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})'\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right] \\ &= e_{11} + e_{12} \text{ say.} \end{aligned}$$

It follows from the fact that $E[u_k] = 0$ that $e_{11} = 0$. Now consider e_{12},

$$\begin{aligned} e_{12} &= E\left[h_{i-1}u_i h_{k-1}u_k\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})'\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right] \\ &= E\left[h_{i-1}u_i h_{k-1}u_k\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})'\boldsymbol{P}^{l-k-1}(\boldsymbol{P}\boldsymbol{Y}_{k-1} + \boldsymbol{\varepsilon}_k)\right] \\ &= E\left[h_{i-1}u_i E\left[h(\boldsymbol{Y}_{k-1})u(\varepsilon_k)\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})'\boldsymbol{P}^{l-k-1}(\boldsymbol{P}\boldsymbol{Y}_{k-1} + \boldsymbol{\varepsilon}_k) \mid \mathcal{F}_{k-1}\right]\right] \\ &= E[h_{i-1}u_i g(\boldsymbol{Y}_{k-1})] \text{ say,} \end{aligned}$$

where,

$$g(\boldsymbol{Y})$$

$$= \int h(\boldsymbol{Y})\, u(\varepsilon_k)\, \boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl}(\boldsymbol{Y}))' \boldsymbol{P}^{l-k-1}(\boldsymbol{PY} + \boldsymbol{\varepsilon}_k)\, f(\varepsilon_k) \cdots f(\varepsilon_{l-1})\, d\varepsilon_k \cdots d\varepsilon_{l-1}$$

$$= \int u(\varepsilon_k)\, \boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl}(\boldsymbol{Y}))' \boldsymbol{P}^{l-k-1}(\boldsymbol{P}h(\boldsymbol{Y})\boldsymbol{Y} + h(\boldsymbol{Y})\boldsymbol{\varepsilon}_k)\, dP \text{ say.}$$

Recall that W4 and W5 imply that $\boldsymbol{D}\varphi$, $h(\boldsymbol{Y})\boldsymbol{Y}$, and h are uniformly bounded. Therefore, the bound on F, the finite second moment assumption on ε_k, and Lemma 3.4.2 can be used to show the following,

$$|g(\boldsymbol{Y})| \leq K|\lambda_{\mathrm{M}}|^{l-k-1},$$

where K depends on the bounds given by W4 and W5, F, and σ_ε. Note that in the case of $i = k-1$ the sum in $E_{n21}(5)$ reduces to $O(n^3)$. Since our assumptions imply that $h_{i-1}u_i$ is bounded, $|g(\boldsymbol{Y}_{k-1})| \leq K|\lambda_{\mathrm{M}}|^{l-k-1}$, and we are dividing by n^3 it follows that $E_{n21}(5) = o(1)$ when $i = k-1$. Thus, consider the case when $i < k-1$. First, consider the following function of $\boldsymbol{Y}$,

$$G(\boldsymbol{Y}) = \boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl}(\boldsymbol{Y}))' \boldsymbol{P}^{l-k-1}(\boldsymbol{P}h(\boldsymbol{Y})\boldsymbol{Y} + h(\boldsymbol{Y})\boldsymbol{\varepsilon}_k).$$

Note that assumption W3 and Lemma 3.4.3 imply that G is differentiable with respect to $\boldsymbol{Y}$. Furthermore, direct computations can be used to show the following,

$$\begin{aligned} \boldsymbol{D}G(\boldsymbol{Y}) = {} & \left(\frac{d}{d\boldsymbol{Y}}\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl}(\boldsymbol{Y}))\right) \boldsymbol{P}^{l-k-1}(\boldsymbol{P}h(\boldsymbol{Y})\boldsymbol{Y} + h(\boldsymbol{Y})\boldsymbol{\varepsilon}_k) + \\ & \boldsymbol{P}^{l-k-1}(\boldsymbol{PD}h(\boldsymbol{Y})\boldsymbol{Y}' + \boldsymbol{PI}h(\boldsymbol{Y}) + \boldsymbol{D}h(\boldsymbol{Y})\boldsymbol{\varepsilon}_k')\, \boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl}(\boldsymbol{Y})). \end{aligned}$$

Now, assuming interchangeability of differentiation and integration it follows from the differentiability of G that g is differentiable. Thus, we will let dg_i and dG_i denote the i^{th} components of $\boldsymbol{D}g$ and $\boldsymbol{D}G$ respectively. That is,

$$dg_i = \int u(\varepsilon_k)\, dG_i dP.$$

Now, using Lemmas 3.4.3 and 3.4.2 along with the bounds given by W4, W5, F, and σ_ε one can show the following,

$$|dg_i| \leq K |\lambda_{\mathrm{M}}|^{l-k-1},$$

where K depends on the bounds for φ, h, $\boldsymbol{D}\varphi$, $\boldsymbol{D}h$, σ_ε, and the constants given in Lemmas 3.4.3 and 3.4.2. Thus, it follows that $\boldsymbol{D}g$ is also bounded. Let us now return to the problem at hand. Recall that we are considering E_{n21} (5) under the case when $i < k-1$. Using the fact that $\boldsymbol{Y}_{k-1} = \boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i + \boldsymbol{\Gamma}_{i+1}^{k-1}$ and the multivariate Mean Value Theorem on g and independence yields the following,

$$\begin{aligned}
e_{12} &= E\left[h_{i-1}u_i g\left(\boldsymbol{Y}_{k-1}\right)\right] \\
&= E\left[h_{i-1}u_i g\left(\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i + \boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right] \\
&= E\left[h_{i-1}u_i\left(g\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right) + \boldsymbol{D}g\left(\boldsymbol{\xi}_{ik}\right)'\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right)\right] \\
&= E\left[h_{i-1}\right]E\left[u_i\right]E\left[g\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right] + E\left[h_{i-1}u_i\boldsymbol{D}g\left(\boldsymbol{\xi}_{ik}\right)'\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right] \\
&= e_{121} + e_{122} \text{ say.}
\end{aligned}$$

Again, $E[u_i] = 0$ implies that $e_{121} = 0$. Thus, consider e_{122}. The bounds on h, u, and $\boldsymbol{D}g$; the finite second moment of the process; and Lemma 3.4.2 can be used

to show the following,

$$
\begin{aligned}
|e_{122}| &\leq K_1 E\left[\left|\boldsymbol{D}g\left(\boldsymbol{\xi}_{ik}\right)' \boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right|\right] \\
&= K_1 E\left[\left|\sum_{a=1}^{p}\sum_{b=1}^{p} dg_a P_{ab}^{(k-i-1)} X_{i-b}\right|\right] \\
&\leq K_1 \sum_{a=1}^{p}\sum_{b=1}^{p}\left|dg_a P_{ab}^{(k-i-1)}\right| E\left[|X_{i-b}|\right] \\
&\leq K_2 \sum_{a=1}^{p}\sum_{b=1}^{p} |\lambda_{\mathrm{M}}|^{l-k-1} |\lambda_{\mathrm{M}}|^{k-i-1} E\left[|X_{i-b}|\right] \\
&\leq |\lambda_{\mathrm{M}}|^{k-i-1} |\lambda_{\mathrm{M}}|^{l-k-1} p^2 K_2 \sigma_X \\
&\leq |\lambda_{\mathrm{M}}|^{k-i-1} |\lambda_{\mathrm{M}}|^{l-k-1} K \text{ say.}
\end{aligned}
$$

Putting the pieces together we obtain the following,

$$
\begin{aligned}
|E_{n21}(5)| &\leq \frac{4}{n^3}\sum_{i<k<l<j} |e_1| \\
&= \frac{4}{n^3}\sum_{i<k<l<j} |e_{11} + e_{121} + e_{122}| \\
&\leq \frac{4}{n^3}\sum_{i<k<l<j} |\lambda_{\mathrm{M}}|^{l-k-1} |\lambda_{\mathrm{M}}|^{k-i-1} K \\
&= O\left(n^{-3}\right) O\left(n^2\right) \\
&= o(1),
\end{aligned}
$$

since $|\lambda_{\mathrm{M}}| < 1$. This completes the proof of $E_{n21}(5) = o(1)$. To complete the proof we need to consider the following,

$$
\begin{aligned}
E_{n22}(5) &= \frac{4}{n^3}\sum_{i<k<l<j} e_2 \\
&= \frac{4}{n^3}\sum_{i<k<l<j} E\left[h_{i-1}u_i h_{k-1}\psi_{lk}\varphi_{l-1}\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\right)' \boldsymbol{P}^{j-l-1}\boldsymbol{Y}_l\right].
\end{aligned}
$$

Consider only the expectation given in $E_{n22}(5)$,

e_2

$$= E\left[h_{i-1}u_i h_{k-1}\psi_{lk}\varphi_{l-1}\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\right)'\boldsymbol{P}^{j-l-1}\boldsymbol{Y}_l\right]$$
$$= E\left[h_{i-1}u_i h_{k-1}\psi_{lk}\varphi\left(\boldsymbol{Y}_{l-1}\right)\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\left(\boldsymbol{Y}_{l-1}\right)\right)'\boldsymbol{P}^{j-l-1}\left(\boldsymbol{P}\boldsymbol{Y}_{l-1}+\boldsymbol{\varepsilon}_l\right)\right]$$
$$= E\left[h_{i-1}u_i h_{k-1}E\left[\psi_{lk}\varphi\left(\boldsymbol{Y}_{l-1}\right)\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\left(\boldsymbol{Y}_{l-1}\right)\right)'\boldsymbol{P}^{j-l-1}\left(\boldsymbol{P}\boldsymbol{Y}_{l-1}+\boldsymbol{\varepsilon}_l\right)\mid\mathcal{F}_{l-1}\right]\right]$$
$$= E\left[h_{i-1}u_i h_{k-1}g\left(\varepsilon_k;\boldsymbol{Y}_{l-1}\right)\right]\text{ say,}$$

where,

$$g\left(\varepsilon_k;\boldsymbol{Y}\right)$$
$$=\int\psi\left(\varepsilon_k;\varepsilon_l\right)\varphi\left(\boldsymbol{Y}\right)\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\left(\boldsymbol{Y}\right)\right)'\boldsymbol{P}^{j-l-1}\left(\boldsymbol{P}\boldsymbol{Y}+\boldsymbol{\varepsilon}_l\right)f\left(\varepsilon_l\right)\cdots f\left(\varepsilon_{j-1}\right)d\varepsilon_l\cdots d\varepsilon_{j-1}$$
$$=\int\psi\left(\varepsilon_k;\varepsilon_l\right)\varphi\left(\boldsymbol{Y}\right)\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\left(\boldsymbol{Y}\right)\right)'\boldsymbol{P}^{j-l-1}\left(\boldsymbol{P}\boldsymbol{Y}+\boldsymbol{\varepsilon}_l\right)dP$$
$$=\int\psi\left(\varepsilon_k;\varepsilon_l\right)G\left(\boldsymbol{Y}\right)dP\text{ say.}$$

Next, multivariate differentiation can be used to show the following,

$$\begin{aligned}\boldsymbol{D}G\left(\boldsymbol{Y}\right) &= \boldsymbol{D}\varphi\left(\boldsymbol{Y}\right)\left(\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\right)'\boldsymbol{P}^{j-l-1}\left(\boldsymbol{P}\boldsymbol{Y}+\boldsymbol{\varepsilon}_l\right)\right)+\\ &\quad\varphi\left(\boldsymbol{Y}\right)\left(\left(\frac{d}{d\boldsymbol{Y}}\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\right)\right)\boldsymbol{P}^{j-l-1}\left(\boldsymbol{P}\boldsymbol{Y}+\boldsymbol{\varepsilon}_l\right)+\boldsymbol{P}^{j-l-1}\boldsymbol{P}\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\right)\right).\end{aligned}$$

Now, assuming interchangeability of differentiation and integration one can use the bounds on F, φ, and $\boldsymbol{D}\varphi$ along with Lemma 3.4.2 to show the following,

$$E\left[\left|g\left(\varepsilon_k;\boldsymbol{Y}_{l-1}\right)\right|\mid\mathcal{F}_{l-1}\right]\ \le\ \left|\lambda_{\mathrm{M}}\right|^{j-l-1}\left(K_1\left\|\boldsymbol{Y}_{l-1}\right\|+K_2\right),$$

where K_1 and K_2 depend on the bounds for F, φ, $\boldsymbol{D}\varphi$, σ_ε, and the constant given in Lemma 3.4.2. Similarly, one can use the bounds on F, φ, and $\boldsymbol{D}\varphi$ along with Lemmas 3.4.2 and 3.4.3 to show the following,

$$E\left[\left\|\boldsymbol{D}g\left(\varepsilon_k;\boldsymbol{Y}_{l-1}\right)\right\|\mid\mathcal{F}_{l-1}\right]\ \le\ \left|\lambda_{\mathrm{M}}\right|^{j-l-1}\left(K_1\left\|\boldsymbol{Y}_{l-1}\right\|+K_2\right),$$

where K_1 and K_2 depend on the bounds for F, φ, $\boldsymbol{D}\varphi$, σ_ε, and the constants given in Lemmas 3.4.2 and 3.4.3. Note that the above two results imply that $E\left[|g(\varepsilon_k;\boldsymbol{Y}_{l-1})|\right]$ and $E\left[||\boldsymbol{D}g(\varepsilon_k;\boldsymbol{Y}_{l-1})||\right]$ can be bounded by a constant that depends on K_1, K_2, and σ_X. Now note that when $k = l-1$ the sum in $E_{n22}(5)$ reduces to $O(n^3)$. Since h and u are bounded, $E\left[|g(\varepsilon_k;\boldsymbol{Y}_{l-1})|\right] \leq K\left|\lambda_{\mathrm{M}}\right|^{j-l-1}$, and we are dividing by n^3 it follows that $E_{n22}(5) = o(1)$ when $k = l-1$. Hence, consider the case when $k < l-1$. Using the fact that $\boldsymbol{Y}_{l-1} = \boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}$ in conjunction with the multivariate Mean Value Theorem on $g(\cdot;\boldsymbol{Y})$ and independence yields the following,

$$
\begin{aligned}
e_2 &= E\left[h_{i-1}u_i h_{k-1} g(\varepsilon_k;\boldsymbol{Y}_{l-1})\right] \\
&= E\left[h_{i-1}u_i h_{k-1} g\left(\varepsilon_k;\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right] \\
&= E\left[h_{i-1}u_i h_{k-1}\left(g\left(\varepsilon_k;\boldsymbol{\Gamma}_{k+1}^{l-1}\right) + \boldsymbol{D}g(\varepsilon_k;\boldsymbol{\xi}_{kl})'\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right)\right] \\
&= E\left[h_{i-1}u_i h_{k-1}\right]E\left[g\left(\varepsilon_k;\boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right] + E\left[h_{i-1}u_i h_{k-1}\boldsymbol{D}g(\varepsilon_k;\boldsymbol{\xi}_{kl})'\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right] \\
&= e_{21} + e_{22} \text{ say.}
\end{aligned}
$$

Consider e_{22} first. To begin with, note the following,

$$
\begin{aligned}
&||\boldsymbol{D}G(\boldsymbol{\xi}_{kl})|| \\
&\leq \left\|\boldsymbol{D}\varphi(\boldsymbol{\xi}_{kl})\left(\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\right)'\boldsymbol{P}^{j-l-1}(\boldsymbol{P}\boldsymbol{\xi}_{kl} + \varepsilon_l)\right)\right\| + \\
&\quad \left\|\varphi(\boldsymbol{\xi}_{kl})\left(\frac{d}{d\boldsymbol{Y}}\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\right)\boldsymbol{P}^{j-l-1}(\boldsymbol{P}\boldsymbol{\xi}_{kl} + \varepsilon_l) + \boldsymbol{P}^{j-l-1}\boldsymbol{P}\boldsymbol{D}\varphi\left(\boldsymbol{\xi}_{lj}\right)\right)\right\| \\
&\leq K_1\left\|\boldsymbol{P}^{j-l-1}\right\|\,||\boldsymbol{P}\boldsymbol{\xi}_{kl} + \varepsilon_l|| + K_2\left\|\boldsymbol{P}^{j-l-1}\right\|\,||\boldsymbol{P}\boldsymbol{\xi}_{kl} + \varepsilon_l|| + K_3\left\|\boldsymbol{P}^{j-l-1}\right\| \\
&\leq K\left|\lambda_{\mathrm{M}}\right|^{j-l-1}\left(||\boldsymbol{P}\boldsymbol{\xi}_{kl} + \varepsilon_l|| + 1\right),
\end{aligned}
$$

where K depends on the bounds for φ, $\boldsymbol{D}\varphi$, $\boldsymbol{P}$, and the constants given in Lemmas 3.4.2 and 3.4.3. Now recall that $\boldsymbol{\xi}_{kl}$ lies on the line segment connecting $\boldsymbol{Y}_{l-1} = \boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}$ and $\boldsymbol{\Gamma}_{k+1}^{l-1}$. That is,

$$\begin{aligned}
\boldsymbol{\xi}_{kl} &= (1-\lambda)\left(\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{l-1}\right) + \lambda\boldsymbol{\Gamma}_{k+1}^{l-1} \\
&= \boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k + \boldsymbol{\Gamma}_{k+1}^{l-1} - \lambda\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k \\
&= \boldsymbol{Y}_{l-1} - \lambda\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k,
\end{aligned}$$

for some $\lambda \in (0,1)$. The above two results now imply the following,

$$\begin{aligned}
\|\boldsymbol{D}G(\boldsymbol{\xi}_{kl})\| &\leq K\,|\lambda_{\mathrm{M}}|^{j-l-1}\left(\|\boldsymbol{P}\boldsymbol{\xi}_{kl} + \boldsymbol{\varepsilon}_l\| + 1\right) \\
&\leq K\,|\lambda_{\mathrm{M}}|^{j-l-1}\left(\left\|\boldsymbol{P}\boldsymbol{Y}_{l-1} - \lambda\boldsymbol{P}^{l-k}\boldsymbol{Y}_k + \boldsymbol{\varepsilon}_l\right\| + 1\right) \\
&\leq K\,|\lambda_{\mathrm{M}}|^{j-l-1}\left(\|\boldsymbol{Y}_{l-1}\| + \|\boldsymbol{Y}_k\| + \|\boldsymbol{\varepsilon}_l\| + 1\right),
\end{aligned}$$

where the bounds for $\boldsymbol{P}$ and $\boldsymbol{P}^{l-k}$ are absorbed into the constant K. Let us now return to $\boldsymbol{D}g$,

$$\begin{aligned}
\|\boldsymbol{D}g(\varepsilon_k;\boldsymbol{\xi}_{kl})\| &\leq \int |\psi_{kl}|\,\|\boldsymbol{D}G(\boldsymbol{\xi}_{kl})\|\,dP \\
&\leq K\int \|\boldsymbol{D}G(\boldsymbol{\xi}_{kl})\|\,dP \\
&\leq K\int K\,|\lambda_{\mathrm{M}}|^{j-l-1}\left(\|\boldsymbol{Y}_{l-1}\| + \|\boldsymbol{Y}_k\| + \|\boldsymbol{\varepsilon}_l\| + 1\right)dP \\
&\leq K\,|\lambda_{\mathrm{M}}|^{j-l-1}\left(\|\boldsymbol{Y}_{l-1}\| + \|\boldsymbol{Y}_k\| + \sigma_\varepsilon + 1\right),
\end{aligned}$$

since dP depends on ε_l, ε_{l+1}, $\ldots$, ε_{j-1} and we are working under the case where $i < k < l < j$. This result now implies the following,

$$\left|\boldsymbol{D}g(\varepsilon_k;\boldsymbol{\xi}_{kl})'\,\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right|$$

$$\leq \|\boldsymbol{D}g(\varepsilon_k;\boldsymbol{\xi}_{kl})\| \left\|\boldsymbol{P}^{l-k-1}\right\| \|\boldsymbol{Y}_k\|$$

$$\leq K|\lambda_{\mathrm{M}}|^{j-l-1}|\lambda_{\mathrm{M}}|^{l-k-1}(\|\boldsymbol{Y}_{l-1}\| + \|\boldsymbol{Y}_k\| + \sigma_\varepsilon + 1)\|\boldsymbol{Y}_k\|$$

$$= K|\lambda_{\mathrm{M}}|^{j-l-1}|\lambda_{\mathrm{M}}|^{l-k-1}\left(\|\boldsymbol{Y}_{l-1}\|\|\boldsymbol{Y}_k\| + \|\boldsymbol{Y}_k\|^2 + \sigma_\varepsilon\|\boldsymbol{Y}_k\| + \|\boldsymbol{Y}_k\|\right).$$

Note that the finite second moment assumption on the process and the Cauchy Schwarz Inequality can be used to show the following,

$$E\left[\left|\boldsymbol{D}g(\varepsilon_k;\boldsymbol{\xi}_{kl})'\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right|\right] \leq K|\lambda_{\mathrm{M}}|^{j-l-1}|\lambda_{\mathrm{M}}|^{l-k-1},$$

where K now also depends on σ_X. Let us now return to e_{22}. The bound on h and the above result imply the following,

$$\begin{aligned} |e_{22}| &\leq E\left[\left|h_{i-1}u_ih_{k-1}\boldsymbol{D}g(\varepsilon_k;\boldsymbol{\xi}_{kl})'\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_k\right|\right] \\ &\leq KE\left[\left|\boldsymbol{D}g(\varepsilon_k;\boldsymbol{\xi}_{kl})'\boldsymbol{P}^{l-k-1}\boldsymbol{Y}_{kk}\right|\right] \\ &\leq K|\lambda_{\mathrm{M}}|^{j-l-1}|\lambda_{\mathrm{M}}|^{l-k-1}. \end{aligned}$$

Now recall that,

$$e_{21} = E[h_{i-1}u_ih_{k-1}]E\left[g\left(\varepsilon_k;\boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right].$$

