12-2002

Robust Residuals and Diagnostics in Autoregressive Time Series

Kirk W. Anderson
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Statistical Methodology Commons, and the Statistical Models Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1156

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
ROBUST RESIDUALS AND DIAGNOSTICS
IN AUTOREGRESSIVE TIME SERIES

by

Kirk W. Anderson

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 Statistics

Western Michigan University
Kalamazoo, Michigan
December 2002
One of the goals of model diagnostics is outlier detection. In particular, we would like to use the residuals, appropriately standardized, to "flag" outliers. Hopefully, our (robust) procedure has yielded a fit that resists undue influence by outlying points, while simultaneously drawing attention to these interesting points via residual analysis.

In this study we consider several different methods of standardizing the residuals resulting from autoregression. A large sample approximation for the variance of rank-based first order autoregressive time series residuals is developed. This provides studentized residuals, specific to the time series model and estimation procedure.

Simulation studies are presented that illustrate outlier detection ability among different standardization methods, and differences in fits among estimation procedures in the presence of innovation and additive outliers.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®
ACKNOWLEDGMENTS

I would like to thank Joe McKeen for his guidance, insight, and most of all, his patience. Jeff Terpstra’s help was also invaluable.

Kirk W. Anderson
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ ii  
LIST OF TABLES ................................................................................................................. v  
LIST OF FIGURES ................................................................................................................ vi  
CHAPTER  

I. INTRODUCTION ................................................................................................................. 1  
  1.1 Residuals in Time Series Estimation ................................................................. 1  
    1.1.1 The Autoregressive Time Series Model .................................................. 1  
    1.1.2 Assumptions ......................................................................................... 2  
    1.1.3 Dependence Structure ........................................................................ 3  
    1.1.4 The No Intercept Model ....................................................................... 5  
    1.1.5 The Asymptotic Representation of the Estimate .................................. 6  

II. VARIANCE OF TIME SERIES RESIDUALS ................................................................. 9  
  2.1 Expectation of $\hat{\varepsilon}_k$ ........................................................................ 9  
  2.2 Variance of $\hat{\varepsilon}_k$ ............................................................................. 12  
  2.3 Estimation of $\text{Var}(\hat{\varepsilon}_k)$ ................................................................. 41  

III. CLASSIFICATION RATES FOR DETECTING OUTLIERS ...................................... 43  
  3.1 Various Standardizations of Residuals .............................................................. 43  
  3.2 Simulation Studies ............................................................................................. 48  
    3.2.1 AR(1) Simulation Study ......................................................................... 50  

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table of Contents—Continued

CHAPTER

3.2.2 AR(2) Simulation Study ............................................. 57
3.2.3 Higher Order AR Simulation Study ........................... 60

IV. DIAGNOSTICS COMPARING ESTIMATES .......................... 69

4.1 Estimating Procedures ............................................................ 69
4.2 TDBETAS Analysis ................................................................. 70
4.3 Residential Extensions Data ............................................... 72
4.4 Outliers in Autoregressive Time Series ................................. 80
  4.4.1 Types of Outliers .......................................................... 80
  4.4.2 Generated Examples ................................................... 81
4.5 A Monte Carlo Study .............................................................. 84

V. CONCLUSIONS ..................................................................... 91

5.1 Summary of Work Presented .................................................. 91
5.2 Further Research ................................................................. 91
  5.2.1 Extensions of Results .................................................... 91
  5.2.2 Application to Quality Control Charts ..................... 92

REFERENCES ........................................................................... 95
LIST OF TABLES

1. Possible Outcomes for Outlier Detection ........................................... 50
2. Parameter Values Used for AR(2) Simulation ........................................ 57
3. Parameter Values Used for Higher Order AR Simulation .................. 63
4. Parameter Estimates for Residential Extensions Data ..................... 76
5. TDBETAS Values for Residential Extensions Data ............................. 77
6. Parameter Estimates for Generated Examples .................................... 87
7. TDBETAS Values for Generated Examples ......................................... 87
8. Frequencies out of 1000 that TDBETAS exceeded the benchmark ...... 90
LIST OF FIGURES

1. Type I error rates for AR(1) simulation, contaminated variance is 16. . 52
2. Type I error rates for AR(1) simulation, contaminated variance is 100. 53
3. Power rates for AR(1) simulation, contaminated variance is 16. ....... 54
4. Power rates for AR(1) simulation, contaminated variance is 100. ....... 55
5. Type I error rates for AR(2) simulation, contaminated variance is 16. . 58
6. Type I error rates for AR(2) simulation, contaminated variance is 100. 59
7. Power rates for AR(2) simulation, contaminated variance is 16. ....... 61
8. Power rates for AR(2) simulation, contaminated variance is 100. ....... 62
9. Type I error rates for Higher Order AR simulation, contaminated variance is 16. ................................................................. 64
10. Type I error rates for Higher Order AR simulation, contaminated variance is 100. ................................................................. 65
11. Power rates for Higher Order AR simulation, contaminated variance is 16. ................................................................. 66
12. Power rates for Higher Order AR simulation, contaminated variance is 100. ................................................................. 67
13. Time series plot and lag 1 scatterplot with fits for the seasoned Residential Extensions data. .................................................. 75

vi
List of Figures—Continued

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.</td>
<td>Lag 1 scatterplot with fits for the seasoned Residential Extensions data. Only the main cluster of points is pictured.</td>
<td>76</td>
</tr>
<tr>
<td>15.</td>
<td>Residuals over time for the Residential Extensions data.</td>
<td>79</td>
</tr>
<tr>
<td>16.</td>
<td>Time series plots for four outlier cases.</td>
<td>85</td>
</tr>
<tr>
<td>17.</td>
<td>Lag 1 scatterplots with fits for four outlier cases.</td>
<td>86</td>
</tr>
<tr>
<td>18.</td>
<td>Boxplots of the 1000 estimates of $p = 0.5$ with a reference line drawn to check bias.</td>
<td>90</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

1.1 Residuals in Time Series Estimation

In this chapter we consider the residuals obtained from a rank-based first-order autoregression. A large sample approximation for the variance of the residuals is developed. This is done for the purpose of studentizing the residuals, specific to the time series model and the estimation procedure.

Typically, time series residuals are standardized with a denominator of $\hat{\sigma}$, an estimate of the error distribution standard deviation. This will hereafter be referred to as the "naive" standardization, as it does not take into account the position in time.

1.1.1 The Autoregressive Time Series Model

Let $X_1, X_2, \ldots, X_n$ denote random variables pertaining to the realizations of a stationary autoregressive model of order $p = 1$, denoted by $AR(1)$. The model can be written as

$$X_i = \alpha + \rho X_{i-1} + \varepsilon_i, \quad i = 1, \ldots, n,$$  \hspace{1cm} (1.1)
where the random errors $\varepsilon_1, \ldots, \varepsilon_n$ are assumed independent and identically distributed according to some distribution function, $F$. We also require $E(\varepsilon_1) = 0$ and $E(\varepsilon_1^2) < \infty$. 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. $X_0$ is an observable random variable independent of the errors. We will also assume $E(X_0^4) < \infty$. To insure stationarity of the process, we require $\rho \in (-1, 1)$. The estimate of $\rho$ will be the value that minimizes Jaeckel's (1972) dispersion function

$$D(\rho) = \sum_{i=1}^{n} \varphi \left( \frac{R(\varepsilon_i)}{n+1} \right) \varepsilon_i,$$

where $R(\varepsilon_i)$ denotes the rank of $\varepsilon_i$ among $\varepsilon_1, \ldots, \varepsilon_n$. The score function, $\varphi(u)$, is defined on $(0, 1)$, is nondecreasing, square-integrable, bounded, and differentiable.

In addition, we require

$$\int_0^1 \varphi(u)du = 0 \text{ and } \int_0^1 \varphi^2(u)du = 1.$$

In practice, we will employ Wilcoxon scores, which are given by $\varphi(u) = \sqrt{12}(u - \frac{1}{2})$.

1.1.2 Assumptions

It will be helpful now to list the assumptions that we have made on the model, errors, and scores for future reference. An "M" represents an assumption on the model, an "E" represents an assumption on the error terms, and an "S" represents an assumption on the score function.

**M1.** $X_i = \alpha + \rho X_{i-1} + \varepsilon_i$ where $i = 1, \ldots, n$ and $|\rho| < 1$. 
M2. $X_0$ is an observable random variable independent of $\varepsilon_1, \ldots, \varepsilon_n$ with $E(X_0^4) < \infty$.

E1. $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. $F$ random variables with $E(\varepsilon_1) = 0$ and $E(\varepsilon_1^4) < \infty$.

E2. $F$ has an absolutely continuous density, $f$, where $f > 0$ a.e., $f$ has finite Fisher information, and $f'$ is uniformly bounded.

S1. The score function, $\varphi(u)$, is defined on $(0, 1)$, is nondecreasing, square-integrable, bounded, and differentiable.

S2. $\int_0^1 \varphi(u)du = 0$ and $\int_0^1 \varphi^2(u)du = 1$.

1.1.3 Dependence Structure

Consider the following definition concerning $\sigma$-fields (Ash, 1972).

**Definition 1.1.1.** Let $\mathcal{F}$ be a collection of subsets of a set $\Omega$. Then $\mathcal{F}$ is called a $\sigma$-field iff $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under complementation and countable union.

Let $\mathcal{F}(X_0, \varepsilon_1, \ldots, \varepsilon_i)$, $i = 1, \ldots, n$ denote the smallest $\sigma$-field generated by $\{X_0, \varepsilon_1, \ldots, \varepsilon_i\}$. Conditioning on a $\sigma$-field will essentially allow us to consider any random variables depending on the set $\{X_0, \varepsilon_1, \ldots, \varepsilon_i\}$ as fixed. It will be necessary to condition on "past" information for some expectations later. For our
An AR(1) model, then, we will have the following:

\[
X_1 = \alpha + \rho X_0 + \varepsilon_1 \implies X_1 \in \mathcal{F}(X_0, \varepsilon_1)
\]

\[
X_2 = \alpha + \rho X_1 + \varepsilon_2 = \alpha + \rho(\alpha + \rho X_0 + \varepsilon_1) + \varepsilon_2 = \alpha + \rho\alpha + \rho^2 X_0 + \rho\varepsilon_1 + \varepsilon_2 \implies X_2 \in \mathcal{F}(X_0, \varepsilon_1, \varepsilon_2)
\]

\[
X_3 = \alpha + \rho X_2 + \varepsilon_3 = \alpha + \rho(\alpha + \rho\alpha + \rho^2 X_0 + \rho\varepsilon_1 + \varepsilon_2) + \varepsilon_3 = \alpha + \rho\alpha + \rho^2 X_0 + \rho^2 \varepsilon_1 + \rho\varepsilon_2 + \varepsilon_3 \implies X_3 \in \mathcal{F}(X_0, \varepsilon_1, \varepsilon_2, \varepsilon_3)
\]

\[
X_i = \sum_{k=0}^{i-1} \rho^k \alpha + \rho^i X_0 + \sum_{k=1}^{i} \rho^{i-k} \varepsilon_k \implies X_i \in \mathcal{F}(X_0, \varepsilon_1, \ldots, \varepsilon_i)
\]

\[
X_n = \sum_{k=0}^{n-1} \rho^k \alpha + \rho^n X_0 + \sum_{k=1}^{n} \rho^{n-k} \varepsilon_k \implies X_n \in \mathcal{F}(X_0, \varepsilon_1, \ldots, \varepsilon_n)
\]

Thus we have:

\(X_0 \in \mathcal{F}(X_0)\) is independent of \(\varepsilon_1, \ldots, \varepsilon_n\).

\(X_1 \in \mathcal{F}(X_0, \varepsilon_1)\) is independent of \(\varepsilon_2, \ldots, \varepsilon_n\).

\(X_2 \in \mathcal{F}(X_0, \varepsilon_1, \varepsilon_2)\) is independent of \(\varepsilon_3, \ldots, \varepsilon_n\).

\(X_i \in \mathcal{F}(X_0, \varepsilon_1, \ldots, \varepsilon_i)\) is independent of \(\varepsilon_{i+1}, \ldots, \varepsilon_n\).

\(X_{n-1} \in \mathcal{F}(X_0, \varepsilon_1, \ldots, \varepsilon_{n-1})\) is independent of \(\varepsilon_n\).
$X_n \in \mathcal{F}(X_0, \varepsilon_1, \ldots, \varepsilon_n)$ is not independent of any errors.

Here we introduce some notation (Terpstra, 1997) that will be very useful. For $u \leq v$, define the random variable $\Gamma_u^v = \sum_{r=0}^{v-u} \rho^r \varepsilon_{v-r}$, a finite linear combination of error terms. Our model (1.1) can be written as

$$X_i = \alpha \sum_{j=0}^{i-1} \rho^j + \sum_{j=0}^{i} \rho^{i-j} \varepsilon_j \text{, where } \varepsilon_0 \equiv X_0.$$  

By geometric series $\sum_{j=0}^{i-1} \rho^j = \frac{1-\rho^i}{1-\rho}$. Substituting $r = i - j$ we can write

$$\sum_{j=0}^{i} \rho^{i-j} \varepsilon_j = \sum_{r=0}^{i-0} \rho^r \varepsilon_{i-r} = \sum_{r=0}^{i-0} \rho^r \varepsilon_{i-r} = \Gamma_i^0.$$  

This results in the following representation of the model (1.1):

$$X_i = \alpha \frac{1 - \rho^i}{1 - \rho} + \Gamma_0^i, \; i = 1, \ldots, n.$$  \hspace{1cm} (1.3)

1.1.4 The No Intercept Model

Under assumptions M1 and E1, the mean of the process is $\mu_X = E(X_1) = \frac{\alpha}{1-\rho}$. If we define the "centered" process as $X_i^c = X_i - \mu_X$, then (1.1) gives us

$$X_i^c + \mu_X = \alpha + \rho(X_{i-1}^c + \mu_X) + \varepsilon_i$$

$$= \alpha + \rho \frac{\alpha}{1 - \rho} + \rho X_{i-1}^c + \varepsilon_i$$

$$= \frac{(1 - \rho)\alpha + \rho \alpha}{1 - \rho} + \rho X_{i-1}^c + \varepsilon_i$$

$$X_i^c + \frac{\alpha}{1 - \rho} = \frac{\alpha}{1 - \rho} + \rho X_{i-1}^c + \varepsilon_i$$

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

Of course, the $\rho$ that appears in the centered model is the same $\rho$ as in the non-centered model. Since the estimation of $\rho$ is our primary concern, let us assume
without loss of generality that the process has zero mean, i.e. $E(X_1) = 0$. For convenience we drop the $X_i^c$ notation and simply write

$$X_i = \rho X_{i-1} + \varepsilon_i, \quad i = 1, \ldots, n. \quad (1.4)$$

From now on we will employ this “no intercept” model. Let’s re-state our first model assumption:

**M1.** $X_i = \rho X_{i-1} + \varepsilon_i$ where $i = 1, \ldots, n$. $|\rho| < 1$, and $E(X_i) = 0$.

In practice, we center the series as an initial step in the estimation procedure. Using an unbiased estimate of the process mean, we can simply subtract it from each observation and use the centered series in the subsequent analysis. If an estimate of $\alpha$ is desired, one can do the following. Fit (1.4) to the centered series to get an estimate of $\rho$. The residuals are defined as $\hat{\varepsilon}_i = X_i - \hat{\rho} X_{i-1}$. Assumption M1 implies $X_i - \rho X_{i-1} = \alpha + \varepsilon_i$, so we can fit the location model (Hettmansperger & McKean, 1998. pg. 2) $\hat{\varepsilon}_i = \alpha + \varepsilon_i$ to determine $\hat{\alpha}$.

1.1.5 The Asymptotic Representation of the Estimate

With the no intercept model, (1.3) becomes $X_i = \Gamma_0^i$. The following lemma gives us a very useful relation, one that will be prominent in the derivations to follow.

**Lemma 1.1.1.** For $i < k$, $X_k = \Gamma_{k+1}^k + \rho^{k-i} X_i$.
Proof.

\[ X_k = \Gamma_k^k - \sum_{r=0}^{k-0} \rho^r \varepsilon_{k-r} \]

\[ = \sum_{r=0}^{k-(i+1)} \rho^r \varepsilon_{k-r} + \sum_{r=k-(i+1)+1}^{k} \rho^r \varepsilon_{k-r} \]

\[ = \Gamma_{i+1}^k + \sum_{r=k-i}^{k} \rho^r \varepsilon_{k-r} \]

\[ = \Gamma_{i+1}^k + \rho^{k-i} \varepsilon_{k-(k-i)} + \rho^{k-i+1} \varepsilon_{k-(k-i+1)} + \cdots + \rho^k \varepsilon_{k-k} \]

\[ = \Gamma_{i+1}^k + \rho^{k-i} \varepsilon_i + \rho^{k-i+1} \varepsilon_{i-1} + \cdots + \rho^k \varepsilon_0 \]

\[ = \Gamma_{i+1}^k + \rho^{k-i} (\varepsilon_i + \rho \varepsilon_{i-1} + \rho^2 \varepsilon_{i-2} + \cdots + \rho^i \varepsilon_0) \]

\[ = \Gamma_{i+1}^k + \rho^{k-i} (\rho^i X_0 + \rho^{i-1} \varepsilon_1 + \rho^{i-2} \varepsilon_2 + \cdots + \rho \varepsilon_{i-1} + \varepsilon_i) \]

\[ = \Gamma_{i+1}^k + \rho^{k-i} X_i \]

\[ \square \]

The gradient, defined a. e., of the dispersion function (1.2) is

\[ S(\rho) = -\nabla D(\rho) = -\frac{d}{d\rho} \sum_{i=1}^{n} \varphi \left( \frac{R(\varepsilon_i)}{n+1} \right) \varepsilon_i \]

\[ = -\frac{d}{d\rho} \sum_{i=1}^{n} \varphi \left( \frac{R(X_i - \rho X_{i-1})}{n+1} \right) (X_i - \rho X_{i-1}) \]

\[ = -\sum_{i=1}^{n} \varphi \left( \frac{R(\varepsilon_i)}{n+1} \right) (-X_i) \]

\[ = \sum_{i=1}^{n} X_{i-1} \varphi \left( \frac{R(\varepsilon_i)}{n+1} \right) \]

(Hettmansperger & McKean, 1998, pg. 148). The R-estimate of \( \rho \) can be viewed as an approximate solution of \( S(\rho) = 0 \). From Koul and Saleh (1993), we know

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
that
\[
\sqrt{n}(\hat{\rho}_n - \rho) = \tau_\varphi \sigma_{\tilde{X}}^{-2} \sum_{i=1}^{n} X_{i-1} \varphi(F(\varepsilon_i)) + o_p(1),
\]
where \(\hat{\rho}_n\) is the asymptotic representation of the estimate, \(\sigma_{\tilde{X}}^2 = Var(X_1)\), and the scale parameter \(\tau_\varphi\) is given by
\[
\tau_\varphi^{-1} = \int \varphi(u) \varphi_f(u) du,
\]
where \(\varphi_f(u)\), the optimal score function, is
\[
\varphi_f(u) = \frac{f'(F^{-1}(u))}{f(F^{-1}(u))}
\]

Note that for Wilcoxon scores,
\[
\tau_\varphi^{-1} = \sqrt{\frac{12}{3}} \int f^2(t) dt.
\]
This leads us to the asymptotic representation of the R-estimate.
\[
\hat{\rho}_n = \rho + \tau_\varphi \sigma_{\tilde{X}}^{-2} \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \varphi(F(\varepsilon_i)) + o_p \left( \frac{1}{\sqrt{n}} \right)
\]
For now on, we will use this expression to denote the (Wilcoxon) R-estimate:
\[
\hat{\rho}_R = \rho + \tau_\varphi \sigma_{\tilde{X}}^{-2} \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \varphi(F(\varepsilon_i)).
\]

In all that follows, we will use the letter "R" to denote any estimate, residual or procedure that is rank-based. This is general notation, but for our purposes, the meaning will be specifically the R-estimates of Koul and Saleh (1993) using Wilcoxon scores. No other R-estimates are considered.
CHAPTER II

VARIANCE OF TIME SERIES RESIDUALS

2.1 Expectation of $\hat{\varepsilon}_k$

Once an estimate of $\rho$ has been obtained, we will have the residuals

$$\hat{\varepsilon}_k = X_k - \rho_R X_{k-1}, \quad k = 1, \ldots, n. \quad (2.1)$$

