



Graduate Theses, Dissertations, and Problem Reports

2002

Three essays on data contaminants, outliers and macroeconomic time series

Joseph Michael Palardy
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Palardy, Joseph Michael, "Three essays on data contaminants, outliers and macroeconomic time series" (2002). *Graduate Theses, Dissertations, and Problem Reports*. 1686.
<https://researchrepository.wvu.edu/etd/1686>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Three Essays on Data Contaminants, Outliers and
Macroeconomic Time Series

Joseph Michael Palardy

Dissertation
submitted to the
College of Business and Economics
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Economics

Ronald Balvers, Ph.D., Chair
Douglas Mitchell, Ph.D.
Subhayu Bandyopadhyay, Ph.D.
Stratford Douglas, Ph.D.
George Hammond, Ph.D.

Department of Economics

Morgantown, West Virginia
2002

Abstract

Three Essays on Data Contaminants, Outliers and Macroeconomic Time Series

Joseph Michael Palardy

This thesis illustrates the effects of outliers on forecasting and on simulated general equilibrium models. The first essay demonstrates that using robust-estimation techniques may greatly improve long-run autoregressive forecasts if the model to be estimated is generated from a heavy-tailed distribution. The second essay continues with a similar theme. It shows that in the context of random contamination models, automatic autoregressive selection criteria may be severely inaccurate. If the estimated model is misspecified, forecasting accuracy may be considerably reduced. Once again, robust-estimation techniques limit the influence of the aberrant data – in this case from the contaminating distribution. The third essay demonstrates that calibrations, and thus simulation results, for International Real Business Cycle models can change if aberrant data is accounted for by using robust-estimation techniques. In the particular case examined in this essay, there is more evidence for higher levels of persistence in technology than indicated by OLS.

In memory of my advisor, Jon Vilasuso.

Table of Contents

Table of Contents	iv
List of Tables	vi
List of Figures	viii
1 Introduction	1
2 An Introduction to Outliers, Contaminants, and Extreme Values	5
2.1 What are Outliers, Contaminants, and Extreme Values	5
2.2 Outliers in the Context of Models	7
2.3 Contaminants in Time Series	8
2.3.1 Standard specification	9
2.3.2 Tsay's specification	11
2.3.3 Generating Time Series Outliers	12
2.4 Conclusion	16
3 Forecasting Heavy-Tailed Autoregressive Models	17
3.1 Introduction	17
3.2 Robust Estimation	21
3.3 Monte Carlo Simulations	27
3.3.1 Relation to Forecasts	30
3.4 Interest Rate Forecasts	31
3.5 Conclusion	34
4 Autoregressive Lag Selection Criteria and Random Contamination Models	42
4.1 Introduction	42
4.2 Innovation vs. Additive Contaminants	44
4.3 Model Selection Criteria	47
4.4 Effects of Contamination on Selection Criteria	50
4.5 Robust Selection Criteria	53
4.6 Simulations	56
4.7 Simulation Results	57

4.8	Example with Oil Prices	62
4.9	Conclusions and Directions for Future Work	64
5	International Real Business Cycle Models and Decisions before Shocks	86
5.1	Introduction	86
5.2	Previous Literature	90
5.2.1	International Real Business Cycles	90
5.2.2	Outliers in Macroeconomic Data	92
5.3	General Model	94
5.3.1	Decision Before Shock	97
5.3.2	Calibration and Solution Method	99
5.4	Results	100
5.5	Modifications and Sensitivity	103
5.5.1	Habit Persistence Model	103
5.5.2	Solow Residual Sensitivity	104
5.5.3	MM-based Technology Specification	107
5.5.4	Elasticity of Substitution	108
5.6	Conclusions	109
A	Technical Appendix for Chapter 5	121
A.1	The Model	121
A.2	Deterministic Steady State	122
A.3	Linear Quadratic Form	126
A.3.1	Quadratic Approximation	126
A.3.2	Equation of Motion	127
A.4	Solution	128
B	An Outlier Robust Estimation Package for OX	131
B.1	Program robest.ox	131
B.2	Program rscale.ox	135
B.3	robest.ox code	137
B.4	rscale.ox code	156
	Bibliography	171

List of Tables

3.1	Percent Bias of Estimated Autoregressive Parameters, n=250	37
3.2	Percent Bias of Estimated Autoregressive Parameters, n=500	38
3.3	Federal Funds Rate Unit Root Test	39
3.4	Robust Federal Funds Rate Unit Root Test	39
3.5	AR(3) Parameter Estimates for Federal Funds Rate Changes	40
3.6	AR(3) Out-of-sample Federal Funds Rate Changes Forecast Summary	41
4.1	Parametrization of Innovation Contamination and Additive Contam- ination Models	68
4.2	Contamination Types	68
4.3	Distributions of Selection Criteria for Model 1: $X_t - 2 - .5X_{t-1} = \varepsilon_t$ with Innovation Contamination	69
4.4	Distributions of Selection Criteria for Model 1: $X_t - 2 - .5X_{t-1} = \varepsilon_t$ with Additive Contamination	70
4.5	Distributions of Selection Criteria for Model 2: $X_t - 2 - .9X_{t-1} = \varepsilon_t$ with Innovation Contamination	71
4.6	Distributions of Selection Criteria for Model 2: $X_t - 2 - .9X_{t-1} = \varepsilon_t$ with Additive Contamination	72
4.7	Distributions of Selection Criteria for Model 3: $X_t - 2 - .3X_{t-1} -$ $.1X_{t-2} = \varepsilon_t$ with Innovation Contamination	73
4.8	Distributions of Selection Criteria for Model 3: $X_t - 2 - .3X_{t-1} -$ $.1X_{t-2} = \varepsilon_t$ with Additive Contamination	74
4.9	Distributions of Selection Criteria for Model 4: $X_t - 2 - .5X_{t-1} -$ $.35X_{t-2} = \varepsilon_t$ with Innovation Contamination	75
4.10	Distributions of Selection Criteria for Model 4: $X_t - 2 - .5X_{t-1} -$ $.35X_{t-2} = \varepsilon_t$ with Additive Contamination	76
4.11	Forecasting Costs of Misspecification for Model 1: $X_t - 2 - .5X_{t-1} = \varepsilon_t$ Under Various Contaminations	77
4.12	Forecasting Costs of Misspecification for Model 2: $X_t - 2 - .9X_{t-1} = \varepsilon_t$ Under Various Contaminations	78
4.13	Forecasting Costs of Misspecification for Model 3: $X_t - 2 - .3X_{t-1} -$ $.1X_{t-2} = \varepsilon_t$ Under Various Contaminations	79
4.14	Forecasting Costs of Misspecification for Model 4: $X_t - 2 - .5X_{t-1} -$ $.35X_{t-2} = \varepsilon_t$ Under Various Contaminations	80

4.15	Oil Price Unit Root Test	81
4.16	Oil Price Unit Root Test	82
4.17	Forecasting Changes in Oil Prices with Alternative Selection Criteria	83
4.18	Autoregressive Parameter Estimates for Changes in Real Oil Prices .	84
4.19	Performance Summary of the Various Selection Criteria	85
5.1	Calibration Parameters	117
5.2	Output Variation and Correlations with Output	118
5.3	Cross-Country Correlations and Country-Specific Persistence Levels .	119
5.4	Comparison of OLS and MM VAR estimates for Solow Residuals . . .	120

List of Figures

3.1	Histogram of Federal Fund Rate Changes, 1984-1998	36
4.1	Theoretical Parameter Estimates for Additive Contaminant Model . .	67
5.1	Response of the Home Country to a 1% Increase in the Home Country's Technology: Benchmark Model	111
5.2	Response of the Foreign Country to a 1% Increase in the Home Country's Technology: Benchmark Model	112
5.3	Response of the Home Country to a 1% Increase in the Home Country's Technology: DBS Model	113
5.4	Response of the Foreign Country to a 1% Increase in the Home Country's Technology: DBS Model	114
5.5	Response of the Home Country to a 1% Increase in the Home Country's Technology: DBS Model with Habit Formation	115
5.6	Response of the Foreign Country to a 1% Increase in the Home Country's Technology: DBS Model with Habit Formation	116

Chapter 1

Introduction

Outliers are the bane of statisticians. It is not surprising that the body of research on the effects of outliers and how to handle them is very large. Despite the size of the literature there are still several avenues that need to be explored. One of those avenues is forecasting economic time-series. The field of economics has yet to fully embrace the outlier-robust techniques developed in theoretical statistics. For this reason, many of the statements and inferences made about economic forecasts could be incorrect or misleading if outliers have large effects on those forecasts. Another possible avenue to explore is the effect of outliers on simulated general equilibrium models. These types of models are sensitive to certain calibration parameters that might be incorrectly estimated using traditional statistical techniques instead of more modern outlier-robust techniques. The goal of this thesis is to illustrate the effects of outliers on forecasting and on simulated general equilibrium models.

The second chapter provides a brief introduction to outliers in time-series. It

discusses the definitions of outliers and related events such as contamination and extreme values. Also, this chapter examines the difference between innovation and additive contaminants. Finally, it discusses typical outlier generating models including the widely-used mixture models.

The third chapter contains the first essay which uses heavy-tailed distributions to generate innovation outliers in time-series data.¹ It then evaluates the performance of alternative estimation techniques in the context of forecasting. Ordinary least squares (OLS) poorly estimates the constant term in the simulated models. This is detrimental to long-run forecasting performance because the constant term is equal to the unconditional mean, to which long-term autoregressive (AR) forecasts converge. The robust estimation-techniques provide superior forecasts to OLS in all but the short run.