Similar to the result on $E[|g(\varepsilon_k;\boldsymbol{Y}_{l-1})|]$, it can be show that,

$$E\left[\left|g\left(\varepsilon_k;\boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right|\right] \leq K|\lambda_{\mathrm{M}}|^{j-l-1},$$

where K depends on the bounds for $\boldsymbol{P}$ and σ_ε. Therefore, when $i = k-1$ the sum in $E_{n22}(5)$ reduces to $O(n^3)$. Since the bound on h implies the first expectation in e_{21} is bounded and we are dividing by n^3 the above result can be

used to show that $E_{n22}(5) = o(1)$ when $i = k-1$. Hence, consider the case when $i < k-1$. Consider only the first expectation given in e_{21}. Using the fact that $\boldsymbol{Y}_{k-1} = \boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i + \boldsymbol{\Gamma}_{i+1}^{k-1}$ in conjunction with the multivariate Mean Value Theorem and independence yields the following,

$$\begin{aligned} E\left[h_{i-1}u_i h_{k-1}\right] &= E\left[h_{i-1}u_i h\left(\boldsymbol{Y}_{k-1}\right)\right] \\ &= E\left[h_{i-1}u_i h\left(\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i + \boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right] \\ &= E\left[h_{i-1}u_i\left(h\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right) + \boldsymbol{D}h\left(\boldsymbol{\xi}_{ik}\right)'\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right)\right] \\ &= E\left[h_{i-1}\right]E\left[u_i\right]E\left[h\left(\boldsymbol{\Gamma}_{i+1}^{k-1}\right)\right] + E\left[h_{i-1}u_i\boldsymbol{D}h\left(\boldsymbol{\xi}_{ik}\right)'\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right] \\ &= e_{211} + e_{212} \text{ say.} \end{aligned}$$

It follows from the fact that $E\left[u_i\right] = 0$ that $e_{211} = 0$. Furthermore, the bounds on h and $\boldsymbol{D}h$, the finite second moment assumption on the process, and Lemma 3.4.2 can be used to show the following,

$$\begin{aligned} \left|e_{212}\right| &\leq E\left[\left|h_{i-1}u_i\boldsymbol{D}h\left(\boldsymbol{\xi}_{ik}\right)'\boldsymbol{P}^{k-i-1}\boldsymbol{Y}_i\right|\right] \\ &\leq K\sigma_X\left|\lambda_{\mathrm{M}}\right|^{k-i-1} \\ &= K\left|\lambda_{\mathrm{M}}\right|^{k-i-1} \text{ say.} \end{aligned}$$

Now, combine this result with the result on $E\left[\left|g\left(\varepsilon_k;\boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right|\right]$ to show the following,

$$\begin{aligned} \left|e_{21}\right| &\leq \left|E\left[h_{i-1}u_i h_{k-1}\right]E\left[g\left(\varepsilon_k;\boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right]\right| \\ &\leq \left|e_{211} + e_{212}\right| E\left[\left|g\left(\varepsilon_k;\boldsymbol{\Gamma}_{k+1}^{l-1}\right)\right|\right] \\ &\leq K\left|\lambda_{\mathrm{M}}\right|^{j-l-1}\left|\lambda_{\mathrm{M}}\right|^{k-i-1}. \end{aligned}$$

Putting the results on e_{21} and e_{22} together we obtain the following,

$$\begin{aligned}
|E_{n22}(5)| &\leq \frac{4}{n^3} \sum_{i<k<l<j} |e_2| \\
&\leq \frac{4}{n^3} \sum_{i<k<l<j} |e_{21} + e_{22}| \\
&\leq \frac{4}{n^3} \sum_{i<k<l<j} \left(|\lambda_{\mathrm{M}}|^{j-l-1} |\lambda_{\mathrm{M}}|^{k-i-1} + |\lambda_{\mathrm{M}}|^{j-l-1} |\lambda_{\mathrm{M}}|^{l-k-1}\right) K \\
&= O\left(n^{-3}\right) O\left(n^2\right) \\
&= o(1),
\end{aligned}$$

since $|\lambda_{\mathrm{M}}| < 1$. Since we have shown that both $E_{n21}(5)$ and $E_{n22}(5)$ are $o(1)$ it follows that $E_{n2}(5)$ is $o(1)$. As was mentioned earlier the proofs for $E_{n2}(1)$ and $E_{n2}(3)$ can be treated using similar techniques. This completes the proof of $E\left[S_{n23}^2\right] = o(1)$ which implies that $S_{n23}^2 = o_p(1)$. Therefore, the proof of $S_{n2} = o_p(1)$ is complete. □

Theorem 3.4.2 *Assumptions M1-M3; E1 (with second moment assumptions replaced with fourth moment assumptions); and W1, W4, and W7 imply the following,*

$$\begin{aligned}
S_{n2}(\boldsymbol{\rho}_0) &= \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} \boldsymbol{\lambda}' \boldsymbol{Y}_{i-1} \left(I\left(\varepsilon_j \leq \varepsilon_i\right) - F\left(\varepsilon_i\right)\right) \\
&= o_p(1).
\end{aligned}$$

proof. First, divide $S_{n2}(\boldsymbol{\rho}_0)$ into two sums. The first sum will contain the terms in which $b_{ij} = K$ and the second sum will contain the terms in which $b_{ij} \neq K$. Next, add in and subtract out a sum over $b_{ij} \neq K$ with the b_{ij}'s replaced with a

K. We then have the following,

$$\begin{aligned}
S_{n2}(\boldsymbol{\rho}_0) &= \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\left(I\left(\varepsilon_j \le \varepsilon_i\right) - F\left(\varepsilon_i\right)\right) \\
&= \frac{2K}{n^{\frac{3}{2}}}\sum_{b_{ij}=K} \boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\left(I\left(\varepsilon_j \le \varepsilon_i\right) - F\left(\varepsilon_i\right)\right) + \\
&\quad \frac{2}{n^{\frac{3}{2}}}\sum_{b_{ij}\neq K} b_{ij}\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\left(I\left(\varepsilon_j \le \varepsilon_i\right) - F\left(\varepsilon_i\right)\right) \\
&= \frac{2K}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\left(I\left(\varepsilon_j \le \varepsilon_i\right) - F\left(\varepsilon_i\right)\right) + \\
&\quad \frac{2}{n^{\frac{3}{2}}}\sum_{b_{ij}\neq K} \left(b_{ij}-1\right)\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\left(I\left(\varepsilon_j \le \varepsilon_i\right) - F\left(\varepsilon_i\right)\right) \\
&= S_{n21}(\boldsymbol{\rho}_0) + S_{n22}(\boldsymbol{\rho}_0) \text{ say.}
\end{aligned}$$

Consider $S_{n21}(\boldsymbol{\rho}_0)$ first. Upon examination it is apparent that $S_{n21}(\boldsymbol{\rho}_0) = \sum_{k=1}^{p} T_{nk}$ where,

$$T_{nk} = \frac{2K\lambda_k}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} X_{i-k}\left(I\left(\varepsilon_j \le \varepsilon_i\right) - F\left(\varepsilon_i\right)\right).$$

Hence, in order to show $S_{n21}(\boldsymbol{\rho}_0) = o_p(1)$ it suffices to show that $T_{nk} = o_p(1)$ for $k = 1, 2, \ldots, p$. However, in the proof of Theorem 2.4.2 we showed the following result,

$$\begin{aligned}
S_{n2}(\rho_0) &= \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}X_{i-1}\left(I\left(\varepsilon_j \le \varepsilon_i\right) - F(\varepsilon_i)\right) \\
&= o_p(1).
\end{aligned}$$

To show that $T_{nk} = o_p(1)$ for $k = 1, 2, \ldots, p$ one can use the same technique that was used to prove the above result. The only difference is that the b_{ij}'s are replaced by K and X_{i-1} is replaced by X_{i-k}. Upon doing so we get $S_{n21}(\boldsymbol{\rho}_0) = o_p(1)$. For

an alternative proof one is referred to Koul (1993) and his result on V_n which is essentially $S_{n21}(\boldsymbol{\rho}_0)$.

To complete the proof we must show that $S_{n22}(\boldsymbol{\rho}_0) = o_p(1)$. Under assumptions W1 and W7 we have $b_{ij} = h(\boldsymbol{Y}_{i-1};\boldsymbol{\theta})\,h(\boldsymbol{Y}_{j-1};\boldsymbol{\theta})$ where at most $[n^r]$ ($r < \frac{1}{2}$) of the $h(\boldsymbol{Y}_{i-1};\boldsymbol{\theta})$'s are different from K. That is, the number of elements in $\{h(\boldsymbol{Y}_0;\boldsymbol{\theta}), h(\boldsymbol{Y}_1;\boldsymbol{\theta}), \ldots, h(\boldsymbol{Y}_{n-1};\boldsymbol{\theta})\}$ such that $h(\boldsymbol{Y}_{i-1};\boldsymbol{\theta}) \neq K$ is at most $[n^r]$. This implies that the number of $b_{ij} \neq K$ is at most $N = n[n^r]$. Thus, $S_{n22}(\boldsymbol{\rho}_0)$ contains at most N terms. Now, use W4 to bound b_{ij} by B_1^2 and the bound on F to show the following,

$$\begin{aligned}
|S_{n22}(\boldsymbol{\rho}_0)| &= \left|\frac{2}{n^{\frac{3}{2}}} \sum_{b_{ij}\neq K} (b_{ij}-1)\,\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\left(I(\varepsilon_j \le \varepsilon_i) - F(\varepsilon_i)\right)\right| \\
&\le \frac{2}{n^{\frac{3}{2}}} \sum_{b_{ij}\neq K} |b_{ij}-1|\,|\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}|\,|I(\varepsilon_j \le \varepsilon_i) - F(\varepsilon_i)| \\
&\le \frac{4(B_2^2+1)}{n^{\frac{3}{2}}} \sum_{b_{ij}\neq K} |\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}| \\
&\le \frac{B}{n^{\frac{3}{2}}} \sum_{b_{ij}\neq K} |\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}| \\
&\le \frac{B|\lambda_1|}{n^{\frac{3}{2}}} \sum_{b_{ij}\neq K} |X_{i-1}| + \ldots + \frac{B|\lambda_p|}{n^{\frac{3}{2}}} \sum_{b_{ij}\neq K} |X_{i-p}| \\
&= T_{n1} + T_{n2} + \ldots + T_{np} \text{ say.}
\end{aligned}$$

Therefore, in order to show $S_{n22}(\boldsymbol{\rho}_0) = o_p(1)$ it suffices to show $T_{nk} = o_p(1)$ for $k = 1, 2, \ldots, p$. This will follow from Chebyshev's Inequality if we show $E[T_{nk}^2] = o(1)$. Now, stationarity and the finite second moment assumption imply the following,

$$E\left[T_{nk}^2\right] = E\left[\left(\frac{B|\lambda_k|}{n^{\frac{3}{2}}} \sum_{b_{ij}\neq K} |X_{i-k}|\right)^2\right]$$

$$
\begin{aligned}
&= E\left[\frac{B^2\lambda_k^2}{n^3}\sum_{b_{ij}\neq K} X_{i-k}^2\right] + E\left[\frac{B^2\lambda_k^2}{n^3}\sum_{b_{ij}\neq K}\sum_{b_{uv}\neq K} |X_{i-k}|\,|X_{u-k}|\right] \\
&= \frac{B^2\lambda_k^2}{n^3}\sum_{b_{ij}\neq K} E\left[X_{i-k}^2\right] + \frac{B^2\lambda_k^2}{n^3}\sum_{b_{ij}\neq K}\sum_{b_{uv}\neq K} E\left[|X_{i-k}|\,|X_{u-k}|\right] \\
&\leq \frac{B^2\lambda_k^2 N}{n^3}\sigma_{\mathrm{x}}^2 + \frac{B^2\lambda_k^2}{n^3}\sum_{b_{ij}\neq K}\sum_{b_{uv}\neq K}\sqrt{E\left[X_{i-k}^2\right]E\left[X_{u-k}^2\right]} \\
&= \frac{B^2\lambda_k^2 N}{n^3}\sigma_{\mathrm{x}}^2 + \frac{B^2\lambda_k^2 N^2}{n^3}\sigma_{\mathrm{x}}^2 \\
&= \left(\frac{N(N+1)}{n^3}\right) B^2\lambda_k^2\sigma_{\mathrm{x}}^2 \\
&\leq \left(\frac{1}{n^{2-r}} + \frac{1}{n^{1-2r}}\right) B^2\lambda_k^2\sigma_{\mathrm{x}}^2 \\
&= o(1).
\end{aligned}
$$

The last line follows since $r < \frac{1}{2}$. Thus, by Chebyshev's Inequality, we have that $S_{n22}(\rho_0) = o_p(1)$. This completes the proof of $S_{n2}(\rho_0) = o_p(1)$. □

3.4.3 The Main Result

The main result of this section is now presented in the following theorem.

Theorem 3.4.3 *Under assumptions M1-M3, E1, and W1-W6 we have the following:*

$$
\boldsymbol{S}_n(\boldsymbol{\rho}_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \frac{1}{3}E\left[\boldsymbol{Y}_0 h^2(\boldsymbol{Y}_0)\boldsymbol{Y}_0'\right]\right).
$$

proof. Since $\boldsymbol{S}_n(\boldsymbol{\rho}_0)$ is a vector we will use the Cramer-Wold device and show $\boldsymbol{\lambda}'\boldsymbol{S}_n(\boldsymbol{\rho}_0)$ is asymptotically normal where $\boldsymbol{\lambda} \in \Re^p$ is arbitrary but fixed. To begin, rewrite this linear combination as a sum of two components as follows,

$$
\boldsymbol{\lambda}'\boldsymbol{S}_n(\boldsymbol{\rho}_0) = \boldsymbol{\lambda}'\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\,\mathrm{sgn}\left((X_i - \boldsymbol{\rho}_0'\boldsymbol{Y}_{i-1}) - (X_j - \boldsymbol{\rho}_0'\boldsymbol{Y}_{j-1})\right)\boldsymbol{Y}_{i-1}
$$

$$
\begin{aligned}
&= \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\,\mathrm{sgn}\,(\varepsilon_i-\varepsilon_j)\,\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1} \\
&= \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\,(2I\,(\varepsilon_j\le\varepsilon_i)-1)\,\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1} \text{ a.e.} \\
&= \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\,(2I\,(\varepsilon_j\le\varepsilon_i)-2F(\varepsilon_i)+2F(\varepsilon_i)-1) \\
&= \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\,(2F(\varepsilon_i)-1+2I\,(\varepsilon_j\le\varepsilon_i)-2F(\varepsilon_i)) \\
&= \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\,(2F(\varepsilon_i)-1)+ \\
&\quad\ \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\,(I\,(\varepsilon_j\le\varepsilon_i)-F(\varepsilon_i)) \\
&= S_{n1}(\boldsymbol{\rho}_0)+S_{n2}(\boldsymbol{\rho}_0) \text{ say.}
\end{aligned}
$$

In what follows $S_{n1}(\boldsymbol{\rho}_0)$ will be shown to be asymptotically normal using Theorem 2.4.1 and $S_{n2}(\boldsymbol{\rho}_0)$ will be shown to be $o_p(1)$. Thus, the asymptotic normality of $\boldsymbol{S}_n(\boldsymbol{\rho}_0)$ will follow from an application of Slutsky's Theorem and the Cramer-Wold device (see Serfling (1980, pg. 18–19)). We will start by showing that $S_{n1}(\boldsymbol{\rho}_0)$ is asymptotically normal. For convenience, let $u_i = 2F(\varepsilon_i)-1$. It should be noted that u_i is uniformly distributed over $(-1,1)$ so that $E[u_i] = 0$ and $V[u_i] = \frac{1}{3}$. Now, under W1 we can write $S_{n1}(\boldsymbol{\rho}_0)$ as follows,

$$
\begin{aligned}
S_{n1}(\boldsymbol{\rho}_0) &= \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} h(\boldsymbol{Y}_{i-1})\,h(\boldsymbol{Y}_{j-1})\,\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\,(2F(\varepsilon_i)-1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(\boldsymbol{Y}_{i-1})\,\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}u_i\left(\frac{1}{n}\sum_{j=1}^{n} h(\boldsymbol{Y}_{j-1})\right) \\
&= \bar{h}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}h(\boldsymbol{Y}_{i-1})\,\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}u_i \text{ say.}
\end{aligned}
$$

Before we proceed any further, one should note that $h(\boldsymbol{Y}_{i-1}) = h(\boldsymbol{Y}_{i-1};\boldsymbol{\theta})$. That is, the only random element in h is $\boldsymbol{Y}_{i-1}$. The fact that $h\left(\boldsymbol{Y}_{i-1};\hat{\boldsymbol{\theta}}_n\right)$ can be

replaced by $h(\boldsymbol{Y}_{i-1};\boldsymbol{\theta})$ is shown in the proof of Theorem 3.5.1. Now, the Ergodic Theorem implies $\overline{h} = E[h(\boldsymbol{Y}_0)] + o_p(1)$. Thus, we only need to show the second term is asymptotically normal. To proceed, define

$$Z_{n,t} = \frac{1}{\sqrt{n}} h(\boldsymbol{Y}_{t-1}) \boldsymbol{\lambda}' \boldsymbol{Y}_{t-1} u_t.$$

Then, $Z_{n,t}$ is $\mathcal{F}_{n,t}$ measurable since $\mathcal{F}_{n,t} = \sigma\text{-field}\{\boldsymbol{Y}_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_t\}$. Secondly, using W5 and the fact that $|u_t| \leq 1$ it is easily shown that $E\left[Z_{n,t}^2\right] \leq K < \infty$ where K is a constant that depends only on B_2 and $\boldsymbol{\lambda}$. Next, it follows from the independence of u_t and $\mathcal{F}_{n,t-1}$ and the fact that $E[u_t] = 0$ that,

$$\begin{aligned} E[Z_{n,t} \mid \mathcal{F}_{n,t-1}] &= E\left[\frac{1}{\sqrt{n}} h(\boldsymbol{Y}_{t-1}) \boldsymbol{\lambda}' \boldsymbol{Y}_{t-1} u_t \mid \mathcal{F}_{n,t-1}\right] \\ &= \frac{1}{\sqrt{n}} h(\boldsymbol{Y}_{t-1}) \boldsymbol{\lambda}' \boldsymbol{Y}_{t-1} E[u_t \mid \mathcal{F}_{n,t-1}] \\ &= 0. \end{aligned}$$

Now let,

$$\begin{aligned} S_{n,j} &= \frac{1}{\sqrt{n}} \sum_{t=1}^{j} h(\boldsymbol{Y}_{t-1}) \boldsymbol{\lambda}' \boldsymbol{Y}_{t-1} u_t \\ &= \sum_{t=1}^{j} Z_{n,t}. \end{aligned}$$

Since the three conditions in Definition 2.4.1 are satisfied, $\{S_{n,j}, \mathcal{F}_{n,j}\}$ is a zero-mean square-integrable martingale array with differences $Z_{n,j}$. Thus, we only need to show that the four conditions of Theorem 2.4.1 are satisfied to prove that $S_{n1}(\boldsymbol{\rho}_0)$ is asymptotically normal. Consider the first condition of the theorem.

Using W5 to bound $h(\boldsymbol{Y}_{t-1})\boldsymbol{\lambda}'\boldsymbol{Y}_{t-1}$ and the fact that $|u_t| \leq 1$ we get the following,

$$\begin{aligned}
\max_{1\leq i\leq n} |Z_{n,i}| &= \max_{1\leq i\leq n}\left|\frac{1}{\sqrt{n}}h(\boldsymbol{Y}_{i-1})\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}u_i\right| \\
&\leq \frac{1}{\sqrt{n}}K \\
&= o_p(1),
\end{aligned}$$

where K is a constant depending only on B_2 and $\boldsymbol{\lambda}$. Thus, the first condition of the theorem is satisfied. Consider the second condition next. Using the independence of $\boldsymbol{Y}_{i-1}$ and u_i the Ergodic Theorem can be used to show the following,

$$\begin{aligned}
\sum_{i=1}^{n} Z_{n,i}^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(h(\boldsymbol{Y}_{i-1})\boldsymbol{\lambda}'\boldsymbol{Y}_{i-1}\right)^2 u_i^2 \\
&= E\left[\left(h(\boldsymbol{Y}_0)\boldsymbol{\lambda}'\boldsymbol{Y}_0\right)^2\right]E\left[u_1^2\right] + o_p(1) \\
&= \frac{1}{3}E\left[\boldsymbol{\lambda}'\boldsymbol{Y}_0 h^2(\boldsymbol{Y}_0)\boldsymbol{Y}_0'\boldsymbol{\lambda}\right] + o_p(1) \\
&= \boldsymbol{\lambda}'E\left[\boldsymbol{Y}_0 h^2(\boldsymbol{Y}_0)\boldsymbol{Y}_0'\right]\boldsymbol{\lambda} + o_p(1) \\
&= \eta^2 + o_p(1) \text{ say.}
\end{aligned}$$

This verifies the second condition. Now, the above derivations and stationarity imply the following,

$$\begin{aligned}
E\left[\max_{1\leq i\leq n}\left|Z_{n,i}^2\right|\right] &\leq E\left[\sum_{i=1}^{n} Z_{n,i}^2\right] \\
&= \frac{1}{3}E\left[\boldsymbol{\lambda}'\boldsymbol{Y}_0 h^2(\boldsymbol{Y}_0)\boldsymbol{Y}_0'\boldsymbol{\lambda}\right] \\
&= O(1).
\end{aligned}$$

Hence, the third condition of the theorem is satisfied. For verification of condition four one is referred to the "Remarks" paragraph on page 59 of Hall and

Heyde (1980). Since all four conditions of Theorem 2.4.1 have been verified, we have proven that,

$$S_{n1}(\boldsymbol{\rho}_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \frac{1}{3}\boldsymbol{\lambda}' E\left[\boldsymbol{Y}_0 h^2(\boldsymbol{Y}_0)\boldsymbol{Y}_0'\right]\boldsymbol{\lambda}\right).$$

Since $\overline{h} = E[h(\boldsymbol{Y}_0)] + o_p(1)$ it follows (see Serfling (1980, pg. 19)) that,

$$S_{n1}(\boldsymbol{\rho}_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \frac{1}{3}\boldsymbol{\lambda}' E^2[h(\boldsymbol{Y}_0)]\, E\left[\boldsymbol{Y}_0 h^2(\boldsymbol{Y}_0)\boldsymbol{Y}_0'\right]\boldsymbol{\lambda}\right).$$

As a final note one should note that multiplying the h function in the dispersion function by a positive constant does not change the estimate. Thus, we can assume without loss of generality that $E[h(\boldsymbol{Y}_0)] = 1$ in the above result.