By $\rho_R$ here we mean the (Wilcoxon) R-estimate, but we will be using (1.8), the asymptotic representation. So (2.1) becomes

$$\hat{\varepsilon}_k = X_k - \left( \rho + \tau_\phi \sigma_X^{-2} \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \phi(F(\varepsilon_i)) \right) X_{k-1}$$

$$= \rho X_{k-1} + \varepsilon_k - \rho X_{k-1} - \tau_\phi \sigma_X^{-2} \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \phi(F(\varepsilon_i)) X_{k-1}$$

$$= \varepsilon_k - \tau_\phi \sigma_X^{-2} \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \phi(F(\varepsilon_i)) X_{k-1}, \quad k = 1, \ldots, n. \quad (2.2)$$

**Theorem 2.1.1.** Assume $k$ is fixed. The expected value of $\hat{\varepsilon}_k$ in (2.2) is zero.

**Proof.** If $k$ is considered the "current" point in time, then $i = 1, \ldots, k - 1$ would be the "past", and $i = k + 1, \ldots, n$ would be "future" data points. Keep this in
mind as we consider the following expectation.

\[
E(\varepsilon_k) = E(\varepsilon_k) - \tau\sigma^2 X_n \frac{1}{n} \sum_{i=1}^{n} E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1})
\]

\[
= 0 - \tau\sigma^2 X_n \frac{1}{n} \sum_{i=1}^{k-1} E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1}) - \tau\sigma^2 X_n \frac{1}{n} E(X_{k-1}^2\varphi(F(\varepsilon_k)))
\]

\[
- \tau\sigma^2 X_n \frac{1}{n} \sum_{i=k+1}^{n} E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1})
\]

Splitting up the sum in the above manner is a very useful way to handle the interdependence of the terms. This will be used many more times in the derivations to follow. Consider the \(i < k\) term:

\[
-\tau\sigma^2 X_n \frac{1}{n} \sum_{i=1}^{k-1} E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1})
\]

By lemma 1.1.1 we can also write

\[
X_{k-1} = \rho^{(k-1)-(i-1)}X_{i-1} + \Gamma_k^{i-1}
\]

\[
= \rho^{k-i}X_{i-1} + \Gamma_k^{k-i}.
\]

So we have

\[
E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1}) = E(X_{i-1}\varphi(F(\varepsilon_i))(\rho^{k-i}X_{i-1} + \Gamma_k^{k-1}))
\]

\[
= E(\rho^{k-i}X_{i-1}^2\varphi(F(\varepsilon_i))) + E(X_{i-1}\Gamma_k^{k-1}\varphi(F(\varepsilon_i))).
\]

Note that \(X_{i-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{i-1})\), thus \(X_{i-1}\) is independent of \(\varepsilon_i, \varepsilon_{i+1}, \ldots, \varepsilon_n\). This means we can write

\[
E(\rho^{k-i}X_{i-1}^2\varphi(F(\varepsilon_i))) = \rho^{k-i}E(X_{i-1}^2)E(\varphi(F(\varepsilon_i))).
\]
Now consider $E(\varphi(F(\varepsilon_i))) \equiv E(\varphi(F(\varepsilon_1)))$. Using assumption E1 and the fact that $F$ is strictly increasing, we see that $U = F(\varepsilon_1)$ is a uniform(0,1) random variable (Hogg & Tanis, 1988, pg. 167). Hence $E(\varphi(U)) = \int_0^1 \varphi(u) du = 0$ by assumption S2, and we have

$$E(\rho^{k-i}X^2_{i-1}\varphi(F(\varepsilon_i))) = \rho^{k-i}E(X^2_{i-1}) \cdot 0 = 0.$$ 

We will use this fact to cancel out many terms in the derivations to follow. Now, note that $\Gamma^{k-1}_i = \sum_{r=0}^{k-1-i} \rho^r \varepsilon_{k-1-r}$ contains error terms indexed

$k - 1, k - 2, \ldots, k - 1 - (k - 1 - i) = i$. Since $i < k$, this is $\varepsilon_i, \varepsilon_{i+1}, \ldots, \varepsilon_{k-2}, \varepsilon_{k-1}.$

So $\Gamma^{k-1}_i \in \mathcal{F}(\varepsilon_i, \varepsilon_{i+1}, \ldots, \varepsilon_{k-1})$. Therefore $X_{i-1}$ is independent of $\Gamma^{k-1}_i$ as well as the $\varphi(F(\varepsilon_i))$ term. So we have

$$E(X_{i-1}\Gamma^{k-1}_i \varphi(F(\varepsilon_i))) = E(X_{i-1})E(\Gamma^{k-1}_i \varphi(F(\varepsilon_i))) = 0 \cdot E(\Gamma^{k-1}_i \varphi(F(\varepsilon_i))) = 0.$$ 

Note that $E(X_{i-1}) = 0$ since the process is centered and thus has zero mean.

Hence the $i < k$ term is

$$-\tau_\varphi \sigma^{-2}_x \frac{1}{n} \sum_{i=1}^{k-1} (E(\rho^{k-i}X^2_{i-1})E(\varphi(F(\varepsilon_i))) + E(X_{i-1})E(\Gamma^{k-1}_i \varphi(F(\varepsilon_i))))$$

$$= -\tau_\varphi \sigma^{-2}_x \frac{1}{n} \sum_{i=1}^{k-1} (E(\rho^{k-i}X^2_{i-1}) \cdot 0 + 0 \cdot E(\Gamma^{k-1}_i \varphi(F(\varepsilon_i)))) = 0.$$ 

Now consider the $i = k$ term:

$$-\tau_\varphi \sigma^{-2}_x \frac{1}{n} E(X^2_{k-1} \varphi(F(\varepsilon_k)))$$

Here, $X_{k-1}$ is independent of $\varepsilon_k$. so

$$E(X^2_{k-1} \varphi(F(\varepsilon_k))) = E(X^2_{k-1})E(\varphi(F(\varepsilon_k))) = E(X^2_{k-1}) \cdot 0 = 0,$$
and the $i = k$ term is therefore zero. Lastly, consider the $i > k$ term:

$$-\tau \varphi \sigma_X^{-2} \frac{1}{n} \sum_{i=k+1}^{n} \mathbb{E} (X_{i-1} \varphi(F(\varepsilon_i))X_{k-1})$$

$X_{i-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{i-1})$ and $X_{k-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{k-1})$ are not independent since their respective $\sigma$-fields overlap, but since $i > k$, both $X_{i-1}$ and $X_{k-1}$ are independent of $\varepsilon_i$. So

$$\mathbb{E} (X_{i-1} \varphi(F(\varepsilon_i))X_{k-1}) = \mathbb{E}(X_{i-1}X_{k-1})\mathbb{E}(\varphi(F(\varepsilon_i))) = \mathbb{E}(X_{i-1}X_{k-1}) \cdot 0 = 0,$$

and the $i > k$ term is therefore zero. Thus we have $\mathbb{E}(\hat{\varepsilon}_k) = 0 + 0 + 0 = 0. \quad \Box$

### 2.2 Variance of $\hat{\varepsilon}_k$

Since $\mathbb{E}(\hat{\varepsilon}_k) = 0$, it follows that $\text{Var}(\hat{\varepsilon}_k) = \mathbb{E}(\hat{\varepsilon}_k^2)$. Recall that

$$\hat{\varepsilon}_k = \varepsilon_k - \tau \varphi \sigma_X^{-2} \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \varphi(F(\varepsilon_i))X_{k-1} \varepsilon_k.$$

Squaring this term results in

$$\hat{\varepsilon}_k^2 = \varepsilon_k^2 - 2\tau \varphi \sigma_X^{-2} \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \varphi(F(\varepsilon_i))X_{k-1} \varepsilon_k$$

$$+ \tau^2 \sigma_X^{-4} \frac{1}{n^2} \sum_{i=1}^{n} X_{i-1} \varphi(F(\varepsilon_i))X_{k-1}^2 \sum_{j=1}^{n} X_{j-1} \varphi(F(\varepsilon_j)).$$

The expectation will have three main terms:

$$\mathbb{E}(\hat{\varepsilon}_k^2) \equiv \mathbb{E}(\varepsilon_k^2) - 2\tau \varphi \sigma_X^{-2} \frac{1}{n} \mathbb{E} \left( \sum_{i=1}^{n} X_{i-1} \varphi(F(\varepsilon_i))X_{k-1} \varepsilon_k \right)$$

$$+ \tau^2 \sigma_X^{-4} \frac{1}{n^2} \mathbb{E} \left( \sum_{i=1}^{n} X_{i-1} \varphi(F(\varepsilon_i))X_{k-1}^2 \sum_{j=1}^{n} X_{j-1} \varphi(F(\varepsilon_j)) \right)$$

$$= \mathbb{E}(\varepsilon_1^2) - 2\tau \varphi \sigma_X^{-2} \frac{1}{n} E_1 + \tau^2 \sigma_X^{-4} \frac{1}{n^2} E_2$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Let \( \sigma^2 = E(\varepsilon_i^2) \). \( E_1 \) and \( E_2 \) are shorthand terms used to represent the above expectations.

**Lemma 2.2.1.** Assume \( k \) is fixed. Then

\[
E_1 = \sum_{i=1}^{n} E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1}\varepsilon_k) = \sigma_X^2 \delta_{11},
\]

where \( \sigma_X^2 = E(X_i^2) \) and \( \delta_{11} = E(\varepsilon_1\varphi(F(\varepsilon_1))) \).

*Proof.* We start by splitting up the sum in the \( E_1 \) term like before.

\[
E_1 = \sum_{i=1}^{n} E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1}\varepsilon_k)
\]

\[
= \sum_{i=1}^{k-1} E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1}\varepsilon_k) + E(X_k^2\varphi(F(\varepsilon_k))\varepsilon_k) + \sum_{i=k+1}^{n} E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1}\varepsilon_k)
\]

For the \( i < k \) term, note that \( X_{i-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{i-1}) \) and \( X_{k-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{k-1}) \) are not independent of each other but are both independent of \( \varepsilon_k \). Also \( \varepsilon_k \) and the \( \varphi(F(\varepsilon_i)) \) term are independent since \( i < k \) and \( \varepsilon_1, \ldots, \varepsilon_n \) are iid. Thus we have

\[
\sum_{i=1}^{k-1} E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1}\varepsilon_k) = \sum_{i=1}^{k-1} E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1}) E(\varepsilon_k)
\]

\[
= \sum_{i=1}^{k-1} E(X_{i-1}\varphi(F(\varepsilon_i))X_{k-1}) \cdot 0 = 0.
\]

For the \( i = k \) term, \( X_{k-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{k-1}) \) is independent of \( \varepsilon_k \), so we have

\[
E(X_{k-1}^2\varphi(F(\varepsilon_k))\varepsilon_k) = E(X_{k-1}^2)E(\varepsilon_k\varphi(F(\varepsilon_k)))
\]

\[
= E(X_k^2)E(\varepsilon_1\varphi(F(\varepsilon_1))) = \sigma_X^2 \delta_{11},
\]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
where we define the term \( \delta_{11} = E(\varepsilon_1 \varphi(F(\varepsilon_1))) \). Note that the expectations do not depend on the index \( k \). For the \( i > k \) term, note that \( X_{i-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{i-1}) \) is independent of \( \varepsilon_i \) but not \( \varepsilon_k \), since \( k < i \). \( X_{k-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{k-1}) \) is independent of \( \varepsilon_i \) and \( \varepsilon_k \), but not \( X_{i-1} \). So this term is

\[
\sum_{i=k+1}^{n} E(X_{i-1} \varphi(F(\varepsilon_i))X_{k-1} \varepsilon_k) = \sum_{i=k+1}^{n} E(X_{i-1}X_{k-1} \varepsilon_k)E(\varphi(F(\varepsilon_i)))
\]

\[
= \sum_{i=k+1}^{n} E(X_{i-1}X_{k-1} \varepsilon_k) \cdot 0 = 0.
\]

Thus \( E_1 = 0 + \sigma_X^2 \delta_{11} + 0 = \sigma_X^2 \delta_{11} \). \( \square \)

**Lemma 2.2.2.** Assume \( k \) is fixed. Then

\[
E_2 = E \left( \sum_{i=1}^{n} X_{i-1} \varphi(F(\varepsilon_i)) \sum_{j=1}^{n} X_{j-1} \varphi(F(\varepsilon_j))X_k^2 \right)
\]

\[
= \left( \rho^2 E(X_1^4) + 2\rho \delta_{12} E(X_1^3) + \sigma_X^2 (\delta_{22} - \frac{\sigma^2}{1 - \rho^2}) \left( \frac{\rho^2 - 1 - \rho^{2(k-1)}}{\rho^2 - 1} \right) \right)
\]

\[
+ \sigma^2 \sigma_X^2 \left( \frac{n-1}{1 - \rho^2} \right) + 8\sigma_X^2 \delta_{11} \frac{\rho^2}{(1 - \rho^2)^2} \left( 1 - (k-1)\rho^{2(k-2)} + (k-2)\rho^{2(k-1)} \right)
\]

\[
+ E(X_1^4) + \left( E(X_1^4) - \frac{\sigma^2 \sigma_X^2}{1 - \rho^2} \right) \left( \frac{\rho^2 - \rho^{2(n-k+1)}}{1 - \rho^2} \right).
\]

Before we formally begin the proof of lemma 2.2.2, let's expand the \( E_2 \) term. Splitting up each of the two sums as before.

\[
E_2 = E \left( \sum_{i=1}^{n} X_{i-1} \varphi(F(\varepsilon_i)) \sum_{j=1}^{n} X_{j-1} \varphi(F(\varepsilon_j))X_k^2 \right)
\]

\[
= E \left\{ \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) + X_{k-1} \varphi(F(\varepsilon_k)) \right) + \sum_{i=k+1}^{n} X_{i-1} \varphi(F(\varepsilon_i)) \right\}
\]

\[
\times \left( \sum_{j=1}^{k-1} X_{j-1} \varphi(F(\varepsilon_j)) + X_{k-1} \varphi(F(\varepsilon_k)) + \sum_{j=k+1}^{n} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_k^2 \right\}
\]
Multiplying this out results in an expression consisting of nine terms:

\[
E_2 = E \left\{ \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) \right) \left( \sum_{j=1}^{k-1} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_{k-1}^2 \right. \\
+ \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) \right) X_{k-1} \varphi(F(\varepsilon_k)) X_{k-1}^2 \\
+ \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) \right) \left( \sum_{j=k+1}^{n} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_{k-1}^2 \\
+ X_{k-1} \varphi(F(\varepsilon_k)) \left( \sum_{j=1}^{k-1} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_{k-1}^2 \\
+ X_{k-1} \varphi(F(\varepsilon_k)) X_{k-1} \varphi(F(\varepsilon_k)) X_{k-1}^2 \\
+ X_{k-1} \varphi(F(\varepsilon_k)) \left( \sum_{j=k+1}^{n} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_{k-1}^2 \\
+ \left( \sum_{i=k+1}^{n} X_{i-1} \varphi(F(\varepsilon_i)) \right) \left( \sum_{j=1}^{k-1} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_{k-1}^2 \\
+ \left( \sum_{i=k+1}^{n} X_{i-1} \varphi(F(\varepsilon_i)) \right) X_{k-1} \varphi(F(\varepsilon_k)) X_{k-1}^2 \\
+ \left( \sum_{i=k+1}^{n} X_{i-1} \varphi(F(\varepsilon_i)) \right) \left( \sum_{j=k+1}^{n} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_{k-1}^2 \right\}
\]

Note that the 2\textsuperscript{nd} and 4\textsuperscript{th} terms are equivalent, the 3\textsuperscript{rd} and 7\textsuperscript{th} terms are equivalent, and the 6\textsuperscript{th} and 8\textsuperscript{th} terms are equivalent. Thus we now have an expression with
six terms:

\[
E_2 = E \left\{ \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) \right) \left( \sum_{j=1}^{k-1} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_{k-1}^2 \right\} \\
+ 2E \left\{ \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) \right) X_{k-1}^3 \varphi(F(\varepsilon_k)) \right\} \\
+ 2E \left\{ \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) \right) \left( \sum_{j=k+1}^{n} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_{k-1}^2 \right\} \\
+ E \left( X_{k-1}^4 \varphi^2(F(\varepsilon_k)) \right) \\
+ 2E \left\{ \left( \sum_{i=k+1}^{n} X_{i-1} \varphi(F(\varepsilon_i)) \right) X_{k-1}^3 \varphi(F(\varepsilon_k)) \right\} \\
+ E \left\{ \left( \sum_{i=k+1}^{n} X_{i-1} \varphi(F(\varepsilon_i)) \right) \left( \sum_{j=k+1}^{n} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_{k-1}^2 \right\} \\
= E_{21} + E_{22} + E_{23} + E_{24} + E_{25} + E_{26} \\
\tag{2.4}
\]

This notation will be helpful as we tackle \( E_2 \) piece-by-piece. In fact, we will continue to use this type of notation as we further break up the \( E_2 \) terms into more manageable pieces as necessary. The first of the six terms, \( E_{21} \), will be the most tedious to handle. The following three lemmas will essentially prove lemma 2.2.2.

**Lemma 2.2.3.** Assume \( k \) is fixed. Then

\[
E_{21} = E \left\{ \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) \right) \left( \sum_{j=1}^{k-1} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_{k-1}^2 \right\} \\
= \left( \rho^2 E(X_1^4) + 2\rho \delta_{12} E(X_1^3) + \sigma_X^2 \left( \delta_{22} - \frac{\sigma^2}{1 - \rho^2} \right) \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right) \right) \\
+ \sigma^2 \sigma_X^2 \left( \frac{k-1}{1 - \rho^2} \right) \\
+ 8 \sigma_X^2 \delta_{11}^2 \frac{\rho^2}{(1 - \rho^2)^2} \left( 1 - (k-1)\rho^{2(k-2)} + (k-2)\rho^{2(k-1)} \right),
\]
where \( \delta_{i2} = E(\varepsilon_1\varphi^2(F(\varepsilon_1))) \) and \( \delta_{22} = E(\varepsilon_1^2\varphi^2(F(\varepsilon_1))) \).