The fourth chapter, or the second essay, also evaluates forecasting performance of various estimation techniques. Specifically, it addresses automatic autoregressive model selection techniques. In addition, this essay looks at both innovation type contaminants and additive contaminants by using a random-contamination model for the simulations. The MM-estimator that performed well in the first essay is used to produce contaminant-robust versions of the automatic selection criteria. This essay finds that the accuracy of automatic selection criteria is severely affected by additive contaminants. Innovation contaminants, however, have little effect on the accuracy of the selection criteria. In addition, this essay finds that the selection

¹This essay was originally joint work with advisor, Jon Vilasuso.

criteria based on the MM-estimator perform better than those based on OLS. The poor accuracy of the OLS-based selection criteria affects forecasting accuracy. The mistake of under-selecting the lag order, which occurs in the additive contaminant case, greatly decreases forecasting accuracy.

The fifth chapter is the third essay. It takes a different approach to evaluating the influence of outliers on time-series. This time, outliers are studied in the context of an International Real Business Cycle Model (IRBC). The simulation results of this type of model are highly dependent on the structure of technological innovations that drive the theoretical economy. It is typical to model technology by way of a vector autoregression (VAR) of the Solow residuals. Unfortunately, aberrant data in the output or labor series that are used to generate the Solow residuals may lead to incorrect VAR point estimates. This essay shows that a slightly modified IRBC model with alternative VAR point estimates may generate moments that are more in line with real economic data. More specifically, an IRBC model that incorporates decisions-before-shocks and habit-persistence along with low levels of persistence of technology generates results that are substantially better than those of the benchmark IRBC. A model with high levels of persistence of technology and a small elasticity of substitution between foreign and domestic goods also performs well. Estimating the VAR model via an outlier-robust estimation technique increases the estimated level of persistence of technological innovations.

In conclusion, accounting for contaminants and outliers is important in macroeconomics. This importance is demonstrated for model selection, forecasting, and simulated general equilibrium models. Failing to account for contaminants and outliers may lead to incorrect model specifications, less accurate forecasts, and incorrect model calibrations.

Chapter 2

An Introduction to Outliers, Contaminants, and Extreme Values

2.1 What are Outliers, Contaminants, and Extreme Values

Barnett and Lewis (1994) give two basic definitions for outliers. An *outlier* is 1) an observation that appears to be inconsistent with the remainder of that set of data; 2) an observation that is both surprising – does not appear to fit the data – and meets the criteria of a statistical test for being discordant (i.e. being inconsistent with the rest of the data). An example of the second definition would be labelling any observation that lies outside of two and a half standard deviations from the mean as an outlier. The second definition is an operational subset of the first: it is plausible that an observation that appears to be inconsistent with the data may not meet a statistical criterion to be called an outlier in the second definition. Unless otherwise noted, the term *outlier* will refer to the second definition.¹

¹The exact criteria used will not necessarily be plus or minus two and a half standard deviations.

The terms outlier, contaminant, and extreme value may appear to be synonymous; however, they are not. Important distinctions need to be made between them as indicated by Barnett and Lewis (1994).

A *contaminant* is an observation arising from some other distribution. Suppose F and G are unique density functions. Let F be our ideal distribution from which our data, x_1, \dots, x_T , is drawn. However, suppose we observe $x_1, \dots, y_1, \dots, x_T$ where y_1 is from G ; then y_1 would be a contaminant.

An *extreme value* is simply one that is near the maximum or minimum of the sample. Suppose we observe x_1, \dots, x_T and order them as $x_{(1)}, \dots, x_{(T)}$. The observations $x_{(1)}$ and $x_{(T)}$ are considered extreme values. In practice we can choose some percentage of the largest and smallest observations to be extreme.

Outliers do not have to be contaminants, nor do contaminants have to be outliers. Likewise, extreme values do not have to be outliers, although all outliers are extreme values. It is completely possible to draw numbers from a normal distribution and have one that seems surprising and meets some test of discordancy test—even though there is a low probability that this will happen. This aberrant number is labelled as an outlier even though it is not a contaminant. This occurs more frequently with heavy-tailed – though not necessarily contaminated – distributions. Also, it is possible to have small contaminants that fail to meet a test of discordancy. Thus, there are contaminants that are not outliers. Extreme values also may not meet a discordancy test so they will not necessarily be labelled outliers. However, an outlier

will always be extreme.

When trying to identify outliers, we have to be careful about the assumptions of the true distributions from which we observe the data. For instance, suppose we assume our data are generated from a normal distribution when in fact they come from a distribution with much heavier tails. We are very likely to find outliers, but the extreme observations would not be outliers if we assumed the correct distribution. So heavy-tailed distributions do not inherently generate outliers, but we may observe outliers if we specify an incorrect underlying distribution, which is not uncommon in practice. This makes distinguishing between outliers, contaminants, and extreme values difficult.

2.2 Outliers in the Context of Models

The above definitions are acceptable for unstructured, univariate data, but they need to be modified in order to describe outliers in the context of structured models. An outlier in this context is an observation that is surprising and inconsistent with model. Operationally, outliers in models will be found using the estimated residual series.

It is also useful to make a distinction between outliers in the dependent variable and outliers in the independent variable. Outliers in the dependent variable are still referred to as outliers, but outliers in the independent variable are referred to as *leverage points*.

When glancing at the data, a surprising observation may be consistent with the model, or a seemingly normal observation may be inconsistent with the model. So we have to be careful about how we define surprising. In the context of a model we need to look at how well the observation fits the structure we have assigned. If it does surprisingly poorly at fitting the model, then we might have a candidate for an outlier.

We typically measure how well an observation fits a given model by looking at the estimated residual series. Consider the linear model $y_t = \beta x_t + \epsilon_t$. If the estimated ϵ_i is extremely large and x_i is not a leverage point, then y_i could be considered an outlier. Notice that we did not state that the unobserved ϵ_i was the outlier; it is the actual observation y_i . An obvious question is whether the outlier is in the observation or the unobserved error term. Because we do not observe the true error terms, we cannot conclude if the true error terms are outliers or not.

Model specification is also important in determining what observations are outliers. Suppose we estimate a model that correctly incorporates a large deviation in a particular observation, y_i . This model most likely would not label y_i as an outlier, but a more naive model would.

2.3 Contaminants in Time Series

With time series models, there are correlations between nearby observations. The observation at time period t is correlated to the observation at time period $t + 1$,

which is in turn correlated to the observation at time period $t + 2$, etc. Therefore, it is possible that if there is an outlier at time period t this will follow through to the observation at period $t + 1$ by way of the autocorrelation. If this does occur, it will be referred to as an *innovation outlier*. (An oil price shock that does not dissipate immediately could be an example of an innovation outlier.) However, this does not have to happen. Recall that an outlier is simply an observation that is not consistent with the rest of the data or model. So an observation that does not follow the model's autocorrelation structure can also be considered an outlier. If there is a "blip" at time period t that does not affect nearby values this is referred to as an *additive outlier*. (A short plant strike could be viewed as an additive outlier.)

So, which of these type of outliers is more common? The answer is debatable. Barnett and Lewis (1994) reference a debate between various outlier statisticians. Huber (1972) claims that innovation outliers are more common, but Kleiner, Martin, and Thomson (1979) state that additive outliers are much more common. In the context of macroeconomic data, Balke and Fomby (1994) finds combinations of both types of outliers.

For clarity, I will show two alternative, but equivalent, definitions of time series outliers in the context of autoregressive moving average (ARMA) models.

2.3.1 Standard specification

The following contaminant models are widely used in the literature. Yohai (1986), and more recently Lucas (1995) both use the following models to illustrate differences

between innovation and additive contaminants.

Consider the following ARMA model

$$\phi(L)y_t = \theta(L)u_t \quad (2.1)$$

where $\phi(L)$ and $\theta(L)$ are polynomials in L , the lag operator, and u_t is white noise.

This is the no-contamination model.

In contrast, with an innovation contaminant w_t , which will be zero for most values of t , we have

$$\phi(L)z_t = \theta(L)(u_t + w_t) \quad (2.2)$$

where z_t is the observed series. In terms of the state variable, y_t which follows the ARMA model in equation (2.1), we have

$$z_t = y_t + \frac{\theta(L)}{\phi(L)}w_t. \quad (2.3)$$

In further contrast, with an additive contaminant w_t , again equalling zero for most values of t , we have

$$\phi(L)(z_t - w_t) = \theta(L)u_t \quad (2.4)$$

where z_t is again the observed series. In terms of the state variable, y_t which again follows the ARMA model in equation (2.1), we have

$$z_t = y_t + w_t. \quad (2.5)$$

Notice that according to our definitions in Section 2.1, not all the contaminating observations will be outliers.

2.3.2 Tsay's specification

For Tsay's (1988) specification, let y_t again follow model (2.1). Now assume that z_t follows the model

$$z_t = f(t) + y_t \quad (2.6)$$

where $f(t)$ is a parametric function representing the contaminating disturbances to y_t . $f(t)$ will be assumed to be of the following form

$$f(t) = \omega_0 \frac{\omega(L)}{\delta(L)} \xi_t^{(d)} \quad (2.7)$$

where $\xi_t^{(d)} = 1$ if $t = kd$ for positive integers k and $= 0$ otherwise. So d represents the time of an occurrence of a contaminant. The structure of $f(t)$ will demonstrate what type of outlier we have.

If $\omega_0 = \omega_I$ and $\frac{\omega(L)}{\delta(L)} = \frac{\theta(L)}{\phi(L)}$, then we have an innovative contaminant. This is equivalent to equation (2.3).