To complete the proof we must show that $S_{n2}(\boldsymbol{\rho}_0) = o_p(1)$. However, this follows from Theorem 3.4.1 (or Theorem 3.4.2). It now follows from Slutsky's Theorem and the Cramer-Wold Device (Serfling, 1980, pg. 18–19) that,

$$S_n(\boldsymbol{\rho}_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \frac{1}{3}E^2[h(\boldsymbol{Y}_0)]\, E\left[\boldsymbol{Y}_0 h^2(\boldsymbol{Y}_0)\boldsymbol{Y}_0'\right]\right).$$

One should recall that without loss of generality, $E[h(\boldsymbol{Y}_0)] = 1$. Thus, we have show that,

$$S_n(\boldsymbol{\rho}_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \frac{1}{3}E\left[\boldsymbol{Y}_0 h^2(\boldsymbol{Y}_0)\boldsymbol{Y}_0'\right]\right),$$

which completes the proof. □

3.5 Asymptotic Linearity of the Gradient

Before we state and prove the linearity result we will prove two lemmas that will be needed in the proof of the linearity result. The first lemma pertains

to the component-wise convergence of the matrix $\boldsymbol{C}_n$.

Lemma 3.5.1 *Let $\boldsymbol{C}_n = [C_{nkl}]$ be a $p \times p$ matrix with,*

$$C_{nkl} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\left(X_{j-k} - X_{i-k}\right)\left(X_{j-l} - X_{i-l}\right).$$

Then, under assumptions W1-W4 we have $\frac{1}{n^2}\boldsymbol{C}_n \xrightarrow{p} \boldsymbol{C}$ where

$$C_{kl} = E\left[h\left(\boldsymbol{Y}_0\right)\right]E\left[X_{1-k}h\left(\boldsymbol{Y}_0\right)X_{1-l}\right] - E\left[h\left(\boldsymbol{Y}_0\right)X_{1-k}\right]E\left[h\left(\boldsymbol{Y}_0\right)X_{1-l}\right].$$

proof. Choose k and l and let these values be fixed throughout the proof. Furthermore, let $\widehat{\boldsymbol{\theta}}_n$ denote an estimated parameter vector whose components contain the elements of $\widehat{\boldsymbol{\mu}}_n$ and the upper triangular portion of $\widehat{\boldsymbol{\Sigma}}_n$. Now consider the $(k,l)^{th}$ component of $\boldsymbol{C}_n$,

$$\begin{aligned}
\frac{1}{n^2}C_{nkl} &= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\left(\widehat{\boldsymbol{\theta}}_n\right)\left(X_{j-k} - X_{i-k}\right)\left(X_{j-l} - X_{i-l}\right) \\
&= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \left(b_{ij}\left(\widehat{\boldsymbol{\theta}}_n\right) - b_{ij}\left(\boldsymbol{\theta}\right) + b_{ij}\left(\boldsymbol{\theta}\right)\right)\left(X_{j-k} - X_{i-k}\right)\left(X_{j-l} - X_{i-l}\right) \\
&= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \left(b_{ij}\left(\widehat{\boldsymbol{\theta}}_n\right) - b_{ij}\left(\boldsymbol{\theta}\right)\right)\left(X_{j-k} - X_{i-k}\right)\left(X_{j-l} - X_{i-l}\right) + \\
&\quad \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\left(\boldsymbol{\theta}\right)\left(X_{j-k} - X_{i-k}\right)\left(X_{j-l} - X_{i-l}\right) \\
&= C_{n1} + C_{n2} \text{ say.}
\end{aligned}$$

Under assumptions W1-W4 the Ergodic Theorem and the multivariate Mean Value Theorem can be used to show that $C_{n1} = o_p(1)$ as follows,

$$\begin{aligned}
\left|C_{n1}\right| &\leq \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \left|b_{ij}\left(\widehat{\boldsymbol{\theta}}_n\right) - b_{ij}\left(\boldsymbol{\theta}\right)\right| \left|\left(X_{j-k} - X_{i-k}\right)\left(X_{j-l} - X_{i-l}\right)\right| \\
&\leq \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right\| \left|\left(X_{j-k} - X_{i-k}\right)\left(X_{j-l} - X_{i-l}\right)\right|
\end{aligned}$$

$$
\begin{aligned}
&= \frac{K}{\sqrt{n}}\left\|\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}\right)\right\|\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\left(X_{j-k}-X_{i-k}\right)\left(X_{j-l}-X_{i-l}\right)\right| \\
&\leq \frac{K}{\sqrt{n}}\left\|\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}\right)\right\|\left(E\left[|X_{1-k}X_{1-l}|\right]+E\left[|X_{1-k}|\right]E\left[|X_{1-l}|\right]+o_p(1)\right) \\
&= o(1)O_p(1)O_p(1) \\
&= o_p(1).
\end{aligned}
$$

Since $C_{n1}=o_p(1)$ we only need to show that C_{n2} converges to the desired result. Under assumption W1 the Ergodic Theorem can be used to show the following,

$$
\begin{aligned}
C_{n2} &= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}b_{ij}\left(\boldsymbol{\theta}\right)\left(X_{j-k}-X_{i-k}\right)\left(X_{j-l}-X_{i-l}\right) \\
&= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}h\left(\boldsymbol{Y}_{i-1};\boldsymbol{\theta}\right)h\left(\boldsymbol{Y}_{j-1};\boldsymbol{\theta}\right)\left(X_{j-k}-X_{i-k}\right)\left(X_{j-l}-X_{i-l}\right) \\
&= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}h\left(\boldsymbol{Y}_{i-1}\right)h\left(\boldsymbol{Y}_{j-1}\right)\times \\
&\quad \left(X_{j-k}X_{j-l}-X_{j-k}X_{i-l}-X_{i-k}X_{j-l}+X_{i-k}X_{i-l}\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}h\left(\boldsymbol{Y}_{i-1}\right)h\left(\boldsymbol{Y}_{j-1}\right)X_{j-k}X_{j-l}- \\
&\quad \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}h\left(\boldsymbol{Y}_{i-1}\right)h\left(\boldsymbol{Y}_{j-1}\right)X_{j-k}X_{i-l} \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}h\left(\boldsymbol{Y}_{i-1}\right)\right)\left(\frac{1}{n}\sum_{j=1}^{n}h\left(\boldsymbol{Y}_{j-1}\right)X_{j-k}X_{j-l}\right)- \\
&\quad \left(\frac{1}{n}\sum_{i=1}^{n}h\left(\boldsymbol{Y}_{i-1}\right)X_{i-l}\right)\left(\frac{1}{n}\sum_{j=1}^{n}h\left(\boldsymbol{Y}_{j-1}\right)X_{j-k}\right) \\
&= E\left[h\left(\boldsymbol{Y}_0\right)\right]E\left[X_{1-k}h\left(\boldsymbol{Y}_0\right)X_{1-l}\right]-E\left[h\left(\boldsymbol{Y}_0\right)X_{1-k}\right]E\left[h\left(\boldsymbol{Y}_0\right)X_{1-l}\right]+o_p(1).
\end{aligned}
$$

This proves the result. □

Before we state the next lemma consider the following random variables which are multivariate analogs to those given in section 2.3. For $\boldsymbol{\Delta}\in\Re^p$ define

the following random variable,

$$t_{ij}(\boldsymbol{\Delta}) = \frac{\boldsymbol{\Delta}'}{\sqrt{n}}(\boldsymbol{Y}_{i-1} - \boldsymbol{Y}_{j-1}).$$

Analogous to Corollary 2.3.1 stationarity and a finite second moment assumption imply that,

$$\max_{1 \le i,j \le n} |t_{ij}(\boldsymbol{\Delta})| = o_p(1).$$

Now let $\boldsymbol{\rho}_0$ denote the true autoregressive parameter vector and define the following random variable,

$$\begin{aligned} W_{ij} &= \frac{\operatorname{sgn}\left(\varepsilon_i\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - \varepsilon_j\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right)\right) - \operatorname{sgn}\left(\varepsilon_i(\boldsymbol{\rho}_0) - \varepsilon_j(\boldsymbol{\rho}_0)\right)}{2} \\ &= \begin{cases} -1 & \text{if } 0 < \varepsilon_i - \varepsilon_j < t_{ij}(\boldsymbol{\Delta}) \\ 0 & \text{otherwise} \\ 1 & \text{if } t_{ij}(\boldsymbol{\Delta}) < \varepsilon_i - \varepsilon_j < 0. \end{cases} \end{aligned}$$

Analogous to Lemma 2.3.5 and Corollary 2.3.3 one can show that $E\left[W_{ij}^2\right] = E\left[W_{ij}^4\right] = o(1)$ for all (i,j). Hence, it follows that $|W_{ij}| = W_{ij}^2 = o_p(1)$. We are now ready to state the second lemma.

Lemma 3.5.2 *Let V_i be a random variable such that $E\left[V_i^2\right] \le K$ for all i and let W_{ij} be defined as above. Then,*

$$T_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} V_i |W_{ij}| = o_p(1).$$

proof.

$$E\left[T_n^2\right] = E\left[\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} V_i |W_{ij}|\right)^2\right]$$

$$\begin{aligned}
&= E\left[\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n} V_i^2 W_{ij}^2\right] + E\left[\frac{1}{n^4}\sum_{ij}\sum_{kl} V_i\,|W_{ij}|\,V_k\,|W_{kl}|\right] \\
&= \frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[V_i^2 W_{ij}^2\right] + \frac{1}{n^4}\sum_{ij}\sum_{kl} E\left[V_i\,|W_{ij}|\,V_k\,|W_{kl}|\right] \\
&= T_{n1} + T_{n2} \text{ say.}
\end{aligned}$$

Since $E\left[V_i^2\right] \leq K$, $|W_{ij}| \leq 1$ and we are dividing by n^4 it follows that $T_{n1} = O(n^{-4})O(n^2) = o(1)$. Now consider T_{n2}. The sum in T_{n2} is either $O(n^3)$ or $O(n^4)$. When the sum is $O(n^3)$ the Cauchy Schwarz Inequality and the bounds on $E\left[V_i^2\right]$ and W_{ij} imply that $T_{n2} = O(n^{-1})O(1) = o(1)$. Now consider the case when the sum is $O(n^4)$. The bounds on $E\left[V_i^2\right]$ and W_{ij} and the fact that $W_{ij}^2 = o_p(1)$ can be applied to Lemma 2.5.5 to show that $E\left[V_i^2 W_{ij}^2\right] = o(1)$ for all (i,j). Thus, the Cauchy Schwarz Inequality implies that we essentially have an average of terms which are all $o(1)$. Hence, the case when the sum is $O(n^4)$ is also $o(1)$. Since $T_{n1} = T_{n2} = o(1)$, Chebyshev's Inequality implies the result. □

We are now ready to state and prove the main result of this chapter; the asymptotic linearity result. First, define $\boldsymbol{S}_n(\boldsymbol{\rho}) = \frac{1}{n^{\frac{3}{2}}}\boldsymbol{S}(\boldsymbol{\rho})$ and let $\boldsymbol{C}_n$ and $\boldsymbol{C}$ be defined as in Lemma 3.5.1. Then we can prove the following theorem.

Theorem 3.5.1 *Under assumptions M1-M3, E1-E2, and W1-W5 the following holds for all* $\boldsymbol{\Delta} \in \Re^p$*:*

$$\boldsymbol{S}_n\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - \boldsymbol{S}_n(\boldsymbol{\rho}_0) + 2\tau\left(\frac{1}{n^2}\boldsymbol{C}_n\right)\boldsymbol{\Delta} \;=\; \boldsymbol{o}_p(1).$$

proof. Choose a $\boldsymbol{\Delta} \in \Re^p$ and let this $\boldsymbol{\Delta}$ be fixed throughout the proof. The first thing to note is the following,

$$\begin{aligned}
&\boldsymbol{S}_n\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - \boldsymbol{S}_n(\boldsymbol{\rho}_0) + 2\tau\left(\frac{1}{n^2}\boldsymbol{C}_n\right)\boldsymbol{\Delta} \\
&= \boldsymbol{S}_n\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - \boldsymbol{S}_n(\boldsymbol{\rho}_0) + 2\tau\left(\frac{1}{n^2}\boldsymbol{C}_n - \boldsymbol{C} + \boldsymbol{C}\right)\boldsymbol{\Delta} \\
&= \boldsymbol{S}_n\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - \boldsymbol{S}_n(\boldsymbol{\rho}_0) + 2\tau\boldsymbol{C}\boldsymbol{\Delta} + 2\tau\left(\frac{1}{n^2}\boldsymbol{C}_n - \boldsymbol{C}\right)\boldsymbol{\Delta}.
\end{aligned}$$

Hence, by Lemma 3.5.1 it suffices to show,

$$\boldsymbol{S}_n\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - \boldsymbol{S}_n(\boldsymbol{\rho}_0) + 2\tau\boldsymbol{C} = \boldsymbol{o}_p(1),$$

in order to prove the result. One should now note that the weights given in the above equation depend on $\widehat{\boldsymbol{\theta}}_n$, which contains the location and scatter estimates for the sequence, $\{\boldsymbol{Y}_{i-1}\}$. In order to properly apply Theorem 3.3.3 the weights can only be a function of $(\boldsymbol{Y}_{i-1}, \boldsymbol{Y}_{j-1})$. Thus, consider the following difference,

$$\begin{aligned}
&\boldsymbol{S}_n\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - \boldsymbol{S}_n(\boldsymbol{\rho}_0) \\
&= \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\left(\widehat{\boldsymbol{\theta}}_n\right)\boldsymbol{Y}_{i-1} \times \\
&\quad \left(\operatorname{sgn}\left(\varepsilon_i\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - \varepsilon_j\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right)\right) - \operatorname{sgn}\left(\varepsilon_i(\boldsymbol{\rho}_0) - \varepsilon_j(\boldsymbol{\rho}_0)\right)\right) \\
&= \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\left(\widehat{\boldsymbol{\theta}}_n\right)\boldsymbol{Y}_{i-1}W_{ij} \\
&= \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(b_{ij}\left(\widehat{\boldsymbol{\theta}}_n\right) - b_{ij}(\boldsymbol{\theta}) + b_{ij}(\boldsymbol{\theta})\right)\boldsymbol{Y}_{i-1}W_{ij} \\
&= \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(b_{ij}\left(\widehat{\boldsymbol{\theta}}_n\right) - b_{ij}(\boldsymbol{\theta})\right)\boldsymbol{Y}_{i-1}W_{ij} + \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}(\boldsymbol{\theta})\boldsymbol{Y}_{i-1}W_{ij} \\
&= \boldsymbol{T}_{n1} + \boldsymbol{T}_{n2} \text{ say.}
\end{aligned}$$

Now, assumptions W1-W4, stationarity, and the finite second moment assumption along with Lemma 3.5.2 imply the following result,

$$\begin{aligned}
\|\boldsymbol{T}_{n1}\| &= \left\| \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(b_{ij}\left(\widehat{\boldsymbol{\theta}}_n\right) - b_{ij}(\boldsymbol{\theta}) \right) \boldsymbol{Y}_{i-1} W_{ij} \right\| \\
&\leq \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| b_{ij}\left(\widehat{\boldsymbol{\theta}}_n\right) - b_{ij}(\boldsymbol{\theta}) \right| \|\boldsymbol{Y}_{i-1}\| \, |W_{ij}| \\
&\leq \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} K \left\| \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \right\| \|\boldsymbol{Y}_{i-1}\| \, |W_{ij}| \\
&= 2K \left\| \sqrt{n} \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \right) \right\| \left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\boldsymbol{Y}_{i-1}\| \, |W_{ij}| \right) \\
&= O_p(1) o_p(1) \\
&= o_p(1).
\end{aligned}$$

Therefore, in order to prove the linearity result it suffices to show that,

$$\boldsymbol{T}_{n2} = -2\tau \boldsymbol{C} \boldsymbol{\Delta} + \boldsymbol{o}_p(1).$$

However, since vector convergence holds if and only if component-wise convergence holds (see Serfling (1980, pg. 52)) it suffices to show that the k^{th} component of $\boldsymbol{T}_{n2}$ converges to $-2\tau \boldsymbol{C}_k \boldsymbol{\Delta}$ where $\boldsymbol{C}_k$ denotes the k^{th} row of the matrix $\boldsymbol{C}$. Denoting the k^{th} component of $\boldsymbol{T}_{n2}$ by t_{nk} yields the following,

$$\begin{aligned}
&t_{nk} \\
&= \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}(\boldsymbol{\theta}) X_{i-k} W_{ij} \\
&= \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}(\boldsymbol{\theta}) X_{i-k} \times \\
&\quad \left(\frac{\operatorname{sgn}\left(\varepsilon_i \left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}} \right) - \varepsilon_j \left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}} \right) \right) - \operatorname{sgn}\left(\varepsilon_i(\boldsymbol{\rho}_0) - \varepsilon_j(\boldsymbol{\rho}_0) \right)}{2} \right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}(\boldsymbol{\theta})X_{i-k} \times \\
&\quad \left(I\left(\varepsilon_j\left(\boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) \le \varepsilon_i\left(\boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right)\right) - I\left(\varepsilon_j(\boldsymbol{\rho}_0) \le \varepsilon_i(\boldsymbol{\rho}_0)\right)\right) \text{ a.e.} \\
&= \frac{2}{n}\sum_{i=1}^{n} h(\boldsymbol{Y}_{i-1})X_{i-k}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} h(\boldsymbol{Y}_{j-1})\times\right. \\
&\quad \left.\left(I\left(\varepsilon_j\left(\boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) \le \varepsilon_i\left(\boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right)\right) - I\left(\varepsilon_j(\boldsymbol{\rho}_0) \le \varepsilon_i(\boldsymbol{\rho}_0)\right)\right)\right).
\end{aligned}
$$

Now, adding in and subtracting out $I\left(\varepsilon_j(\boldsymbol{\rho}_0) \le \varepsilon_i\left(\boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right)\right)$ into the indicator portion of the above result yields the following,

$$
\begin{aligned}
&t_{nk} \\
&= \frac{2}{n}\sum_{i=1}^{n} h(\boldsymbol{Y}_{i-1})X_{i-k}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} h(\boldsymbol{Y}_{j-1})\times\right. \\
&\quad \left.\left(I\left(\varepsilon_j\left(\boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) \le \varepsilon_i\left(\boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right)\right) - I\left(\varepsilon_j(\boldsymbol{\rho}_0) \le \varepsilon_i\left(\boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right)\right)\right)\right) \\
&+ \frac{2}{n}\sum_{i=1}^{n} h(\boldsymbol{Y}_{i-1})X_{i-k}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} h(\boldsymbol{Y}_{j-1})\times\right. \\
&\quad \left.\left(I\left(\varepsilon_j(\boldsymbol{\rho}_0) \le \varepsilon_i\left(\boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right)\right) - I\left(\varepsilon_j(\boldsymbol{\rho}_0) \le \varepsilon_i(\boldsymbol{\rho}_0)\right)\right)\right) \\
&= t_{nk1} + t_{nk2} \text{ say.}
\end{aligned}
$$

Now apply the definition of $W_\gamma(x,\boldsymbol{\rho})$ of Theorem 3.3.3 with $\gamma(\boldsymbol{Y}_{j-1}) = h(\boldsymbol{Y}_{j-1})$ to show the following,

$$
\begin{aligned}
t_{nk1} &= \frac{2}{n}\sum_{i=1}^{n} h(\boldsymbol{Y}_{i-1})X_{i-k} \times \\
&\quad \sqrt{n}\left(W_h\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}, \boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - W_h\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}, \boldsymbol{\rho}_0\right)\right).
\end{aligned}
$$

Next, use the fact that $I(A) = 1 - I(A^c)$, the definition of $W_\gamma(x,\boldsymbol{\rho})$ with $\gamma(\boldsymbol{Y}_{i-1}) = h(\boldsymbol{Y}_{i-1})X_{i-k}$, and reverse the order of summation to show the fol-

lowing,

$$t_{nk2} = -\frac{2}{n}\sum_{j=1}^{n} h(\boldsymbol{Y}_{j-1})\sqrt{n}\left(W_{hx}\left(\varepsilon_j, \boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - W_{hx}(\varepsilon_j, \boldsymbol{\rho}_0)\right).$$

Consider t_{nk1} first. Adding in and subtracting out the following quantity,

$$\boldsymbol{\Delta}'\frac{1}{n}\sum_{j=1}^{n} h(\boldsymbol{Y}_{j-1})\boldsymbol{Y}_{j-1} f\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}\right),$$

yields the following result,

$$\begin{aligned}
& t_{nk1} \\
&= \frac{2}{n}\sum_{i=1}^{n} h(\boldsymbol{Y}_{i-1})X_{i-k} \times \\
&\quad \sqrt{n}\left(W_h\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}, \boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - W_h\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}, \boldsymbol{\rho}_0\right)\right) \\
&= \frac{2}{n}\sum_{i=1}^{n} h(\boldsymbol{Y}_{i-1})X_{i-k} \times \\
&\quad \left[\sqrt{n}\left(W_h\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}, \boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - W_h\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}, \boldsymbol{\rho}_0\right)\right) - \right. \\
&\quad \left. \boldsymbol{\Delta}'\frac{1}{n}\sum_{j=1}^{n} h(\boldsymbol{Y}_{j-1})\boldsymbol{Y}_{j-1} f\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}\right)\right] \\
&+ \frac{2}{n}\sum_{i=1}^{n} h(\boldsymbol{Y}_{i-1})X_{i-k}\left[\boldsymbol{\Delta}'\frac{1}{n}\sum_{j=1}^{n} h(\boldsymbol{Y}_{j-1})\boldsymbol{Y}_{j-1} f\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}\right)\right] \\
&= t_{nk11} + t_{nk12} \text{ say.}
\end{aligned}$$