**Proof.** Let's consider the inter-dependence of the terms, keeping in mind that \( i < k \) and \( j < k \). \( X_{i-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{i-1}) \), \( X_{j-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{j-1}) \), and \( X_{k-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{k-1}) \) are all dependent on each other. \( X_{i-1} \) is independent of \( \varepsilon_i \), but not \( \varepsilon_j \). \( X_{j-1} \) is independent of \( \varepsilon_j \), but not \( \varepsilon_i \). \( \varepsilon_i \) and \( \varepsilon_j \) are independent, unless \( i = j \). To better get around this dependence, let's rewrite the \( E_{21} \) term, separating the double sum.

\[
E_{21} = \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} X_{i-1}X_{j-1}\varphi(F(\varepsilon_i))\varphi(F(\varepsilon_j))X_{k-1}^2
\]

\[
= \sum_{i=1}^{k-1} X_{i-1}^2\varphi^2(F(\varepsilon_i))X_{k-1}^2 + \sum_{i \neq j} X_{i-1}X_{j-1}\varphi(F(\varepsilon_i))\varphi(F(\varepsilon_j))X_{k-1}^2
\]

\[
= \sum_{i=1}^{k-1} X_{i-1}^2\varphi^2(F(\varepsilon_i))X_{k-1}^2 + 2\sum_{i<j} X_{i-1}X_{j-1}\varphi(F(\varepsilon_i))\varphi(F(\varepsilon_j))X_{k-1}^2
\]

\[
= \sum_{i=1}^{k-1} X_{i-1}^2\varphi^2(F(\varepsilon_i))X_{k-1}^2
\]

\[
+ 2\sum_{i<j} X_{i-1}X_{j-1}\varphi(F(\varepsilon_i))\varphi(F(\varepsilon_j))X_{k-1}^2
\]

\[
= E_{211} + E_{212}
\]

In the following derivations, we will continue to use lemma 1.1.1 and the method of re-writing \( X \) terms with larger indices in terms of \( X \)'s with smaller indices, given by (2.3). In this expression, the "i" is arbitrary, and of course is chosen to be the same \( i \) as the one already in the sum. Squaring (2.3) gives us

\[
X_{k-1}^2 = \rho^{2(k-1)}X_{i-1}^2 + 2\rho^{k-i}X_{i-1}\Gamma_i^{k-1} + (\Gamma_i^{k-1})^2,
\]
which we now work into the $E_{211}$ term.

$$E_{211} = E\left(\sum_{i=1}^{k-1} X_i^{-1} \varphi^2(F(\varepsilon_i)) \left(\rho^{2(k-i)} X_i^{-2} + 2\rho^{k-i} X_i^{-1} \Gamma_i^{-1} + \left(\Gamma_i^{-1}\right)^2\right)\right)$$

\[
= \sum_{i=1}^{k-1} \rho^{2(k-i)} E\left(X_i^{-1} \varphi^2(F(\varepsilon_i))\right) + 2 \sum_{i=1}^{k-1} \rho^{k-i} E\left(X_i^{-1} \varphi^2(F(\varepsilon_i))\Gamma_i^{-1}\right) \\
+ \sum_{i=1}^{k-1} E\left(X_i^{-1} \varphi^2(F(\varepsilon_i))\left(\Gamma_i^{-1}\right)^2\right) \\
= E_{2111} + E_{2112} + E_{2113}
\]

Again we will handle each piece of the expression one at a time.

$$E_{2111} = \sum_{i=1}^{k-1} \rho^{2(k-i)} E(X_i^{-1}) E\left(\varphi^2(F(\varepsilon_i))\right) = \sum_{i=1}^{k-1} \rho^{2(k-i)} E(X_i^4) E\left(\varphi^2(F(\varepsilon_i))\right)$$

Note again that the expectations do not depend on the index, $i$. For the $E(\varphi^2(F(\varepsilon_i)))$ term, consider the following. Recall that $U = F(\varepsilon_1) \sim \text{uniform}(0,1)$. Thus $E(\varphi^2(U)) = \int_0^1 \varphi^2(u) du = 1$, by assumption S2. For the $\sum_{i=1}^{k-1} \rho^{2(k-i)}$ term, recall the geometric series: $\sum_{i=0}^{n-1} x^i = \frac{1-x^n}{1-x}, \ x \neq 1$.

$$\sum_{i=1}^{k-1} \rho^{2(k-i)} = \sum_{i=1}^{k-1} \rho^{2k} \rho^{-2i} = \rho^{2k} \sum_{i=1}^{k-1} \left(\frac{1}{\rho^2}\right)^i = \rho^{2k} \left(\sum_{i=0}^{k-1} \left(\frac{1}{\rho^2}\right)^i - \left(\frac{1}{\rho^2}\right)^k\right) = \rho^{2k} \left(\frac{1 - \left(\frac{1}{\rho^2}\right)^k}{1 - \frac{1}{\rho^2}} - 1\right) = \frac{\rho^{2k} - 1}{(\rho^2 - 1)/\rho^2} - \rho^{2k} = \rho^2 \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2k}$$

Thus

$$E_{2111} = \sum_{i=1}^{k-1} \rho^{2(k-i)} E(X_1^4) \cdot 1 = E(X_1^4) \left(\frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2k}\right).$$

The next term is

$$E_{2112} = 2 \sum_{i=1}^{k-1} \rho^{k-i} E\left(X_{i-1}^3 \varphi^2(F(\varepsilon_i))\Gamma_i^{-1}\right).$$
Here, \( X_{i-1} \in \mathcal{F}(\varepsilon_0, \ldots, \varepsilon_{i-1}) \) and \( \Gamma_{i-1}^k = \sum_{r=0}^{k-1-i} \rho^r \varepsilon_{k-1-r} \in \mathcal{F}(\varepsilon_i, \ldots, \varepsilon_{k-1}) \) are independent. The only dependence here is between \( \Gamma_{i-1}^k \) and \( \varepsilon_i \), just for one term in the \( \Gamma_{i-1}^k \) sum. So therefore let's write

\[
\Gamma_{i-1}^k = \sum_{r=0}^{k-1-i} \rho^r \varepsilon_{k-1-r} \\
= \rho^0 \varepsilon_{k-1-0} + \rho^1 \varepsilon_{k-1-1} + \cdots + \rho^{k-1-(i+1)} \varepsilon_{k-1-(k-1-(i+1))} + \rho^{k-1-i} \varepsilon_{k-1-(k-1-i)} \\
= \sum_{r=0}^{k-1-i} \rho^r \varepsilon_{k-1-r} + \rho^{k-1-i} \varepsilon_i = \Gamma_{i+1}^k + \rho^{k-1-i} \varepsilon_i.
\]

Thus now we have

\[
E_{2112} = 2 \sum_{i=1}^{k} \rho^{k-i} E \left( X_{i-1}^3 \varphi^2(F(\varepsilon_i)) (\Gamma_{i+1}^k + \rho^{k-1-i} \varepsilon_i) \right) \\
= 2 \sum_{i=1}^{k} \rho^{k-i} \left( E(X_{i-1}^3) E \left( \varphi^2(F(\varepsilon_i)) \right) E(\Gamma_{i+1}^k) + E(X_{i-1}^3) E \left( \varphi^2(F(\varepsilon_i)) \varepsilon_i \right) \rho^{k-1-i} \right) \\
= 2 \sum_{i=1}^{k} \rho^{k-i} \left( E(X_{i-1}^3) E \left( \varphi^2(F(\varepsilon_i)) \right) \right) \cdot 0 + E(X_{i-1}^3) E \left( \varepsilon_i \varphi^2(F(\varepsilon_i)) \right) \rho^{k-1-i} \\
= 2 \sum_{i=1}^{k} \rho^{2(k-i)-1} E(X_{i}^3) E \left( \varepsilon_i \varphi^2(F(\varepsilon_i)) \right) \\
= \frac{2}{\rho} \sum_{i=1}^{k} \rho^{2(k-i)} E(X_{i}^3) \delta_{12} \\
= \frac{2}{\rho} E(X_{1}^3) \delta_{12} \left( \frac{\rho^2 \rho^{2k} - 1}{\rho^2 - 1} - \rho^{2k} \right) \\
= 2E(X_{1}^3) \delta_{12} \rho \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right).
\]

Note that \( E(\Gamma_{i+1}^k) = \sum_{r=0}^{k-1-(i+1)} \rho^r E(\varepsilon_{k-1-r}) = \sum_{r=0}^{k-1-(i+1)} \rho^r \cdot 0 = 0 \). Also here we define the term \( \delta_{12} = E(\varepsilon_1 \varphi^2(F(\varepsilon_1))) \).
For \( E_{2113} \), we use the same approach that worked for \( E_{2112} \).

\[
E_{2113} = \sum_{i=1}^{k-1} E \left( X_{i-1}^2 \varphi^2(F(\epsilon_i))(\Gamma_{i-1}^{k-1})^2 \right)
\]

\[
= \sum_{i=1}^{k-1} E \left( X_{i-1}^2 \varphi^2(F(\epsilon_i))(\Gamma_{i+1}^{k-1} + \rho^{k-1-i}\epsilon_i)^2 \right)
\]

\[
= \sum_{i=1}^{k-1} E \left\{ X_{i-1}^2 \varphi^2(F(\epsilon_i)) \left( (\Gamma_{i+1}^{k-1})^2 + 2\rho^{k-1-i}\epsilon_i\Gamma_{i+1}^{k-1} + \rho^2(k-1-i)\epsilon_i^2 \right) \right\}
\]

\[
= \sum_{i=1}^{k-1} E(X_{i-1}^2) E(\varphi^2(F(\epsilon_i))) E((\Gamma_{i+1}^{k-1})^2)
\]

\[
+ 2\sum_{i=1}^{k-1} \rho^{k-1-i}E(X_{i-1}^2) E(\epsilon_i \varphi^2(F(\epsilon_i))) E(\Gamma_{i+1}^{k-1})
\]

\[
+ \sum_{i=1}^{k-1} \rho^2(k-1-i)E(X_{i-1}^2) E(\epsilon_i^2 \varphi^2(F(\epsilon_i)))
\]

Consider the \( E((\Gamma_{i+1}^{k-1})^2) \) term.

\[
(\Gamma_{i+1}^{k-1})^2 = \left( \sum_{r=0}^{k-1-(i+1)} \rho^r \epsilon_{k-1-r} \right) \left( \sum_{s=0}^{k-1-(i+1)} \rho^s \epsilon_{k-1-s} \right)
\]

\[
= \sum_{r=0}^{k-1-(i+1)} \sum_{s=0}^{k-1-(i+1)} \rho^r \rho^s \epsilon_{k-1-r} \epsilon_{k-1-s}
\]

\[
= \sum_{r=0}^{k-1-(i+1)} \rho^{2r} \epsilon_{k-1-r}^2 + \sum_{r \neq s} \rho^r \rho^s \epsilon_{k-1-r} \epsilon_{k-1-s}
\]

\[
E((\Gamma_{i+1}^{k-1})^2) = \sum_{r=0}^{k-1-(i+1)} \rho^{2r} E(\epsilon_{k-1-r}^2) + \sum_{r \neq s} \rho^r \rho^s E(\epsilon_{k-1-r} \epsilon_{k-1-s})
\]

\[
= \sum_{r=0}^{k-1-(i+1)} \rho^{2r} E(\epsilon_{k-1}^2) + \sum_{r \neq s} \rho^r \rho^s \cdot 0 = \sum_{r=0}^{k-1-(i+1)} \rho^{2r} \sigma^2
\]

\[
= \sigma^2 \frac{1 - (\rho^2)^{k-1-(i+1)+1}}{1 - \rho^2} = \sigma^2 \frac{1 - (\rho^2)^{k-i-1}}{1 - \rho^2}
\]

Note that \( E(\epsilon_{k-1-r} \epsilon_{k-1-s}) = 0 \) due to the independence of the errors. Continuing,
we have

\[ E_{2113} = \sum_{i=1}^{k-1} E(X_i^2) E(\varphi^2(F(\varepsilon_1))) \sigma^2 \frac{1 - \rho^2(k-i)}{1 - \rho^2} \]

\[ + 2 \sum_{i=1}^{k-1} E(X_i^2) E(\varepsilon_1 \varphi^2(F(\varepsilon_1))) \cdot 0 \]

\[ + \sum_{i=1}^{k-1} \rho^2(k-i) \frac{1}{\rho^2} E(X_i^2) E(\varepsilon_1^2 \varphi^2(F(\varepsilon_1))) \]

\[ = \sum_{i=1}^{k-1} \sigma_X^2 \cdot 1 \cdot \sigma^2 \frac{1}{1 - \rho^2} (1 - \rho^2(k-i) \frac{1}{\rho^2}) + \frac{1}{\rho^2} \sum_{i=1}^{k-1} \rho^2(k-i) \sigma_X^2 \delta_{22}, \]

and here we define the term \( \delta_{22} = E(\varepsilon_1^2 \varphi^2(F(\varepsilon_1))) \). Continuing,

\[ E_{2113} = \sigma^2 \sigma_X^2 \frac{1}{1 - \rho^2} \sum_{i=1}^{k-1} (1 - \rho^2(k-i) \frac{1}{\rho^2}) + \sigma^2 \delta_{22} \frac{1}{\rho^2} \sum_{i=1}^{k-1} \rho^2(k-i) \]

\[ = \sigma^2 \sigma_X^2 \frac{1}{1 - \rho^2} \left( (k - 1) - \frac{1}{\rho^2} \sum_{i=1}^{k-1} \rho^2(k-i) \right) + \sigma^2 \delta_{22} \frac{1}{\rho^2} \left( \frac{\rho^2}{\rho^2 - 1} - \rho^2 \right) \]

\[ = \sigma^2 \sigma_X^2 \frac{1}{1 - \rho^2} \left( k - 1 - \frac{1}{\rho^2} (\frac{\rho^2}{\rho^2 - 1} - \rho^2) \right) \]

\[ + \sigma^2 \delta_{22} \left( \frac{\rho^2}{\rho^2 - 1} - \rho^2 \right) \]

\[ = \sigma^2 \sigma_X^2 \left( \frac{k - 1}{1 - \rho^2} \right) - \sigma^2 \sigma_X^2 \left( \frac{1}{1 - \rho^2} \right) \left( \frac{\rho^2}{\rho^2 - 1} - \rho^2 \right) \]

\[ + \sigma^2 \delta_{22} \left( \frac{\rho^2}{\rho^2 - 1} - \rho^2 \right) \]

\[ = \sigma^2 \sigma_X^2 \left( \frac{k - 1}{1 - \rho^2} \right) + \sigma^2 \left( \delta_{22} - \frac{\sigma^2}{1 - \rho^2} \right) \left( \frac{\rho^2}{\rho^2 - 1} - \rho^2 \right), \]
and therefore the $E_{211}$ term is now

$$E_{211} = E(X_i^4) \left( \frac{\rho^2 \rho^{2k} - 1}{\rho^2 - 1} - \rho^{2k} \right) + 2E(X_i^3) \delta_{12} \rho \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right) + \sigma^2 \sigma_X^2 \left( \frac{k - 1}{1 - \rho^2} \right) + \sigma_X^2 \left( \delta_{22} - \frac{\sigma^2}{1 - \rho^2} \right) \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right)

= E(X_i^4) \rho^2 \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right) + 2E(X_i^3) \delta_{12} \rho \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right) + \sigma^2 \sigma_X^2 \left( \frac{k - 1}{1 - \rho^2} \right) + \sigma_X^2 \left( \delta_{22} - \frac{\sigma^2}{1 - \rho^2} \right) \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right)

= \left( E(X_i^4) \rho^2 + 2E(X_i^3) \delta_{12} \rho + \sigma_X^2 \left( \delta_{22} - \frac{\sigma^2}{1 - \rho^2} \right) \right) \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right) + \sigma^2 \sigma_X^2 \left( \frac{k - 1}{1 - \rho^2} \right)

= \left( \rho^2 E(X_i^4) + 2 \rho \delta_{12} E(X_i^3) + \sigma_X^2 \left( \delta_{22} - \frac{\sigma^2}{1 - \rho^2} \right) \right) \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right) + \sigma^2 \sigma_X^2 \left( \frac{k - 1}{1 - \rho^2} \right).

Next we have

$$E_{212} = 2E \left( \sum \sum X_{i-1}X_{j-1} \varphi(F(\epsilon_i)) \varphi(F(\epsilon_j)) X_{j-1}^2 \right)

= 2E \left( \sum \sum X_{i-1}X_{j-1} \varphi(F(\epsilon_i)) \varphi(F(\epsilon_j))(\rho^{k-1}X_{i-1} + \Gamma_i^{k-1})^2 \right)

= 2E \left( \sum \sum X_{i-1}X_{j-1} \varphi(F(\epsilon_i)) \varphi(F(\epsilon_j))(\rho^{2(k-1)}X_{i-1}^2 + 2\rho^{k-1}X_{i-1} \Gamma_i^{k-1} + \Gamma_i^{k-1})^2 \right)

= 2 \sum \sum \rho^{2(k-1)} E \left( X_{i-1}^3X_{j-1} \varphi(F(\epsilon_i)) \varphi(F(\epsilon_j)) \right)

+ 4 \sum \sum \rho^{k-1} E \left( X_{i-1}^2X_{j-1} \varphi(F(\epsilon_i)) \varphi(F(\epsilon_j)) \Gamma_i^{k-1} \right)

+ 2 \sum \sum E \left( X_{i-1}X_{j-1} \varphi(F(\epsilon_i)) \varphi(F(\epsilon_j)) \Gamma_i^{k-1})^2 \right)

= E_{2121} + E_{2122} + E_{2123}.$
again breaking the term up into three pieces. The first term is equal to zero:

\[
E_{2121} = 2 \sum_{i<j} \rho^{2(k-i)} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \right) E \left( \varphi(F(\varepsilon_j)) \right)
\]

\[
= 2 \sum_{i<j} \rho^{2(k-i)} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \right) \cdot 0 = 0
\]

As we move on to \(E_{2122}\), consider the \(\Gamma^k\) term, and remember that \(i < j < k\).

\[
\Gamma^k = \sum_{r=0}^{k-1-i} \rho^r \varepsilon_{k-1-r}
\]

\[
= \rho^0 \varepsilon_{k-1-0} + \rho^1 \varepsilon_{k-1-1} + \cdots + \rho^{k-1-(i+1)} \varepsilon_{k-1-(k-1-(i+1))} + \rho^{k-1-i} \varepsilon_{k-1-(k-1-i)}
\]

\[
= \rho^{k-1-i} \varepsilon_i + \rho^{k-1-(i+1)} \varepsilon_{i+1} + \cdots + \rho^{k-1-(j+1)} \varepsilon_{j-1} + \rho^{k-1-j} \varepsilon_j + \rho^{k-1-(j+1)} \varepsilon_{j+1} + \cdots + \rho^0 \varepsilon_{k-1}
\]

\[
= \rho^{k-1-i} \varepsilon_i + \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r + \rho^{k-1-j} \varepsilon_j + \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r
\]

With \(\Gamma_i^k\) written this way, notice the dependence structure of the four pieces, in particular the two sums:

\[
\sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r \in \mathcal{F}(\varepsilon_{i+1}, \ldots, \varepsilon_{j-1}), \quad \text{and} \quad \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r \in \mathcal{F}(\varepsilon_{j+1}, \ldots, \varepsilon_{k-1}).
\]
Now we can isolate some independent terms within $E_{2122}$.

\[
E_{2122} = 4 \sum_{i < j} \sum_{i < j} \rho^{k-i} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \varphi(F(\varepsilon_j)) \Gamma_i^{j-1} \right)
\]

\[
= 4 \sum_{i < j} \sum_{i < j} \rho^{k-i} \rho^{k-1-i} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \varphi(F(\varepsilon_j)) \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r \right)
\]

\[
+ 4 \sum_{i < j} \sum_{i < j} \rho^{k-1} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r \right) E \left( \varphi(F(\varepsilon_j)) \right)
\]

\[
+ 4 \sum_{i < j} \sum_{i < j} \rho^{2k-1-j-1} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \right) E \left( \varepsilon_j \varphi(F(\varepsilon_j)) \right)
\]

\[
+ 4 \sum_{i < j} \sum_{i < j} \rho^{k-1} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \right) E \left( \varphi(F(\varepsilon_j)) \right) \left( \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r \right)
\]

\[
= 4 \sum_{i < j} \sum_{i < j} \rho^{k-i} \rho^{k-1-i} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \right) \cdot 0
\]

\[
+ 4 \sum_{i < j} \sum_{i < j} \rho^{k-1} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \right) \cdot 0
\]

\[
+ 4 \sum_{i < j} \sum_{i < j} \rho^{2k-1-j-1} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \right) E \left( \varepsilon_j \varphi(F(\varepsilon_j)) \right)
\]

\[
+ 4 \sum_{i < j} \sum_{i < j} \rho^{k-1} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \right) \cdot 0 \cdot 0
\]

\[
= 4 \sum_{i < j} \sum_{i < j} \rho^{2k-1-j-1} E \left( X_{i-1}^2 X_{j-1} \varphi(F(\varepsilon_i)) \right) \delta_{i1}
\]

Once again we resort to lemma 1.1.1, this time re-writing $X_{j-1}$ like so:

\[
X_{j-1} = \rho^{j-i} X_{i-1} + \Gamma_i^{j-1}, \text{ where } i < j.
\]
Continuing, we have

\[ E_{2122} = 4 \sum_{i<j} \sum_{i<\theta} \rho^{2k-i-j-1} E(X^2_{i-1}(\rho^{i}X_{i-1} + \Gamma_i^{j-1})) \varphi(F(\varepsilon_{i}))) \delta_{11} \]

\[ = 4\delta_{11} \sum_{i<j} \sum_{i<\theta} \rho^{2k-i-j-1}(\rho^{i}E(X^3_{i-1}\varphi(F(\varepsilon_{i})))) + E(X^2_{i-1}\varphi(F(\varepsilon_{i})))\Gamma_i^{j-1}). \]

At this point, it will be useful to separate the \( i \)th error from the \( \Gamma_i^{j-1} \) term.

\[ \Gamma_i^{j-1} = \sum_{r=0}^{j-1-i} \rho^r \varepsilon_{j-1-r} \]

\[ = \rho^0 \varepsilon_{j-1-0} + \cdots + \rho^{j-1-(i+1)} \varepsilon_{j-1-(j-1-(i+1))} + \rho^{j-1-i} \varepsilon_{j-1-(j-1-i)} \]

\[ = \sum_{r=0}^{j-1-(i+1)} \rho^r \varepsilon_{j-1-r} + \rho^{j-1-i} \varepsilon_i = \Gamma_{i+1}^{j-1} + \rho^{j-1-i} \varepsilon_i \]

Continuing.

\[ E_{2122} = 4\delta_{11} \sum_{i<j} \sum_{i<\theta} \rho^{2k-i-j-1}(\rho^{i}E(X^3_{i-1})E(\varphi(F(\varepsilon_{i})))) + E(X^2_{i-1}\varphi(F(\varepsilon_{i})))\Gamma_{i+1}^{j-1} + \rho^{j-1-i} \varepsilon_i) \]

\[ = 4\delta_{11} \sum_{i<j} \sum_{i<\theta} \rho^{2k-i-j-1}(\rho^{i}E(X^3_{i-1}) \cdot 0 + E(X^2_{i-1}\varphi(F(\varepsilon_{i})))\Gamma_{i+1}^{j-1} + \rho^{j-1-i} \varepsilon_i) \]

\[ = 4\delta_{11} \sum_{i<j} \sum_{i<\theta} \rho^{2k-i-j-1}(E(X^2_{i-1})E(\varphi(F(\varepsilon_{i})))E(\Gamma_{i+1}^{j-1}) + \rho^{j-1-i} E(X^2_{i-1})E(\varepsilon_i\varphi(F(\varepsilon_{i})))) \]

Since both \( E(\varphi(F(\varepsilon_{i}))) \) and \( E(\Gamma_{i+1}^{j-1}) \) are zero, and \( E(\varepsilon_i\varphi(F(\varepsilon_{i})))) \) = \( E(\varepsilon_i\varphi(F(\varepsilon_{i})))) = \]
\[ E_{2122} = 4\delta_{11} \sum_{i<j} \rho^{2k-2i-2} \sigma_x^2 \delta_{11} \]
\[ = 4\sigma_x^2 \delta_{11} \sum_{i<j} \rho^{-2i} \rho^{2k-2} \]
\[ = 4\sigma_x^2 \delta_{11} \rho^{2(k-1)} \sum_{i<j} \left( \frac{1}{\rho^2} \right)^i. \]

Let's simplify the \( \sum \sum \left( \frac{1}{\rho^2} \right)^i \) term by indexing the double sum explicitly.

\[
\sum \sum \left( \frac{1}{\rho^2} \right)^i = \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \left( \frac{1}{\rho^2} \right)^i = \sum_{i=1}^{k-2} (k-1-i) \left( \frac{1}{\rho^2} \right)^i = k \sum_{i=1}^{k-2} \left( \frac{1}{\rho^2} \right)^i - \sum_{i=1}^{k-2} (i+1) \left( \frac{1}{\rho^2} \right)^i
\]

In general, let us for a moment consider the geometric series \( \sum_{i=0}^{n-1} x^i = \frac{1-x^n}{1-x} \).