If $\omega_0 = \omega_A$ and $\frac{\omega(L)}{\delta(L)} = 1$, then we have an additive contaminant. This is equivalent to equation (2.5).

Similar to the standard specification, we see that Tsay also defines contaminants rather than outliers. However, when he tests for outliers he has set critical values for the size of the disturbance, ω . Therefore, he conforms to our second definition of outliers.

2.3.3 Generating Time Series Outliers

It is useful to describe different ways outliers are generated. These methods are extremely useful when constructing simulations to evaluate the performance of different estimators subject to contamination or outliers. The following methods are frequently used in the outlier literature. It should be noted that these are not the only ways to generate outliers.

Deterministic

It is very easy to just pick a few points and set them equal to extremely large or small values and thus make them outliers. This can be done for either the additive or innovation case. This is one of the more common ways to generate outliers in time series. The size and location of the discordant observations are known with certainty. Unfortunately, the lack of randomness in the size and location of the outliers may lead to incomplete results. Tsay (1988) and Yohai (1986) both use this method to generate outliers for their experiments.

Heavy-Tailed Distributions

If we assume our error terms are from a normal distribution, but in reality our error terms are from some distribution that exhibits excess kurtosis, such as a Student's t or Cauchy distribution, it will appear that we have error terms that are both extreme and surprising. When heavy-tailed distributions are used to generate data based on an ARMA model, this will produce innovation outliers.

Though not specifically referring to outliers, many authors use heavy-tailed distributions to approximate random variables that exhibit excess kurtosis such as stock prices or exchange rates.

Random contamination models

Random contamination models are a useful alternative to the deterministic and heavy-tail methods to construct outliers. The implications of such models for the study of outliers have been studied by Box and Tiao (1968) and Guttman (1973).

Add values from a separate distribution to the series of unobserved state variable values at random times for additive contamination. Add values to the error terms at random times to produce innovation contamination. In both cases, the percentage of contaminants will be controlled by a parameter.

In the context of the standard model, let $w_t \sim \lambda G$ where G is the contaminating distribution and λ is a binomial random variable (0,1) with probabilities $\pi = P(\lambda = 1) = .05$ and $(1 - \pi) = P(\lambda = 0) = .95$.

In the case of innovation contaminants, this mixture type model is simply a heavy-tailed distribution. This can be seen by examining the residual series, ε_t , that is used to generate the contaminated ARMA model. Consider $G = N(0, b\sigma^2)$ where $b > 1$ and $\varepsilon_t = u_t + w_t$ and σ^2 is the variance of u_t . u_t is the non-contaminated residual and w_t is the contaminant. Therefore

$$\varepsilon_t \sim (1 - \lambda)N(0, \sigma^2) + \lambda[N(0, \sigma^2) + N(0, b\sigma^2)]. \quad (2.8)$$

The variance of ε_t is $(1 + b\pi)\sigma^2$ and the fourth moment of ε_t is $3[(1 - \pi) * 1^2 + \pi(1 + b)^2] * 3\sigma^4$. Therefore the kurtosis – computed as the ratio of the fourth moment and the square of the variance – of ε_t is equal to

$$3 * \frac{1 + 2b\pi + b^2\pi}{(1 + b\pi)^2} \quad (2.9)$$

which is greater than the kurtosis for a standard normal distribution, 3, because $\pi < 1$.

We can use our knowledge of a normal distribution to get an estimate of how many residual outliers we would expect in this random contamination model. This will be done by estimating the residuals of the observed series, z_t , using the true model of the state variable, y_t . Note that the actual number of contaminants remains fixed at $\pi\%$, but the whether or not the contaminant is an outlier will change between models.

Consider the simple and conservative discordancy rule declaring any element ε_t an outlier if it falls outside 3 standard deviations from the mean. Also, suppose we know the true mean and variance of the non-contaminated distribution, μ and σ^2 respectively. Also let λ be drawn from a binomial distribution with $P(\lambda = 1) = .05$ and let $b = 9$.

For innovation outliers, the estimated residuals are $\varepsilon_t = u_t + w_t$ where u_t and w_t are as defined above. We also know that $\varepsilon_t \sim (1 - \lambda)N(0, \sigma^2) + \lambda[N(0, \sigma^2) + N(0, 9\sigma^2)]$ from equation (2.8). 95% of the residuals are not contaminated and .3% of those will be outside of plus-or-minus 3σ from μ . 5% of the residuals will be contaminated and

those points will have a variance of $(1 + 9)\sigma^2$. Therefore 34.3% of the contaminated points will be outside of plus-or-minus 3σ from μ . Combining the two sets of outliers we have $.95 * .3\% + .05 * 34.3\% = 2\%$ of the residuals will be outliers.

Additive outliers are a little trickier. We must now take into account the model parameters in determining what percentage of the contaminants are outliers. We will find that higher levels of persistence in the autoregressive terms relative to the moving average terms generates higher levels of additive outliers. Consider the additive contamination model (2.4) using the same parameters as in the innovation case. Rewriting the model, we get

$$\frac{\phi(L)}{\theta(L)}z_t = u_t + \frac{\phi(L)}{\theta(L)}w_t. \quad (2.10)$$

If we know the parameters with certainty, our estimated residuals, ε , are simply

$$\varepsilon = u_t + \Psi(L)w_t \quad (2.11)$$

where $\Psi(L) = \frac{\phi(L)}{\theta(L)} = 1 - \psi_1L - \psi_2L^2 - \dots$. Once again we will use our simple decision rule that if any element of ε_i falls outside three of the standard deviations, σ^2 , from the mean then it is an outlier. The non-contaminated residuals have a variance of σ^2 , while the contaminated residuals have a variance of $[(1 + \psi_1^2 + \psi_2^2 + \dots) * 9\sigma^2 + \sigma^2]$. As is the case with the innovation outliers, the higher the variance of the contaminating distribution, the higher the number of contaminants will be outliers. As the square of the parameters (ψ_1, ψ_2, \dots) increases, so will the variance of the contaminating part thus increasing the percentage of the estimated residuals that will be termed outliers. If we have an autoregressive model of order 1 with $\phi_1 = .5$ the variance

of contaminated points increases to $12.25\sigma^2 = (1 + 0.5^2) * 9\sigma^2 + \sigma^2$. In this case, 39% of the contaminated residuals fall outside of plus-or-minus 3σ from the mean. Therefore, the number of outliers will jump to $.95 * .3\% + .05 * 39\% = 2.24\%$ because more of the contaminants create outliers.

Additive outliers also have the problem of becoming leverage points – outliers in the independent variables – in subsequent periods. For instance if y_t is an outlier, y_t also is a right-hand side variable in k periods where k is the AR length of the model. Therefore, in an autoregressive (AR) model, AR(k), one outlier actually corrupts infinitely many different data positions.

2.4 Conclusion

This chapter provided a brief introduction to the vast literature on outliers, in particular, time-series outliers. For a more in-depth description of time-series outliers, refer to Martin (1979) or to Barnett and Lewis (1994).

The definitions and descriptions of outliers and outlier models presented in this chapter will be used throughout this thesis. Any alternative definitions or descriptions that are used will be discussed accordingly within each essay.

Chapter 3

Forecasting Heavy-Tailed Autoregressive Models

3.1 Introduction

The purpose of this essay is to evaluate the forecasting performance of autoregressive models estimated by ordinary least squares (OLS) and robust estimation techniques when the disturbance term is heavy-tailed.¹ Autoregressive (AR) models represent a simple and inexpensive way to forecast time series, and in many cases, the forecasting performance of AR models stacks up well against large-scale econometric models. An AR model is routinely identified using the techniques outlined by Box and Jenkins (1976), and forecasts are based on a linear model fit to sample data where parameter estimates are obtained using ordinary OLS. In the case of a heavy-tailed distribution of the disturbance term, however, OLS parameter estimates may

¹In this context, robustness refers to qualitative robustness which is concerned with the behavior of an estimator when the shape of the underlying distribution deviates from the ideal model (Huber 1977).

be unduly influenced by unexpected observations, or outliers.² Consequently, the accuracy of forecasts can be seriously compromised. Robust estimation techniques, in contrast, are less susceptible to outlying observations, and as a result, offer potential gains in forecast accuracy.³

It is common in applied work to detect evidence of heavy tails in high-frequency economic and financial time series such as stock returns (Fama 1965), futures prices (Hall, Brorsen, and Irwin 1989), and exchange rates (Boothe and Glassman 1987). In a recent study, Hamilton (1996) reports evidence of heavy tails for interest rates, citing the existence of “spectacular outliers” in the federal funds rate series. A histogram of weekly federal funds rate changes for the 1984-1998 period is shown in Figure 3.1. A normal distribution with matching values for the mean and variance is included for comparison. The typical characteristics of a heavy-tailed distribution are evident: a relatively large number of observations are still concentrated around the mean, but a substantial probability mass is associated with the tails. If we were to assume that changes in the federal funds rate are normally distributed, then a 100-basis point change would be expected to occur once every 268 years or so. The normal distribution, however, is a poor description of the data as a 100-basis point change is not a rare event; over the 14 year sample period, a 100-basis point change

²Outliers from heavy-tailed distributions usually arise from assuming an incorrect distribution, such as assuming a normal distribution when the data is really from a Cauchy distribution. Observations that would be expected from a Cauchy distribution appear to be unexpected if a Normal distribution was assumed. This is in contrast to contamination that is from another distribution.