Now consider t_{nk11}. Using W5 to bound $h(\boldsymbol{Y}_{i-1})X_{i-k}$ by a constant K that only depends on B_2 we can now show the following,

$$\begin{aligned}
& |t_{nk11}| \\
&\leq \frac{2}{n}\sum_{i=1}^{n} |h(\boldsymbol{Y}_{i-1})X_{i-k}| \times
\end{aligned}$$

$$
\begin{aligned}
&\left|\sqrt{n}\left(W_h\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}, \boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - W_h\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}, \boldsymbol{\rho}_0\right)\right) - \right. \\
&\left. \boldsymbol{\Delta}'\frac{1}{n}\sum_{j=1}^{n} h\left(\boldsymbol{Y}_{j-1}\right)\boldsymbol{Y}_{j-1} f\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}\right)\right| \\
\leq\ & \frac{2}{n}\sum_{i=1}^{n}\left|h\left(\boldsymbol{Y}_{i-1}\right)X_{i-k}\right| \times \sup_{x\in\Re, \|\boldsymbol{\Delta}\|\leq b} \\
&\left|\sqrt{n}\left(W_h\left(x, \boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - W_h\left(x, \boldsymbol{\rho}_0\right)\right) - \boldsymbol{\Delta}'\frac{1}{n}\sum_{j=1}^{n} h\left(\boldsymbol{Y}_{j-1}\right)\boldsymbol{Y}_{j-1} f\left(x\right)\right| \\
\leq\ & 2K \times \sup_{x\in\Re, \|\boldsymbol{\Delta}\|\leq b} \\
&\left|\sqrt{n}\left(W_h\left(x, \boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - W_h\left(x, \boldsymbol{\rho}_0\right)\right) - \boldsymbol{\Delta}'\frac{1}{n}\sum_{j=1}^{n} h\left(\boldsymbol{Y}_{j-1}\right)\boldsymbol{Y}_{j-1} f\left(x\right)\right|.
\end{aligned}
$$

Hence, Theorem 3.3.3 implies that $t_{nk11} = o_p(1)$. Now consider t_{nk12}. Adding in and subtracting out the quantity $f\left(\varepsilon_i\right)$ yields the following,

$$
\begin{aligned}
&t_{nk12} \\
=\ & \frac{2}{n}\sum_{i=1}^{n} h\left(\boldsymbol{Y}_{i-1}\right)X_{i-k}\left[\boldsymbol{\Delta}'\frac{1}{n}\sum_{j=1}^{n} h\left(\boldsymbol{Y}_{j-1}\right)\boldsymbol{Y}_{j-1} f\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}\right)\right] \\
=\ & \frac{2}{n}\sum_{i=1}^{n} h\left(\boldsymbol{Y}_{i-1}\right)X_{i-k}\left(f\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}\right) - f\left(\varepsilon_i\right)\right)\left(\boldsymbol{\Delta}'\frac{1}{n}\sum_{j=1}^{n} h\left(\boldsymbol{Y}_{j-1}\right)\boldsymbol{Y}_{j-1}\right) \\
+\ & \left(\frac{2}{n}\sum_{i=1}^{n} h\left(\boldsymbol{Y}_{i-1}\right)X_{i-k} f\left(\varepsilon_i\right)\right)\left(\boldsymbol{\Delta}'\frac{1}{n}\sum_{j=1}^{n} h\left(\boldsymbol{Y}_{j-1}\right)\boldsymbol{Y}_{j-1}\right) \\
=\ & t_{nk121} + t_{nk122} \text{ say.}
\end{aligned}
$$

Now consider t_{nk121}. One can use the bound given in W5 to show the following,

$$
\begin{aligned}
&\left|t_{nk121}\right| \\
\leq\ & \frac{2}{n}\sum_{i=1}^{n}\left|h\left(\boldsymbol{Y}_{i-1}\right)X_{i-k}\right|\left|f\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}\right) - f\left(\varepsilon_i\right)\right|\left(\frac{1}{n}\sum_{j=1}^{n}\left|h\left(\boldsymbol{Y}_{j-1}\right)\boldsymbol{\Delta}'\boldsymbol{Y}_{j-1}\right|\right) \\
\leq\ & \max_{1\leq i\leq n}\left|f\left(\varepsilon_i - \frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}\right) - f\left(\varepsilon_i\right)\right|\left(\frac{2}{n}\sum_{i=1}^{n}\left|h\left(\boldsymbol{Y}_{i-1}\right)X_{i-k}\right|\right) \times
\end{aligned}
$$

$$
\begin{aligned}
&\left(\frac{1}{n}\sum_{j=1}^{n}|h(\boldsymbol{Y}_{j-1})\boldsymbol{\Delta}'\boldsymbol{Y}_{j-1}|\right)\\
\leq\ & 2K\max_{1\leq i\leq n}\left|f\left(\varepsilon_i-\frac{\boldsymbol{\Delta}'}{\sqrt{n}}\boldsymbol{Y}_{i-1}\right)-f(\varepsilon_i)\right|,
\end{aligned}
$$

where K depends only on B_2 and $\boldsymbol{\Delta}$. It now follows from Corollary 3.3.1 that $t_{nk121}=o_p(1)$. Consider t_{nk122} next. A direct application of the Ergodic Theorem to the two sums given in t_{nk122} and independence of ε_i and $\boldsymbol{Y}_{i-1}$ yield the following result,

$$
t_{nk122} \;=\; 2\tau\boldsymbol{\Delta}'E\left[h(\boldsymbol{Y}_0)X_{1-k}\right]E\left[h(\boldsymbol{Y}_0)\boldsymbol{Y}_0\right]+o_p(1).
$$

Thus, combining the results on t_{nk11}, t_{nk121} and t_{nk122} yields the following result,

$$
t_{nk1} \;=\; 2\tau\boldsymbol{\Delta}'E\left[h(\boldsymbol{Y}_0)X_{1-k}\right]E\left[h(\boldsymbol{Y}_0)\boldsymbol{Y}_0\right]+o_p(1).
$$

Let us now consider t_{nk2}. Adding in and subtracting out the following quantity,

$$
\boldsymbol{\Delta}'\frac{1}{n}\sum_{i=1}^{n}h(\boldsymbol{Y}_{i-1})X_{i-k}\boldsymbol{Y}_{i-1}f(\varepsilon_j),
$$

yields the following result,

$$
\begin{aligned}
&t_{nk2}\\
&= -\frac{2}{n}\sum_{j=1}^{n}h(\boldsymbol{Y}_{j-1})\sqrt{n}\left(W_{hx}\left(\varepsilon_j,\boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right)-W_{hx}(\varepsilon_j,\boldsymbol{\rho}_0)\right)\\
&= -\frac{2}{n}\sum_{j=1}^{n}h(\boldsymbol{Y}_{j-1})\times\left[\sqrt{n}\left(W_{hx}\left(\varepsilon_j,\boldsymbol{\rho}_0+\frac{\boldsymbol{\Delta}}{\sqrt{n}}\right)-W_{hx}(\varepsilon_j,\boldsymbol{\rho}_0)\right)-\right.\\
&\qquad\left.\boldsymbol{\Delta}'\frac{1}{n}\sum_{i=1}^{n}h(\boldsymbol{Y}_{i-1})X_{i-k}\boldsymbol{Y}_{i-1}f(\varepsilon_j)\right]+\\
&\qquad -\frac{2}{n}\sum_{j=1}^{n}h(\boldsymbol{Y}_{j-1})\left[\boldsymbol{\Delta}'\frac{1}{n}\sum_{i=1}^{n}h(\boldsymbol{Y}_{i-1})X_{i-k}\boldsymbol{Y}_{i-1}f(\varepsilon_j)\right]\\
&= t_{nk21}+t_{nk22}\ \text{say.}
\end{aligned}
$$

Consider t_{nk21} first. Using W4 to bound $h(\boldsymbol{Y}_{j-1})$ by a constant K that only depends on B_1 we can now show the following,

$$
\begin{aligned}
&|t_{nk21}| \\
&\leq \frac{2}{n}\sum_{j=1}^{n}|h(\boldsymbol{Y}_{j-1})| \times \left|\sqrt{n}\left(W_{hx}\left(\varepsilon_j, \boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - W_{hx}(\varepsilon_j, \boldsymbol{\rho}_0)\right) - \right. \\
&\qquad \left. \boldsymbol{\Delta}'\frac{1}{n}\sum_{i=1}^{n} h(\boldsymbol{Y}_{i-1})X_{i-k}\boldsymbol{Y}_{i-1} f(\varepsilon_j)\right| \\
&\leq 2K \times \sup_{x\in\Re, ||\boldsymbol{\Delta}||\leq b} \\
&\qquad \left|\sqrt{n}\left(W_{hx}\left(x, \boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - W_{hx}(x, \boldsymbol{\rho}_0)\right) - \boldsymbol{\Delta}'\frac{1}{n}\sum_{i=1}^{n} h(\boldsymbol{Y}_{i-1})X_{i-k}\boldsymbol{Y}_{i-1} f(x)\right|.
\end{aligned}
$$

Hence, Theorem 3.3.3 implies that $t_{nk21} = o_p(1)$. Now consider t_{nk22}. A direct application of the Ergodic Theorem to the two sums given in t_{nk22} and the fact that ε_j is independent of $\boldsymbol{Y}_{j-1}$ yields the following result,

$$
\begin{aligned}
t_{nk22} &= -\frac{2}{n}\sum_{j=1}^{n} h(\boldsymbol{Y}_{j-1}) f(\varepsilon_j)\left[\boldsymbol{\Delta}'\frac{1}{n}\sum_{i=1}^{n} h(\boldsymbol{Y}_{i-1})X_{i-k}\boldsymbol{Y}_{i-1}\right] \\
&= -2\tau\boldsymbol{\Delta}' E\left[h(\boldsymbol{Y}_0)\right] E\left[h(\boldsymbol{Y}_0)X_{1-k}\boldsymbol{Y}_0\right] + o_p(1).
\end{aligned}
$$

Thus, combining the results on t_{nk21} and t_{nk22} yields the following,

$$
t_{nk2} = -2\tau\boldsymbol{\Delta}' E\left[h(\boldsymbol{Y}_0)\right] E\left[h(\boldsymbol{Y}_0)X_{1-k}\boldsymbol{Y}_0\right] + o_p(1).
$$

Finally, if we combine the results on t_{nk1} and t_{nk2} we get the following,

$$
\begin{aligned}
&t_{nk} \\
&= 2\tau\boldsymbol{\Delta}' E\left[h(\boldsymbol{Y}_0)X_{1-k}\right] E\left[h(\boldsymbol{Y}_0)\boldsymbol{Y}_0\right] - \\
&\quad 2\tau\boldsymbol{\Delta}' E\left[h(\boldsymbol{Y}_0)\right] E\left[h(\boldsymbol{Y}_0)X_{1-k}\boldsymbol{Y}_0\right] + o_p(1)
\end{aligned}
$$

$$= -2\tau\boldsymbol{\Delta}'(E[h(\boldsymbol{Y}_0)]E[h(\boldsymbol{Y}_0)X_{1-k}\boldsymbol{Y}_0]-$$

$$E[h(\boldsymbol{Y}_0)X_{1-k}]E[h(\boldsymbol{Y}_0)\boldsymbol{Y}_0]) + o_p(1)$$

$$= -2\tau\boldsymbol{\Delta}'\boldsymbol{C}_k + o_p(1).$$

Hence, we have shown that,

$$S_{nk}\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - S_{nk}(\boldsymbol{\rho}_0) + 2\tau\boldsymbol{\Delta}'\boldsymbol{C}_k = o_p(1).$$

Lastly, since the above result holds for $k = 1, 2, \ldots, p$ we have that,

$$\boldsymbol{S}_n\left(\boldsymbol{\rho}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}\right) - \boldsymbol{S}_n(\boldsymbol{\rho}_0) + 2\tau\left(\frac{1}{n^2}\boldsymbol{C}_n\right)\boldsymbol{\Delta} = \boldsymbol{o}_p(1),$$

which proves the result. □

3.6 The Asymptotic Distribution of the Estimate

3.6.1 Asymptotic Uniform Linearity and Uniform Quadraticity

From Chapter II we see that asymptotic uniform linearity and uniform quadraticity results need to be established before one can proceed with the asymptotic distribution of the estimate. The development here will be analogous to that found in Section 2.6.3. The only difference is that we are now in a multivariate setting.

To begin, let $\boldsymbol{\rho}_0$ denote the true parameter vector for the AR(p) and let $\boldsymbol{\Delta} \in \Re^p$. One will recall from Section 3.1 that $D(\boldsymbol{\rho})$ represents the dispersion function to be minimized and $\boldsymbol{S}(\boldsymbol{\rho})$ is the negative of its' gradient. Now define

the following functions of $\mathbf{\Delta}$, which are multivariate analogs to those functions found in Section 2.6.3,

$$D_n(\mathbf{\Delta}) = \frac{1}{n} D\left(\boldsymbol{\rho}_0 + \frac{\mathbf{\Delta}}{\sqrt{n}}\right)$$

$$\boldsymbol{S}_n(\mathbf{\Delta}) = -\frac{d}{d\mathbf{\Delta}} D_n(\mathbf{\Delta}) = \frac{1}{n^{\frac{3}{2}}} \boldsymbol{S}\left(\boldsymbol{\rho}_0 + \frac{\mathbf{\Delta}}{\sqrt{n}}\right)$$

$$Q_n(\mathbf{\Delta}) = D_n(\mathbf{0}) - \boldsymbol{S}_n'(\mathbf{0})\,\mathbf{\Delta} + \tau \mathbf{\Delta}' \boldsymbol{C} \mathbf{\Delta}.$$

One can now follow a multivariate version of the heuristic discussion found in Section 2.6.3 to motivate the following results.

We will say that we have asymptotic linearity (AL) if for all $\mathbf{\Delta} \in \Re^p$ the following result holds,

$$\boldsymbol{S}_n(\mathbf{\Delta}) - \boldsymbol{S}_n(\mathbf{0}) + 2\tau \boldsymbol{C}\mathbf{\Delta} \xrightarrow{p} \mathbf{0}.$$

One should note that this result is given in Theorem 3.5.1. Secondly, we will say we have asymptotic uniform linearity (AUL) if for all $c > 0$ the following result holds,

$$\sup_{\|\mathbf{\Delta}\| \le c} \|\boldsymbol{S}_n(\mathbf{\Delta}) - \boldsymbol{S}_n(\mathbf{0}) + 2\tau \boldsymbol{C}\mathbf{\Delta}\| \xrightarrow{p} 0.$$

Lastly, we will say that we have asymptotic uniform quadraticity (AUQ) if for all $c > 0$ the following result holds,

$$\sup_{\|\mathbf{\Delta}\| \le c} |D_n(\mathbf{\Delta}) - Q_n(\mathbf{\Delta})| \xrightarrow{p} 0.$$

In their 1988 paper Heiler and Willers (1988) show that AL, AUL, and AUQ are equivalent in the context of linear regression. In Section 2.6.3 we presented their proof in the context of the AR(1). However, the proof for $p > 1$ is straight forward; just replace univariate convergence with vector convergence. In fact, the result given in Heiler and Willers is this multivariate version. For the sake of completeness we state this result in the following theorem.

Theorem 3.6.1 *Under M1-M3, E1-E2, and W1-W5 AL, AUL, and AUQ are equivalent.*

proof. See the proof of Lemma 3.1 in Heiler and Willers (1988). □

With these results established we can now proceed to the next step, the asymptotic distribution of $\hat{\boldsymbol{\rho}}_n$.

3.6.2 Asymptotic Normality of the Estimate

In this section it will be convenient to change notation by letting $\boldsymbol{\Delta} = \sqrt{n}\,(\boldsymbol{\rho} - \boldsymbol{\rho}_0)$ where $\boldsymbol{\rho}_0$ still represents the true parameter for the AR(p). Substituting this value of $\boldsymbol{\Delta}$ into the functions defined in Section 3.6.1 and thinking of these functions as functions of $\boldsymbol{\rho}$ instead of $\boldsymbol{\Delta}$ we obtain the following,

$$
\begin{aligned}
D_n(\boldsymbol{\rho}) &= \frac{1}{n} D(\boldsymbol{\rho}) \\
\boldsymbol{S}_n(\boldsymbol{\rho}) &= \frac{1}{n^{\frac{3}{2}}} \boldsymbol{S}(\boldsymbol{\rho})
\end{aligned}
$$

$$
Q_n(\boldsymbol{\rho}) = D_n(\boldsymbol{\rho}_0) - \sqrt{n}\boldsymbol{S}_n'(\boldsymbol{\rho}_0)(\boldsymbol{\rho} - \boldsymbol{\rho}_0) + n\tau(\boldsymbol{\rho} - \boldsymbol{\rho}_0)'\boldsymbol{C}(\boldsymbol{\rho} - \boldsymbol{\rho}_0).
$$

Now, taking the derivative of $Q_n(\boldsymbol{\rho})$ with respect to $\boldsymbol{\rho}$ yields,

$$D Q_n(\boldsymbol{\rho}) = -\sqrt{n}\boldsymbol{S}_n(\boldsymbol{\rho}_0) + 2n\tau\boldsymbol{C}(\boldsymbol{\rho} - \boldsymbol{\rho}_0).$$

If $\tilde{\boldsymbol{\rho}}_n$ is such that $DQ_n(\tilde{\boldsymbol{\rho}}_n) = \mathbf{0}$ then $\tilde{\boldsymbol{\rho}}_n$ denotes the value of $\boldsymbol{\rho}$ for which $Q_n(\boldsymbol{\rho})$ is minimized. It follows that,

$$\tilde{\boldsymbol{\rho}}_n = \boldsymbol{\rho}_0 + \frac{1}{2\sqrt{n}\tau}\boldsymbol{C}^{-1}\boldsymbol{S}_n(\boldsymbol{\rho}_0).$$

Equivalently, $\tilde{\boldsymbol{\rho}}_n$ is such that,

$$\sqrt{n}(\tilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) = \frac{1}{2\tau}\boldsymbol{C}^{-1}\boldsymbol{S}_n(\boldsymbol{\rho}_0). \tag{3.2}$$

Because $\tilde{\boldsymbol{\rho}}_n$ depends on the true value of the process, $\tilde{\boldsymbol{\rho}}_n$ it is not a statistic. However, we can still derive its asymptotic distribution. In Section 3.4 we showed that,

$$\boldsymbol{S}_n(\boldsymbol{\rho}_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \frac{1}{3}E\left[\boldsymbol{Y}_0 h^2(\boldsymbol{Y}_0)\boldsymbol{Y}_0'\right]\right). \tag{3.3}$$

It follows (see Serfling (1980, pg. 19)) from (3.2) and (3.3) that,

$$\sqrt{n}(\tilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \frac{1}{12\tau^2}\boldsymbol{C}^{-1}E\left[\boldsymbol{Y}_0 h^2(\boldsymbol{Y}_0)\boldsymbol{Y}_0'\right]\boldsymbol{C}^{-1}\right). \tag{3.4}$$

Now consider the estimate, $\widehat{\boldsymbol{\rho}}_n$. We want to determine the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0)$. Adding in and subtracting out $\tilde{\boldsymbol{\rho}}_n$ within this expression yields,

$$\sqrt{n}(\widehat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) = \sqrt{n}(\widehat{\boldsymbol{\rho}}_n - \tilde{\boldsymbol{\rho}}_n) + \sqrt{n}(\tilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0).$$

If we can show that $\sqrt{n}\,(\widehat{\boldsymbol{\rho}}_n - \tilde{\boldsymbol{\rho}}_n) = \boldsymbol{o}_p(1)$ then it will follow (see Serfling (1980, pg. 19)) from (3.4) that,

$$\sqrt{n}\,(\widehat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \frac{1}{12\tau^2}\boldsymbol{C}^{-1}E\left[\boldsymbol{Y}_0 h^2(\boldsymbol{Y}_0)\boldsymbol{Y}_0'\right]\boldsymbol{C}^{-1}\right).$$

The fact that $\sqrt{n}\,(\widehat{\boldsymbol{\rho}}_n - \tilde{\boldsymbol{\rho}}_n) = \boldsymbol{o}_p(1)$ is given in the following theorem. The proof of this theorem is due to Jaeckel (1972). In Section 2.7 we presented Jaeckel's proof in the context of the AR(1). However, since Jaeckel's proof is a multivariate version, this result will also hold for the AR(p). For the sake of completeness, we present Jaeckel's result in the context of the AR(p).

Theorem 3.6.2 *Under the conditions of Theorem 3.4.3 we have,*

$$\sqrt{n}\,(\widehat{\boldsymbol{\rho}}_n - \tilde{\boldsymbol{\rho}}_n) = \boldsymbol{o}_p(1).$$

proof. See Jaeckel (1972). □

CHAPTER IV

EXAMPLES

We will consider four estimates throughout this chapter, three of which are rank-based. The first estimate will be the classical least squares fit (LS) while the second estimate will be the Wilcoxon R-estimator (WIL). This corresponds to the GR estimate of Chapter II with $b_{ij} \equiv 1$. It is well-known that both of these estimates are sensitive to bad leverage points in the classical linear regression model (e.g. Naranjo and Hettmansperger (1994)). It will be demonstrated that these estimates can still be sensitive to "bad" leverage points caused from additive outliers in the stationary autoregressive model. The third estimate will be the positive breakdown bounded influence rank-based estimate (GR) of Naranjo and Hettmansperger (1994) while the last estimate calculated is a high breakdown bounded influence rank-based estimate (HBR) proposed by Chang (1996). Theory pertaining to the autoregressive model and the GR estimate is given in Chapter III. However, no theory is given for the HBR estimate. It is simply presented for comparison purposes. These two estimates are less sensitive to bad leverage points in the classical linear regression setting since they both have a totally bounded influence function. Thus, for the autoregressive model, it should follow that these estimates will be less sensitive to "bad" leverage points caused by additive outliers.

This will be demonstrated in the examples to follow.

The GR estimate will be computed using the same weighting scheme as that found in Chang (1996). Specifically, a Mallows type weighting scheme, $b_{ij} = h_i h_j$, with

$$h_i = \min\left\{1, \left[\frac{b}{(\boldsymbol{Y}_{i-1} - \boldsymbol{\mu}_n)'\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{Y}_{i-1} - \boldsymbol{\mu}_n)}\right]^{k/2}\right\} \tag{4.1}$$

will be used to calculate the weights. Here, $\boldsymbol{Y}_{i-1} = (X_{i-1}, X_{i-2}, \ldots, X_{i-p})'$ denotes the design point, and the terms $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ are the associated minimum volume ellipsoid (MVE) measures of location and dispersion proposed by Rousseeuw and van Zomeren (1991). They will be computed using the algorithm proposed by Stromberg and Hawkins (1993). The cutoff point, b, will be set at the 95^{th} percentile of the $\chi^2(p)$ distribution while the parameter k will be set at 2. It should be noted that one can control the severity of a downweighted observation by increasing or decreasing the value of k. Additionally, a value of $k = 0$ gives the Wilcoxon R-estimator.