Differentiating both sides, with respect to \( x \), gives

\[
\sum_{i=1}^{n-1} ix^{i-1} = \frac{(1-x)(-nx^{n-1}) - (1-x^n)(-1)}{(1-x)^2} = \frac{-nx^{n-1} + nx^n + 1 - x^n}{(1-x)^2} - \frac{(n-1)x^n - nx^{n-1} + 1}{(1-x)^2}.
\]

Going back to our notation, and substituting \( j = i + 1 \), we have

\[
\sum \sum \left( \frac{1}{\rho^2} \right)^i = k \sum_{i=1}^{k-2} \left( \frac{1}{\rho^2} \right)^i - \sum_{j=2}^{k-1} \left( \frac{1}{\rho^2} \right)^{j-1} \]
\[ = k \left( \sum_{i=0}^{k-2} \left( \frac{1}{\rho^2} \right)^i - 1 \right) - \left( \sum_{j=1}^{k-1} \left( \frac{1}{\rho^2} \right)^{j-1} - 1 \right) \]
\[ = k \left( \frac{1 - \left( \frac{1}{\rho^2} \right)^{k-1}}{1 - \frac{1}{\rho^2}} - 1 \right) - \left( \frac{(k-1)(\frac{1}{\rho^2})^k - k(\frac{1}{\rho^2})^{k-1} + 1}{(1 - \frac{1}{\rho^2})^2} - 1 \right) \]
\[ = k\rho^2 \left( 1 - \rho^{-2(k-1)} \right) - k - \rho^4 \left( (k-1)\rho^{-2k} - k\rho^{-2(k-1)} + 1 \right) \]
\[ + 1. \]
Getting back to the $E_{2122}$ term:

$$E_{2122} = 4\sigma_X^2 \delta_{11}^2 \rho^{2(k-1)} \left( \frac{k \rho^2 (1 - \rho^{-2(k-1)})}{\rho^2 - 1} - (k - 1) \right)$$

$$- \frac{\rho^4 \left( (k - 1) \rho^{-2k} - k \rho^{-2(k-1)} + 1 \right)}{(\rho^2 - 1)^2}$$

$$= 4\sigma_X^2 \delta_{11}^2 \left( \frac{k \rho^2 (1 - \rho^{2(k-1)})}{\rho^2 - 1} - (k - 1) \rho^{2(k-1)} \right)$$

$$- \frac{\rho^4 \left( (k - 1) \rho^{-2} - k + \rho^{2(k-1)} \right)}{(\rho^2 - 1)^2}$$

$$= 4\sigma_X^2 \delta_{11}^2 \left( \frac{k \rho^2 (1 - \rho^{2(k-1)})}{(1 - \rho^2)^2} \right) - \frac{(k - 1) \rho^{2(k-1)} (1 - \rho^2)^2}{(1 - \rho^2)^2}$$

$$- \frac{\rho^2 (k - 1 - k \rho^2 + \rho^2)}{(1 - \rho^2)^2}$$

$$= 4\sigma_X^2 \delta_{11}^2 \frac{\rho^2}{(1 - \rho^2)^2} \left( k - k \rho^2 - k \rho^{2(k-1)} + k \rho^2 \right)$$

$$- (k \rho^{2(k-2)} - \rho^{2(k-2)}) (1 - 2 \rho^2 + \rho^4) - k + 1 + k \rho^2 - \rho^{2k}$$

$$= 4\sigma_X^2 \delta_{11}^2 \frac{\rho^2}{(1 - \rho^2)^2} \left( k \rho^{2k} - k \rho^{2(k-1)} - k \rho^{2(k-2)} + \rho^{2(k-2)} + 2k \rho^{2(k-1)} \right)$$

$$- 2 \rho^{2(k-1)} - k \rho^{2k} + \rho^2 + 1 - \rho^{2k}$$

$$= 4\sigma_X^2 \delta_{11}^2 \frac{\rho^2}{(1 - \rho^2)^2} \left( k \rho^{2(k-1)} - (k - 1) \rho^{2(k-2)} - 2 \rho^{2(k-1)} + 1 \right)$$

$$= 4\sigma_X^2 \delta_{11}^2 \frac{\rho^2}{(1 - \rho^2)^2} \left( 1 - (k - 1) \rho^{2(k-2)} + (k - 2) \rho^{2(k-1)} \right)$$

Now consider the third part of $E_{212}$.

$$E_{2123} = 2 \sum_{i \neq j} E \left( X_{i-1} X_{j-1} \varphi(F(\xi_i)) \varphi(F(\xi_j)) (\Gamma_{i-1}^k)^2 \right)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Here we need to square the $\Gamma_i^{k-1}$ term, first writing it in four pieces like before.

\[
(\Gamma_i^{k-1})^2 = \left( \rho^{k-1-i} \varepsilon_i + \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r + \rho^{k-1-j} \varepsilon_j + \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r \right)
\times \left( \rho^{k-1-i} \varepsilon_i + \sum_{s=i+1}^{j-1} \rho^{k-1-s} \varepsilon_s + \rho^{k-1-j} \varepsilon_j + \sum_{s=j+1}^{k-1} \rho^{k-1-s} \varepsilon_s \right)
\]

\[
= \rho^{2(k-1-i)} \varepsilon_i^2 + \rho^{k-1-i} \varepsilon_i \sum_{s=i+1}^{j-1} \rho^{k-1-s} \varepsilon_s + \rho^{k-1-i} \varepsilon_i \rho^{k-1-j} \varepsilon_j
\]

\[
+ \rho^{k-1-i} \varepsilon_i \sum_{s=j+1}^{k-1} \rho^{k-1-s} \varepsilon_s + \rho^{k-1-i} \varepsilon_i \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r
\]

\[
+ \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r \sum_{s=i+1}^{j-1} \rho^{k-1-s} \varepsilon_s + \rho^{k-1-j} \varepsilon_j \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r
\]

\[
+ \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r \sum_{s=j+1}^{k-1} \rho^{k-1-s} \varepsilon_s + \rho^{k-1-j} \varepsilon_j \rho^{k-1-i} \varepsilon_i
\]

\[
+ \rho^{k-1-j} \varepsilon_j \sum_{s=i+1}^{j-1} \rho^{k-1-s} \varepsilon_s + \rho^{2(k-1-j)} \varepsilon_j^2 + \rho^{k-1-j} \varepsilon_j \sum_{s=j+1}^{k-1} \rho^{k-1-s} \varepsilon_s
\]

\[
+ \rho^{k-1-i} \varepsilon_i \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r + \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r \sum_{s=i+1}^{j-1} \rho^{k-1-s} \varepsilon_s
\]

\[
+ \rho^{k-1-j} \varepsilon_j \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r + \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r \sum_{s=j+1}^{k-1} \rho^{k-1-s} \varepsilon_s
\]

\[
= \rho^{2(k-1-i)} \varepsilon_i^2 + 2 \rho^{k-1-i} \varepsilon_i \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r + 2 \rho^{k-1-i} \rho^{k-1-j} \varepsilon_i \varepsilon_j
\]

\[
+ 2 \rho^{k-1-i} \varepsilon_i \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r + \left( \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r \right)^2
\]

\[
+ 2 \rho^{k-1-j} \varepsilon_j \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r + 2 \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r
\]

\[
+ \rho^{2(k-1-j)} \varepsilon_j^2 + 2 \rho^{k-1-j} \varepsilon_j \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r + \left( \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r \right)^2
\]
Now let's work this expression into $E_{2123}$.

$$E_{2123} = 2 \sum_{i<j} \left\{ \rho^{2(k-1-i)} E \left( X_{i-1} X_{j-1} \varepsilon_i \varepsilon_j \varphi(F(\varepsilon_i)) \right) E \left( \varphi(F(\varepsilon_j)) \right) \\
+ 2 \rho^{k-1-i} E \left( X_{i-1} X_{j-1} \varepsilon_i \varphi(F(\varepsilon_i)) \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r \right) E \left( \varphi(F(\varepsilon_j)) \right) \\
+ 2 \rho^{k-1-i} \rho^{k-1-j} E \left( X_{i-1} X_{j-1} \varepsilon_i \varphi(F(\varepsilon_i)) \right) E \left( \varepsilon_j \varphi(F(\varepsilon_j)) \right) \\
+ 2 \rho^{k-1-i} E \left( X_{i-1} X_{j-1} \varepsilon_i \varphi(F(\varepsilon_i)) \right) E \left( \varphi(F(\varepsilon_j)) \right) E \left( \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r \right) \\
+ E \left( X_{i-1} X_{j-1} \varepsilon_i \varphi(F(\varepsilon_i)) \right) E \left( \left( \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r \right)^2 \right) E \left( \varphi(F(\varepsilon_j)) \right) \\
+ 2 \rho^{k-1-j} E \left( X_{i-1} X_{j-1} \varphi(F(\varepsilon_i)) \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r \right) E \left( \varepsilon_j \varphi(F(\varepsilon_j)) \right) \\
+ 2 E \left( X_{i-1} X_{j-1} \varphi(F(\varepsilon_i)) \sum_{r=i+1}^{j-1} \rho^{k-1-r} \varepsilon_r \right) E \left( \varphi(F(\varepsilon_j)) \right) E \left( \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r \right) \\
+ \rho^{2(k-1-j)} E \left( X_{i-1} X_{j-1} \varphi(F(\varepsilon_i)) \right) E \left( \varepsilon_j^2 \varphi(F(\varepsilon_j)) \right) \\
+ 2 \rho^{k-1-j} E \left( X_{i-1} X_{j-1} \varphi(F(\varepsilon_i)) \right) E \left( \varepsilon_j \varphi(F(\varepsilon_j)) \right) E \left( \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r \right) \\
+ E \left( X_{i-1} X_{j-1} \varphi(F(\varepsilon_i)) \right) E \left( \varphi(F(\varepsilon_j)) \right) E \left( \left( \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r \right)^2 \right) \right\}$$

Note that all of the above $E \left( \varphi(F(\varepsilon_j)) \right) = E \left( \varphi(F(\varepsilon_i)) \right)$ terms are equal to zero.

Also, $E \left( \sum_{r=j+1}^{k-1} \rho^{k-1-r} \varepsilon_r \right) = \sum_{r=j+1}^{k-1} \rho^{k-1-r} E(\varepsilon_r) = 0$, thus canceling out the 1st, 2nd, 4th, 5th, 7th, 9th, and 10th pieces of the above equation. We're left with the
following.

\[
E_{2123} = 2 \sum_{i<j} \left\{ 2\rho^{2k-i-j-2}E(X_{i-1}X_{j-1}\epsilon_i\varphi(F(\epsilon_i)))E(\epsilon_1\varphi(F(\epsilon_1))) \\
+2\rho^{k-1-j}E\left(X_{i-1}X_{j-1}\varphi(F(\epsilon_i))\sum_{r=i+1}^{j-1}\rho^{k-1-r}\epsilon_r\right)E(\epsilon_1\varphi(F(\epsilon_1))) \\
+\rho^{2(k-1-j)}E(X_{i-1}X_{j-1}\varphi(F(\epsilon_i)))E(\epsilon_1^2\varphi(F(\epsilon_1))) \right\}
\]

\[
= 4 \sum_{i<j} \rho^{2k-i-j-2}E(X_{i-1}X_{j-1}\epsilon_i\varphi(F(\epsilon_i)))\delta_{11} \\
+4 \sum_{i<j} \rho^{k-1-j}E\left(X_{i-1}X_{j-1}\varphi(F(\epsilon_i))\sum_{r=i+1}^{j-1}\rho^{k-1-r}\epsilon_r\right)\delta_{11} \\
+2 \sum_{i<j} \rho^{2(k-1-j)}E(X_{i-1}X_{j-1}\varphi(F(\epsilon_i)))\delta_{21}
\]

\[
= E_{21231} + E_{21232} + E_{21233}
\]

Note that \(\delta_{21} = E(\epsilon_1^2\varphi(F(\epsilon_1)))\) is defined above. Continuing, again we use (2.5).

\[
E_{21231} = 4\delta_{11} \sum_{i<j} \rho^{2k-i-j-2}E(X_{i-1}X_{j-1}\epsilon_i\varphi(F(\epsilon_i)))
\]

\[
= 4\delta_{11} \sum_{i<j} \rho^{2k-i-j-2}E(X_{i-1}(\rho^{j-i}X_{i-1} + \Gamma_i^{j-1})\epsilon_i\varphi(F(\epsilon_i)))
\]

\[
= 4\delta_{11} \sum_{i<j} \rho^{2(k-1)}\rho^{-(i+j)}(\rho^{j-i}E(X_{i-1}^2\varphi(F(\epsilon_i)))
\]

\[
+ E(X_{i-1}\epsilon_i\varphi(F(\epsilon_i))\Gamma_i^{j-1})
\]

\[
= 4\delta_{11} \sum_{i<j} \rho^{2(k-1)}\rho^{-(i+j)}(\rho^{j-i}E(X_{i-1}^2\varphi(F(\epsilon_i)))
\]

\[
+ 0 \cdot E(\epsilon_i\varphi(F(\epsilon_i))\Gamma_i^{j-1})
\]

\[
= 4\delta_{11}\rho^{2(k-1)} \sum_{i<j} \rho^{-2i}E(X_{i-1}^2)E(\epsilon_1\varphi(F(\epsilon_1)))
\]

\[
= 4\sigma^2 \rho^{-2i} \sum_{i<j} \left(\frac{1}{\rho^2}\right)^i
\]
Recall that this is identical to the $E_{2122}$ term that we dealt with previously. Thus

$$E_{21231} = 4\sigma_X^2 \delta_{i1}^2 \frac{\rho^2}{(1 - \rho^2)^2} \left( 1 - (k - 1)\rho^{2(k-2)} + (k - 2)\rho^{2(k-1)} \right).$$

The second and third part of $E_{2123}$ are equal to zero:

$$E_{21232} = 4\delta_{i1} \sum_{i<j} \sum_{j=1}^{J-1} \rho^{k-1-j} E \left( X_{i-1} X_{j-1} \varphi(F(\epsilon_i)) \sum_{r=i+1}^{j-1} \rho^{k-1-r} \epsilon_r \right)$$

$$= 4\delta_{i1} \sum_{i<j} \sum_{j=1}^{J-1} \rho^{k-1-j} E \left( X_{i-1} (\rho^{j-i} X_{i-1} + \Gamma_i^{j-1}) \varphi(F(\epsilon_i)) \sum_{r=i+1}^{j-1} \rho^{k-1-r} \epsilon_r \right)$$

$$= 4\delta_{i1} \sum_{i<j} \sum_{j=1}^{J-1} \rho^{k-1-j} \left\{ \rho^{j-i} E \left( X_{i-1}^2 \varphi(F(\epsilon_i)) \sum_{r=i+1}^{j-1} \rho^{k-1-r} \epsilon_r \right) + E \left( X_{i-1} \varphi(F(\epsilon_i)) \Gamma_i^{j-1} \sum_{r=i+1}^{j-1} \rho^{k-1-r} \epsilon_r \right) \right\}$$

$$= 4\delta_{i1} \sum_{i<j} \sum_{j=1}^{J-1} \rho^{k-1-j} \left\{ \rho^{j-i} E(X_{i-1}^2) \cdot 0 \cdot 0 + 0 \cdot E \left( \varphi(F(\epsilon_i)) \Gamma_i^{j-1} \sum_{r=i+1}^{j-1} \rho^{k-1-r} \epsilon_r \right) \right\} = 0$$

$$E_{21233} = 2\delta_{21} \sum_{i<j} \sum_{j=1}^{J-1} \rho^{2(k-1-j)} E (X_{i-1} X_{j-1} \varphi(F(\epsilon_i)))$$

$$= 2\delta_{21} \sum_{i<j} \sum_{j=1}^{J-1} \rho^{2(k-1-j)} E (X_{i-1} (\rho^{j-i} X_{i-1} + \Gamma_i^{j-1}) \varphi(F(\epsilon_i)))$$

$$= 2\delta_{21} \sum_{i<j} \sum_{j=1}^{J-1} \rho^{2(k-1-j)} \left\{ \rho^{j-i} E(X_{i-1}^2) E(\varphi(F(\epsilon_i))) + E(X_{i-1}) E(\varphi(F(\epsilon_i)) \Gamma_i^{j-1}) \right\}$$

$$= 2\delta_{21} \sum_{i<j} \sum_{j=1}^{J-1} \rho^{2(k-1-j)} \left\{ \rho^{j-i} E(X_{i-1}^2) \cdot 0 + 0 \cdot E(\varphi(F(\epsilon_i))\Gamma_i^{j-1}) \right\} = 0$$
Thus

\[ E_{2123} = E_{21231} + 0 + 0 \]

\[ = 4\sigma_X^2 \delta_{11}^2 \frac{\rho^2}{(1 - \rho^2)^2} \left( 1 - (k - 1)\rho^{2(k-2)} + (k - 2)\rho^{2(k-1)} \right). \]

Recall that \( E_{2121} = 0 \). Since \( E_{2122} \) and \( E_{2123} \) are identical, we have

\[ E_{212} = 8\sigma_X^2 \delta_{11}^2 \frac{\rho^2}{(1 - \rho^2)^2} \left( 1 - (k - 1)\rho^{2(k-2)} + (k - 2)\rho^{2(k-1)} \right). \]