³Recent studies have explored the effects of outliers on testing for unit roots and cointegration (Franses and Haldrup 1994; Lucas 1995), non-linearities (Van Dijk et. al. 1999a), and conditional heteroskedasticity (Van Dijk et. al. 1999b).

occurs seven times.

A number of heavy-tailed distributions and processes have been advanced in the literature to account for observed excess kurtosis. Early work by Mandelbrot (1963) and Fama (1965) stressed heavy-tailed, stable distributions that possess convenient statistical properties. These include domains of attraction according to the generalized central limit theorem, and stability under addition. Alternatives to stable distributions such as a Students-t distribution with finite variance (Blattberg and Gonedes 1974) and a mixture of normals (Kon 1984)⁴ have also been proposed to capture the salient features of the data. Another heavy-tail model commonly used in empirical studies is the class of autoregressive conditional heteroscedasticity (ARCH) models popularized by Engle (1982) and Bollerslev (1987).

Robust estimation techniques are designed to limit the influence of outlying observations. The class of M-estimators introduced by Huber (1964) selects parameter estimates based on minimizing an objective function other than the sum of squares. That is, robust M-estimators are constructed using functions of the residuals that attach less weight to outlying observations. The S-estimator developed by Rousseeuw and Yohai (1984) – described in the next section – is even more resilient to extreme observations but lacks efficiency should the disturbances actually follow a normal distribution. To remedy this, the MM-estimator (Yohai 1986) and τ -estimator (Yohai and Zamar 1988) – both described in the next section – have

⁴Unlike the other distributions mentioned, the mixture of normals can be thought of as a contaminated distribution as opposed to simply a heavy-tailed distribution

more recently been introduced to be both highly efficient when the disturbances have a normal distribution and highly robust to outlying observations. The influence of outlying observations for autoregressive moving average (ARMA) models is studied by Denby and Martin (1979), Martin (1979), and Bustos and Yohai (1986). In this essay we extend this work and consider robust estimators that do not sacrifice efficiency should the actual data follow a normal distribution. We also extend their work by addressing specific forecasting applications of the robust estimators. In this paper, we compare the properties of this set of robust estimators and OLS in a series of Monte Carlo exercises.

The Monte Carlo simulations in this paper consider heavy-tailed distributions and processes commonly used to model high-frequency time series. Overall, the Monte Carlo results point to substantial gains associated with robust estimation methods. In all cases, robust estimation yields more accurate AR parameter estimates than OLS. And in the majority of cases, the accuracy of the parameter estimates is improved by more than 75%. In an effort to gauge whether large gains are likely to be realized in practice, we evaluated out-of-sample forecasts for federal funds rate changes. Forecast results suggest that the answer is indeed yes, especially at long forecast horizons.

The remainder of the paper is organized as follows. The next section provides an outline of robust estimation, examining both the sensitivity to outlying observations and efficiency. In Section 3, we present Monte Carlo evidence on the performance of

OLS and robust estimation for alternative heavy-tailed distributions and processes. Section 4 applies robust estimation methods and OLS to short-term interest rate data. The final section concludes and discusses avenues of future research.

3.2 Robust Estimation

Consider a time series x_t for $t = 1, \dots, n$ generated by an AR process:

$$\Phi(L)(x_t - \mu) = \varepsilon_t \tag{3.1}$$

where ε_t is independently and identically distributed with symmetric density function $f(\bullet)$ and distribution function $F(\bullet)$, μ is the mean of x_t , and $\Phi(L)$ is a polynomial in the lag operator L given by $\Phi(L) = 1 - \Phi_1 L - \dots - \Phi_p L^p$ with all roots lying outside the unit circle. Also, define the parameter vector $\lambda = (\Phi, \mu)'$ where $\Phi = (\Phi_1, \dots, \Phi_p)$. Our goal is to evaluate estimators of λ in terms of robustness and efficiency when ε_t follows a heavy-tailed distribution.

The breakdown point is a convenient measure of robustness. Consider an estimator defined by the functional T . Also, let P_ϵ be an ϵ -neighborhood of the ideal model distribution F^* . The maximum bias of T is then given by

$$\sup_{F \in P_\epsilon} |T(F) - T(F^*)| \tag{3.2}$$

where F is the true underlying distribution. The *breakdown point* of T at F is defined as the largest proportion of observations that can be arbitrarily replaced such that (3.2) is finite (Huber 1977). Note that a high breakdown point – thus

defined – does not necessarily mean that parameter bias is negligible. Intuitively, a breakdown point can be thought of as the percentage of bad observations with which an estimator can “cope.”

At this point it is convenient to specify the continuous real-valued function ρ having the following properties:

Assumption 1. (i) $\rho(0) = 0$, (ii) $\rho(-u) = \rho(u)$, (iii) $0 \leq u \leq v$ implies $\rho(u) \leq \rho(v)$, (iv) letting $a = \sup \rho(u)$ such that $0 < a < \infty$, if $\rho(u) < a$ and $0 \leq u < v$, then $\rho(u) < \rho(v)$.

Following the research of Huber (1964), an M-estimator minimizes the objective function

$$\sum_{t=1}^T \rho(\varepsilon_t, \lambda) \tag{3.3}$$

or equivalently satisfies the $(p + 1)$ system of equations

$$\sum_{t=1}^T \varphi(\varepsilon_t, \lambda) = 0 \tag{3.4}$$

where $\varphi(\varepsilon_t, \lambda) = (\partial/\partial\lambda)\rho(\varepsilon_t, \lambda)$. Inspection of (3.3) reveals that OLS is an M-estimator where $\rho(u) = u^2$. When $f(\bullet)$ is Gaussian, the OLS estimate λ corresponds to the maximum likelihood estimate and is efficient as its covariance matrix attains the Cramer-Rao lower bound. Researchers have also shown that OLS is consistent if $f(\bullet)$ is non-Gaussian with finite variance (Mann and Wald 1943) and even if $f(\bullet)$ is a non-Gaussian stable distribution with infinite variance (Kanter and Steiger 1974). OLS, however, may be highly inefficient in these cases. Let $\hat{\Phi}_{LS}$ and $\hat{\mu}_{LS}$ be the OLS estimates of Φ and μ , respectively. Bustos and Yohai (1986) explain that if $\hat{\Phi}_{LS}$ is non-Gaussian, then the covariance matrix of $\sqrt{T}(\hat{\Phi}_{LS} - \Phi)$ is independent of $f(\bullet)$,

but the variance of $\sqrt{T}(\hat{\mu}_{LS} - \mu)$ can be significantly inflated if $f(\bullet)$ has heavy tails because it depends on $E_f(\varepsilon^2)$.

Because OLS is sensitive to outlying observations which can be reasonably expected to occur if $f(\bullet)$ is heavy-tailed, robust M-estimators have been advanced in the literature. Unlike the case for OLS, however, a solution for (3.4) is not generally invariant with respect to scale,⁵ s_n , implying that ε_t must be standardized. Huber (1964) proposed the minimax M-estimator:

$$\rho_H(u_t) = \begin{cases} u_t^2/2 & |u_t| < k_H \\ k_H|u_t| - k_H^2/2 & |u_t| \geq k_H \end{cases} \quad (3.5)$$

and

$$\varphi_H(u_t) = \min(k_H, \max(u_t, -k_H)) \quad (3.6)$$

where $u_t = \varepsilon_t/s_n$ is the standardized residual and k_H is a “tuning” constant where $0 < k_H < \infty$. Huber’s M-estimator corresponds to the maximum likelihood estimator for the least favorable distribution in the minimax sense (Hampel et. al. 1986). Note that if $k_H = \infty$, then Huber’s estimator reduces to the OLS estimator. In the empirical work to follow we set $k_H = 1.654$ and use $s_n = 1.483 * \text{median}\{|u_t - \text{median}(u_t)|\}$ which is a robust estimate of scale.⁶ Another popular M-estimator is

⁵The term *scale* refers to a general measure of variation in data.

⁶The tuning constants, k_H and k_B are chosen such that the M-estimator based on the least favorable distribution in the minimax sense is efficient. See Hampel et. al. 1986 for more information on how to choose tuning constants.

based on the bisquare function

$$\rho_B(u_t) = \begin{cases} u_t^2/2 - u_t^4/2 + u_t^6/6 & |u_t| < k_B \\ 1/6 & |u_t| \geq k_B \end{cases} \quad (3.7)$$

and

$$\varphi_B(u_t) = \begin{cases} u_t(1 - u_t^2)^2 & |u_t| < k_B \\ 0 & |u_t| \geq k_B \end{cases}. \quad (3.8)$$

For this specification, we use $k_B = 4.685$ and s_n as defined above. One advantage of the bisquare function is that it has higher efficiencies than Huber’s formulation should $f(\bullet)$ be “extremely heavy-tailed” (Martin 1979).

Huber (1977) proves that estimates from an M-estimator are asymptotically normally distributed with mean zero and variance

$$\int IF(x; F, \mathbb{T})^2 dF(x) \quad (3.9)$$

where IF is the influence function. The IF describes the effect of adding an additional observation x on \mathbb{T} , given the distribution F . Denby and Martin (1979) show in a Monte Carlo study that M-estimators perform well in terms of efficiency, but estimates of λ can be seriously biased.

Our interest focuses on innovation outliers that satisfy the AR model given by (3.1). Additive outliers, in contrast, do not follow the underlying model (see Chapter 2 for a comparison between innovation and additive outliers). While innovation outliers may not seriously compromise the performance of robust M-estimators, additive outliers may seriously compromise that performance. In fact, an M-estimator’s

breakdown point is no better than $1/T$ in the case of additive outliers. In practice, one may not be able to discern which type of outlier is in fact present, although some guidance is offered by Martin (1979). Consequently, we take a conservative approach and explore the use of high breakdown point estimators.