For the HBR estimate we will again follow Chang (1996) and use the following weights,

$$b_{ij} = \psi\left[\left|\frac{c m_i m_j}{(e_i(\hat{\boldsymbol{\rho}}_0)/\hat{\sigma}_n)(e_j(\hat{\boldsymbol{\rho}}_0)/\hat{\sigma}_n)}\right|\right]; \tag{4.2}$$

where $\psi(t) = 1, t, \text{or} -1$ according to $t \geq 1$, $-1 < t < 1$, or $t \leq -1$ and

$$m_i = \psi\left[\frac{b}{(\boldsymbol{Y}_{i-1} - \boldsymbol{\mu}_n)'\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{Y}_{i-1} - \boldsymbol{\mu}_n)}\right],$$

where $\boldsymbol{Y}_{i-1}$, $\boldsymbol{\mu}_n$, $\boldsymbol{\Sigma}_n$, and b are the same as those defined in (4.1). The tuning constant, c, in (4.2) will be set at $[\mathrm{med}\{a_i\} + 3\mathrm{MAD}\{a_i\}]^2$ where $a_i = e_i(\hat{\boldsymbol{\rho}}_0)/(\hat{\sigma}_n m_i)$ and

$$\hat{\sigma}_n = \mathrm{MAD} = 1.483\mathrm{med}_i|e_i(\hat{\boldsymbol{\rho}}_0) - \mathrm{med}_j\{e_j(\hat{\boldsymbol{\rho}}_0)\}|.$$

Lastly, $e_i(\hat{\boldsymbol{\rho}}_0)$ denotes the i^{th} residual of some initial estimate. In the examples to follow, $\hat{\boldsymbol{\rho}}_0$ is taken to be the LMS estimate, computed using the algorithm written by Stromberg (1993).

The Wilcoxon R-estimate will be computed using the algorithm discussed in Kapenga et al. (1996) while similar Gauss-Newton type algorithms will be used to obtain the weighted R-estimates (Chang et al., 1996). In all cases the estimate of an intercept will be the median of the residuals.

4.1 A Simulated Example

To illustrate the effectiveness of the proposed estimate, the first example discussed deals with a simulated time series. Specifically, 340 observations were simulated from the following AR(1) model,

$$X_i = 50.0 + 0.3X_{i-1} + \varepsilon_i, \quad \text{with } \varepsilon_i \text{ iid } N(0,10).$$

The first 300 observations were used to establish the process "past" while the last 40 observations represented a realization of the process. The simulated observations, read from left to right and from top to bottom in time, are displayed in

Table 2

Simulated AR(1) With Gaussian Errors

69.8	71.4	70.4	72.9	71.5	70.2	70.9	71.4
69.6	70.0	70.5	69.4	71.5	71.3	69.1	71.0
73.1	72.0	72.8	71.4	73.0	71.5	71.9	70.7
70.2	69.2	70.7	70.7	70.3	70.1	71.2	71.2
71.8	72.7	71.0	70.4	72.7	71.6	70.8	70.5

Table 3

Parameter Estimates for the Simulated AR(1)

	Original Series		Original Series with Outlier	
Estimate	$\hat{\alpha}_n$	$\hat{\rho}_n$	$\hat{\alpha}_n$	$\hat{\rho}_n$
LS	53.63	0.246	71.16	-0.024
WIL	53.28	0.250	70.87	0.005
GR	55.29	0.222	51.94	0.268
HBR	55.29	0.222	51.63	0.273

Table 2. A time series plot and lag one scatter plot (with fits) of the data are given in Figure 3. Since the error distribution is Gaussian, we would expect the LS fit to be good. However, we would also like the rank-based estimates to perform well under the normality assumption. As is seen in Table 3, as well as Figure 3, the rank-based estimates are close to the least squares estimate. However, all of the estimates seem to be biased. Now suppose there was a recording error, for instance, a misplaced decimal point. Specifically, suppose the 15^{th} observation,

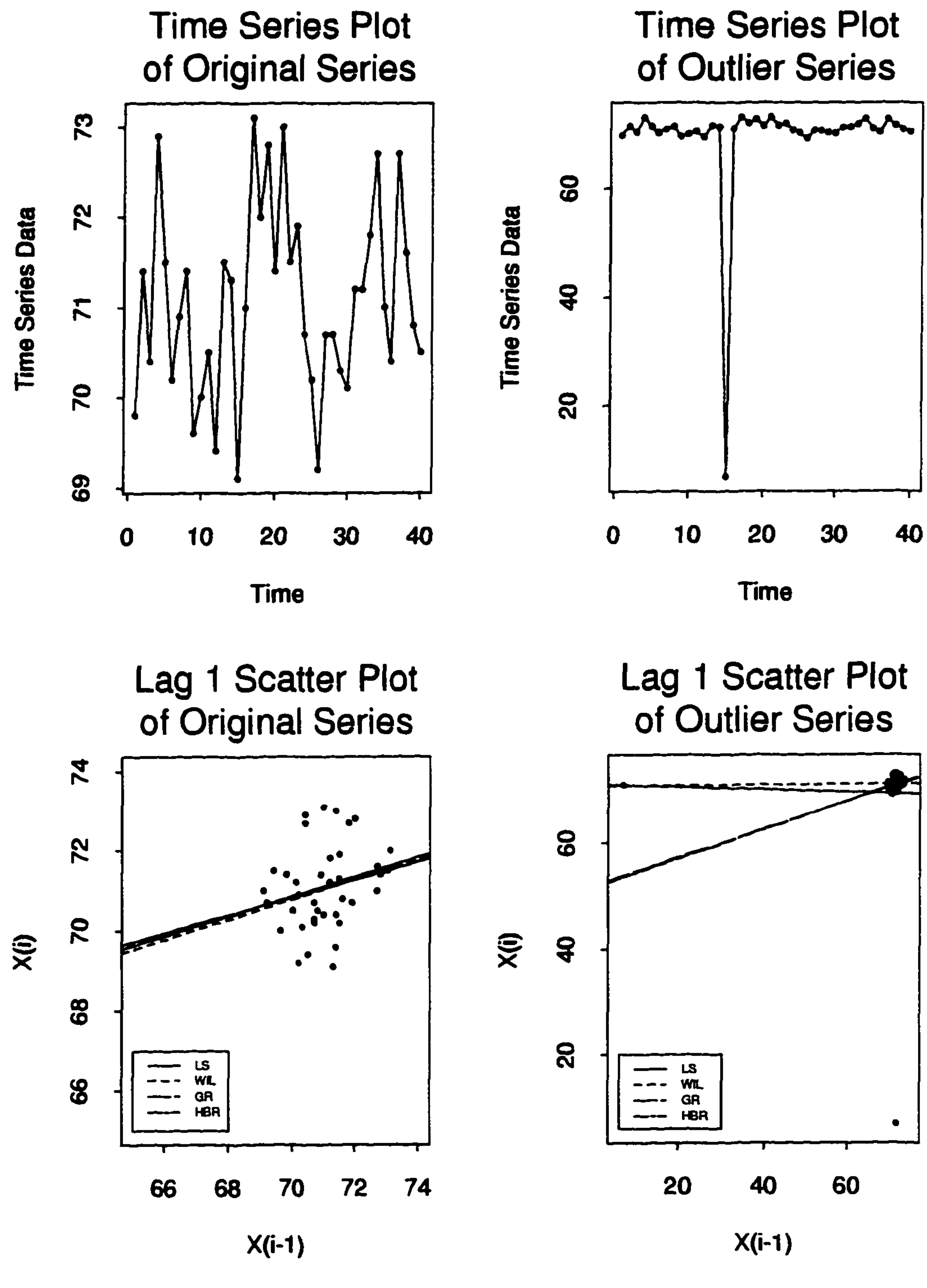

Figure 3. Time Series Plots and Lag 1 Scatter Plots With Fits for Both the Original Series and the Series With One Outlier.

69.1, was recorded as 6.91. This type of error is representative of an additive outlier. The time series plot in Figure 3 for this series clearly demonstrates the outlier and its' additive effect. The lag one scatter plot in Figure 3 identifies two points that are different from the rest of the data. The point $(71.3, 6.91)$ is an outlier in the residual, while the point $(6.91, 71.0)$ is a bad leverage point. It is well-known that the LS and WIL fits are sensitive to "bad" leverage points (e.g. Naranjo and Hettmansperger (1994)). Thus, these fits should be biased towards zero. This is easily verified in the lag one scatter plot of Figure 3 as well as the parameter estimates of Table 3. However, as illustrated in Figure 3 and Table 3, the GR and HBR estimates are less sensitive to these outlying points. Thus, they are more consistent with the bulk of the data. In fact, as is seen in Table 3, the GR and HBR estimates are closer to the true parameter value of 0.3 than with the original data. Hence, since it is an unfortunate reality that recording errors, as well as other additive effects occur in practice, the necessity of the generalized rank-based estimates is clearly demonstrated.

4.2 Residential Extensions Data

The next example deals with a monthly time series (RESX) which originated at Bell Canada. The data describes the installation of residential telephone extensions in a fixed geographical area from January 1966 to May 1973 (Rousseeuw & Leroy, 1987, pg. 278–280) and can be found in Rousseeuw and Leroy (1987, pg.

280). Upon examination of the data set, it is quite obvious that the installation data for November, 1972 and December, 1972 are outliers. However, these two outlying observations have a known cause. November, 1972 represents a bargain month in which residential extensions could be requested for free. Due to the large number of requests in November, not all orders could be filled. Hence, there was a carry over effect for December (Rousseeuw & Leroy, 1987, pg. 281).

Now consider the following seasonal differenced series,

$$X_t \;=\; \mathrm{RESX}_{t+12} - \mathrm{RESX}_t, \quad t = 1, 2, \ldots, 77.$$

According to Rousseeuw and Leroy (1987, pg. 281) Brubacher (1974) found that this series was stationary with a zero mean (except for the outliers) and could be modeled as an AR(p) without an intercept. One should recall that the mean of an autoregressive process is zero if and only if the intercept term is zero. A time series plot of the seasoned data is given in Figure 4 along with a lag one scatter plot with fits. The presence of the above mentioned outliers is clearly evident in these two plots. Following Rousseeuw and Leroy (1987), both the AR(1) and the AR(2) models will be fit to this seasoned series.

To begin with, consider the AR(1) model. Since the rank-based estimates are invariant to location, these estimation procedures are forced to fit a model which includes an intercept term. If the non-intercept model is desired one needs to first fit the model with a dummy intercept and then project the estimate of ρ into the correct space. For a more thorough discussion see Dixon and McK-

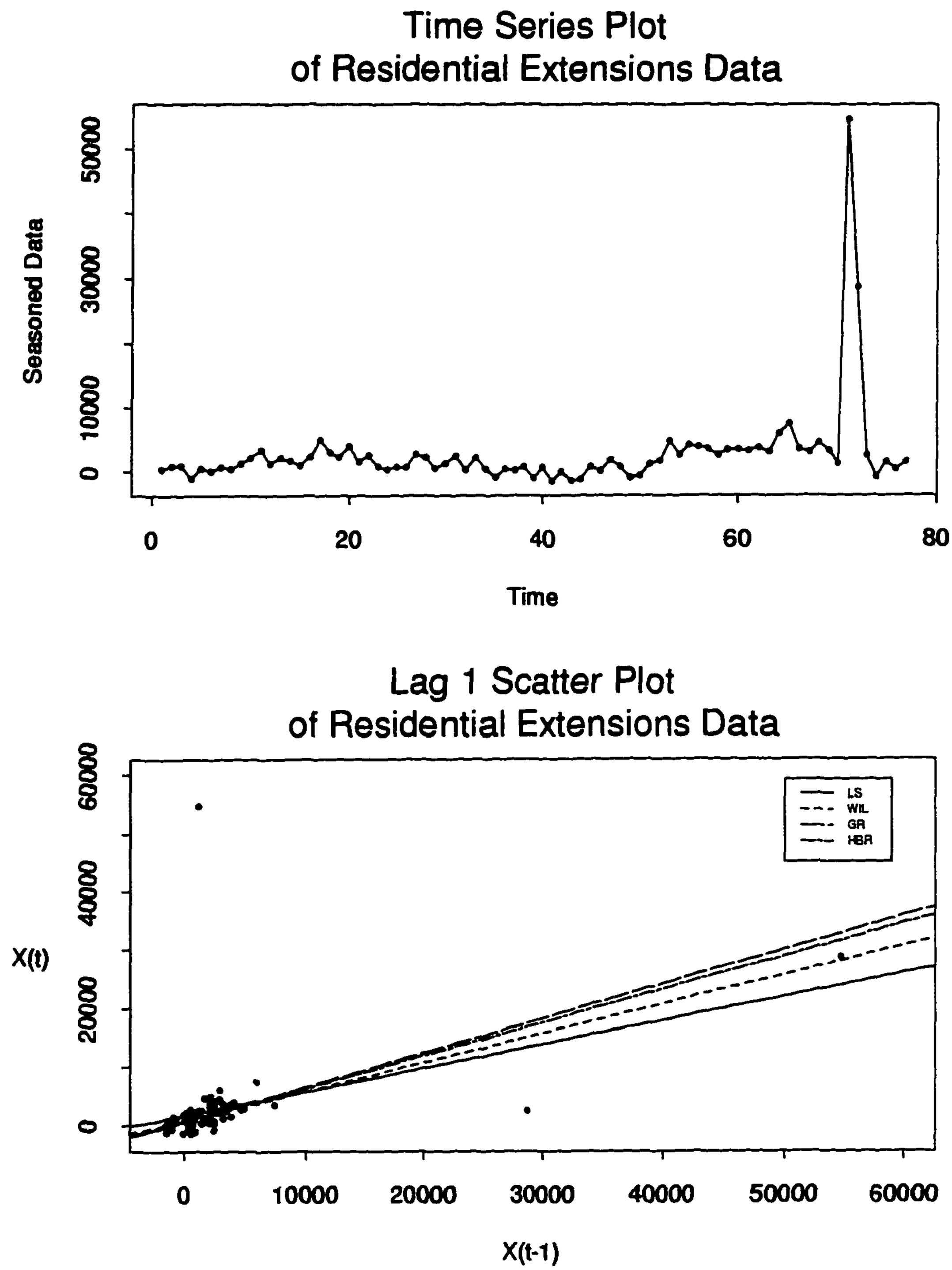

Figure 4. Time Series Plot and Lag 1 Scatter Plot With Fits for the Seasoned RESX Data.

Table 4

Outliers for the Seasoned Residential Extensions Data

AR(1)			AR(2)		
Case Number	Point	Type of Outlier	Case Number	Point	Type of Outlier
70	(1215, 54671)	Residual	69	(3165, 1215, 54671)	Residual
71	(54671, 28619)	Leverage	70	(1215, 54671, 28619)	Leverage
72	(28619, 2478)	Leverage	71	(54671, 28619, 2478)	Leverage
			72	(28619, 2478, -890)	Leverage

ean (1996). However, since we did not do this, the following model will be assumed,

$$X_t = \alpha + \rho X_{t-1} + \varepsilon_t, \quad t = 2, 3, \ldots, 77.$$

Now, since we suspect two outliers in the data set, there will be three points of interest. Information pertaining to these points is given in Table 4. A point is an outlier in residual space if the response itself is outlying. On the other hand, a point will be called a leverage point if the outlier rests in design space. The lag one scatter plot given in Figure 4 clearly demonstrates the sensitivity of both the LS and WIL fits to the leverage points. Since case number 71 is outlying in both residual and design space, this case results in a "good" leverage point and has little effect on the fits. Hence, it appears that case number 72, which contains only the outlier in design space, has the most influence on the LS and WIL fits. However, as evidenced in the plot, the GR and HBR estimates are not affected

by the outliers. Thus, these estimates portray a more accurate description of the bulk of the data.

Now consider the AR(2) model. Before we begin we should point out that analogous comments pertaining to the intercept in the AR(1) apply here as well. Thus, the following model is considered,

$$X_t = \alpha + \rho_1 X_{t-1} + \rho_2 X_{t-2} + \varepsilon_t, \quad t = 3, 4, \ldots, 77.$$

Since we are fitting the AR(2) model, the two suspect observations will effect four cases. Information pertaining to these points is also presented in Table 4. To get a visual display of the effects of these points on the estimates we need a plot in three dimensions. However, there is another way to view things. Consider the plot of the standardized residuals and time for the four fits given in Figure 5. The LS and WIL fits only identify one point, the residual outlier given by case number 69. The three leverage points given in Table 4 are "masked" from these analyses. However, both of the residual plots for the GR and HBR estimates clearly identify cases 69 and 71 as containing outliers. The plots also exhibit some evidence that cases 70 and 72 contain outliers. As was done in the AR(1) model, and for similar resons, these two points may be classified as "good" leverage points. Regardless of "good" or "bad" leverage points, the robustness of both the GR and HBR estimates to these points is evident. Hence, as with the AR(1) fits, these estimates provide a better description of the bulk of the data.

Other authors have also considered this data set. Rousseeuw and

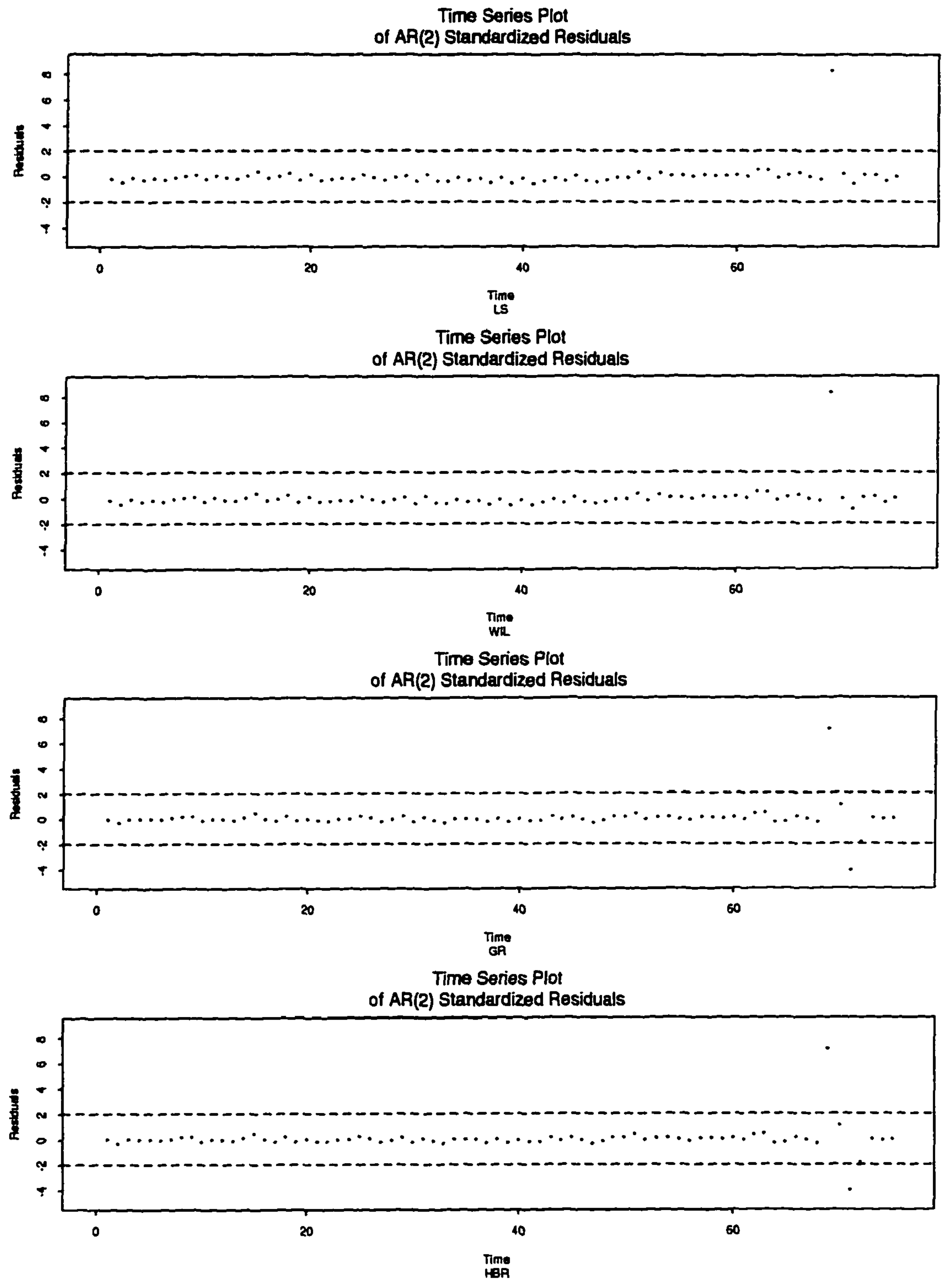

Figure 5. Residual Plots for the Four Fits of the AR(2).

Table 5

Parameter Estimates for the Residential Extensions Data

Estimate	AR(1) $\hat{\rho}_n$	AR(2) $\hat{\rho}_{n1}$	$\hat{\rho}_{n2}$
LS	0.482	0.533	-0.106
LMS	0.535	0.393	0.674
RLS	0.546	0.412	0.501
BRUB	NA	0.530	0.370
M	0.498	0.503	-0.147
GM	NA	0.510	0.380
WIL	0.497	0.503	-0.151
GR	0.586	0.363	0.399
HBR	0.563	0.357	0.387

NA - Not Available

Leroy (1987, pg. 283) discuss the least squares (LS), least median squares (LMS), and LMS-based reweighted least squares (RLS) estimates. Brubacher's (1974) estimates utilize an interpolation technique which requires that one specify the outlier positions in advance (Rousseeuw & Leroy, 1987, pg. 283). Pre-determining outlier positions would be similar to using the non-random weights discussed in Chapter II. Lastly, Martin (1980) applied a GM-estimate to this data set. For the sake of comparison, a M-estimate and the rank-based estimates of this chapter are computed. The various estimates for both the AR(1) and AR(2) models are given in Table 5. The various estimates yield similar results for the AR(1) model. However, there is some variation among the estimates for the AR(2) model. To begin with, the LS, M, and WIL fits yield negative parameter estimates for ρ_2.

This is due to the fact that these estimates do not have a bounded influence function, and thus, are sensitive to bad leverage points. Secondly, there is discrepancy between the estimates that are not sensitive to bad leverage points. Rousseeuw and Leroy (1987, pg. 283) offer as an explanation the high correlation between X_{t-1} and X_{t-2} (r=0.634). This is also apparent in the scatter plot of Figure 4. Thus, it is possible that a collinearity situation exists and this somehow effects the various estimates in different ways. To avoid this problem, one could certainly consider only the AR(1) model where the estimates are more stable. However, Martin (1980) found that the AR(2) model is more suitable for this situation using a robustified version of Akaike's AIC function for determining the order of an autoregression (Rousseeuw & Leroy, 1987, pg. 284).