Recall that

\[ E_{211} = \left( \rho^2 E(X_1^4) + 2\rho \delta_{12} E(X_1^3) + \sigma_X^2 \left( \delta_{22} - \frac{\sigma^2}{1 - \rho^2} \right) \right) \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right) \]

\[ + \sigma^2 \sigma_X^2 \left( \frac{k - 1}{1 - \rho^2} \right). \]

Thus we finally have

\[ E_{21} = \left( \rho^2 E(X_1^4) + 2\rho \delta_{12} E(X_1^3) + \sigma_X^2 \left( \delta_{22} - \frac{\sigma^2}{1 - \rho^2} \right) \right) \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right) \]

\[ + \sigma^2 \sigma_X^2 \left( \frac{k - 1}{1 - \rho^2} \right) \]

\[ + 8\sigma_X^2 \delta_{11}^2 \frac{\rho^2}{(1 - \rho^2)^2} \left( 1 - (k - 1)\rho^{2(k-2)} + (k - 2)\rho^{2(k-1)} \right). \]

\[ \Box \]

**Lemma 2.2.4.** Recall expression (2.4). Assume \( k \) is fixed. Then \( E_{22} = 0, E_{23} = 0, E_{24} = E(X_1^4), \) and \( E_{25} = 0. \)
Proof.

\[ E_{22} = 2E \left( \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) \right) X_{k-1}^3 \varphi(F(\varepsilon_k)) \right) \]
\[ = 2E \left( \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) \right) X_{k-1}^3 \right) E(\varphi(F(\varepsilon_k))) \]
\[ = 2E \left( \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) \right) X_{k-1}^3 \right) \cdot 0 = 0 \]

\[ E_{23} = 2E \left( \left( \sum_{i=1}^{k-1} X_{i-1} \varphi(F(\varepsilon_i)) \right) \left( \sum_{j=k+1}^{n} X_{j-1} \varphi(F(\varepsilon_j)) \right) X_{k-1}^2 \right) \]
\[ = 2E \left( \sum_{i=1}^{k-1} \sum_{j=k+1}^{n} X_{i-1} X_{j-1} X_{k-1}^2 \varphi(F(\varepsilon_i)) \varphi(F(\varepsilon_j)) \right) \]
\[ = 2 \sum_{i=1}^{k-1} \sum_{j=k+1}^{n} E \left( X_{i-1} X_{j-1} X_{k-1}^2 \varphi(F(\varepsilon_i)) \varphi(F(\varepsilon_j)) \right) \]
\[ = 2 \sum_{i=1}^{k-1} \sum_{j=k+1}^{n} E \left( X_{i-1} X_{j-1} X_{k-1}^2 \varphi(F(\varepsilon_i)) \right) \cdot 0 = 0 \]

\[ E_{24} = E \left( X_{k-1}^4 \varphi^2(F(\varepsilon_k)) \right) \]
\[ = E(X_{k-1}^4) E(\varphi^2(F(\varepsilon_k))) \]
\[ = E(X_{k-1}^4) E(\varphi^2(F(\varepsilon_1))) \]
\[ = E(X_{k-1}^4) \cdot 1 = E(X_{k-1}^4) \]

\[ E_{25} = 2E \left( \left( \sum_{i=k+1}^{n} X_{i-1} \varphi(F(\varepsilon_i)) \right) X_{k-1}^3 \varphi(F(\varepsilon_k)) \right) \]
\[ = 2E \left( \sum_{i=k+1}^{n} X_{i-1} X_{k-1}^3 \varphi(F(\varepsilon_k)) \varphi(F(\varepsilon_i)) \right) \]
\[ = 2 \sum_{i=k+1}^{n} E \left( X_{i-1} X_{k-1}^3 \varphi(F(\varepsilon_k)) \right) E(\varphi(F(\varepsilon_i))) \]
\[ = 2 \sum_{i=k+1}^{n} E \left( X_{i-1} X_{k-1}^3 \varphi(F(\varepsilon_k)) \right) \cdot 0 = 0 \]
Lemma 2.2.5. Assume $k$ is fixed. Then

$$E_{26} = E \left( \left( \sum_{i=k+1}^{n} X_{i-1} \varphi (F(\varepsilon_i)) \right) \left( \sum_{j=k+1}^{n} X_{j-1} \varphi (F(\varepsilon_j)) \right) X_{k-1}^2 \right)$$

$$= (n-k) \frac{\sigma_2^2 \chi_k^2}{1-\rho^2} + \rho^2 - \rho^{2(n-k+1)} \left( E(X_1^4) - \frac{\sigma_2^2 \chi_k^2}{1-\rho^2} \right).$$

Proof. This last part of $E_2$ will require some of the methods we used to solve $E_{21}$.

$$E_{26} = E \left( \left( \sum_{i=k+1}^{n} X_{i-1} \varphi (F(\varepsilon_i)) \right) \left( \sum_{j=k+1}^{n} X_{j-1} \varphi (F(\varepsilon_j)) \right) X_{k-1}^2 \right)$$

$$= E \left( \sum_{i=k+1}^{n} \sum_{j=k+1}^{n} X_{i-1} X_{j-1} X_{k-1}^2 \varphi (F(\varepsilon_i)) \varphi (F(\varepsilon_j)) \right)$$

$$= E \left( \sum_{i=k+1}^{n} X_{i-1}^2 X_{k-1}^2 \varphi^2 (F(\varepsilon_i)) \right)$$

$$+ 2 \sum_{i<j} \sum_{i=k+1}^{n} X_{i-1} X_{j-1} X_{k-1}^2 \varphi (F(\varepsilon_i)) \varphi (F(\varepsilon_j))$$

$$= \sum_{i=k+1}^{n} E(X_{i-1}^2 X_{k-1}^2) E(\varphi^2 (F(\varepsilon_i)))$$

$$+ 2 \sum_{i<j} \sum_{i=k+1}^{n} E (X_{i-1} X_{j-1} X_{k-1}^2 \varphi (F(\varepsilon_i))) \cdot 0$$

$$= \sum_{i=k+1}^{n} E(X_{i-1}^2 X_{k-1}^2) \cdot 1$$

$$= \sum_{i=k+1}^{n} E (\rho_{i-k}^2 X_{k-1} + \Gamma_{i-1}^2 X_{k-1}^2)$$

$$= \sum_{i=k+1}^{n} E (\rho^{2(i-k)} X_{k-1}^2 + 2\rho_{i-k}^2 X_{k-1} \Gamma_{i-1}^2 + (\Gamma_{i-1}^2) X_{k-1}^2)$$

$$= \sum_{i=k+1}^{n} E (\rho^{2(i-k)} X_{k-1}^4 + 2\rho_{i-k}^2 X_{k-1}^3 \Gamma_{i-1}^2 + X_{k-1}^2 (\Gamma_{i-1}^2))$$

$$= \sum_{i=k+1}^{n} (\rho^{2(i-k)} E(X_{k-1}^4) + 2\rho_{i-k}^2 E(X_{k-1}^3) E(\Gamma_{i-1}^2) + E(X_{k-1}^2 E((\Gamma_{i-1}^2)))$$
Consider the expectations of the gamma terms:

\[
E(\Gamma_k^{i-1}) = E\left( \sum_{r=0}^{i-1-k} \rho^r \varepsilon_{i-1-r} \right) = \sum_{r=0}^{i-1-k} \rho^r E(\varepsilon_{i-1-r}) = \sum_{r=0}^{i-1-k} \rho^r \cdot 0 = 0
\]

\[
(\Gamma_k^{i-1})^2 = \left( \sum_{r=0}^{i-1-k} \rho^r \varepsilon_{i-1-r} \right) \left( \sum_{s=0}^{i-1-k} \rho^s \varepsilon_{i-1-s} \right)
= \sum_{r=0}^{i-1-k} \sum_{s=0}^{i-1-k} \rho^r \rho^s \varepsilon_{i-1-r} \varepsilon_{i-1-s}
= \sum_{r=0}^{i-1-k} \rho^{2r} \varepsilon_{i-1-r}^2 + \sum_{r \neq s} \rho^r \rho^s \varepsilon_{i-1-r} \varepsilon_{i-1-s}
\]

\[
E\left((\Gamma_k^{i-1})^2\right) = \sum_{r=0}^{i-1-k} \rho^{2r} E(\varepsilon_{i-1-r}^2) + \sum_{r \neq s} \rho^r \rho^s E(\varepsilon_{i-1-r} \varepsilon_{i-1-s})
= \sum_{r=0}^{i-1-k} \rho^{2r} \sigma^2 + \sum_{r \neq s} \rho^r \rho^s \cdot 0
= \sigma^2 \sum_{r=0}^{i-1-k} (\rho^2)^r = \sigma^2 \frac{1 - \rho^{2(i-k)}}{1 - \rho^2}
\]

So now we have

\[
E_{26} = \sum_{i=k+1}^{n} \left( \rho^{2(i-k)} E(X_{1}^4) + 2\rho^{i-k} E(X_{1}^3) \cdot 0 + E(X_{1}^2) \sigma^2 \frac{1 - \rho^{2(i-k)}}{1 - \rho^2} \right)
= E(X_{1}^4) \sum_{i=k+1}^{n} \rho^{2(i-k)} + E(X_{1}^2) \sigma^2 \frac{1 - \rho^{2(i-k)}}{1 - \rho^2} \sum_{i=k+1}^{n} (1 - \rho^{2(i-k)})
= E(X_{1}^4) \rho^{-2k} \sum_{i=k+1}^{n} (\rho^2)^i + \frac{\sigma^2 \sigma_{X}^2}{1 - \rho^2} \left( n - k - \rho^{-2k} \sum_{i=k+1}^{n} (\rho^2)^i \right).
\]

The above sum is another geometric series:

\[
\sum_{i=k+1}^{n} (\rho^2)^i = \sum_{i=0}^{n} (\rho^2)^i - \sum_{i=0}^{k} (\rho^2)^i
= \frac{1 - (\rho^2)^{n+1}}{1 - \rho^2} - \frac{1 - (\rho^2)^{k+1}}{1 - \rho^2} = \frac{\rho^{2(k+1)} - \rho^{2(n+1)}}{1 - \rho^2}
\]
Finally for $E_{26}$ we have

$$E_{26} = \frac{E(X^4_1)\rho^{-2k}\rho^{2(k+1)} - \rho^{2(n+1)}}{1 - \rho^2}$$

$$+ E(X^2_1) \frac{\sigma^2}{1 - \rho^2} \left( n - k - \rho^{-2k}\rho^{2(k+1)} - \rho^{2(n+1)} \right)$$

$$= E(X^4_1) \frac{\sigma^2}{1 - \rho^2} + (n - k) \frac{\sigma^2}{1 - \rho^2} E(X^2_1)$$

$$- E(X^2_1) \frac{\sigma^2}{1 - \rho^2} \rho^2 - \rho^{2(n-k+1)}$$

$$= (n - k) \frac{\sigma^2 \sigma_X^2}{1 - \rho^2} + \frac{\rho^2 - \rho^{2(n-k+1)}}{1 - \rho^2} \left( E(X^4_1) - \frac{\sigma^2 \sigma_X^2}{1 - \rho^2} \right).$$

Now we have all the pieces required to prove lemma 2.2.2.

**Proof.**

$$E_2 = E_{21} + E_{22} + E_{23} + E_{24} + E_{25} + E_{26}$$

$$= \left( \rho^2 E(X^4_1) + 2 \rho \delta_{12} E(X^3_1) + \sigma_X^2 (\delta_{22} - \frac{\sigma^2}{1 - \rho^2}) \right) \left( \frac{\rho^{2k} - 1}{\rho^2} - \rho^{2(k-1)} \right)$$

$$+ \sigma_X^2 \frac{k - 1}{1 - \rho^2} + 8 \sigma_X^2 \delta_{11} \frac{\rho^2}{(1 - \rho^2)^2} \left( 1 - (k - 1) \rho^{2(k-2)} + (k - 2) \rho^{2(k-1)} \right)$$

$$+ 0 + 0 + E(X^4_1) + 0$$

$$+ \sigma_X^2 \left( \frac{n - k}{1 - \rho^2} \right) + \left( E(X^4_1) - \frac{\sigma^2 \sigma_X^2}{1 - \rho^2} \right) \left( \frac{\rho^2 - \rho^{2(n-k+1)}}{1 - \rho^2} \right)$$

$$= \left( \rho^2 E(X^4_1) + 2 \rho \delta_{12} E(X^3_1) + \sigma_X^2 (\delta_{22} - \frac{\sigma^2}{1 - \rho^2}) \right) \left( \frac{\rho^{2k} - 1}{\rho^2} - \rho^{2(k-1)} \right)$$

$$+ \sigma_X^2 \left( \frac{n - 1}{1 - \rho^2} \right) + 8 \sigma_X^2 \delta_{11} \frac{\rho^2}{(1 - \rho^2)^2} \left( 1 - (k - 1) \rho^{2(k-2)} + (k - 2) \rho^{2(k-1)} \right)$$

$$+ E(X^4_1) + \left( E(X^4_1) - \frac{\sigma^2 \sigma_X^2}{1 - \rho^2} \right) \left( \frac{\rho^2 - \rho^{2(n-k+1)}}{1 - \rho^2} \right).$$
Theorem 2.2.1. Assume $k$ is fixed. The variance of $\hat{\varepsilon}_k$ in (2.1) can be expressed as

\[ \text{Var}(\hat{\varepsilon}_k) = \sigma^2 - 2\tau_\sigma \frac{1}{n} \delta_{11} \]

\[ + \tau_\sigma^2 (1 - \rho^2)^2 \frac{1}{n^2} \left\{ \frac{6\rho^4}{(1 - \rho^2)(1 - \rho^4)} + \frac{\rho^2}{1 - \rho^2} K_\varepsilon \right. \]

\[ + 2\rho \delta_{12} \frac{1}{\sigma(1 - \rho^2)} S_\varepsilon + \frac{1}{\sigma^2(1 - \rho^2)} (\delta_{22} - \frac{\sigma^2}{1 - \rho^2}) \]

\[ \times \left( \frac{\rho^{2k} - 1 - \rho^{2(k-1)}}{\rho^2 - 1 - \rho^{2(k-1)}} \right) + \frac{n - 1}{(1 - \rho^2)^2} \]

\[ + 8\delta_{11}^2 \frac{\rho^2}{\sigma^2(1 - \rho^2)^3} (1 - (k - 1)\rho^{2(k-2)} + (k - 2)\rho^{2(k-1)}) \]

\[ + \frac{6\rho^2}{(1 - \rho^2)(1 - \rho^4)} + \frac{1}{1 - \rho^2} K_\varepsilon \]

\[ + \left( \frac{6\rho^2}{(1 - \rho^2)(1 - \rho^4)} + \frac{1}{1 - \rho^2} K_\varepsilon - \frac{1}{(1 - \rho^2)^2} \right) \]

\[ \times \left( \frac{\rho^2 - \rho^{2(n-k+1)}}{1 - \rho^2} \right) \]

(2.6)

where

\[ S_\varepsilon = \frac{E(\varepsilon_1^3)}{\sigma^3} \text{ and } K_\varepsilon = \frac{E(\varepsilon_1^4)}{\sigma^4}. \]

Proof. Recall that

\[ \text{Var}(\hat{\varepsilon}_k) = E(\hat{\varepsilon}_k^2) = \sigma^2 - 2\tau_\sigma \sigma_X^{-2} \frac{1}{n} E_1 + \tau_\sigma^2 \sigma_X^{-4} \frac{1}{n^2} E_2, \]
and also recall that \( E_1 = \sigma_X^2 \delta_{11} \).

\[
\text{Var}(\varepsilon_k) = \sigma^2 - 2\tau_\sigma \sigma_X^{-2} \frac{1}{n} \sigma_X^2 \delta_{11}
\]

\[
+ \tau_\sigma \sigma_X^{-4} \frac{1}{n^2} \left\{ \left( \rho^2 E(X_1^4) + 2\rho \delta_{12} E(X_3^4) + \sigma_X^2 (\delta_{22} - \frac{\sigma^2}{1 - \rho^2}) \right) \times \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right) + \sigma_X^2 \left( \frac{k - 1}{1 - \rho^2} \right) \right. \\
+ 8\sigma_X^2 \delta_{11} \frac{\rho^2}{(1 - \rho^2)^2} (1 - (k - 1)\rho^{2(k-2)} + (k - 2)\rho^{2(k-1)}) \\
+ E(X_4) + \sigma_X^2 \left( \frac{n - k}{1 - \rho^2} \right) \\
+ \left( E(X_1^4) - \frac{\sigma_X^2}{1 - \rho^2} \right) \left( \frac{\rho^2 - \rho^{2(n-k+1)}}{1 - \rho^2} \right) \right\} \\
= \sigma^2 - 2\tau_\sigma \frac{1}{n} \delta_{11}
\]

\[
+ \tau_\sigma \sigma_X^{-4} \frac{1}{n^2} \left\{ \left( \rho^2 E(X_1^4) + 2\rho \delta_{12} E(X_3^4) + \sigma_X^2 (\delta_{22} - \frac{\sigma^2}{1 - \rho^2}) \right) \times \left( \frac{\rho^{2k} - 1}{\rho^2 - 1} - \rho^{2(k-1)} \right) + \sigma_X^2 \left( \frac{n - k}{1 - \rho^2} \right) \right. \\
+ 8\sigma_X^2 \delta_{11} \frac{\rho^2}{(1 - \rho^2)^2} (1 - (k - 1)\rho^{2(k-2)} + (k - 2)\rho^{2(k-1)}) \\
+ E(X_4) + \left( E(X_1^4) - \frac{\sigma_X^2}{1 - \rho^2} \right) \left( \frac{\rho^2 - \rho^{2(n-k+1)}}{1 - \rho^2} \right) \right\} \\
= \sigma^2 - 2\tau_\sigma \frac{1}{n} \delta_{11}
\]

Since we make more assumptions on the errors than we do about the time series, \( \text{Var}(\varepsilon_k) \) should be in terms of "\( \varepsilon \)" and not "\( X \)". Starting with the second moment:

\[
E(X_1^2) = E((\rho X_0 + \varepsilon_1)^2)
\]

\[
= E(\rho^2 X_0^2 + 2\rho X_0 \varepsilon_1 + \varepsilon_1^2)
\]

\[
= \rho^2 E(X_0^2) + 2\rho E(X_0) E(\varepsilon_1) + E(\varepsilon_1^2)
\]

Note that \( X_0 \) and \( \varepsilon_1 \) are independent and their individual expectations are zero.
Also note that the expectations of the above terms do not depend on the index of \( X \) or \( \varepsilon \). Therefore the above expression implies \( E(X^2) = \rho^2 E(X^2) + E(\varepsilon^2) \), which gives us \( E(X^2) = E(\varepsilon^2)/(1 - \rho^2) \). Now, since \( E(X^2) = Var(X) + (E(X))^2 = \sigma_X^2 + 0 \), and \( E(\varepsilon^2) = Var(\varepsilon) + (E(\varepsilon))^2 = \sigma^2 + 0 \), we have \( \sigma_X^2 = \sigma^2/(1 - \rho^2) \). For the third moment, we have

\[
E(X_3^3) = E((\rho X_0 + \varepsilon_1)^3)
\]

\[
= E((\rho X_0 + \varepsilon_1)(\rho^2 X_0^2 + \rho^2 X_0^2 + 2 \rho X_0 \varepsilon_1 + \varepsilon_1^2))
\]

\[
= E(\rho^3 X_0^3 + 3 \rho^2 X_0^2 \varepsilon_1 + 3 \rho X_0 \varepsilon_1^2 + \varepsilon_1^3)
\]

\[
= \rho^3 E(X_0^3) + 3 \rho^2 E(X_0^2) E(\varepsilon_1) + 3 \rho E(X_0) E(\varepsilon_1^2) + E(\varepsilon_1^3)
\]

This implies \( E(X^3) = \rho^3 E(X^3) + E(\varepsilon^3) \), which gives us \( E(X^3) = E(\varepsilon^3)/(1 - \rho^3) \).