Rousseeuw and Yohai (1984) introduce the high breakdown point S-estimator based on minimizing a scale estimate. The S-estimator is defined as the solution to

$$\text{minimize } s(\varepsilon_1(\lambda), \dots, \varepsilon_T(\lambda)) \quad (3.10)$$

subject to

$$\frac{1}{T} \sum_{t=1}^T \rho(\varepsilon_t/s(\bullet)) = K \quad (3.11)$$

and ρ satisfies Assumption 1. The constant K is set equal to $E_{\Theta}[\rho(u)]$ where Θ denotes the standard normal distribution. The estimate $s_n(\bullet)$ is just an M-estimator of the scale parameter. The function ρ associated with the S-estimator is given by (3.7), and Rousseeuw and Yohai (1984) show that the S-estimator has a 50% breakdown point for $k_s = 1.547$ and $K = 0.1995$.

S-estimators are asymptotically normal and have the usual convergence rate of $T^{-\frac{1}{2}}$. Rousseeuw and Yohai (1984) also compute the asymptotic efficiency for the Gaussian model. For a 50% breakdown point, asymptotic efficiency is only 28.7%. Efficiency can be improved by choosing a larger value of k_s , but at the cost of reducing the breakdown point.

The MM-estimator introduced by Yohai (1986) achieves both a high breakdown

point and high efficiency if the errors conform to a normal distribution. The MM-estimator is defined in three steps. In the first step, a high breakdown point estimator is used to compute an initial estimate of λ . Yohai explains that in the first step the estimator does not have to be efficient. In the second step, an M-estimate of scale is computed from the residuals obtained in the first step. That is, for some function ρ_0 that fulfills Assumption 1, a scale estimate $s(\varepsilon_1(\lambda), \dots, \varepsilon_T(\lambda))$ is obtained that satisfies

$$\frac{1}{T} \sum_{t=1}^T \rho_0(\varepsilon_t/s(\bullet)) = b \quad (3.12)$$

where $b = E_{\Theta}[\rho(u)]$. In the final step, the MM-estimates are the solution to (3.3) for some function ρ_1 that satisfies Assumption 1 – essentially another M estimate using the scale estimate from the first step.

Yohai (1986) suggests choosing ρ_0 and ρ_1 as described in (3.7) for tuning constants k_0 and k_1 , respectively. In the empirical work to follow, we set $k_0 = 1.56$, $b = 0.0833$ and $k_1 = 4.68$ which corresponds to 95% efficiency for normally distributed errors.

The last robust estimator we consider is the τ -estimator developed by Yohai and Zamar (1988). Like the MM-estimator, the τ -estimator has a high breakdown point and is highly efficient in the case of Gaussian errors. For functions ρ_0 and ρ_1 that satisfy Assumption 1 and where s_n is an M-estimate of scale as shown in (3.12), the τ scale is

$$\tau^2(\varepsilon_t) = s_n^2(\varepsilon_t) \frac{1}{T} \sum_{t=1}^T \rho_1\left(\frac{\varepsilon_t}{s_n(\varepsilon_t)}\right). \quad (3.13)$$

The τ -estimator of λ is then found from $\min_{\lambda} \tau_n[\varepsilon(\lambda)]$. In the work to follow, ρ_0 and ρ_1 are given by (3.7) and $k_0 = 1.56$ ($b = 0.203$ in expression (3.12)) and $k_1 = 6.08$, which achieves 95% efficiency.

3.3 Monte Carlo Simulations

The performances of M, S, MM, τ , and OLS are examined via Monte Carlo simulations. The simulations focus on an AR(1) $X_t = \Phi_0 + \Phi_1 X_{t-1} + \varepsilon_t$ process where $\Phi_0 = \mu(1 - \Phi_1)$ with four heavy-tailed⁷ descriptions of the disturbance term, ε_t (as indicated in the introduction, each of these distributions have been advanced to account for excess kurtosis):

- i** A Student's- t distribution with three degrees of freedom
- ii** A mixture of normals where a draw is from $N(0,1)$ with probability $p = 0.85$ or is from $N(0,8.5)$ with probability $(1 - p)$ ⁸
- iii** An ARCH process such that $\varepsilon_t = u_t[1 - 0.5\varepsilon_{t-1}]^{\frac{1}{2}}$ and $u_t \sim N(0, 1)$
- iv** A mean-zero, symmetric, infinite variance sum-stable distribution with characteristic exponent $\alpha = 1.24$ (see McCulloch 1998).

Random draws are conducted using the methods outlined in McCulloch (1998).

Two parameterizations of the AR(1) model are considered: a moderate persistence model that has $\Phi_0 = 2.0$ and $\Phi_1 = 0.5$, denoted model 1, and a high persistence

⁷The focus of this simulation is on innovation outliers rather than additive outliers.

⁸This distribution has a variance of 1.46 and a kurtosis of 7.76 which are similar to those of stock returns examined in Kon (1989).

model that has $\Phi_0 = 2.0$ and $\Phi_1 = 0.9$, denoted model 2. The simulation exercises use 1,000 replications for each of the sample sizes $n = 250, 500$.

The performance of the estimators is judged in terms of the absolute value of the average percent bias for each simulation exercise. For example, suppose that $\hat{\Phi}_i$ is the average estimated value of parameter Φ_i for the 1,000 simulations. The absolute value of the average percent bias is then given by $100 \left| \frac{\hat{\Phi}_i - \Phi_i}{\Phi_i} \right|$. We also tabulate the interquartile range in an effort to measure the dispersion of the estimated parameter.

Table 3.1 reports results for sample size $n = 250$. Beginning with Model 1 shown in the top panel, robust estimation yields a more accurate estimate of the AR parameters compared to OLS. The sensitivity of OLS parameter estimates to outlying observations is evident, and is most apparent for the infinite variance sum-stable model as this model possesses the “thickest tails.” One notable finding shown in Table 3.1 is that outlying observations exert more influence on the OLS estimate of the intercept. This is consistent with the work of Bustos and Yohai (1986) which suggests that outliers have more influence on μ than Φ_1 . Overall, substantial improvement in the accuracy of AR parameter estimates can be achieved if a robust estimator is used, and, in the majority of cases, the extent of parameter bias is reduced by more than 75% compared to OLS.

Results for Model 2 are summarized in the bottom panel of Table 3.1. The difference in this case concerns the degree of persistence exhibited by the time series (as measured by Φ_1). Results suggest that persistence affects the performance of

OLS parameter estimates, particularly the intercept term. In the sum-stable model, for example, the percent parameter bias is at least 15 times greater for OLS compared to robust estimation. In forecasting applications, the influence of persistence is likely to be greater when the forecast horizon is longer, as forecasts tend towards the unconditional mean (μ).

Estimates of the interquartile range, shown in parentheses, also point to the superiority of robust estimation. With the exception of a few cases, robust estimators are associated with narrower interquartile ranges compared to OLS for both Φ_0 and Φ_1 . The exception is the S-estimator, that likely traces to the selection of the tuning constant as it is set to place greater weight on the breakdown point than on efficiency.

Results shown in Table 3.1 find that the robust estimators all perform well. This is likely due to focusing exclusively on innovation outliers. In the case of additive outliers, high breakdown point estimators that do not sacrifice efficiency should outperform other robust estimators. Recent work by Lucas (1995) suggests that at least in the case of the Nelson-Plosser data, outliers are more likely to be of the innovation variety. In any event, a conservative estimation strategy is to use either the MM-estimator or the τ -estimator as one may be unable to differentiate between innovation and additive outliers in practice.

As the sample size increases to $n = 500$, the extent of parameter bias decreases for the majority of cases as shown in Table 3.2. Overall, robust estimation results

in more precise estimation of the parameters, and again a 75% reduction in bias is achieved relative to OLS in the majority of cases. Consistent with the smaller sample size results, robust estimators are also associated with narrower interquartile range values. In sum, robust estimation appears to afford good protection against innovation outliers.

3.3.1 Relation to Forecasts

To see why an unbiased estimate of the constant term⁹ is important for forecasting, consider forecasting the following AR(1) model

$$x_{t+1} = \Phi_0 + \Phi_1 x_t + \varepsilon_t \quad (3.14)$$

with

$$\hat{x}_{t+1} = \hat{\Phi}_0 + \hat{\Phi}_1 x_t \quad (3.15)$$

where $\hat{\Phi}_1$ is unbiased, but $\hat{\Phi}_0$ is biased such that $|\hat{\Phi}_0 - \Phi_0| = B$.

The one-period-ahead mean squared forecasting error is

$$E[(x_{t+1} - \hat{x}_{t+1})^2] = B^2 + \sigma^2 \quad (3.16)$$

where σ^2 is the variance of ε . Maintaining the same parameter estimates and iterating forward in time, the k-period-ahead mean squared forecasting error is

$$E[(x_{t+k} - \hat{x}_{t+k})^2] = (1 + \hat{\Phi}_1 + \hat{\Phi}_1^2 + \dots + \hat{\Phi}_1^{k-1})^2 B^2 + (1 + \hat{\Phi}_1^2 + \hat{\Phi}_1^4 + \dots + \hat{\Phi}_1^{2(k-1)}). \quad (3.17)$$

Notice that the bias in the constant term accumulates as the forecast horizon, k, increases. The problem is more severe when Φ_1 is large.