4.3 Missing Observations

Consider first the classical linear regression model with possible missing observations in the response and or the design variables. In many statistical packages missing observations in this setting are handled simply by deleting all cases which contain missing information from the data set. However, in time series analysis, this technique would disrupt the "nature" of the time series. Hence, other methods need to be considered when a time series consists of missing observations. Such procedures usually involve maximizing a likelihood function. For instance, if missing values are present, the S-plus statistical software package computes the

likelihood using the Kalman filter applied to a state space representation of the likelihood (Sciences, 1993, pg. 16-22). One is referred to Kohn and Ansley (1986) for the state space representation and Bell and Hillmer (1987) for the method of initializing the filter recursions. If there are not too many missing values one can use indicator variables and an ordinary maximum likelihood program to construct the maximum likelihood estimators (Fuller, 1996, pg. 459). This technique is illustrated with an example given in Fuller (1996, pg. 459).

As an alternative, consider the GR estimate with non-random weights. Specifically, define the weights by $b_{ij} = h_i h_j$, where $h_i = 0$ if the i^{th} case contains a missing observation in either the response and or the design and $h_i = 1$ otherwise. This essentially eliminates the cases that contain missing observations, similar to what is done in most statistical packages with classical linear regression. However, the use of weights preserves the structure of the time series. The advantage this method has over maximum likelihood methods is that a model for the innovations does not have to be specified, and hence, does not depend on any likelihood function. This technique, as well as the maximum likelihood methods mentioned, are illustrated in the following example.

Consider the data found in Table 8.7.1 of Fuller (1996, pg. 459). The data represents computer generated observations from an AR(2) model with Gaussian errors. Thus, we consider the following model,

$$X_i = \alpha + \rho_1 X_{i-1} + \rho_2 X_{i-2} + \varepsilon_i, \quad i = 1, 2, \ldots, 100, \qquad (4.3)$$

where it is assumed that observations 40, 57, and 58 are missing at random. If the i^{th} case is defined by the tuplet (X_{i-1}, X_{i-2}, X_i), then it follows that cases 1, 2, 40, 41, 42, 57, 58, 59, and 60 contain missing observations. These missing observations appear either in the response (e.g. Case 40), the design (e.g. Case 41), or both (e.g. Case 58). To analyze this data set using the GR estimate with non-random weights, let $h_i = 0$ for $i = 1, 2, 40, 41, 42, 57, 58, 59, 60$ and let $h_i = 1$ otherwise. Also, since the program used to calculate the GR estimates assumes non-missing observations we replace the missing observations with the sample mean of the non-missing data. This should have little effect on the estimate since these cases are basically being disregarded from the analysis. For comparison purposes, estimates from S-plus and those of Fuller were also obtained. It should be noted that the parameterization given in Fuller yields the negatives of the autoregressive parameters given in (4.3). If $(\hat{\rho}_{n1}, \hat{\rho}_{n2})'$ denotes the estimate of the autoregressive parameter vector, the S-plus, Fuller, and GR estimates are given by $(1.338, -0.847)'$, $(1.377, -0.899)'$, and $(1.340, -0.859)'$ respectively. Hence, the GR estimate is comparable to those obtained in S-plus and by Fuller. However, this should not be surprising since $b_{ij} = 1$ yields the Wilcoxon estimate and it is well-known that the Wilcoxon estimate has 95.5% efficiency when the innovations are normal (e.g. Hettmansperger (1984, pgs. 72,243)). With only 9 of the h_i's set to zero, 9% of the b_{ij}'s are different from one. Thus, this estimate should be close to the Wilcoxon estimate, and hence, the maximum likelihood estimates. Again,

the advantage the GR estimate has over the maximum likelihood procedures is that the error distribution does not have to be specified. However, the effect of using the GR estimate to account for missing observations may have some effect on inference. Since the intent of this example is to illustrate an application of the non-random weight theory of Chapter II, this will not be pursued.

CHAPTER V

A MONTE CARLO STUDY

In this chapter the behavior of the proposed estimate is studied via Monte Carlo. For the sake of simplicity and computation time only the AR(1) is simulated. Furthermore, since the primary interest is the behavior of the autoregressive parameter and not the intercept, only the zero mean AR(1) is considered. Thus, the "core" process in this chapter will be denoted by,

$$X_i \;=\; \rho X_{i-1} + \varepsilon_i, \quad i = 2, 3, \ldots, n,$$

where $\rho \in (-1, 1)$ and the ε_i are iid F.

The "observed" process, defined in Section 1.3, is given by,

$$X_i^* \;=\; X_i + \nu_i, \quad i = 2, 3, \ldots, n.$$

When $\nu_i = 0$ the "observed" process reduces to the "core" process. This is the situation that gives rise to the Innovation Outlier (IO) model. When $\nu_i \neq 0$ the classical Additive Outlier (AO) model is obtained. In this chapter the ν_i will have a distribution given by $(1-\gamma)\delta_0(\cdot) + \gamma G(\cdot)$ where γ denotes the proportion of contamination, δ_0 is a point mass at zero, and G is some contaminating distribution function. As mentioned in Section 1.3 the ν_i do not necessarily have to be independent.

As in Chapter IV we consider the least squares estimate (LS) and the three rank-based estimates; WIL, GR, and HBR. One will recall that the WIL, GR, and HBR estimates are determined by their weighting schemes. The weighting schemes are defined by $b_{ij} = 1$, (4.1), and (4.2) respectively. The following Monte Carlo study will simulate, compute, and compare the four estimates of ρ under both the IO and AO models for different settings of γ, G, and F. In all of the simulations 1000 realizations of size 100 were generated and the estimates were computed using the algorithms discussed in Chapter IV. However, due to a processing error in a few of the cases it was not possible to obtain 1000 realizations all of the time. The worst case scenario was 994 realizations. Thus, when applicable, we will denote the actual number of simulations by $nsims$.

Evaluation of the four estimates under the different situations will be done via comparison boxplots. The length of the box is determined by the first and third quartiles. The box also exhibits a line denoting the median. The whiskers for the boxplot are drawn to the nearest point not beyond 1.5 times the interquartile range. Furthermore, the widths of the boxes are proportional to the empirical mean square error, $\widehat{\text{MSE}}$. That is, the widths are proportional to,

$$\widehat{\text{MSE}} = \frac{1}{nsims} \sum_{i=1}^{nsims} (\hat{\rho}_{ni} - \rho)^2.$$

Lastly, a line at the true value of ρ is also displayed. Hence, these plots allow for comparison between the four estimates' distributions in regards to shape, variability, accuracy, and bias.

5.1 Innovation Outliers

5.1.1 IO Setup

Under model IO, $\nu_i = 0$ so that the "observed" process reduces to the "core" process. Hence, the model in this section is,

$$X_i = \rho X_{i-1} + \varepsilon_i, \qquad i = 2, 3, \ldots, n,$$

where $\rho \in (-1, 1)$ and the ε_i are iid F. The values of ρ considered throughout the study are 0, 0.5, and 0.9. For each value of ρ four different F's were considered. The error distributions considered were the Standard Normal, Logistic, and Contaminated Normal. One Contaminated Normal had a contaminating variance of 16 (10% of the time) while the second Contaminated Normal had a contaminating variance of 100 (25% of the time). The Standard Normal and Logistic were chosen since these are the respective distributions where the LS and WIL estimates are optimal (e.g. Hampel et al. (1986, pg. 112)). The Contaminated Normal parameter settings were chosen to reflect "mild" contamination and "extreme" contamination.

5.1.2 IO Simulation Results

The results for $\rho = 0, 0.5$, and 0.9 and the four error distributions are given in Figure 6, Figure 7, and Figure 8 respectively. As expected, the LS estimate does the best when the error distribution is Gaussian. However, as evidenced in

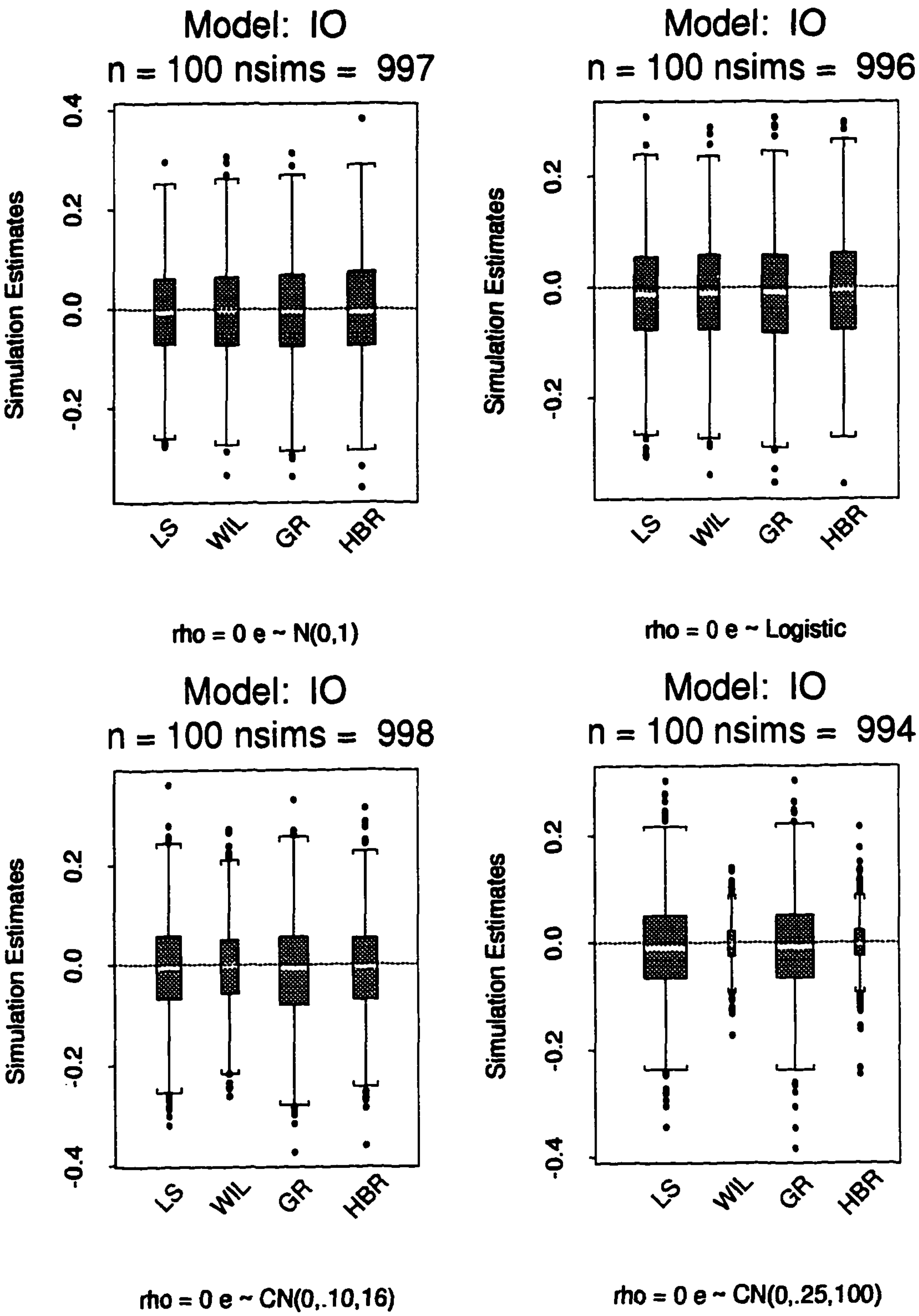

Figure 6. Simulations for Model IO Under $\rho = 0$.

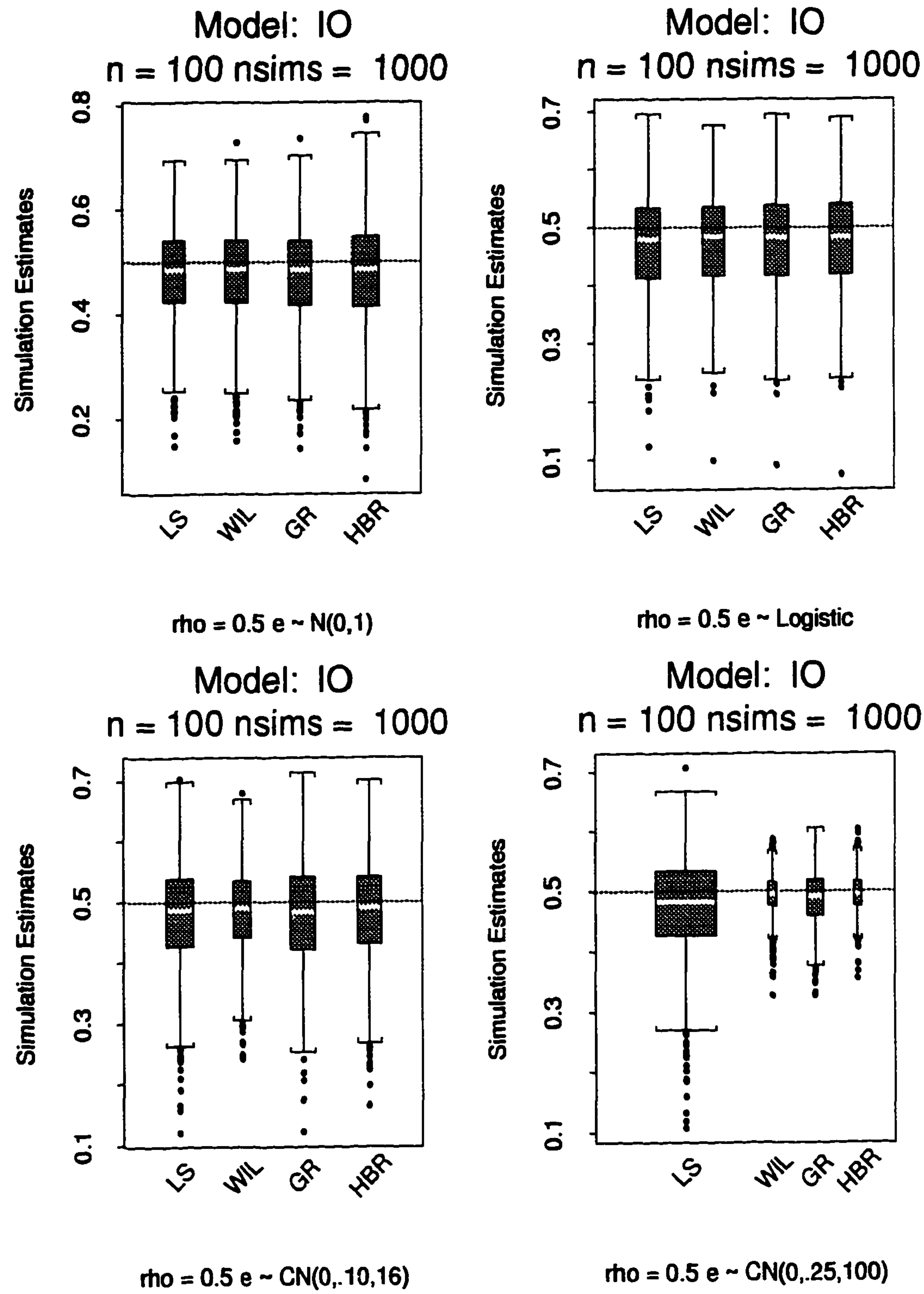

Figure 7. Simulations for Model IO Under $\rho = 0.5$.

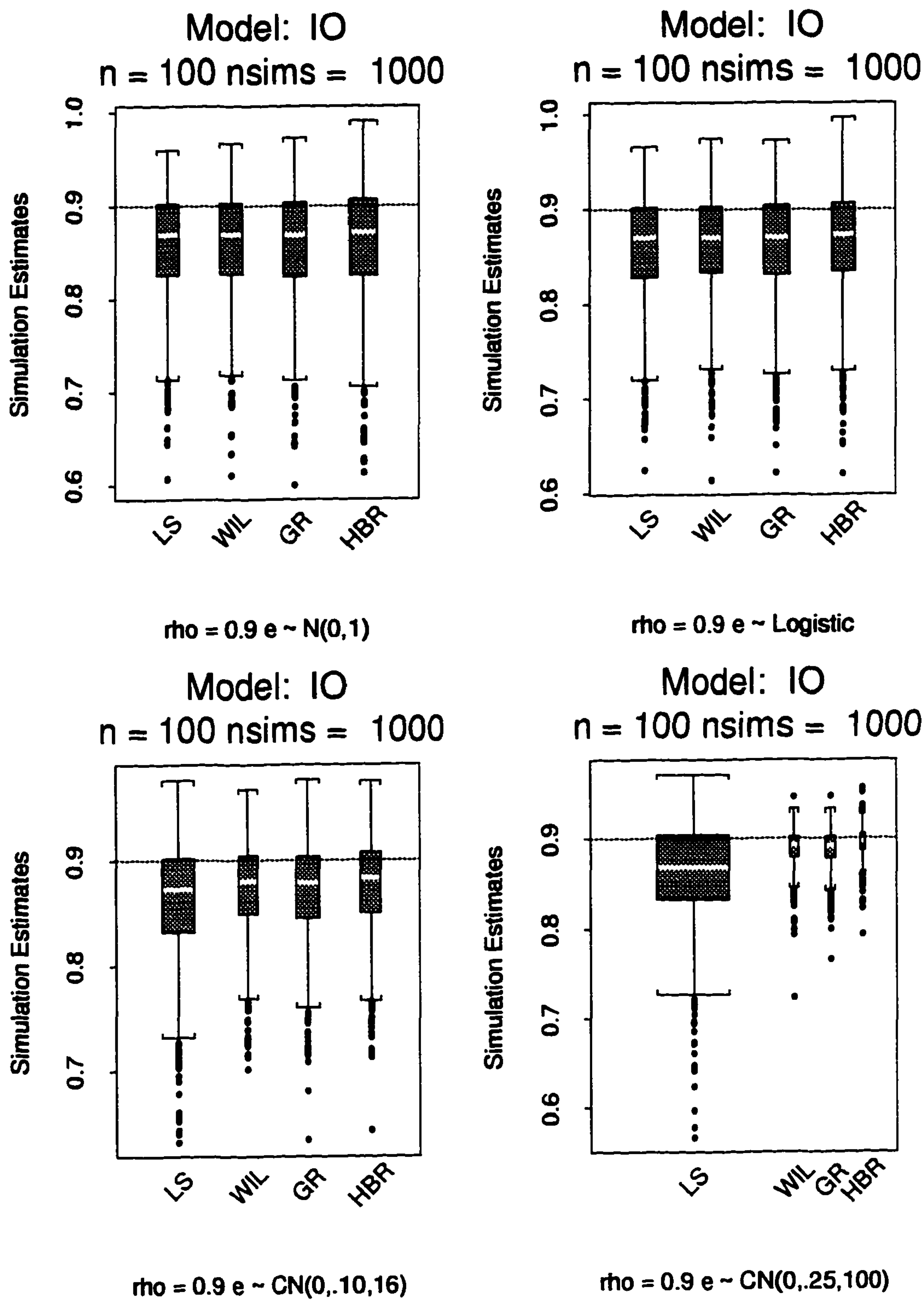

Figure 8. Simulations for Model IO Under $\rho = 0.9$.

the plots, the distributions for the rank-based estimates are quite close to that of least squares, except for a small increase in variability. Furthermore, it is also worth noting that as ρ increases, the bias in all four estimates also increases.

Next, consider the situation when the error distribution is Logistic. This is the situation in which the WIL estimate is optimal (Hampel et al. (1986, pg. 112)). The GR and HBR, as well as the LS estimate, do quite well compared to the optimal estimate. Additionally, the bias still appears to be an increasing function of ρ.

Now consider the situations where the error terms are Contaminated Normals. It is obvious from the plots that the least squares estimate does not perform very well. However, the GR also seems to be having difficulty, although not to the extent of the LS estimate. Consider the following explanation. Under model IO outliers are introduced through the error distribution. Due to the structure of the AR(1) these errors get introduced into future observations of the process, and hence, produce a sequence of "good" leverage points. Although the LS estimate will tend to be bias towards these leverage points, it will still be sensitive to outliers in residual space. However, the rank-based estimates have bounded influence functions in residual space, and thus, are not effected by outliers in the vertical direction. The difference between the WIL, GR, and HBR estimates is directly related to the associated weighting schemes. The WIL estimate is obtained by letting the weights be constant. Thus, the "good" leverage points are given the

same weight as the other design points. This tends to strengthen the estimate. Since the HBR estimate distinguishes between "good" and "bad" leverage points, the HBR will be similar to the WIL, as portrayed in all of the figures. The GR does not distinguish between "good" and "bad" leverage points. Thus, the GR downweights both "good" and "bad" leverage points, and as a consequence, tends to lose efficiency and have a larger variance. However, it does appear that the performance of the GR improves as ρ increases. This should not be surprising since it is well-known that $\text{Corr}(X_i, X_{i-1}) = \rho$ (Abraham & Ledolter, 1983, pg. 200). The GR essentially throws away the leverage points and bases the estimate of ρ on the "good" cluster. But, as ρ increases, the "good" cluster gets tighter and tighter, hence improving the estimate. As a final remark it should be pointed out that once again the bias seems to be an increasing function of ρ. Thus, taking into consideration the above comments it would appear that under model IO the WIL estimate is superior.

The superiority of the WIL estimate to that of the GR under model IO can be shown analytically as well. Although the following ideas should extend to the general case, a few additional assumptions are added to simplify the discussion. The weights will be of the form, $b_{ij} = h(X_{i-1})h(X_{j-1})$. Since multipling by a constant ($K = E^{-2}[h(X)]$) does not change the value of the estimate, one can assume without loss of generality, that $E[h(X)] = 1$. Furthermore, assume that the error distribution is such that the distribution of X is symmetric. Then,

assuming h is an even function, it follows immediately that $E[h(X)X] = 0$. Now we are ready to proceed with the discussion.