The fourth moment will be handled similarly.

\[
E(X_4^4) = E((\rho X_0^2 + 2 \rho X_0 \varepsilon_1 + \varepsilon_1^2)^2)
\]

\[
= E(\rho^4 X_0^4 + 4 \rho^3 X_0^3 \varepsilon_1 + 6 \rho^2 X_0^2 \varepsilon_1^2 + 4 \rho X_0 \varepsilon_1^3 + \varepsilon_1^4)
\]

\[
= \rho^4 E(X_0^4) + 4 \rho^3 E(X_0^3) \cdot 0 + 6 \rho^2 \sigma_X^2 \sigma^2 + 4 \rho \cdot 0 \cdot E(\varepsilon_1^3) + E(\varepsilon_1^4)
\]

\[
= \rho^4 E(X_0^4) + 6 \rho^2 \sigma_X^2 \sigma^2 + E(\varepsilon_1^4)
\]

This leads to

\[
E(X^4) = \rho^4 E(X^4) + 6 \rho^2 \left( \frac{\sigma^2}{1 - \rho^2} \right) \sigma^2 + E(\varepsilon^4)
\]

\[
(1 - \rho^4) E(X^4) = \frac{6 \rho^2 \sigma^4}{1 - \rho^2} + E(\varepsilon^4)
\]

\[
E(X^4) = \frac{6 \rho^2 \sigma^4}{(1 - \rho^2)(1 - \rho^4)} + \frac{E(\varepsilon^4)}{1 - \rho^4}.
\]
Now we can write (2.7) as

\[ \text{Var}(\hat{\epsilon}_k) = \sigma^2 - 2\tau_\varphi \frac{1}{n} \delta_{11} \]

\[ + \tau_\varphi^2 \frac{(1 - \rho^2)^2}{\sigma^4} \frac{1}{n^2} \left\{ \frac{\rho^2}{(1 - \rho^2)(1 - \rho^4)} \left( \frac{6\rho^2\sigma^4}{1 - \rho^4} + \frac{E(\epsilon_1^4)}{1 - \rho^4} \right) + 2\rho\delta_{12} \frac{E(\epsilon_1^4)}{1 - \rho^3} \right\} \]

\[ + \frac{\sigma^2}{1 - \rho^2} \left( \frac{\sigma^2}{\rho^2 - 1} - \frac{\sigma^2}{1 - \rho^2} \right) \left( \frac{\rho^2 - 1}{\rho^2 - 1} - \rho^2(k-1) \right) \]

\[ + \frac{\rho^2}{1 - \rho^2} \left( \frac{6\rho^2\sigma^4}{(1 - \rho^2)(1 - \rho^4)} + \frac{E(\epsilon_1^4)}{1 - \rho^4} \right) \]

\[ + \left( \frac{6\rho^2\sigma^4}{(1 - \rho^2)(1 - \rho^4)} + \frac{E(\epsilon_1^4)}{1 - \rho^4} - \frac{\sigma^2}{1 - \rho^2} \right) \frac{\sigma^2}{1 - \rho^2} \]

\[ \times \left( \frac{\rho^2 - \rho^2(n-k+1)}{1 - \rho^2} \right) \}

\[ = \sigma^2 - 2\tau_\varphi \frac{1}{n} \delta_{11} \]

\[ + \tau_\varphi^2 (1 - \rho^2)^2 \frac{1}{n^2} \left\{ \frac{6\rho^4}{(1 - \rho^2)(1 - \rho^4)} + \frac{\rho^2}{1 - \rho^4} \frac{E(\epsilon_1^4)}{\sigma^4} \right\} \]

\[ + 2\rho\delta_{12} \frac{1}{1 - \rho^3} \frac{E(\epsilon_1^4)}{\sigma^3} + \frac{1}{\sigma^2(1 - \rho^2)} \left( \delta_{22} - \frac{\sigma^2}{1 - \rho^2} \right) \]

\[ \times \left( \frac{\rho^2 - \rho^2(n-k+1)}{1 - \rho^2} \right) \]

Finally, terms for skewness, \( S_\epsilon = E(\epsilon_1^3)/\sigma^3 \), and kurtosis, \( K_\epsilon = E(\epsilon_1^4)/\sigma^4 \), of the
errors are incorporated into the variance expression.

\[
\text{Var}(\hat{\epsilon}_k) = \sigma^2 - 2\tau_\varphi \frac{1}{n} \delta_{11} \\
+ \tau_\varphi^2 (1 - \rho^2)^2 \frac{1}{n^2} \left\{ \left[ \frac{6\rho^4}{(1-\rho^2)(1-\rho^4)} + \frac{\rho^2}{1-\rho^4}K_\epsilon \right] + 2\rho\delta_{12} \frac{1}{\sigma(1-\rho^4)} S_\epsilon + \frac{1}{\sigma^2(1-\rho^2)} (\delta_{22} - \frac{\sigma^2}{1-\rho^2}) \right\} \\
\times \left( \frac{\rho^2 - 1}{\rho^2 - 1 - \rho^{2(k-1)}} \right) + \frac{n-1}{(1-\rho^2)^2} \\
+ 8\delta_{11}^2 \frac{\rho^2}{\sigma^2(1-\rho^2)^3} (1 - (k-1)\rho^{2(k-2)} + (k-2)\rho^{2(k-1)}) \\
+ \frac{6\rho^2}{(1-\rho^2)(1-\rho^4)} + \frac{1}{1-\rho^4}K_\epsilon \\
+ \left( \frac{6\rho^2}{(1-\rho^2)(1-\rho^4)} + \frac{1}{1-\rho^4}K_\epsilon - \frac{1}{(1-\rho^2)^2} \right) \\
\times \left( \frac{\rho^2 - \rho^{2(n-k+1)}}{1-\rho^2} \right) \right\} 
\]

\[\square\]

2.3 Estimation of \text{Var}(\hat{\epsilon}_k)

For estimation purposes, we use the (Wilcoxon) R-estimate \(\hat{\rho}_R\) resulting from the autoregression for \(\rho\), and the median absolute deviation (\(MAD\)) for \(\sigma\).

\[MAD = 1.483 \text{median} |\hat{\epsilon}_i - \bar{\epsilon}| , \text{where } \bar{\epsilon} \text{ is the median of } \hat{\epsilon}_1, \ldots, \hat{\epsilon}_n.\]

This is a consistent estimate of \(\sigma\) if the errors have a normal distribution (Hettmansperger & McKean, 1998, pg. 199). For \(\tau_\varphi\), we use (3.7.6) of Hettmansperger & McKean (1998). The skewness and kurtosis are estimated like so:

\[
\hat{S}_\epsilon = \frac{\text{median}(\hat{\epsilon}_i - \bar{\epsilon})^3}{MAD^3} , \text{and } \hat{K}_\epsilon = \frac{\text{median}(\hat{\epsilon}_i - \bar{\epsilon})^4}{MAD^4}.
\]
Standard moment estimators of $\delta_{11}, \delta_{12},$ and $\delta_{22}$ are given by

$$
\hat{\delta}_{ij} = \frac{1}{n} \sum_{k=1}^{n} \hat{\epsilon}_{(k)} \varphi^i \left( \frac{k}{n} \right),
$$

where $\hat{\epsilon}_{(k)}$ denotes the $k^{th}$ ordered residual. Let's denote the estimated standard error of the time series residuals as

$$
\hat{SE}(\hat{\epsilon}_k) = \sqrt{\hat{Var}(\hat{\epsilon}_k)}.
$$

Here we define the studentized time series R-residual, denoted by

$$
STSR_k = \frac{\hat{\epsilon}_k}{\hat{SE}(\hat{\epsilon}_k)}.
$$

This standardization accounts for the variance, skewness and kurtosis of the error distribution, as well as the position in time, $k$. The parameters $\tau, \delta_{11}, \delta_{12},$ and $\delta_{22}$ are correction factors due to using the rank-score function, $\varphi$ (Terpstra, 1997).
CHAPTER III

CLASSIFICATION RATES FOR DETECTING OUTLIERS

3.1 Various Standardizations of Residuals

One of the goals of model diagnostics is outlier detection. In particular, we would like to use the residuals, appropriately standardized, to "flag" outliers. Hopefully, our (robust) procedure has yielded a fit that resists undue influence by outlying points, while simultaneously drawing attention to these interesting points via residual analysis.

The goal in this chapter is to explore several different methods for standardizing, or studentizing, the residuals that are obtained from either a least squares or R-fit. We will again focus on the no-intercept \( AR(1) \) model:

\[
X_k = \rho X_{k-1} + \varepsilon_k, \quad k = 1, \ldots, n. \tag{3.1}
\]

Residuals will of course be of arbitrary size depending on the scale of the data, and need to be divided by some sort of standard error in order to make comparisons to a benchmark possible. The term "studentizing" reflects the idea of a unit-free "\( t \)-ratio", that is, the residual, divided by the appropriate standard error, follows the normal (or Student's \( t \)) distribution, at least approximately. This allows us to use a simple benchmark of \( \pm 2 \), which, by the empirical rule, should cover most of
the studentized residuals. Any residual outside that range is flagged as a potential outlier, and can be looked at more closely.

The first method of standardizing we will consider here is the most crude, but perhaps the most commonly used in time series analysis. The denominator is simply taken to be \( \hat{\sigma} \), an estimate of the standard deviation of the error distribution. This standardization was mentioned previously as "naive", as it does not take into account the position in time, \( k \).

For least squares estimation, \( \hat{\sigma} = \sqrt{MSE} \), the square root of the mean square error. We denote these standardized residuals

\[
\text{standLS}_k = \hat{\varepsilon}_k / \hat{\sigma}, \quad k = 1, \ldots, n,
\]

where \( \hat{\varepsilon}_k \) is the \( k \)th residual resulting from the least squares estimation.

For the R-fit, we use the median absolute deviation (MAD) estimator, a robust estimate given by

\[
MAD = 1.483 \text{med}_i \{ |\hat{\varepsilon}_R_i - \text{med}_j \hat{\varepsilon}_R_j | \}
\]

(Hettmansperger & McKean, 1998, pg. 199). These residuals will be denoted

\[
\text{standR}_k = \hat{\varepsilon}_k / MAD, \quad k = 1, \ldots, n,
\]

where \( \hat{\varepsilon}_k \) is the \( k \)th residual resulting from the R-estimation.

The next method of studentizing residuals comes from linear models theory. Consider the simple linear regression model

\[
X_k = \beta Z_k + \varepsilon_k, \quad k = 1, \ldots, n, \quad (3.2)
\]
where $Z_k$ denotes the $k^{th}$ design point. Note that these design points are non-random. For least squares regression, we know that

$$\text{Var}(\hat{\varepsilon}_k) = \sigma^2(1 - h_k),$$

(3.3)

where $h_k$ is the $k^{th}$ diagonal element of the hat matrix, $H = X(X'X)^{-1}X'$, where $X$ is the design matrix (see e.g. Neter, Wasserman & Kutner (1990, pg. 239)).  

For a simple linear regression through the origin like (3.2), the design “matrix” is simply the independent variable vector $Z$, and

$$h_k = \frac{Z_k^2}{\sum_{i=1}^{n} Z_i^2}.$$  

(3.4)

Now consider the AR(1) model (3.1). The design points, $Z_k = X_{k-1}$ for $k = 1, \ldots, n$, are random. A standardization based on (3.3) might not be appropriate. We get around this by thinking of $X_k$ as a random (response) variable, and $X_{k-1}$ as a fixed design point, for all fixed values of $k$. This “conditioning” on the design points allows us the following linear model standardization:

$$SLMLS_k = \frac{\hat{\varepsilon}_k}{\hat{\sigma} \sqrt{1 - h_k}}, \quad k = 1, \ldots, n,$$

where $\hat{\varepsilon}_k$ is the $k^{th}$ least squares residual, and $h_k$ in (3.4) becomes

$$h_k = \frac{X_{k-1}^2}{\sum_{i=1}^{n} X_{i-1}^2}.$$

These are commonly called internally studentized residuals in the linear models context (See e.g. Staudte and Sheather (1990)).
Internally studentized residuals for R-estimation were devised by McKean, Sheather & Hettmansperger (1990). The results are based on a first-order approximation of the R-residual variance. For a simple linear regression through the origin,

\[ \text{Var}(\hat{e}_k) = \sigma^2 \left( 1 - K_R h_k \right), \]

where

\[ K_R = \left( \frac{\tau_\varphi}{\sigma} \right)^2 \left( \frac{2 \delta_{11}}{\tau_\varphi} - 1 \right). \]  

\( \tau_\varphi^{-1}, \varphi_f(u) \) and \( \delta_{11} \) are as given in (1.5), (1.6), and lemma 2.2.1, respectively. See Hettmansperger and McKean (1998, Section 3.9.2). As mentioned in section 1.1.5, we make use of Wilcoxon scores:

\[ \varphi_f(u) = \sqrt{(12)(u - \frac{1}{2})}, \]

which leads to \( \tau_\varphi^{-1} \) as in (1.7). For estimation, we use \( K_R \) as given in section 3.9.25 of Hettmansperger and McKean (1998) and \( \hat{\sigma} = \text{MAD} \). Thus the linear model standardization for R-residuals is

\[ \text{SLMR}_k = \frac{\hat{e}_k}{\text{MAD} \sqrt{1 - K_R h_k}}, \quad k = 1, \ldots, n, \]

where \( \hat{e}_k \) is the \( k^{th} \) R-residual.

For both of the linear model standardized residuals, SLMLS and SLMR, the procedure used for the computations, RGLM, fits an intercept, even for our no-intercept model. The denominator of the internal \( t \) residuals corrects for this intercept (see Kapenga, McKean and Vidmar (1996)). Since we are centering the time series and assuming a no-intercept model, it is possible that this correction is not wanted. “Unadjusted” linear model residuals will be included in this study as well, denoted as USLMLS for least squares and USLMR for R-estimation.
Under the assumption of normal errors (and Wilcoxon scores), an approximate linear model standardization for the R-residuals is also available. With normal errors, we will have \( \tau_{\rho} = \sqrt{\frac{2}{3}} \sigma \), \( \delta_{11} = \sqrt{\frac{3}{\pi}} \sigma \), and \( K_R = 2 - \frac{\pi}{3} \). So we have

\[
SLMRA_k = \frac{\hat{\varepsilon}_k}{MAD \sqrt{1 - (2 - \frac{\pi}{3}) h_k}}, \quad k = 1, \ldots, n,
\]

where \( \hat{\varepsilon}_k \) is the \( k^{th} \) R-residual.

Studentized residuals, specific to the time series model, have been developed for least squares estimation as well (Terpstra et al., 2002). Similar to the STSR residuals devised in Chapter II, these residuals do not require us to "condition" on the design points. Recall (see e.g. Wei (1990, p. 148)) that the least squares estimate of \( \rho \) in (3.1) satisfies

\[
\hat{\rho}_{LS} = \frac{\sum_{i=1}^{n} X_{i-1}X_i}{\sum_{i=1}^{n} X_i^2} = \rho + \sigma_X^2 \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \varepsilon_i + o_p \left( \frac{1}{\sqrt{n}} \right)
\]

where \( \sigma_X^2 = Var(X_1) \). This yields an approximation for the \( k^{th} \) least squares residual.

\[
\hat{\rho}_{LS} = \rho + \sigma_X^2 \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \varepsilon_i
\]

\[
\hat{\rho}_{LS}X_{k-1} = \rho X_{k-1} + \sigma_X^2 \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \varepsilon_i X_{k-1} = X_k - \varepsilon_k + \sigma_X^2 \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \varepsilon_i X_{k-1}
\]

\[
X_k - \hat{\rho}_{LS}X_{k-1} = \varepsilon_k - \sigma_X^2 \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \varepsilon_i X_{k-1}
\]

\[
\hat{\varepsilon}_k = \varepsilon_k - \sigma_X^2 \frac{1}{n} \sum_{i=1}^{n} X_{i-1} \varepsilon_i X_{k-1}
\]
**Theorem 3.1.1.** Under assumptions $M_1$, $M_2$, and $E_1$,

$$
\text{Var}(\hat{\varepsilon}_k) = \frac{\sigma^2}{n^2} \left\{ n^2 - n - 1 + \left( \frac{6\rho^2}{1 + \rho^2} + \frac{1 - \rho^2}{1 + \rho^2} K_\varepsilon \right) \left( 1 + \frac{\rho^2 - \rho^2 k}{1 - \rho^2} + \frac{\rho^2 - \rho^2(n-k+1)}{1 - \rho^2} \right) \\
+ \left( \frac{2(1 - \rho^2)^2}{\rho(1 - \rho^3)} S_\varepsilon^2 - \left( 1 - \frac{1}{\rho^2} \right) K_\varepsilon - \frac{1}{\rho^2} \right) \left( \frac{\rho^2 - \rho^2 k}{1 - \rho^2} \right) - \left( \frac{\rho^2 - \rho^2(n-k+1)}{1 - \rho^2} \right) \\
+ \frac{8\rho^2}{1 - \rho^2} (1 - (k - 1)\rho^2(k-2) + (k - 2)\rho^2(k-1)) \right\} \\
(3.6)
$$

where $\hat{\varepsilon}_k$ is the $k^{th}$ least squares residual, $\sigma^2$, $S_\varepsilon$, and $K_\varepsilon$ are the variance, skewness, and kurtosis of the error distribution, respectively.

See (Terpstra et al., 2002) for the proof. Standard moment estimators are used for $\sigma^2$, $S_\varepsilon$, and $K_\varepsilon$, along with the parameter estimate $\hat{\rho}_{LS}$. This gives us the studentized least squares time series residuals

$$
STSLS_k = \frac{\hat{\varepsilon}_k}{\hat{SE}(\hat{\varepsilon}_k)}, \quad k = 1, \ldots, n,
$$

where $\hat{\varepsilon}_k$ is the $k^{th}$ least squares residual, and $\hat{SE}(\hat{\varepsilon}_k)$ is the estimated standard error based on (3.6).

Lastly, from the work in Chapter II recall

$$
STSR_k = \frac{\hat{\varepsilon}_k}{\hat{SE}(\hat{\varepsilon}_k)}, \quad k = 1, \ldots, n,
$$

where $\hat{\varepsilon}_k$ is the $k^{th}$ R-residual, and $\hat{SE}(\hat{\varepsilon}_k)$ is the estimated standard error based on (2.6).

### 3.2 Simulation Studies

To gauge how well each of our different standardization schemes flag outliers in the time series, a simulation study is presented here. The simplicity of
the autoregressive time series model allows us to easily simulate the behavior of
time series data in the presence of different types of outliers. “Innovation” (see
Chapter IV for a detailed definition) outliers can be introduced into the error
distribution during the construction of the series. This gives us the ability to note
which error terms stand out as larger than the others, and then see if the corre­
sponding residual reflects that. The errors will follow the contaminated normal
distribution with proportion of contamination $p_c$ and contaminated variance $\sigma_c^2$.
That is,
\[
\varepsilon_i \sim (1 - p_c)N(0, 1) + p_c N(0, \sigma_c^2).
\]
A standard benchmark of $\pm 2$ will be used to decide if an error term (and later, a
standardized residual) should be considered an outlier or not. The nomenclature
of hypothesis testing is employed here in the following fashion. The $k^{th}$ error term,
$\varepsilon_k$, is the population parameter of interest. It is estimated by the $k^{th}$ standardized
residual,
\[
\hat{\varepsilon}_k = \frac{\hat{\varepsilon}_k}{SE(\hat{\varepsilon}_k)},
\]
which can be considered both the sample statistic and the test statistic, since we
will base the decision rule on its magnitude. The null hypothesis is $H_0 : |\varepsilon_k| \leq 2$,
i.e. the $k^{th}$ error term is not an outlier. Alternatively, we have $H_1 : |\varepsilon_k| > 2$, i.e.
the $k^{th}$ error term is an outlier. The decision rule is

Reject $H_0$ if $|\hat{\varepsilon}_k^*| > 2$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Thus since we have knowledge of the error terms, the number of times that $|\varepsilon_k| \leq 2$ and $|\hat{\varepsilon}_k^*| > 2$ gives us an empirical type I error rate: $P(|\hat{\varepsilon}_k^*| > 2 \text{ given } |\varepsilon_k| \leq 2) = P(\text{residual is falsely declared an outlier } | H_0 \text{ is true}) = P(\text{type I error})$. Similarly, the number of times that $|\varepsilon_k| > 2$ and $|\hat{\varepsilon}_k^*| \leq 2$ gives us an empirical type II error rate, i.e. $P(|\hat{\varepsilon}_k^*| \leq 2 \text{ given } |\varepsilon_k| > 2) = P(\text{residual is not identified as an outlier } | H_0 \text{ is false}) = P(\text{type II error})$. Further, the frequency of observing $|\varepsilon_k| > 2$ and $|\hat{\varepsilon}_k^*| > 2$ is an empirical measure of power: $P(|\hat{\varepsilon}_k^*| > 2 \text{ given } |\varepsilon_k| > 2) = P(\text{residual is identified as an outlier } | H_0 \text{ is false}) = \text{"power"}$. This is summarized in table 1.