⁹Because OLS does well at estimating the slope terms, our focus will be on the constant term.

3.4 Interest Rate Forecasts

In this section we evaluate out-of-sample forecasts of the weekly federal funds rate using OLS and robust estimation. Data were obtained from the St. Louis Federal Reserve data files. The sample runs from March 7, 1984 to December 30, 1998 (early dates were not included to avoid a possible level shift with a Federal Reserve regime change.) Autoregressive models are fit to the March 7, 1984 to December 26, 1996 period which leaves two years of data in which to evaluate forecasts.

There are numerous extreme movements in the daily federal funds rate over this period. Hamilton (1996) argues that such outliers should not be discarded as these observations contain useful information. Regressing the federal funds rate on a constant and dummies for two extreme observations yields an R^2 of .40.

In empirical work we model federal funds rate changes because, consistent with numerous studies, we are unable to reject an I(1) null hypothesis for the federal fund rate (see Tables 3.3 and 3.4). Also, an AR(3) model is selected based on the Bayesian Information Criterion computed in the OLS framework.¹⁰ Parameter estimates are shown in Table 3.5 along with standard errors in parentheses. OLS parameter estimates find substantial persistence as the sum of the AR coefficients is -0.65. This degree of persistence, however, is not shared by the robust estimators as the sum of the coefficients ranges from -0.32 to -0.47. As is apparent in the Monte

¹⁰The work of Basci and Zaman (1998) as well as the results from chapter 4 indicate that the BIC performs well in the event that there are only innovation outliers, which is the focus of this paper.

Carlo simulations, the robust methods offer very different estimates of the intercept term compared to OLS.

Next, we compute out-of-sample interest rate forecasts. The first forecast is based on estimating an AR(3) model up through date t_0 (December 26, 1996) and computing a dynamic forecast until date $t_0 + k$ where k is the forecast horizon ($k > 1$). This forecast, generated using method i , is denoted $\hat{X}_{t_0+k}^i$, and the resultant forecast error is given by $u_{t_0+k}^i = \hat{X}_{t_0+k}^i - X_{t_0+k}$. The second forecast is based on estimating the AR(3) up through date $t_0 + 1$, and computing and computing a dynamic forecast until date $t_0 + k + 1$. The process continues until date $T - (t_0 + k)$ which yields a sequence of $N_k = T - (t_0 + k) + 1$ forecasts and forecast errors.

Forecast accuracy is assessed using a mean square error (MSE) criterion. Because we are primarily interested in determining how forecasts based on robust estimation compare to OLS, we compute the ratio of the MSE associated with robust estimation to the OLS MSE.¹¹ If this ratio is less than one, then forecasts based on robust estimation are more accurate. If this ratio is greater than one, then forecasts based on robust estimation are less accurate.

We use the test developed by Diebold and Mariano (1995) to determine whether the ratio of MSEs is statistically different from one. For each horizon k , the mean

¹¹It is known that a linear projection of Y_{t+1} on X_t produces the smallest mean squared forecasting error for the class of linear forecasts. Under ideal conditions, OLS provides a consistent estimate of the linear projection. However, under heavy-tailed distributions OLS gives an inaccurate estimate of the true parameter values (as seen in Table 2), and thus is an inaccurate estimate of the linear projection. Therefore robust estimation techniques that are more accurate may provide superior estimates of the linear projection, and thus a lower mean squared forecasting error.

forecast error difference is given by

$$\bar{d} = N_k^{-1} \sum_{t=t_0+k}^T [(u_t^i)^2 - (u_t^{LS})^2] \quad (3.18)$$

for robust method i where u_t^{LS} is the forecast error associated with OLS. The Diebold and Mariano test statistic

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\phi\hat{f}_d(0)}{N_k}}} \quad (3.19)$$

is asymptotically distributed as a standard normal under the null hypothesis that there is no difference in forecast accuracy. The term $\hat{f}_d(0)$ refers to the spectral density function of the forecast error difference evaluated at the zero frequency. One complication is that for $k > 1$, the forecast errors exhibit $(k - 1)th$ order serial correlation. As a result, we use the Newey and West (1987) covariance estimator with the bandwidth parameter set $k - 1$ to construct a consistent estimate of $f_d(0)$.

The performance of out-of-sample forecasts is summarized in Table 3.6. Table entries correspond to robust MSE values relative to OLS. The associated DM statistic is shown in parentheses. For one-period ahead forecasts, there is no significant difference in forecast accuracy. But as the forecast horizon increases, forecasts based on robust estimation are superior to OLS forecasts. At a 16-week horizon, improvements in MSE range from 31.8% (=1-0.682) for Huber's minimax estimator to 64.61% (=1-0.354) for the τ -estimator. Substantial improvements in forecast accuracy are also apparent at the 26-week forecast horizon. The gains are most likely due to superior estimates of the constant term. As shown in section 3.3, errors in the constant term accumulate as the forecasting horizon increases. Overall, forecast

accuracy using AR models can be improved by using robust estimation, particularly for long forecast horizons.

3.5 Conclusion

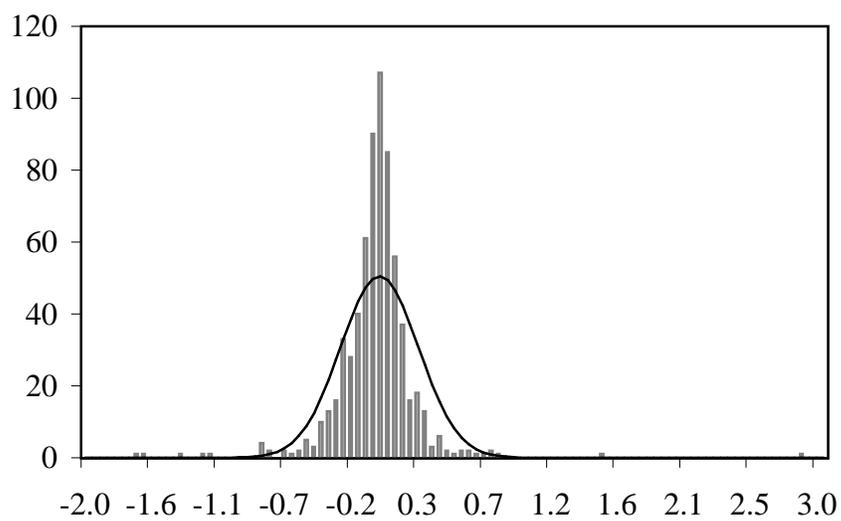
This essay demonstrates that substantial gains in forecast accuracy can be achieved by employing robust estimation when the regression error in an AR model follows a heavy-tailed distribution. By limiting the influence of outlying observations on the parameter estimates, robust estimation methods are able to offer more precise parameter estimates compared to OLS.

A secondary message of this essay is that the problems resulting from the presence of outlying observations should not be ignored as many economic and financial time series display excessive kurtosis compared to the normal distribution. It is suggested that any temptation to simply remove an outlier from the sample and proceed with OLS estimation should also be avoided. However, an exact comparison of the two techniques is for future research. Outliers may contain important information and consequently should not be discarded. This is primarily why robust estimation techniques can be so useful. They limit the influence of outliers, but they do not remove it entirely. Therefore, parameter estimates can be improved but not by completely removing important information. This point is stressed by Hamilton (1996) for the case of the federal funds rate.

The results of this essay can be applied or extended in a number of ways. For

example, robust estimation may offer insight as to whether the forward exchange rate predicts future spot rates or whether the term structure of interest rates predicts future short-term rates, as evidence of heavy tails is typically reported. One possible extension involves moving beyond linear models. For instance, Resnick (1998) argues that linear models are in many cases unable to adequately capture the dependency structure of heavy-tailed data, and nonlinear models such as the bilinear model represent viable candidates.

Figure 3.1: Histogram of Federal Fund Rate Changes, 1984-1998



The bars in this graph are the Federal Fund Rate Changes. The line is a normal distribution with mean and variance equal to that of the Federal Fund Rate Changes

Table 3.1: Percent Bias of Estimated Autoregressive Parameters, n=250
 Percent bias of Φ_0

DGP	Percent bias of Φ_0						Percent bias of Φ_1					
	OLS	Huber Ψ	Bisquare Ψ	S	MM	τ	OLS	Huber Ψ	Bisquare Ψ	S	MM	τ
Model 1: $x_t - 2 - .5x_{t-1} = \varepsilon_t$												
Student-t	1.8 (0.31)	0.78 (0.26)	0.61 (0.22)	0.66 (0.30)	0.01 (0.23)	0.3 (0.22)	1.73 (0.07)	0.98 (0.05)	0.19 (0.05)	0.37 (0.07)	0.46 (0.05)	0.31 (0.05)
Mixture	2.45 (0.40)	0.45 (0.17)	0.19 (0.15)	0.07 (0.13)	0.08 (0.14)	0.20 (0.18)	1.96 (0.07)	0.41 (0.04)	0.04 (0.04)	0.1 (0.03)	0.05 (0.03)	0.14 (0.03)
ARCH	3.45 (0.47)	1.54 (0.43)	0.93 (0.47)	0.26 (0.61)	1.05 (0.43)	0.91 (0.48)	3.43 (0.11)	1.7 (0.09)	1.8 (0.11)	1.03 (0.16)	1.72 (0.09)	1.05 (0.10)
S(=1.25)	6.53 (0.81)	0.35 (0.18)	0.27 (0.17)	0.18 (0.34)	0.46 (0.16)	0.81 (0.16)	1.57 (0.04)	0.24 (0.02)	0.2 (0.02)	0.13 (0.02)	0.21 (0.02)	0.09 (0.02)
Model 2: $x_t - 2 - .9x_{t-1} = \varepsilon_t$												
Student-t	10.27 (0.68)	1.86 (0.41)	4.64 (0.41)	1.94 (0.59)	1.15 (0.31)	1.35 (0.35)	1.24 (0.04)	0.06 (0.02)	0.47 (0.02)	0.22 (0.03)	0.2 (0.02)	0.21 (0.02)
Mixture	12.1 (0.77)	1.21 (0.29)	1.32 (0.24)	0.4 (0.34)	0.36 (0.23)	1.59 (0.26)	1.45 (0.04)	0.12 (0.02)	0.14 (0.02)	0.28 (0.02)	0.1 (0.02)	0.12 (0.02)
ARCH	12.67 (0.72)	1.5 (0.58)	1.33 (0.57)	0.44 (0.91)	1.23 (0.53)	0.82 (-0.6)	1.44 (0.04)	0.23 (0.03)	1.03 (0.03)	0.2 (0.05)	0.33 (0.03)	0.15 (0.03)
S(=1.25)	14.77 (1.01)	0.56 (0.24)	0.75 (0.24)	0.36 (0.49)	0.2 (0.29)	0.88 (0.22)	1.31 (0.03)	0.1 (0.02)	0.63 (0.02)	0.04 (0.03)	0.07 (0.01)	0.43 (0.01)