In Section 3.4 it was shown that

$$\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{12\tau^2} C^{-1} E[h(X)] E\left[h^2(X)X^2\right] C^{-1}\right) \quad \text{where} \tag{5.1}$$

$$C = E[h(X)] E[Xh(X)X] - E^2[h(X)X]. \tag{5.2}$$

The simplifying assumptions imply that (5.1) and (5.2) reduce to

$$\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{12\tau^2} C^{-1} E\left[h^2(X)X^2\right] C^{-1}\right) \quad \text{where} \tag{5.3}$$

$$C = E[Xh(X)X] = \mathrm{COV}[X, h(X)X] = \rho_{X,h(X)X} \sigma_X \sigma_{h(X)X}. \tag{5.4}$$

In the above expression $\rho_{X,h(X)X}$ denotes the correlation between X and $h(X)X$ while σ_X and $\sigma_{h(X)X}$ denote the standard deviations for X and $h(X)X$ respectively. Following Denby and Martin (1979) let

$$V_{IO} = \frac{1}{12\tau^2} C^{-1} E\left[h^2(X)X^2\right] C^{-1} \tag{5.5}$$

denote the asymptotic variance of the estimate. Inserting (5.4) into (5.5) and applying algebra yields,

$$\begin{aligned} V_{IO} &= \left(\frac{1-\rho^2}{\rho^2_{X,h(X)X}\sigma^2_\varepsilon}\right)\left(\frac{1}{12\tau^2}\right) \\ &= \left(\frac{1-\rho^2}{\rho^2_{X,h(X)X}\sigma^2_\varepsilon}\right) V_{LOC}, \end{aligned}$$

where ρ denotes the AR(1) parameter, $\sigma^2_\varepsilon = E[\varepsilon^2]$, and $V_{LOC} = \frac{1}{12\tau^2}$ represents the asymptotic variance of the Wilcoxon estimator of location (Hettmansperger,

1984, pg. 76). Next, assuming only that $\sigma_\varepsilon^2 < \infty$, the asymptotic form of the Cramer-Rao bound for estimates of ρ under model IO can be shown to be (Denby & Martin, 1979)

$$V_{CR} = \frac{1-\rho^2}{I(f)\sigma_\varepsilon^2},$$

where

$$I(f) = E\left[\left(\frac{f'(\varepsilon)}{f(\varepsilon)}\right)^2\right]$$

denotes the Fisher Information for f. Hence, the asymptotic efficiency of the GR estimate based on h and f is given by

$$\begin{aligned} \mathrm{EFF}_{GR}(h, f) &= \frac{V_{CR}}{V_{IO}} \\ &= \rho^2_{X,h(X)X}\left(\frac{12\tau^2}{I(f)}\right) \\ &= \rho^2_{X,h(X)X}\mathrm{EFF}_{LOC}(f), \end{aligned}$$

where $\mathrm{EFF}_{LOC}(f)$ denotes the asymptotic efficiency of the Wilcoxon estimate of location. Now, because $\rho_{X,h(X)X}$ is the correlation between X and $h(X)X$ it is immediate that the maximum of $\mathrm{EFF}_{GR}(h, f)$ with respect to h is obtained when $h = 1$. That is, the asymptotic efficiency of the GR is greatest when the GR reduces to the WIL estimate ($b_{ij} = 1$). As a consequence, one should never use the GR estimate if model IO is expected to hold. However, as our simulations will show, the GR can be quite useful in situations where model IO does not hold.

5.2 Additive Outliers

5.2.1 AO Setup

Under model AO, $\nu_i \neq 0$. In the simulations that follow ν_i will have a distribution given be $(1-\gamma)\delta_0(\cdot) + \gamma G(\cdot)$ where γ denotes the proportion of contamination, δ_0 is a point mass at zero, and G is $\mathcal{N}(0, \sigma_\nu^2)$. However, the ν_i do not have to be independent. Thus, the model for the "observed" process is,

$$\begin{aligned} X_i^* &= X_i + \nu_i \\ &= \rho X_{i-1} + \varepsilon_i + \nu_i, \quad i = 2, 3, \ldots, n. \end{aligned}$$

The values of ρ where set at 0, 0.5, and 0.9. In all of the AO simulations the ε_i were distributed as $\mathcal{N}(0, 1-\rho^2)$. This implies that the "core" process has a Standard Normal distribution. The ν_i were assumed independent of the ε_i and were distributed according to $(1-\gamma)\delta_0(\cdot) + \gamma\mathcal{N}(0, \sigma_\nu^2)$, but the ν_i were not necessarily independent. A dependence structure was obtained by generating a sequence of binary random variables, γ_i, according to

$$\gamma_i \sim \begin{cases} \text{bernoulli}(p_1) & \text{if } \gamma_{i-1} = 0 \\ \text{bernoulli}(p_2) & \text{if } \gamma_{i-1} = 1 \end{cases}$$

where $p_1 \leq p_2$. When $p_1 = p_2$ the sequence is iid. However, when $p_1 < p_2$ the sequence will not be independent. In the simulations to follow, "patches" of outliers were obtained by letting $p_1 \leq 0.25$ and $p_2 = 0.80$. For each of the three values of ρ there were four situations. These situations are given by $(\gamma, \sigma_\nu^2) =$

(0.10, 16) and $(\gamma, \sigma_\nu^2) = (0.25, 100)$ for both the iid and dependent ν_i.

5.2.2 AO Simulation Results

The results for $\rho = 0, 0.5,$ and 0.9 along with the four AO situations are given in Figure 9, Figure 10, and Figure 11 respectively. As one may expect, the boxplots for the situation when $\rho = 0$ closely resemble those presented in Figure 6. This is obvious when comparing row 2 in Figure 6 with row 1 of Figure 9. This follows because in both the IO and AO models the "observed" process is just a sequence of random noise when $\rho = 0$.

However, when $\rho \neq 0$ the results are quite different from the IO model results. The first thing to note is that in all of the plots (where $\rho \neq 0$) both the GR and HBR estimates are superior to the LS and WIL estimates. This follows since the LS and WIL estimates do not have a bounded influence function in design space. One should recall that model AO is the situation that gives rise to "bad" leverage points. Secondly, in the majority of the cases the GR seems to perform somewhat better than the HBR estimate in regards to bias and mean square error. This brings us to the obvious feature that is present in all of the plots. That is, all of the estimates are bias, and in some situations the bias is quite large. This is due to the fact that the true AR(1) model is no longer valid. In fact, under our AO setup the AR(1) holds $(1-\gamma)100\%$ of the time while $\gamma 100\%$ of the time it is something different. It appears that the bias may be a function

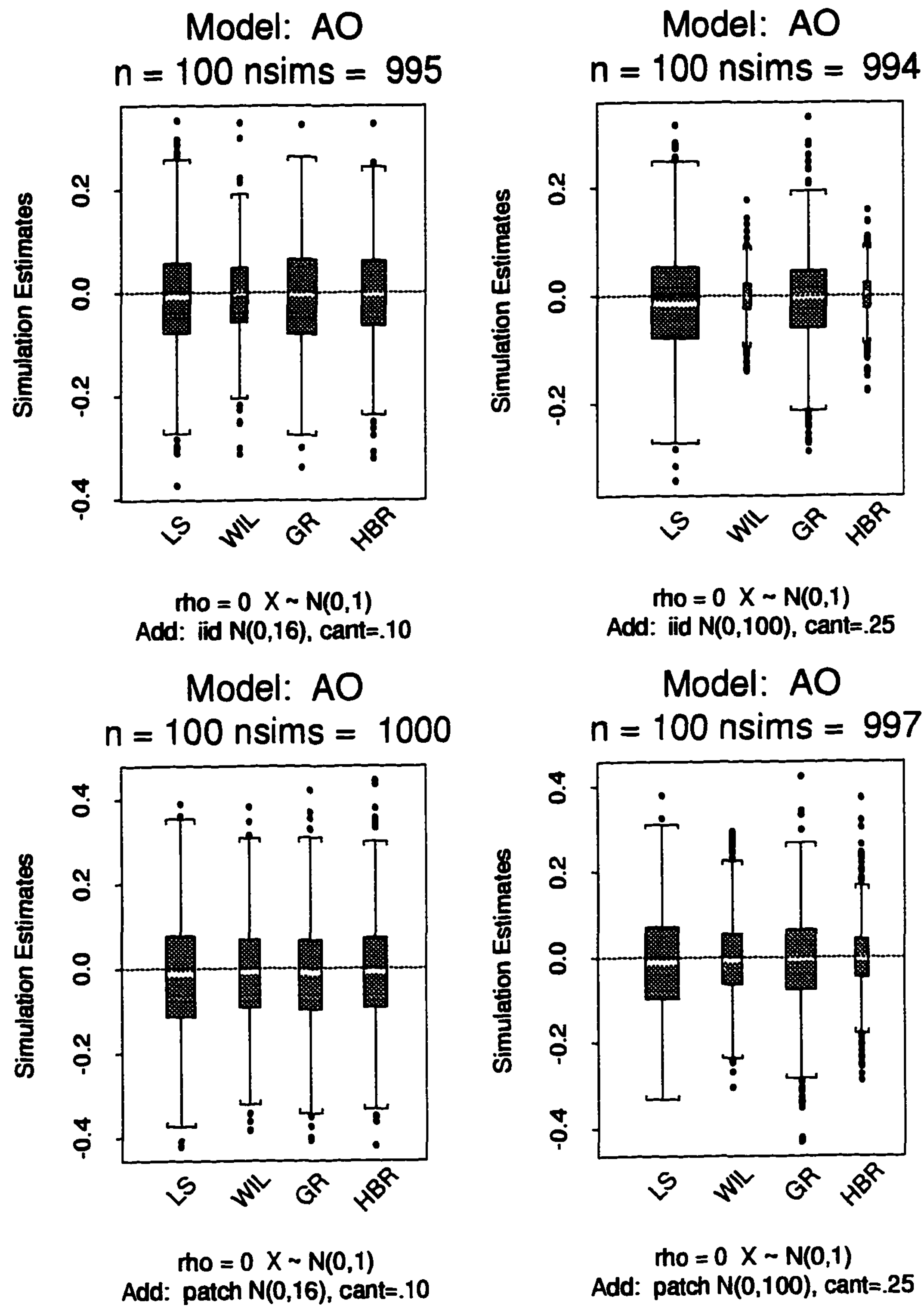

Figure 9. Simulations for Model AO Under $\rho = 0$.

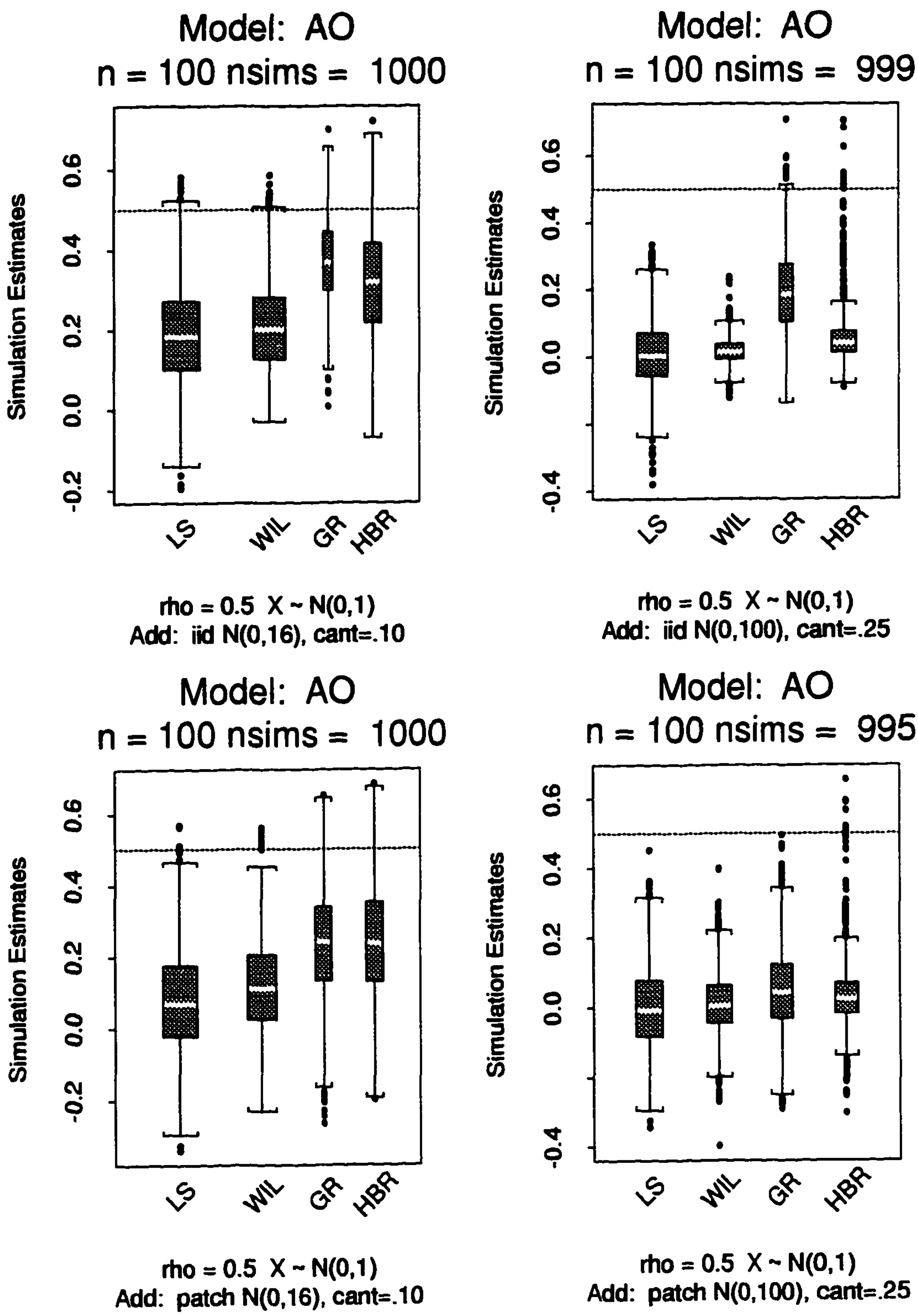

Figure 10. Simulations for Model AO Under $\rho = 0.5$.

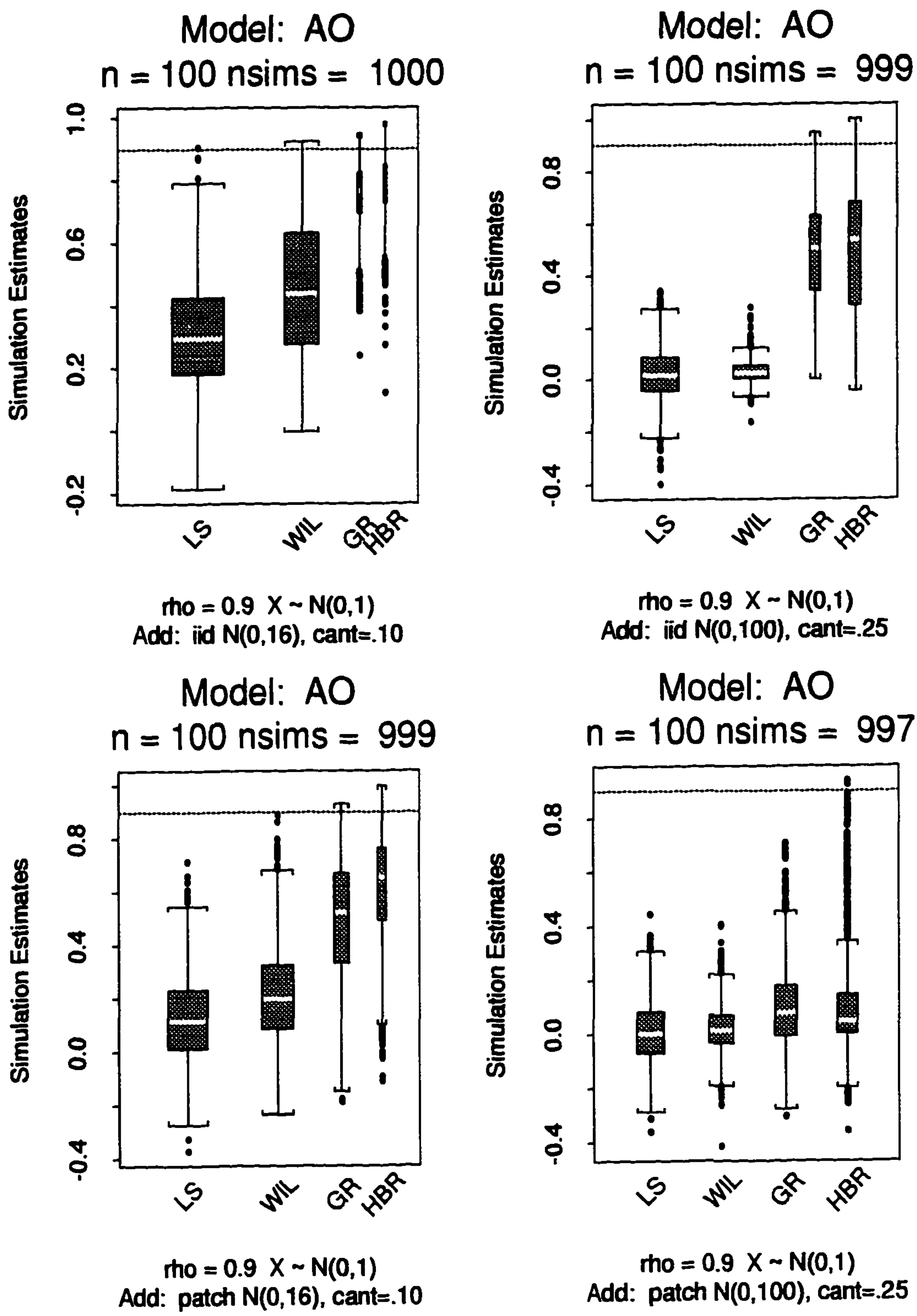

Figure 11. Simulations for Model AO Under $\rho = 0.9$.

of the proportion and amount of contamination. Smaller values of γ and σ_ν^2 are associated with a smaller bias while larger values of these parameters yield a larger bias. This is apparent in row 1 of both Figure 10 and Figure 11. Furthermore, this feature is consistent under both the iid and dependent scenarios. As a final note concerning these figures consider the case of dependent ν_i where γ and σ_ν^2 are large. For both $\rho = 0.5$ and $\rho = 0.9$, the distributions of the GR and HBR estimates are centered close to zero. Although this bias is a direct consequence of the AR(1) being invalid, there may be another factor that contributes to the severity of the bias. Contamination under the dependent situation when $\gamma = 0.25$ produces "patches" of outliers 25% of the time. Thus, for a sample size of 100 one would expect at least 25 outlying observations, and because of the dependency, probably more. Hence, it is possible that the actual proportion of contamination may be approaching the estimates' breakdown point. As a consequence, the bias may be magnified.

To further investigate the relationship between bias and amount of contamination additional simulations (with smaller amounts of contamination) where obtained. Five situations with $\rho = 0.5$, $X \sim \mathcal{N}(0,1)$, and ν_i iid were considered. The situations are given by $(\gamma, \sigma_\nu^2) = (0.02, 16)$, $(0.02, 100)$, $(0.05, 16)$, $(0.05, 100)$, and $(0.10, 100)$ respectively. The results are given in Figure 12 and Figure 13. It is immediate upon comparing Figure 10 with Figure 12 that the bias has reduced. Figure 12 and Figure 13 also indicate additional evidence that the GR performs

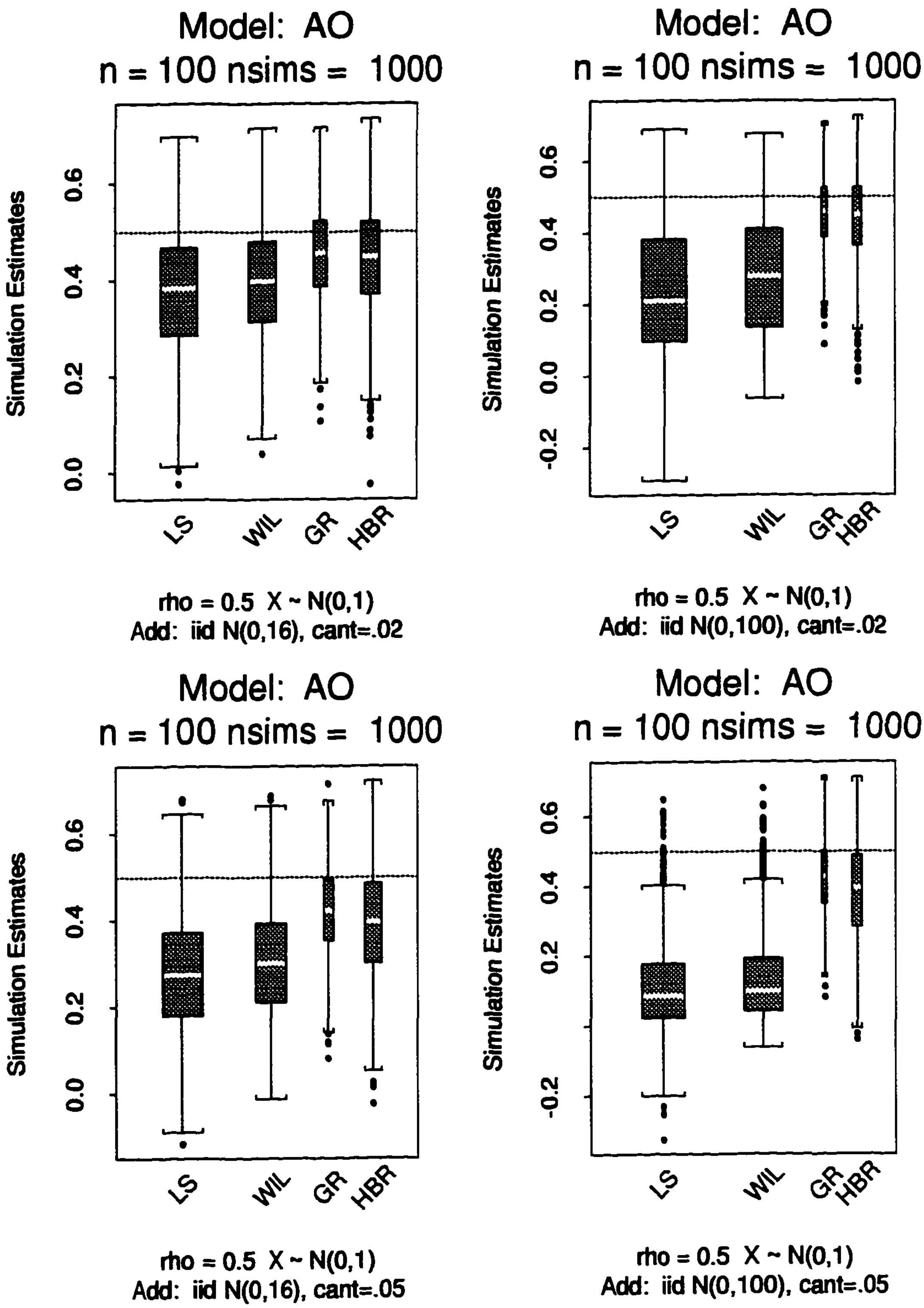

Figure 12. Additional Simulations for Model AO Under $\rho = 0.5$.