Table 1

<table>
<thead>
<tr>
<th>Truth about error term</th>
<th>Decision made based on standardized residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\varepsilon_k</td>
</tr>
<tr>
<td>$</td>
<td>\varepsilon_k</td>
</tr>
</tbody>
</table>

3.2.1 AR(1) Simulation Study

We first study the outlier detection ability of the various standardization schemes for the AR(1) model. Recall (3.1):

$$X_k = \rho X_{k-1} + \varepsilon_k, \quad k = 1, \ldots, n.$$
This series is generated 1000 times for 144 different cases. The cases result from the combinations of parameter inputs. The coefficient term takes on the values $\rho \in \{\pm 0.1, \pm 0.5, \pm 0.9\}$. The error distribution is generated for six levels of contamination: $p_c \in \{0, .05, .10, .15, .20, .25\}$, and two levels of contaminated variance: $\sigma^2_c \in \{16, 100\}$. Finally, we consider two values of the sample size, $n = 25$ and $n = 100$. For each case, least squares and R-estimates are obtained. The residuals are then standardized in the variety of ways discussed in section 3.1. The results consist of classification rates. Each type of standardized residual, for each case, has an empirical type I and type II error rate, and an empirical measure of power, as pictured in table 1, calculated from the 1000 simulations.

The rates are illustrated in figures 1-4, with each plot showing all combinations of $\rho$ and $p_c$, and comparing the two levels of $n$. Figure 1 shows the empirical type I error rates for $\sigma^2_c = 16$, and figure 2 shows the type I error rates when $\sigma^2_c = 100$. It will be sufficient to include plots of the empirical power results but not type II error rates, as these are compliments of each other. Figures 3 and 4 show the empirical power results for $\sigma^2_c = 16$ and $\sigma^2_c = 100$, respectively.

The first aspect of figure 1 that stands out is the higher type I error rates when $n = 25$ for the R-estimates. This is not true for the least squares estimates. The exception is USLMLS when $\rho = 0.9$. The highest type I error is for SLMR, but it rarely exceeds .05, the traditional significance level. Note that the highest values for type I error occur when $p_c = .05$ or no contamination at all. In general,
Type I error rates for AR(1) simulation, contaminated variance is 16.
AR(1) Results for Type I Error, var=100

<table>
<thead>
<tr>
<th></th>
<th>var=100.0</th>
<th>var=100.0</th>
<th>var=100.0</th>
<th>var=100.0</th>
<th>var=100.0</th>
<th>var=100.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=25</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
</tr>
<tr>
<td>n=100</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
</tbody>
</table>

Figure 2

Type I error rates for AR(1) simulation, contaminated variance is 100.
Figure 3

Power rates for AR(1) simulation, contaminated variance is 16.
Figure 4

Power rates for AR(1) simulation, contaminated variance is 100.
for both least squares and R-estimates and both levels of $n$, the rates decrease as $p_c$ increases. Figure 2 shows the same behavior for the most part, with more pronounced differences. SLMR exceeds .05 for all values of $p_c$ when $\rho = \pm 0.9$. This is perhaps due to the parameter being close to the stationarity bound. So we are observing lower least squares type I error rates than for R-estimates, which holds true over all levels of $n$ and $\sigma^2$. Keeping the standard 5% significance level in mind, the rates based on R-estimation are still acceptable, however.

In figure 3, the power rates are greater for $n = 100$ when $p_c = 0$. This is more true for least squares than R-estimates. As $p_c$ increases, the $n = 25$ and $n = 100$ values move closer to each other for both LS and R. For $p_c > 0$, the LS power rates are all about the same. Similarly, the R power results cluster together. The residuals based on R-estimation consistently give higher power results than least squares, with one exception. When $\rho = \pm 0.9$, the STSR power is lower for $n = 25$. Proximity to the stationarity bound, along with the smaller sample size, could be the culprit. Similar observations are in order for figure 4. Overall for the AR(1) simulation, the residuals from R-estimation are superior to those from least squares. Empirical power rates range from about .7 to .8 for R, while LS power ranges from roughly .3 to .5. Caution should be taken in regard to possible type I misclassification with SLMR residuals when $\rho$ is close to $\pm 1$, $n$ is small, and the contaminated variance is large. Caution also should be taken with STSR residuals giving high type II error rates when $\rho$ is close to $\pm 1$ and $n$ is small.
3.2.2 AR(2) Simulation Study

Next we will explore the outlier detection success rates under a second order autoregressive model.

\[ X_k = \rho_1 X_{k-1} + \rho_2 X_{k-2} + \varepsilon_k, \quad k = 1, \ldots, n. \]  

(3.7)

Eight realizations of model (3.7) are considered, and are listed in table 2. Parameter values were chosen to cover a variety of possible AR(2) models, while ensuring stationarity. (See e.g. Abraham and Ledolter (1983) or Box, Jenkins and Reinsel (1994).) The simulation here employs the same levels of \( p_c, \sigma^2_e \) and \( n \) that were used before. Note that time series standardizations (STSLS and STSR) are yet to be derived for the AR(2) model, so these residual types are not included in the study.

Similar to the AR(1) simulation, figure 5 shows higher type I error rates

---

### Table 2

Parameter Values Used for AR(2) Simulation

<table>
<thead>
<tr>
<th>Model</th>
<th>( \rho_1 )</th>
<th>( \rho_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>B</td>
<td>-0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>C</td>
<td>-0.9</td>
<td>-0.5</td>
</tr>
<tr>
<td>D</td>
<td>-0.7</td>
<td>-0.3</td>
</tr>
<tr>
<td>E</td>
<td>0.1</td>
<td>0.8</td>
</tr>
<tr>
<td>F</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>G</td>
<td>0.3</td>
<td>-0.7</td>
</tr>
<tr>
<td>H</td>
<td>0.8</td>
<td>-0.5</td>
</tr>
</tbody>
</table>
AR(2) Results for Type I error, var=16

Figure 5

Type I error rates for AR(2) simulation, contaminated variance is 16.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
AR(2) Results for Type I error, var=100

Type I error rates for AR(2) simulation, contaminated variance is 100.
for the R-residuals than LS, with a noticeable separation between the two levels of $n$ for R. Smaller sample size results in higher type I error rates for R. SLMR is near .7 most of the time. The highest is USLMLS at .8 when $p_c = 0$. Note that the rates don't decrease as $p_c$ increases nearly as much as they did for the AR(1) simulation. Figure 6 looks almost the same as the previous plot with $\sigma^2_c = 16$, with differences among the R-residuals more obvious. SLMR attains a type I error of .08 quite often, especially for model E. The least squares residuals appear to win out over their R counterparts here.

The power results look quite a bit like they did for the AR(1) simulation. Figures 7 and 8 show the least squares residuals clumping together with less power than the R-residuals, which are also close to each other. Note that the power is consistent for all values of $p_c > 0$ for R, but the LS values decrease slightly as $p_c$ increases. The R-residuals win again in terms of power detecting outliers. The R power rates are greater than .7 almost always, and near .8 often. Least squares gives power rates falling between .3 and .5.

3.2.3 Higher Order AR Simulation Study

Lastly, we consider a few higher order autoregressive models of the form

$$X_k = \rho_1 X_{k-1} + \rho_2 X_{k-2} + \cdots + \rho_q X_{k-q} + \epsilon_k, \quad k = 1, \ldots, n.$$  (3.8)

We will generate two autoregressive models of order three, and one each of orders 4, 5, 6, 7, and 8. The parameter values, chosen such that all models are
**Figure 7**

Power rates for AR(2) simulation, contaminated variance is 16.
Figure 8

Power rates for AR(2) simulation, contaminated variance is 100.
stationary, are given in Table 3. The same levels of $p_c$ and $\sigma^2_e$ are used as before, but the sample sizes are larger here due to the larger orders of the models. The actual $n$ that ends up in the design is $n - q$, so we have increased $n$ to 50 and 200.

Figures 9 and 10 show the type I error rates for the higher order AR simulation, comparing the two levels of $n$. The least squares residuals are very consistent, with low type I error for all values of $p_c$. There is a slight increase noticeable when $n = 50$ for the higher order models. This is much more so for the R-residuals. When $n = 50$ we see higher type I error rates than for $n = 200$ in all cases, but especially as the order increases. Within each model, there is not much change as $p_c$ increases. The highest type I error rate observed here is for SLMR and USLMR, between .10 and .15 for model AR8. Least squares clearly is more conservative here, falsely declaring outliers much less often than their rank-based counterparts.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>$\rho_3$</th>
<th>$\rho_4$</th>
<th>$\rho_5$</th>
<th>$\rho_6$</th>
<th>$\rho_7$</th>
<th>$\rho_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR3a</td>
<td>0.2</td>
<td>0.45</td>
<td>-0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR3b</td>
<td>-0.3</td>
<td>0.64</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR4</td>
<td>0.4</td>
<td>0.96</td>
<td>-0.26</td>
<td>-0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR5</td>
<td>0.8</td>
<td>0.80</td>
<td>-0.64</td>
<td>-0.10</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR6</td>
<td>0.4</td>
<td>1.01</td>
<td>-0.32</td>
<td>-0.27</td>
<td>0.04</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR7</td>
<td>0.4</td>
<td>1.33</td>
<td>-0.53</td>
<td>-0.48</td>
<td>0.19</td>
<td>0.05</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>AR8</td>
<td>0.6</td>
<td>1.28</td>
<td>-0.74</td>
<td>-0.45</td>
<td>0.23</td>
<td>0.06</td>
<td>-0.02</td>
<td>-0.002</td>
</tr>
</tbody>
</table>
Type I error rates for Higher Order AR simulation, contaminated variance is 16.

Figure 9
Type I error rates for Higher Order AR simulation, contaminated variance is 100.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Higher Order AR Results for Power, var=16

Figure 11

Power rates for Higher Order AR simulation, contaminated variance is 16.
Higher Order AR Results for Power, var=100

<table>
<thead>
<tr>
<th>var 100.0</th>
<th>var 100.0</th>
<th>var 100.0</th>
<th>var 100.0</th>
<th>var 100.0</th>
<th>var 100.0</th>
<th>var 100.0</th>
<th>var 100.0</th>
<th>var 100.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>model AP2</td>
<td>model AP2</td>
<td>model AP3</td>
<td>model AP3</td>
<td>model AP4</td>
<td>model AP4</td>
<td>model AP4</td>
<td>model AP4</td>
<td>model AP7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 12

Power rates for Higher Order AR simulation, contaminated variance is 100.
Figures 11 and 12 show the empirical power rates. Not much difference is evident between the two sample sizes except for the LS residuals when $p_c = 0$. The smaller sample size yields less power. Again, like the previous studies, the power results cluster together among least squares, as do the R-residuals. The R-residuals are showing a superior ability to correctly detect outliers. There's not much change as $p_c$ increases among the R rates, but the LS power decreases somewhat. Not much change is noticeable as model order increases. Power results for the R-residuals are typically between .7 and .8, while the LS power varies between .3 and .4.

So we can conclude that LS is conservative, R is liberal. The LS residuals rarely make the mistake of falsely declaring outliers, but miss them often. The R-residuals excel at detecting outliers, but sometimes cry wolf. Not much can be said comparing residual types within LS, or within R. For the time series standardizations in the AR(1) study, the conclusion may be the following. STSLS and STSR are comparable, but not superior, to the conditional linear models versions in terms of outlier detection. Therefore, we suggest using the linear model residuals, since they are commonly available in many statistical software packages, and less computationally expensive.
CHAPTER IV

DIAGNOSTICS COMPARING ESTIMATES

4.1 Estimating Procedures

In addition to least squares and rank-based methods using Wilcoxon scores, there is another class of estimates of interest to us here. Instead of the dispersion function given in (1.2) that was minimized in order to obtain an R-estimate, consider the following variability measure:

\[ D(p) = \sum_{i<j} b_{ij} |\varepsilon_i - \varepsilon_j|, \quad 1 \leq i < j \leq n, \]  

(4.1)

where \( b_{ij} \) are weights for the \( (i, j)^{th} \) comparison. The weights depend on the design points. The purpose is to downweight outlying design points by making \( b_{ij} \) small. Estimates based on (4.1) were first proposed by Sievers (1983) and are called generalized rank, or GR, estimates. Note that for constant weights, i.e. \( b_{ij} \equiv 1 \), (4.1) reduces to (1.2) with Wilcoxon scores.

GR estimates were further developed by Naranjo and Hettmansperger (1994), showing that, for a special case of the weights, the estimates have bounded influence in both \( X \) and \( Y \) space, along with positive breakdown. Note that the R-estimates primarily focused on in this manuscript have bounded influence only in \( Y \) space.
Another special case of the weights that not only depend on the design points but also depend on the response result in another class of estimates. These high-breakdown (HBR) estimates, named for their 50% breakdown property, were proposed by Chang (1995) and were further developed by Chang et al (1999).

These estimates should prove to be quite robust in the presence of time series outliers, particularly "additive" (see definition in section 4.4.1) outliers that can adversely affect LS and R estimates.

A first order approximation of the standard error of GR residuals was derived by Naranjo et al (1994) in the linear models context. This gives analogous standardized residuals to SLMLS and SLMR, which will be denoted SLMGR. Similarly, Chang et al (1999) provides a first order approximation of the standard error of linear model HBR residuals. The standardized residuals based on this method are denoted SLMHBR.

4.2 TDBETAS Analysis

Different estimating procedures can give quite different fitted models, especially in the presence of outliers. This is true both in the linear models context and for time series data. For a given data set, fitting more than one estimate can reveal much about the behavior of the data, particularly when the fits differ substantially.

A measure to quantify the difference between fits has been developed by
McKean et al (1996a) and (1996b). In particular, these papers exploit the differences that can occur between the highly efficient Wilcoxon R-estimate and the highly robust GR-estimate for linear models.

Recall the no-intercept AR(1) model:

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

(4.2)

where again we condition on the design points in order to treat the autoregression like a lag one linear model. Suppose that both R and GR methods are used to produce estimates of $\rho$. A measure of the difference in fits (due to the down-weighting of high-leverage points) is $|\hat{\rho}_R - \hat{\rho}_{GR}|$. Of course, this difference needs to be appropriately standardized in order to be useful. Details of the asymptotic distribution of $|\hat{\rho}_R - \hat{\rho}_{GR}|$ can be found in McKean et al (1996a); just the results are reported here. Define

$$TDBETAS_{R,GR} = \frac{(\hat{\rho}_R - \hat{\rho}_{GR})^2}{\hat{\text{Var}}(\hat{\rho}_R)},$$

a statistic that measures the total difference in fits. $\hat{\text{Var}}(\hat{\rho}_R)$ is used in the denominator instead of the intuitive choice $\hat{\text{Var}}(\hat{\rho}_R - \hat{\rho}_{GR})$ since, for close estimates, this can tend to zero. A benchmark for $TDBETAS_{R,GR}$ was proposed by McKean et al (1996a) and is essentially an extrapolation of the benchmark of $2\sqrt{(p+1)/n}$ suggested by Belsley, Kuh and Welsch (1980) for $DFFITS_i$, a least squares diagnostic that measures the influence of the $i^{th}$ observation on its own fitted value. The proposed TDBETAS benchmark is $4(p + 1)^2/n$. This is meant to be used
informally; it is not a strict critical value as such. When high leverage outliers are present, the TDBETAS diagnostic can be quite large, sometimes exceeding the benchmark more than a hundredfold. When comparing several fits, one may wish to note large gaps in size that separate TDBETAS statistics among comparisons, in addition to using the proposed benchmark.

TDBETAS diagnostics that compare traditional least squares fits to R and GR fits are explored in McKean et al (1999). These ideas can be extrapolated to include comparisons with the HBR estimates as well, allowing us to compare fits (two at a time) among four estimation procedures: LS, R, GR and HBR. It has been suggested by the authors to use $\hat{\text{Var}}(\hat{\beta}_R)$ in the denominator for any TDBETAS comparison, even those that don’t involve the R-estimate.

4.3 Residential Extensions Data

As a motivating example, consider the monthly time series data from Bell Canada concerning the installation of residential telephone extensions in a fixed geographic area from January 1966 to May 1973. The time series, which can be found in Rousseeuw and Leroy (1987, pg. 280), consists of the number of residential extensions that were installed for each of the 89 months. The data set has two obvious outliers, occurring at November and December 1972. It turns out that installations of residential extensions were free in November 1972, and in fact more orders were placed than could be completed. These orders spilled
over into the next month, causing two outliers. These data were first analyzed by Brubacher (1974). The original series was found to have a yearly pattern, and thus was seasonally adjusted by differencing at lag 12. That is, the following series

\[ Y_t = X_{t+12} - X_t \]

was constructed, where \( X \) represents the original series, and \( Y \) represents the “seasoned” data. Brubacher found the seasoned data to be stationary with zero mean (except for the outliers). The outliers are at cases \( t = 71 \) and \( t = 72 \) for the seasoned data. This is seen quite clearly in figure 13, which shows the time series plot for the seasoned residential extensions data. For our example here, we will fit an AR(1) model, given by (1.1), to these data using our four estimation procedures. This design results in large leverage values for the two outlying observations: \( h_{71} = 0.745 \) and \( h_{72} = 0.196 \), while \( h_t \leq 0.020 \) for all other cases. The two outliers cause three \( (Y_{t-1}, Y_t) \) pairs of interest, namely \( (Y_{70}, Y_{71}) = (1215, 54671) \), \( (Y_{71}, Y_{72}) = (54671, 28619) \), and \( (Y_{72}, Y_{73}) = (28619, 2478) \). As can be seen in the lag one scatterplot of figure 13, \( (Y_{70}, Y_{71}) \) is an outlier in \( Y \) space but not \( X \) space, and hence is not a high leverage point. \( (Y_{71}, Y_{72}) \) is an outlier in both dimensions, and does not lie far off the path of the main cluster of points. This point is considered a “good” leverage point. \( (Y_{72}, Y_{73}) \) is an outlier in factor space only, and is considered a “bad” leverage point.

Table 4 gives the estimates for the four fits. Note that the legend for figure 13 presents the labels for the four fitted lines in a top-to-bottom manner as
The GR fit seems to model the main cluster of points best, while ignoring all of the outliers. The R fit is obviously drawn to the good leverage point, and is slightly dragged down by the bad leverage point. The HBR fit seems to be drawn to the good leverage point, but is more likely simply modeling the core of the data cluster. The LS fit is clearly dragged down by the bad leverage point, and possibly has a higher intercept than the other three fits due to \((Y_{70}, Y_{71}) = (1215, 54671)\).

Figure 14 shows a close-up of the scatterplot without the three outliers. This allows us to more closely examine how well each of the four fits follow the main cluster of points. Note that the legend for this plot presents the labels for the fits in a top-to-bottom manner as LS, R, HBR, and GR, as one looks at the left side of the graph. Again, the GR fit looks best, and the LS fit is off. The R and HBR fits are almost indistinguishable. Indeed, the TDBETAS analysis backs this up.

Table 5 contains the TDBETAS values for the residential extensions data. Note that all TDBETAS values exceed the benchmark except the R vs. HBR comparison. So according to the TDBETAS analysis, all the fits differ except for R and HBR. The benchmark for this example is \(4(p + 1)^2/n = 4(2 + 1)^2/76 = 0.4737\), where \(p\) is the order of the design of the autoregression, namely one for the order of the AR plus one for the intercept.

It is perhaps worthwhile here to point out a key difference between tradi-
Figure 13

Time series plot and lag 1 scatterplot with fits for the seasoned Residential Extensions data.
Figure 14. Lag 1 scatterplot with fits for the seasoned Residential Extensions data. Only the main cluster of points is pictured.