Notes: Table entries correspond to the absolute value of the percentage bias of estimated AR(1) coefficients and are based on 1,000 replications of sample size n=250. Numbers in parentheses are interquartile ranges of the parameter estimates. DGP refers to the data generating process of for Models 1 and 2. The column headings refer to the estimation method: OLS is ordinary least squares, Huber Ψ is Huber's (1964) M-estimator, Bisquare Ψ is an M-estimator that uses a bisquare function, S is the S-Estimator of Rousseeuw and Yohai (1984), MM is MM-Estimator of Yohai (1986), and τ is the τ -Estimator of Yohai and Zamar (1988).

Table 3.2: Percent Bias of Estimated Autoregressive Parameters, n=500

DGP	Percent bias of Φ_0						Percent bias of Φ_1					
	OLS	Huber Ψ	Bisquare Ψ	S	MM	τ	OLS	Huber Ψ	Bisquare Ψ	S	MM	τ
Model 1: $x_t - 2 - .5x_{t-1} = \varepsilon_t$												
Student-t	1.1 (0.22)	0.13 (0.20)	0.65 (0.17)	0.39 (0.20)	0.39 (0.17)	0.03 (0.17)	1.12 (0.05)	0.25 (0.03)	0.10 (0.04)	0.08 (0.03)	0.53 (0.04)	0.29 (0.04)
Mixture	1.2 (0.30)	0.42 (0.13)	0.02 (0.10)	0.02 (0.14)	0.20 (0.11)	0.24 (0.09)	1.06 (0.05)	0.45 (0.02)	0.06 (0.02)	0.27 (0.02)	0.07 (0.01)	0.11 (0.02)
ARCH	1.40 (0.34)	0.82 (0.31)	0.55 (0.30)	0.43 (0.56)	0.50 (0.29)	1.39 (0.32)	1.22 (0.07)	0.99 (0.07)	0.80 (0.07)	0.97 (0.14)	0.79 (0.06)	1.87 (0.08)
S(=1.25)	6.12 (0.69)	0.09 (0.13)	0.36 (0.13)	0.03 (0.13)	0.38 (0.12)	0.06 (0.12)	1.37 (0.03)	0.04 (0.02)	0.66 (0.02)	0.01 (0.02)	0.04 (0.01)	0.15 (0.01)
Model 2: $x_t - 2 - .9x_{t-1} = \varepsilon_t$												
Student-t	5.58 (0.48)	1.33 (0.37)	2.00 (0.34)	1.38 (0.49)	1.41 (0.33)	0.61 (0.34)	0.64 (0.03)	0.02 (0.02)	0.22 (0.02)	0.16 (0.02)	0.15 (0.02)	0.03 (0.02)
Mixture	5.57 (0.55)	0.50 (0.22)	1.15 (0.19)	0.91 (0.27)	0.51 (0.20)	1.05 (0.19)	0.63 (0.03)	0.08 (0.02)	0.1 (0.02)	0.10 (0.02)	0.01 (0.01)	0.09 (0.01)
ARCH	6.64 (0.55)	1.92 (0.51)	1.43 (0.46)	0.69 (0.71)	1.28 (0.44)	1.16 (0.47)	0.74 (0.03)	0.24 (0.02)	0.03 (0.02)	0.05 (0.04)	0.19 (0.02)	0.17 (0.02)
S(=1.25)	11.72 (0.87)	0.47 (0.17)	0.40 (0.16)	0.78 (0.15)	0.28 (0.17)	0.5 (0.16)	0.76 (0.02)	0.08 (0.01)	0.02 (0.01)	0.04 (0.01)	0.03 (0.01)	0.01 (0.01)

Notes: Table entries correspond to the absolute value of the percentage bias of estimated AR(1) coefficients and are based on 1,000 replications of sample size n=500. Numbers in parentheses are interquartile ranges of the parameter estimates. DGP refers to the data generating process of for Models 1 and 2. The column headings refer to the estimation method: OLS is ordinary least squares, Huber Ψ is Huber's (1964) M-estimator, Bisquare Ψ is an M-estimator that uses a bisquare function, S is the S-Estimator of Rousseeuw and Yohai (1984) , MM is MM-Estimator of Yohai (1986), and τ is the τ -Estimator of Yohai and Zamar (1988).

Table 3.3: Federal Funds Rate Unit Root Test

Estimated autoregression using OLS
$fr_t = \frac{0.05161}{(0.02993)} + \frac{0.9907}{(0.004521)} fr_{t-1}$
Dicky-Fuller Statistic t-test
$(\hat{\rho} - 1)/\sigma = (0.9907 - 1)/0.004521 = -2.05$
5% Critical Value for Dicky-Fuller t-test
-3.51
Therefore the null of a unit root cannot be rejected.

Notes: fr refers to level of the federal funds rate. Parameter estimates are based on the sample March 1984 - December 1998.

Table 3.4: Robust Federal Funds Rate Unit Root Test

Estimated autoregression using MM-estimator
$fr_t = \frac{0.043645}{(0.02000)} + \frac{0.9971}{(0.003021)} fr_{t-1}$
Dicky-Fuller Statistic t-test
$(\hat{\rho} - 1)/\sigma = (0.9971 - 1)/0.003021 = -0.9732$
5% Critical Value for Dicky-Fuller t-test
-3.51
Therefore the null of a unit root cannot be rejected.

Notes: fr refers to level of the federal funds rate. Parameter estimates are based on the sample March 1984 - December 1998. Standard errors are computed according to Yohai (1986). Under assumption of no contamination, the critical values for the MM-based Dicky-Fuller test are nearly identical to those of the OLS-based Dicky-Fuller test when the number of observations is greater than 400 (Lucas 1995).

Table 3.5: AR(3) Parameter Estimates for Federal Funds Rate Changes

Parameter	OLS	Huber Ψ	Bisquare Ψ	S	MM	τ
Φ_0	-1.05 (0.96)	-0.77 (0.64)	-0.38 (0.61)	0.50 (0.76)	-0.41 (0.61)	-0.40 (0.58)
Φ_1	-0.33 (0.04)	-0.33 (0.06)	-0.30 (0.05)	-0.34 (0.06)	-0.3 (0.05)	-0.30 (0.05)
Φ_2	-0.12 (0.04)	-0.02 (0.05)	0.01 (0.00)	0.00 (0.04)	0.02 (0.03)	0.02 (0.03)
Φ_3	-0.20 (0.04)	-0.12 (0.03)	-0.09 (0.02)	-0.06 (0.03)	-0.09 (0.03)	-0.09 (0.02)
s_n	24.77	14.11	14.10	13.98	14.30	14.10

Notes: Parameter estimates are based on the sample March 1984 - December 1996. The column headings refer to the estimation method: OLS is ordinary least squares, Huber Ψ is Huber's (1964) M-estimator, Bisquare Ψ is an M-estimator that uses a bisquare function, S is the S-Estimator of Rousseeuw and Yohai (1984), MM is MM-Estimator of Yohai (1986), and τ is the τ -Estimator of Yohai and Zamar (1988). Standard errors are in parentheses.

Table 3.6: AR(3) Out-of-sample Federal Funds Rate Changes Forecast Summary
Forecast

Horizon	Huber Ψ /LS	Bisquare Ψ /LS	S/LS	MM/LS	τ /LS
1	1.149 (1.67)	1.059 (0.62)	1.07 (0.62)	1.074 (0.78)	1.034 (0.37)
4	0.846 (1.81)	0.579* (4.85)	0.401* (4.29)	0.594* (4.80)	0.565* (4.86)
16	0.682* (3.25)	0.358* (8.41)	0.451* (4.32)	0.364* (8.42)	0.354* (8.48)
26	0.668* (4.04)	0.320* (8.03)	0.493* (4.36)	0.326* (7.99)	0.316* (8.09)

Notes: Table entries are the ratio of out-of-sample mean squared forecast errors of robust estimation to ordinary least squares estimation (LS) for the forecasting period January 1997 - December 1998. Robust estimation methods are: Huber Ψ is Huber's (1964) M-estimator, Bisquare Ψ is an M-estimator that uses a bisquare Ψ function, S is S-Estimator of Rousseeuw and Yohai (1984), MM is MM-Estimator of Yohai (1986), and τ is the τ -Estimator of Yohai and Zamar (1988). The absolute value of the Diebold and Mariano (1995) statistic is in parentheses. These statistics use a Bartlett kernel with a bandwidth parameter of $(k - 1)$ to estimate the spectral density function at the zero frequency. An asterisks indicates statistical significance at the 1-percent level.