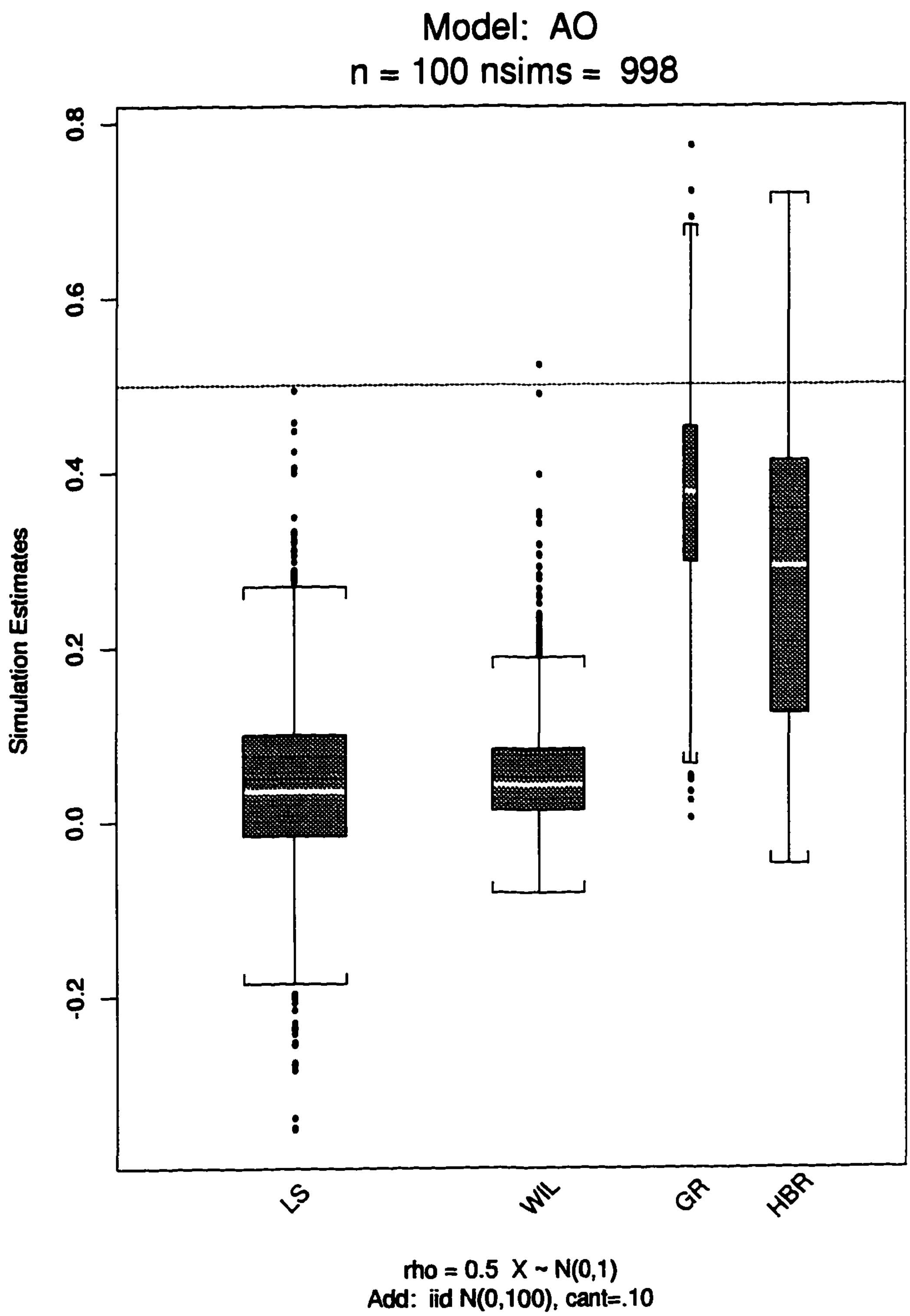

Figure 13. One More Simulation for Model AO Under $\rho = 0.5$.

slightly better than the HBR estimate.

5.3 Summary of Monte Carlo Results

To summarize, the Monte Carlo study provides evidence supporting the superiority of the WIL estimate to that of the GR estimate under model IO. This was shown analytically as well. However, the performance of the HBR resembles that of the WIL under model IO.

Under model AO, the GR and HBR estimates are superior to the LS and WIL estimates, although bias. The bias seems to be directly related to the proportion and degree of contamination. There is also some evidence that suggests the GR dominates the HBR, although this evidence is not strong enough to make a definitive statement.

Because the true model is never really known in practice and the HBR seems to perform well under both the IO and AO situations, it would appear that the HBR would be most suitable for AR(p) estimation. However, since the HBR estimate is not the subject of this paper, this will have to be considered in future research.

CHAPTER VI

CONCLUSIONS

6.1 Concluding Remarks

Throughout this manuscript we considered the following model,

$$\begin{aligned} X_i &= \alpha + \rho_1 X_{i-1} + \rho_2 X_{i-2} + \cdots + \rho_p X_{i-p} + \varepsilon_i \\ &= \alpha + \boldsymbol{Y}'_{i-1}\boldsymbol{\rho} + \varepsilon_i; \qquad i = 1, 2, \ldots, n, \end{aligned}$$

where $p \geq 1$, $\boldsymbol{Y}'_{i-1} = (X_{i-1}, X_{i-2}, \ldots, X_{i-p})$, and $\boldsymbol{\rho}' = (\rho_1, \rho_2, \ldots, \rho_p)$. $\boldsymbol{Y}_0$ was assumed to be an observable random vector independent of ε_1, ε_2, $\ldots$, ε_n such that $E\left[\|\boldsymbol{Y}_0\|^k\right] < \infty$ for $k = 2$ or 4. Furthermore, the ε_i were assumed iid F with $E\left[\varepsilon_i\right] = 0$ and $E\left[\varepsilon_i^k\right] < \infty$ for $k = 2$ or 4. Lastly, the solutions to the following equation,

$$X^p - \rho_1 X^{p-1} - \rho_2 X^{p-2} - \cdots - \rho_p = 0,$$

were assumed to lie in the interval $(-1, 1)$. These conditions implied that the process followed a stationary p^{th} order autoregressive time series, denoted by AR(p).

The primary objective was the robust estimation of the autoregressive parameter vector, $\boldsymbol{\rho}$. Specifically, the proposed estimate of $\boldsymbol{\rho}$ was defined to be a

minimum of the following dispersion function,

$$\begin{aligned} D(\boldsymbol{\rho}) &= \sum_{1\le i<j\le n} b_{ij}\,|\varepsilon_i - \varepsilon_j| \\ &= \sum_{1\le i<j\le n} b_{ij}\,|(X_i - X_j) - (\boldsymbol{Y}_{i-1} - \boldsymbol{Y}_{j-1})'\boldsymbol{\rho}|\,, \end{aligned}$$

where the b_{ij} denoted weights used for the $(i,j)^{th}$ comparison. Alternatively, one can view the estimate of $\boldsymbol{\rho}$ as an approximate solution of the equation $\boldsymbol{S}(\boldsymbol{\rho}) = \boldsymbol{0}$ where,

$$\begin{aligned} \boldsymbol{S}(\boldsymbol{\rho}) &= -\boldsymbol{\nabla} D(\boldsymbol{\rho}) \\ &= 2\sum_{1\le i<j\le n} b_{ij}\,(\boldsymbol{Y}_{j-1} - \boldsymbol{Y}_{i-1})\left(\varphi(\varepsilon_i(\boldsymbol{\rho}), \varepsilon_j(\boldsymbol{\rho})) - \frac{1}{2}\right). \end{aligned}$$

This estimate was first considered by Sievers (1983) in the context of the linear regression model. In the case where the weights are constant, Hettmansperger and McKean (1978a) have shown that this dispersion function is equivalent to Jaeckel's (1972) dispersion function using Wilcoxon scores. Since then, this estimate has been referred to as a generalized (or modified) rank-based estimate (GR for short).

In Chapter II we considered the case where the weights were non-random. This, of course, covers the situation when the weights are considered to be constant. Following traditional rank-based theory, results pertaining to the asymptotic normality of the gradient, $\boldsymbol{S}(\boldsymbol{\rho})$; asymptotic uniform linearity of the gradient; and asymptotic uniform quadraticity of the dispersion function were proven. These results lead to the asymptotic normality of the proposed estimate. Although

the above results were established in the context of the AR(1), it is conjectured that these results will extend to the AR(p) with a minimal amount of effort.

In Chapter III we considered the case where the weights were random. Specifically, we considered the following weighting scheme,

$$b_{ij} = h\left(\boldsymbol{Y}_{i-1}; \widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n\right) h\left(\boldsymbol{Y}_{j-1}; \widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n\right),$$

where $\widehat{\boldsymbol{\mu}}_n$ and $\widehat{\boldsymbol{\Sigma}}_n$ were robust measures of location and scatter for $\{\boldsymbol{Y}_{i-1}\}$ respectively. Again, results pertaining to the asymptotic normality of the gradient, asymptotic uniform linearity of the gradient, and asymptotic uniform quadraticity of the dispersion function were proven. Once more, these results lead to the asymptotic normality of the proposed estimate. However, unlike the results of Chapter II, these results were proven in the general context of the AR(p).

In terms of the classical linear regression model, the proposed estimate possesses a positive breakdown point and a bounded influence function. One is referred to Naranjo and Hettmansperger (1994) and or Chang (1996) for the details. Although these results are not proven here, they were shown via example and Monte Carlo. In Chapter IV we presented both a simulated and actual example that leads one to believe that these properties will also hold for the autoregressive time series model. Additionally, an example pertaining to missing observations was presented in order to provide another application of the nonrandom weight theory presented in Chapter II.

Lastly, in Chapter V we studied the proposed estimate via Monte Carlo.

Again, there was evidence that suggests that the estimate has both a positive breakdown point and a bounded influence function. More specifically, there was evidence supporting the superiority of the WIL estimate to that of the GR estimate under model IO. The performance of the HBR resembled that of the WIL under model IO. However, under model AO, the GR and HBR estimates were superior to the LS and WIL estimates, although bias. There was also some evidence that suggested the GR dominates the HBR, although this evidence was not strong enough to make a definitive statement.

6.2 Future Research Topics

Recall the weighting scheme used in Chapter III,

$$b_{ij} = h\left(\boldsymbol{Y}_{i-1}; \widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n\right) h\left(\boldsymbol{Y}_{j-1}; \widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n\right).$$

Although this "Mallows type" weighting scheme can probably be justified in most problems, there may be situations when such a "factorization" is not practical. Thus, the need for a more general weighting scheme, such as that defined below,

$$b_{ij} = \psi\left(\boldsymbol{Y}_{i-1}, \boldsymbol{Y}_{j-1}; \widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n\right)$$

is needed. From a theoretical point of view the factorization simplifies many of the proofs needed in order to obtain asymptotic normality. The general weighting scheme should provide an interesting challenge in terms of theoretical development.

In terms of global robustness we saw empirical evidence that indicated the estimate possessed a positive breakdown point, ε^*. For the classical linear regression model and the weights defined in Chapter III Naranjo and Hettmansperger (1994) have show that $0 < \varepsilon^* \leq \frac{1}{3}$. Furthermore, Chang (1996) has shown that the HBR estimate attains the same breakdown point as some initial estimate used to calculate the weights. Therefore, the HBR estimate can have a 50% breakdown point. Due to the linear structure of the AR(p) one may expect these results to hold in this context as well. However, this is not at all clear to this investigator since the mechanism of breakdown plays an important role in determining breakdown points. Thus, it is possible that the breakdown point varies according to the innovation, additive, and or general replacement outlier models.

Our examples and Monte Carlo study also suggested that the influence function for the estimate is bounded. For discussion purposes let $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$, $\ldots$, $\boldsymbol{Z}_n$ denote iid observations from some distribution function, G, and let $\boldsymbol{T}(G)$ denote some functional to be estimated. A natural estimate is $\boldsymbol{T}(G_n)$ where G_n represents the empirical distribution function. Under regularity conditions it has become custom to expect the following,

$$\sqrt{n}\,(\boldsymbol{T}(G_n) - \boldsymbol{T}(G)) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Omega\,(\boldsymbol{Z}_i;\boldsymbol{T},G) + o_p(1),$$

where $\Omega\,(\boldsymbol{Z};\boldsymbol{T},G)$ denotes the influence function of the estimate. One is referred to Hampel (1986) for a thorough discussion. It has also become customary to expect that what holds true for the classical linear regression model will also hold

true for the autoregressive model. Now recall the following result from Chapter III,

$$\begin{aligned}
&\sqrt{n}\,(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) \\
&= \frac{1}{2\tau}\boldsymbol{C}^{-1}\boldsymbol{S}_n(\boldsymbol{\rho}_0) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{2\tau}\boldsymbol{C}^{-1}h(\boldsymbol{Y}_{i-1})\boldsymbol{Y}_{i-1}\left(2F\left(X_i - \boldsymbol{Y}'_{i-1}\boldsymbol{\rho}_0\right) - 1\right) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Omega(\boldsymbol{Z}_i) + o_p(1) \text{ say,}
\end{aligned}$$

where $\boldsymbol{Z}'_i = \left(X_i, \boldsymbol{Y}'_{i-1}\right)$. If we assume that the above representation of the estimate holds for the autoregressive model, the influence function would be given by,

$$\Omega(\boldsymbol{Z}) = \frac{1}{2\tau}\boldsymbol{C}^{-1}h(\boldsymbol{Y})\,\boldsymbol{Y}\,(2F(X - \boldsymbol{Y}'\boldsymbol{\rho}_0) - 1).$$

We note that this is the same result that is obtained by Naranjo and Hettmansperger (1994) in the context of the classical linear regression model. Furthermore, assumptions W3 and W4 of Chapter III imply that the influence function is totally bounded. However, since this heuristic argument is not a proof, a formal proof should be obtained before such a definitive statement is made.

Now consider the topic of inference. Specifically, partition the autoregressive parameter vector as follows, $\boldsymbol{\rho}' = (\boldsymbol{\rho}'_1, \boldsymbol{\rho}'_2)$ and consider testing $H_0 : \boldsymbol{\rho}_2 = 0$ versus $H_1 : \boldsymbol{\rho}_2 \neq 0$ where $\boldsymbol{\rho}_2$ is a $q \times 1$ vector and $q \leq p$. Following Naranjo and Hettmansperger (1994) one can consider a test of these hypotheses based on the gradient function. In the classical linear regression setting the test statistic has an asymptotic Chi-Square distribution with q degrees of freedom. Alternatively,

one may consider a "Wald type" statistic based on the asymptotic distribution of the estimate. Again, because of the dependence structure, these results need to be proven in the context of the AR(p). However, it is conjectured that they will continue to hold in this context.

Now consider the HBR estimate that was used in the examples and Monte Carlo study. This estimate did well under both the IO and AO models. Hence, it appears that this estimate may be more suited for autoregressive estimation. Since the weights for this estimate also depend on the residuals from some initial fit, the asymptotic theory for this estimate will be slightly different from the theory presented in Chapter III and still needs to be obtained. Once this theory is obtained, questions pertaining to the breakdown point, influence function, and inference can also be addressed.

Finally, one may also wish to consider the proposed estimate in other contexts. For instance, moving average, ARMA, and intervention models may be topics for future research. Additionally, one may also want to investigate the classical linear regression model with autoregressive (or ARMA) errors.

REFERENCES

Abraham, B. & Ledolter, J. (1983). *Statistical Methods for Forecasting.* New York: John Wiley and Sons.

Bartle, R. G. (1964). *The Elements of Real Analysis* (Second ed.). New York: John Wiley and Sons.

Bell, W. & Hillmer, S. (1987). "Initializing the Kalman Filter in the Nonstationary Case". Research Report CENSUS/SRC/RR-87/33, Statistical Research Division, Bureau of the Census, Washington, DC, 20233.

Billingsley, P. (1995). *Probability and Measure* (Third ed.). New York: John Wiley and Sons.

Brubacher, S. R. (1974). "Time Series Outlier Detection and Modeling with Interpolation". Technical report, Bell Laboratories.

Bustos, O. H. (1982). "General M-Estimates for Contaminated pth-Order Autoregressive Processes: Consistency and Asymptotic Normality". *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, *59*, 491–504.

Chang, W. H., McKean, J. W., Naranjo, J. D., & Sheather, S. J. (1996). "High Breakdown Rank Regression". *Submitted.*

Coakley, C. W. & Hettmansperger, T. P. (1993). "A Bounded Influence, High Breakdown, Efficient Regression Estimator". *The Journal of the American Statistical Association*, *88*, 872–880.

Denby, L. & Martin, R. D. (1979). "Robust Estimation of the First-Order Autoregressive Parameter". *Journal of the American Statistical Association*, *74*(365), 140–146.

Dixon, S. L. & McKean, J. W. (1996). "Rank-Based Analysis of the Heteroscedastic Linear Model". *The Journal of the American Statistical Association*, *91*(434), 699–712.

Ferretti, N. E., Kelmansky, D. M., & Yohai, V. J. (1991). "Estimators Based on Ranks for ARMA Models". *Communications in Statistics, Theory and Methods*, *20(12)*, 3879–3907.

Fox, A. J. (1972). "Outliers in Time Series". *Journal of the Royal Statistical Society B*, *34*(3), 350–363.

Fuller, W. A. (1996). *Introduction to Statistical Time Series* (Second ed.). New York: John Wiley and Sons.

Goldberg, J. L. (1991). *Matrix Theory with Applications*. New York: McGraw-Hill, Inc.

Hall, P. & Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. New York: Academic Press.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust Statistics The Approach Based on Influence Functions*. New York: John Wiley and Sons.

Heiler, S. & Willers, R. (1988). "Asymptotic Normality of R-Estimates in the Linear Model". *Statistics*, *19*(2), 173–184.

Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks*. New York: John Wiley and Sons.

Hettmansperger, T. P. & McKean, J. W. (1978a). "Statistical Inference Based on Ranks". *Psychometrika*, *43*, 69–79.

Huber, P. J. (1972). "Robust Statistics: A Review". *The Annals of Mathematical Statistics*, *43*, 1041–1067.

Huber, P. J. (1981). *Robust Statistics*. New York: John Wiley and Sons.

Jaeckel, L. A. (1972). "Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals". *The Annals of Mathematical Statistics*, *43*(5), 1449–1458.

Kapenga, J., McKean, J. W., & Vidmar, T. J. (1996). "RGLM: Users Manual". Technical Report 90, Western Michigan University, Department of Mathematics and Statistics.

Kohn, R. & Ansley, C. F. (1986). "Estimation, Prediction, and Interpolation for ARIMA Models with Missing Data". *Journal of the American Statistical Association*, *81*, 751–761.

Koul, H. L. (1991). "A Weak Convergence Result Useful in Robust Autoregression". *Journal of Statistical Planning and Inference*, *29*, 291–308.

Koul, H. L. (1992). "Weighted Empiricals and Linear Models". In R. J. Serfling et al. (Eds.), *Lecture Notes-Monograph Series*, volume 21. Hayward, California: Institute of Mathematical Statistics.

Koul, H. L. & Saleh, A. K. M. E. (1993). "R-Estimation of the Parameters of Autoregressive [AR(p)] Models". *The Annals of Statistics*, *21*(1), 534–551.

Martin, R. D. (1980). "Robust Estimation of Autoregressive Models (with discussion)". In D. R. Brillinger & G. C. Tiao (Eds.), *Directions in Time Series* (pp. 228–254). Hayward, CA: Institute of Mathematical Statistics.

Martin, R. D. & Yohai, V. J. (1991). "Bias Robust Estimation of Autoregression Parameters". In W. Stahel & S. Weisberg (Eds.), *Directions in Robust Statistics and Diagnostics Part I*, volume 33 (pp. 233–246). New York: Springer-Verlag.

Naranjo, J. D. (1989). *Bounded-Influence Regression: A Modified Wilcoxon Procedure.* PhD thesis, Pennsylvania State University.

Naranjo, J. D. & Hettmansperger, T. P. (1994). "Bounded Influence Rank Regression". *Journal of the Royal Statistical Society B*, *56*(1), 209–220.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton, New Jersey: Princeton University Press.

Rousseeuw, P. J. & Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* New York: John Wiley and Sons.

Rousseeuw, P. J. & van Zomeren, B. C. (1991). "Robust Distances: Simulations and Cutoff Values. In W. Stahel & S. Weisberg (Eds.), *Directions in Robust Statistics and Diagnostics Part II*, volume 34 (pp. 195–203). New York: Springer-Verlag.

Royden, H. L. (1988). *Real Analysis* (Third ed.). Englewood Cliffs, New Jersey 07632: Prentice Hall Inc.

Sciences, S. (1993). *S-PLUS Guide to Statistical and Mathematical Analysis*, volume 3.2. Seattle: StatSci, a division of MathSoft, Inc.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley and Sons.

Sievers, G. L. (1983). "A Weighted Dispersion Function for Estimation in Linear Models". *Communications in Statistics, Theory and Methods, 12(10)*, 1161–1179.

Simpson, D. G., Ruppert, D., & Carroll, R. J. (1992). "On One-Step GM Estimates and Stability of Inferences in Linear Regression". *The Journal of the American Statistical Association*, *87*(418), 439–450.

Stromberg, A. J. (1993). "Computing the Exact Least Median of Squares Estimate and Stability Diagnostics in Multiple Linear Regression". *SIAM Journal of Scientific Computing, 14*, 1289–1299.

Stromberg, J. & Hawkins, D. M. (1993). "Algorithm AS 282: High Breakdown Regression and Multivariate Estimation". *Applied Statistics, 42*, 423–432.

Tucker, H. G. (1967). *A Graduate Course in Probability.* New York: Academic Press.