Table 4

Parameter Estimates for Residential Extensions Data

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\phi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>1587.51</td>
<td>0.406045</td>
</tr>
<tr>
<td>R</td>
<td>786.895</td>
<td>0.497351</td>
</tr>
<tr>
<td>GR</td>
<td>599.872</td>
<td>0.584708</td>
</tr>
<tr>
<td>HBR</td>
<td>753.577</td>
<td>0.509285</td>
</tr>
</tbody>
</table>
Table 5

*TD BETAS* Values for Residential Extensions Data

<table>
<thead>
<tr>
<th></th>
<th>LS</th>
<th>R</th>
<th>GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>20.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td>56.3</td>
<td>12.0</td>
<td></td>
</tr>
<tr>
<td>HBR</td>
<td>24.1</td>
<td>0.22</td>
<td>8.97</td>
</tr>
</tbody>
</table>

† Did not exceed benchmark = 0.4737

Traditional least squares analysis and more robust methods. The typical goal behind fitting a robust procedure is to model the majority of the data well, while ignoring outliers. Flagging those outliers for closer inspection is equally important. The statistician’s agenda is not to delete outlying points, but rather to focus on them and possibly determine their cause. Robust procedures aim to achieve the best of both worlds in the sense that the model is not unduly affected by a few high leverage points, and the diagnostics (residuals) draw attention to interesting, if troublesome, observations. Classic least squares methods may often produce the worst case of this scenario, because not only can the fit get thrown off by outliers, but this goes hand-in-hand with the diagnostics not catching those outliers. If the outliers are part of the fit, then the resulting residuals won’t be large. In this way they are “masked” by the LS analysis.

So with this in mind, let’s look at the residuals for the four fits, and see which residuals were “flagged”. Figure 15 shows the residuals over time, with benchmark lines drawn at ±2. The standard linear model studentized residuals are
shown for the four fits, along with the time series standardizations for LS and R, as well as the approximate linear model standardization for the R-residuals. Note that there are 76 pairs of data in this design, so there are 76 residuals plotted. The 70th residual corresponds to \((Y_{70}, Y_{71})\), the 71st residual corresponds to \((Y_{71}, Y_{72})\), and the 72nd residual corresponds to \((Y_{72}, Y_{73})\). We would expect the 70th and 72nd residuals to stand out. The SLMLS and STSLS residual plots look quite similar. At \(t = 70\), there is a clear outlier. The residual at \(t = 72\) stays within the benchmark bounds. The SLMR, STSR, SLMRA, SLMGR and SLMHBR residual plots are very comparable. The outlier at \(t = 70\) is apparent, as is the large negative residual at \(t = 72\). These two residuals stand out clearly, but a closer look reveals two more residuals that only slightly exceed the benchmark at \(t = 63\) and \(t = 64\). Recall from Chapter III that the R-residuals were liberal, declaring outliers falsely some of the time. This may be what we’re seeing here, since we have no knowledge of outliers in the time series other than the ones previously mentioned. More importantly, recall that the residual at \(t = 72\) corresponds to the \((Y_{72}, Y_{73})\) pair, the bad leverage point. Since the LS fit is somewhat drawn to that point, it doesn’t result in a flagged residual.
Figure 15

Residuals over time for the Residential Extensions data.
4.4 Outliers in Autoregressive Time Series

4.4.1 Types of Outliers

Consider a designed experiment that is modeled with linear regression. The design points are fixed; outliers can only occur in the response. Observational studies introduce the possibility of outliers in the (now random) design points. Outliers in the design can have high leverage. Time series data are by definition response, with time acting as the explanatory variable. Of course, we are likely to model such data with an autoregression, which means that any outlying points will appear in both $X$ and $Y^*$ space.

Outliers can be introduced into a time series in at least two ways. Suppose that an outlier enters the series through the error distribution, a situation where a single "innovation" is extreme. It affects not only that particular observation but subsequent observations as well. Innovation outliers (see Fox (1972)), denoted hereafter as IO, occur when the error distribution is heavy tailed, e.g. a contaminated normal. If the error distribution is not heavy tailed, we will still for convenience' sake refer to the model as IO, where the innovations are simply white noise. The effect of an isolated IO lingers in a manner reflective of the autoregressive structure. Eventually, the series returns to its "normal" level. IOs often result in good leverage points in the AR design.

An outlier that is more "artificially" entered into the series, possibly due to a data recording error, one-time only event or other outside influence, such that
a single observation is affected only, is called an additive outlier (see Denby and Martin (1979)), or AO. An isolated AO does not affect future observations. The series returns to its previous state immediately. This transient nature of the AO is unrelated to the autoregressive structure. AO effects are "added" to the core process, and can give rise to bad leverage points. AOs often occur in "patches". Each AO may not be affecting the next, but the outside influence responsible may be present for more than one aberration. This will likely result in both good and bad leverage points.

4.4.2 Generated Examples

In this section, we will simulate an AR(1) time series under a few different outlier schemes. 50 observations were recorded from an AR(1) process with $\rho = 0.5$. Thus we have the following specific realization of model (4.2):

$$X_i = 0.5X_{i-1} + \varepsilon_i, \quad i = 1, \ldots, 50.$$

All four of our estimation procedures (LS, R, GR and HBR) will be fit to the data produced under the various outlier situations. Briefly, the outlier models are the following: case (a): no outliers, case (b): IO, case (c): isolated AO, and case (d): AO patch. Figure 16 shows the time series plots for each of the cases, and figure 17 shows the lag one scatterplots, with the fits overlaid. Table 6 gives the parameter estimates, and table 7 contains the TDBETAS comparing the fits. The benchmark for the TDBETAS statistic here is $4(p + 1)^2/n = 4(2 + 1)^2/49 = 0.73469$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Case (a) is a “well-behaved” time series with no outliers. It was generated with “white noise” innovations from a \( N(0, 1) \) distribution. No outliers in the time series implies no outliers in the lag one scatterplot, hence the four fits are indistinguishable. Note that the TDBETAS values are all quite small, and less than the benchmark.

Case (b) utilized a heavy tailed error distribution, specifically a contaminated normal. This resulted in an IO at \( t = 45 \), with the series returning to its “natural” level by \( t = 48 \). This causes outliers in the design at points \((X_{44}, X_{45}) = (-0.94, 18.2)\), an outlier in response space only, not a high leverage point; \((X_{45}, X_{46}) = (18.2, 10.9)\), a good leverage point; \((X_{46}, X_{47}) = (10.9, 5.14)\), another good leverage point; and \((X_{47}, X_{48}) = (5.14, 2.51)\), which could be considered a good leverage point or not outlying at all. R, GR, and HBR all have similar intercepts, and differ slightly in terms of slope. The TDBETAS analysis tells us that only the GR and HBR fits differ among these three estimates. The IO results in a couple of good leverage points. The R-fit seems to be modeling these better than GR or HBR. LS has a larger intercept and smaller slope than the more robust fits, presumably a result of the outlier at point \((X_{44}, X_{45}) = (-0.94, 18.2)\). The LS fit differs from all three of its robust counterparts according to the TDBETAS statistics. Note that the legend for figure 17, case (b) presents the labels for the four fitted lines in a top-to-bottom manner as HBR, R, GR, and LS, if one looks at the right side of the graph.
Case (c) shows an isolated AO in the time series at \( t = 46 \). This causes outliers in design at \((X_{45}, X_{46}) = (-1.32, -22.3)\), a bad leverage point; and \((X_{46}, X_{47}) = (-22.3, -3.00)\), an outlier in response space only, not a high leverage point. The legend for case (c) in figure 17 presents the labels for the fits top-to-bottom as R, LS, HBR, and GR, if one looks at the left side of the graph. All four fits differ, according to the TDBETAS values. The R and LS fits are both drawn to the high leverage point. The GR and HBR fits are clearly not affected by either outlier and are modeling the main cluster of data much better than R. Note that the LS fit is shifted down a bit from the R fit, no doubt influenced by the outlier in response space at \((X_{46}, X_{47}) = (-22.3, -3.00)\).

Case (d) features a patch of AOs in the range from \( t = 26 \) to \( t = 31 \). Not every point in this range is an outlier, but \( t = 26, 27, \) and 31 are all big jumps from the previous point in time. This patch results in several outliers in the design, about nine that lie apart from the main cluster of points. The legend for case (d) in figure 17 lists the fits as GR, HBR, R, and LS from top-to-bottom, looking at the right side of the graph. All fits differ according to the TDBETAS measure. The LS attempt to model all the points misses the increasing relationship of the majority and gives a negative slope! The R-fit, while a big improvement over LS, has a flatter slope, presumably affected by the bad leverage point at the top left of the graph. Probably the only saving grace of the R-fit is the good leverage point near the top right. The GR fit seems to model the main cluster of points best.
The HBR fit may be even more focused on the group of data at the center of the graph.

Overall, the TDBETAS measure shows differences in fits in the presence of AOs. The IO does result in some differences among fits as well, but the good leverage points seem to keep the R-fit nearby the high-breakdown estimates.

Let’s revisit the residential extensions example for a moment. Recall the two outliers in the time series, due to a bargain month. We could think of this as a (small) patch of two AOs, or an IO that fixes itself rapidly and has returned to the previous level after only two observations. Either way, we end up with one good leverage point, one bad leverage point, and one outlier in $Y$ space only. The TDBETAS statistics showed differences in all fits except R and HBR. This may be a coincidence: the concentrated center of points modeled by HBR just happens to match up with the good leverage point that R is attracted to.

4.5 A Monte Carlo Study

We now turn our attention to a larger scale simulation study involving TDBETAS comparisons under various outlier models. The first “outlier” model is actually the no outlier model, with normal errors. This model is denoted $IO_1$. Next we generate errors from a heavier tailed logistic distribution, which will result in some IOs. This model is called $IO_2$. Thirdly, more IOs are obtained from a contaminated normal distribution with $\sigma_c^2 = 16$, and $p_c = .10$. This case, $IO_3$, is
Figure 16. Time series plots for four outlier cases.
Figure 17

Lag 1 scatterplots with fits for four outlier cases.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 6

Parameter Estimates for Generated Examples

<table>
<thead>
<tr>
<th>Example</th>
<th>Fit</th>
<th>( \hat{\alpha} )</th>
<th>( \hat{\beta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) No Outliers</td>
<td>LS</td>
<td>0.103</td>
<td>0.617</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.092</td>
<td>0.640</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>0.093</td>
<td>0.651</td>
</tr>
<tr>
<td></td>
<td>HBR</td>
<td>0.094</td>
<td>0.659</td>
</tr>
<tr>
<td>(b) IO</td>
<td>LS</td>
<td>0.482</td>
<td>0.484</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.015</td>
<td>0.560</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>-0.010</td>
<td>0.528</td>
</tr>
<tr>
<td></td>
<td>HBR</td>
<td>0.008</td>
<td>0.585</td>
</tr>
<tr>
<td>(c) Isolated AO</td>
<td>LS</td>
<td>-0.828</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>-0.379</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>-0.269</td>
<td>0.399</td>
</tr>
<tr>
<td></td>
<td>HBR</td>
<td>-0.299</td>
<td>0.319</td>
</tr>
<tr>
<td>(d) AO Patch</td>
<td>LS</td>
<td>0.492</td>
<td>-0.098</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>-0.094</td>
<td>0.325</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>-0.174</td>
<td>0.522</td>
</tr>
<tr>
<td></td>
<td>HBR</td>
<td>-0.068</td>
<td>0.407</td>
</tr>
</tbody>
</table>

Table 7

TDBETAS Values for Generated Examples

<table>
<thead>
<tr>
<th>(a) No Outliers</th>
<th>(b) IO</th>
<th>(c) Isolated AO</th>
<th>(d) AO Patch</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS vs. R</td>
<td>0.038</td>
<td>6.189</td>
<td>11.79</td>
</tr>
<tr>
<td>LS vs. GR</td>
<td>0.084</td>
<td>6.236</td>
<td>46.53</td>
</tr>
<tr>
<td>LS vs. HBR</td>
<td>0.128</td>
<td>7.287</td>
<td>25.37</td>
</tr>
<tr>
<td>R vs. GR</td>
<td>0.012</td>
<td>0.430</td>
<td>51.90</td>
</tr>
<tr>
<td>R vs. HBR</td>
<td>0.033</td>
<td>0.222</td>
<td>25.48</td>
</tr>
<tr>
<td>GR vs. HBR</td>
<td>0.005</td>
<td>1.248</td>
<td>4.652</td>
</tr>
</tbody>
</table>

Benchmark = 0.73469
considered a "mild" CN. A "severe" CN is generated with $\sigma^2 = 100$, and $p_c = .25$, called $IO_4$.

The remaining cases will feature AOs. These are generated with a contaminated normal distribution, but unlike the IO cases, here we also incorporate a probability that essentially dictates how long the "patch" of AOs stays in effect. Case $AO_1$ is considered a mild case of isolated AOs, with $\sigma^2 = 16$, and $p_c = .10$. The next outlier model, $AO_2$, is a severe case of isolated AOs, with $\sigma^2 = 100$, and $p_c = .25$. Next, we generate the mild case of AOs, but include a probability of 80% that the AOs will continue to be introduced into the series once they start. Call $AO_3$ the mild AO patch case. Finally, the severe AO case is generated with the same 80% probability for the patch. $AO_4$ is the severe AO patch model.

In all eight cases, our usual no-intercept AR(1) model is generated with $\rho = 0.5$, and $n = 100$. For 1000 simulations, the TDBETAS statistic is calculated to measure the pairwise differences between our four estimation methods. The frequency that TDBETAS exceeded the benchmark out of the 1000 simulations was recorded, and these numbers appear in table 8. The benchmark here is $4(p + 1)^2/n = 4(2 + 1)^2/100 = 0.36$.

Before we focus on the TDBETAS > benchmark frequencies, consider figure 18. Boxplots are shown of the 1000 estimates for each case, in order to illustrate the estimate bias and differences between procedures. The performance of the four estimation methods are quite similar for $IO_1$, $IO_2$, and $IO_3$. The median
\( \hat{\rho} \) value is a bit lower than \( \rho = 0.5 \), even for normal errors, but this may be due to the time series generator. The median \( \hat{\rho} \) value is significantly low for all four AO models. In fact, it's closer to zero for \( AO_2 \), \( AO_3 \), and \( AO_4 \). The variability between estimates is quite similar for \( IO_1 \), \( IO_2 \), \( IO_3 \), \( AO_1 \), and \( AO_3 \). In both of the isolated AO cases, the GR estimates are less biased.

Getting back to the TDBETAS frequencies, consider the comparisons involving least squares. In general, we can say that there are more TDBETAS differences in fits as the models get messier, i.e., from the no outlier model(\( IO_1 \)) to the heavier tailed logistic(\( IO_2 \)), from mild(\( IO_3 \)) to severe CN(\( IO_4 \)), and from mild(\( AO_1 \)) to severe isolated AOs(\( AO_2 \)). Note that for the IOs resulting from the CN distribution, TDBETAS declares differences in fits more often between R and GR than for R vs. HBR. This is perhaps due to what we saw before with the generated examples. The IO produced a good leverage point, this gave R an "advantage" over GR, with R and HBR showing no difference according to TDBETAS. Note further that for all four cases of AOs, TDBETAS shows differences in fits between R and GR (and R vs. HBR) more frequently than for GR vs. HBR. This could be in line with the generated examples also; the GR and HBR fits model the majority of the data, while the R-fit is drawn to bad leverage points.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 18. Boxplots of the 1000 estimates of $\rho = 0.5$ with a reference line drawn to check bias.

Table 8

Frequencies out of 1000 that TDBETAS exceeded the benchmark

<table>
<thead>
<tr>
<th>Model</th>
<th>LSvs.R</th>
<th>LSvs.GR</th>
<th>LSvs.HBR</th>
<th>Rvs.GR</th>
<th>Rvs.HBR</th>
<th>GRvs.HBR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$IO_1$</td>
<td>331</td>
<td>374</td>
<td>497</td>
<td>33</td>
<td>70</td>
<td>126</td>
</tr>
<tr>
<td>$IO_2$</td>
<td>417</td>
<td>467</td>
<td>579</td>
<td>81</td>
<td>86</td>
<td>159</td>
</tr>
<tr>
<td>$IO_3$</td>
<td>693</td>
<td>735</td>
<td>784</td>
<td>359</td>
<td>210</td>
<td>372</td>
</tr>
<tr>
<td>$IO_4$</td>
<td>962</td>
<td>969</td>
<td>972</td>
<td>523</td>
<td>369</td>
<td>560</td>
</tr>
<tr>
<td>$AO_1$</td>
<td>687</td>
<td>947</td>
<td>880</td>
<td>881</td>
<td>603</td>
<td>567</td>
</tr>
<tr>
<td>$AO_2$</td>
<td>954</td>
<td>993</td>
<td>956</td>
<td>949</td>
<td>570</td>
<td>910</td>
</tr>
<tr>
<td>$AO_3$</td>
<td>823</td>
<td>931</td>
<td>945</td>
<td>764</td>
<td>737</td>
<td>628</td>
</tr>
<tr>
<td>$AO_4$</td>
<td>885</td>
<td>921</td>
<td>945</td>
<td>656</td>
<td>598</td>
<td>649</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CHAPTER V

CONCLUSIONS

5.1 Summary of Work Presented

Several different methods of standardizing residuals from an autoregression were considered in this manuscript, including newly developed time series specific studentized residuals. Simulations were undergone to explore the outlier detection ability of the various residuals. The time series residuals, while more mathematically appropriate than their conditional linear models counterparts, did not show any advantage in practice. Since the time series residuals are somewhat computationally complex, it is therefore recommended to simply make use of the linear model residuals commonly included in most any statistical software package.

5.2 Further Research

5.2.1 Extensions of Results

While the simulation studies did not produce encouraging results regarding the performance of the time series residuals, it nonetheless remains to consider further derivations. The large sample approximation for the variance of time series R-residuals could certainly be extended to autoregressive models of order $p$. Also,
time series specific standardized residuals under other estimation methods like GR and HBR might be useful. The above extensions could also be made to the more general autoregressive moving average ARMA\((p,q)\) time series model. Applications to other disciplines where time series data arises may also be explored, as in the next section.

5.2.2 Application to Quality Control Charts

Control charts continue to be a widely used and effective technique for statistical process control, or SPC. Process here refers to any combination of technology, machinery, materials, and manpower employed to attain particular goods or services. Control is a term used to describe the desired state of achieving set standards. A common assumption in Shewhart (1986) or other traditional types of control charts is that observations from the process at different times are independent random variables. This independence assumption is often not reasonable for the processes that arise in many industries. Process observations are frequently autocorrelated, which can have an estimable impact on a control chart that is developed under the independence assumption. If traditional control chart methodology is applied to autocorrelated processes, it typically decreases the in-control average run length (ARL), which can result in biased estimates of parameters, high false out-of-control signal rates, and slow detection of changes in the process (Reynolds & Lu, 1997). The ARL is the usual measure of control
chart performance. It is the expected number of observations until a signal. For an in-control process, any signal on the chart is a false alarm. In this case, a large ARL implies a low false alarm rate, and thus is desirable. Alternately, when there is a significant change in the process, a low ARL is desired so that the shift will be detected quickly.

The "false alarms" of control charts correspond to what we referred to in Chapter III as a type I error. Essentially, an outlier, or out-of-control point, is declared falsely. This may result in an unnecessary pause in the production process, a counter-productive adjustment to the machinery, or other such costly mistake. If autocorrelation is encountered in the process, as will be the case concerning time series data from an autoregressive model, traditional control chart methodology is not recommended (Lu & Reynolds, 1999). Two approaches are generally considered. First, standard charts are used with control limits that have been adjusted in order to account for the autocorrelation. The second method is to fit a time series model to the autocorrelated process, then apply a control chart to the resulting residuals. This latter approach may lead to an interesting application of the time series studentized residuals developed in Chapter II. Exactly how the residuals from the autoregression are standardized may affect the resulting control chart.

Another aspect to consider here is the type of outliers, or out-of-control signals, that may occur in the process. Booth et al (1990) suggest that outlier
type can be related to particular production problems. An in-control process is said to have random, or common cause, variation. This forms a stationary time series. Conversely, assignable, or special cause, variation in a control chart marks an out-of-control process. Points outside of the control limits can differ much in the manner of the IO and AO outlier types discussed in Chapter IV. The one-time-only effect of an AO can be interpreted as a bad production batch. This is perhaps the result of human error, and subsequent batches are not affected. On the other hand, the persistent nature of an IO, where several batches or observations are outside the control limits, is more likely a continuing process problem. Differentiating between type of outlier could be invaluable to the quality manager responsible for the locating and fixing the problem.
REFERENCES


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