Chapter 4

Autoregressive Lag Selection Criteria and Random Contamination Models

4.1 Introduction

Autoregressive (AR) models provide a simple and inexpensive way to forecast time series, and in many cases, the forecasting performance of AR models stacks up well against that of large-scale econometric models. One caveat of using AR models is the need to accurately identify their true lag order. A common approach is to use automatic model selection criteria such as the Akaike Information Criterion (AIC) or the Schwarz Bayesian Information Criterion (BIC). These criteria typically find the minimum of a combination of a goodness-of-fit measure and a model complexity measure. A higher-order lag structure naturally provides a better fit for the data, but comes at the cost of increased complexity. Unfortunately, the goodness-of-fit measure may be unduly influenced by aberrant data such as innovative or additive outliers. Innovative outliers are simply aberrations in the error terms of the model;

an AR model generated from a heavy-tailed distribution would display innovative outliers. Additive outliers, on the other hand, are one-time aberrations that have no inter-temporal effects¹. A clerical error is the classic example of an additive outlier.

While there has been considerable work done examining different methods of choosing AR length, there has been little work done on the effects of outliers on those methods. Basci and Zaman (1998) establish that excess kurtosis, which can be interpreted as non-contamination-based innovation outliers, affects various model selection criteria but only marginally. Le, Raftery, and Martin (1996) argue that additive outliers create downward bias in automatic selection criteria. For instance, they find that under heavy contamination, the BIC selects an AR(0) model instead of the true AR(1) model the majority of the time.

These articles do not include a sensitivity analysis to parameter size and AR length of the true models. In addition, they do not address specific applications to explain the importance of correct AR selection. The goal of this paper is to fill these gaps in the literature by providing both a limited sensitivity analysis to AR parameter size and AR lag length and an application for forecasting contaminated time series. In addition, this essay will also investigate how outlier-robust parameter and scale estimates can limit any bias in the selection criteria.

Through the use of Monte Carlo simulations, this essay shows that innovative

¹Martin(1979) also classifies outliers that follow some autoregressive moving average ARMA process other than the model's to be additive outliers. Tsay (1988) more properly labels these events as transitory or temporary level shifts rather than additive outliers. In the context of this paper, additive outliers do not include these transitory or temporary level shifts.

contaminants² have only a small effect on automatic selection criteria. The effects of additive contaminants are more complex. Additive contaminants do bias the selection criteria, but the degree and direction of that bias depends on the size of the outliers and on the true AR coefficients. In general, those criteria that are affected to a greater degree do worse in forecasting. Robust versions of the automatic selection criteria based on the contaminant-robust MM-estimator (Yohai 1986) are more accurate than their ordinary least squares (OLS) counterparts. However, they still suffer some bias from additive contaminants.

The remainder of the essay will be presented in the following manner. Section 2 gives descriptions and models of additive and innovation type contaminants. Section 3 introduces the automatic selection criteria that will be used. The effects of contamination on the selection criteria are examined in Section 4. Section 5 introduces outlier-robust versions on the selection criteria from Section 3. The simulation setup occurs in Section 6. The simulation results are contained in Section 7. In Section 8, an example using oil prices illustrates the main results of the essay. Section 9 includes conclusions and areas for future research.

4.2 Innovation vs. Additive Contaminants

The essential difference between innovation and additive contaminants is their behavior over time. For clarity, consider Lucas's (1995) innovation and additive models

²The term *contaminant* is used here instead of *outlier* because the data generating process used in the simulations produces contaminants that do not necessarily have to be outliers. The distinction between the two will be discussed in Section 4.2

for an AR process. First, consider the innovation case. Let

$$\Phi(L)(x_t - \mu) = v_t \tag{4.1}$$

where $v_t = \varepsilon_t + w_t$. ε_t is independently and identically distributed with symmetric density function $f(\bullet)$ and distribution function $F(\bullet)$, μ is the mean of x_t , and $\Phi(L)$ is a polynomial in the lag operator L given by $\Phi(L) = 1 - \Phi_1 L - \dots - \Phi_p L^p$ with all roots lying outside the unit circle. Finally, w_t is the contaminating variable that is 0 for most values of t . Notice that the contaminant in this case is only applied to the error term. Therefore, a contaminant, w_t , at time t will still have some effect on the series, x_t at time $t+1$ and beyond depending on the AR parameters.

The additive case is as follows: let

$$\Phi(L)(x_t - \mu - w_t) = \varepsilon_t \tag{4.2}$$

where ε_t is independently and identically distributed with symmetric density function $f(\bullet)$ and distribution function $F(\bullet)$, μ is the mean of x_t , and $\Phi(L)$ is a polynomial in the lag operator L given by $\Phi(L) = 1 - \Phi_1 L - \dots - \Phi_p L^p$ with all roots lying outside the unit circle. Finally, w_t is the contaminating variable that is 0 for most values of t . With this model, the outlier w_t is applied to the series itself only at one specific time. So a contaminant at time t only affects the observed series at time t .

If $w_t = 0 \quad \forall t$, then the two models are identical and equivalent to a standard AR model. In the simulations that follow, w_t will be drawn from a high-variance normal distribution a small percentage of the time and will be zero everywhere else.

Not all values of x_t that have a concurrent contaminant, w_t , have to be outliers. They are only outliers if w_t is large enough to force x_t to fail some model discordancy test.³ It would be simple to define any w_t equal to a ridiculously large or small number to guarantee discordancy for any associated x_t . But in the context of this paper, this is not done. There will exist contaminants that do not force their associated observations to be outliers. Therefore, the model used in the following simulations is more properly labelled as a *random-contamination model* as opposed to an outlier model. Subsequently, I will use the terms innovation contamination and additive contamination as opposed to innovation outliers and additive outliers in the simulation section. However, the term outlier will still be used when referring to previous literature.

The two outlier types have very different effects on time series statistics and it is important to distinguish between them. Outliers of both types create inefficiencies, but additive outliers also create parameter bias.

In the outlier literature, several methods have been proposed to distinguish additive and innovative outliers. Martin and Zeh (1977) proposed a method that examines the differences between M and generalized M (GM) estimators. Tsay (1989) and subsequent work by Chen and Liu (1993) and Balke (1993) have proposed the use of an iterative framework in which they test which type of outlier model fits the data better. Unfortunately, the majority of these identification techniques require

³A simple example of a model discordancy test is to reject the null hypothesis of the model fitting the data if the estimated residual falls outside 2.5 standard deviations from the mean.

a prior model specification. Naturally, if the selection criteria are themselves biased under outliers, these techniques may yield inaccurate results.

4.3 Model Selection Criteria

This essay examines four commonly used automatic selection criteria: the AIC, the BIC, the Hannan-Quinn Criterion HQC, and the sequential F test.

Consider the following time series AR(p) process:

$$\Phi(L)(x_t - \mu) = \varepsilon_t \tag{4.3}$$

with all variables as described in section 2. The goal of any of the selection criteria is to determine p , the true degree of the polynomial $\Phi(L)$. They will do so by searching over a space of possible lag-orders and finding the one that best fits their criterion.

A useful way to look at selection criteria is to see that selection criteria attempt to

$$\max_m P(x_t | AR(m)), \tag{4.4}$$

which is the probability of getting the data given a specified AR model. In general it is not feasible to search over all possible positive integers in order to determine p , therefore it is conventional to determine a prior maximum order, M . The search space for p will then be $m=0,1,\dots,M$.

The computation of $P(x_t | AR(m))$ is not trivial. Several approximations to this probability have been developed in the literature. Schwarz (1978) introduced the

approximation

$$\log p(x_t|AR(m)) \approx \log P(x_t|AR(m), [\hat{\mu}, \hat{\Phi}_1, \dots, \hat{\Phi}_m]) - \frac{m}{2} \log T \quad (4.5)$$

where $\log P(x_t|AR(m), [\hat{\mu}, \hat{\Phi}_1, \dots, \hat{\Phi}_m])$ is equal to the log likelihood function given the model, $AR(m)$ and its estimated parameters. In non-outlier-robust cases, OLS will be used to estimate (μ, Φ) . If we assume that the error terms are i.i.d normal this is equivalent to

$$-\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=0}^T \log(\hat{\sigma}^2) - \frac{1}{2} \sum_{t=0}^T \left(\frac{\hat{\varepsilon}_t}{\hat{\sigma}} \right)^2 \quad (4.6)$$

where the $\hat{\varepsilon}_t$ are the estimated residuals and $\hat{\sigma}$ is the estimated variance of the residuals.

This reduces to

$$-\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\hat{\sigma}^2) - \frac{1}{2}. \quad (4.7)$$

Dropping the constants and multiplying by $-\frac{2}{T}$ the estimate of $\log P(x_t|AR(m))$ becomes

$$\log(\hat{\sigma}^2) + m \frac{\log(T)}{T}. \quad (4.8)$$

This is the widely used Schwarz Bayesian Information Criterion (BIC).

The Akaike Information Criterion (Akaike 1973) is

$$\log(\hat{\sigma}^2) + \frac{2m}{T}, \quad (4.9)$$

which Akaike (1983) later shows is equivalent to the approximation

$$\log P(x_t|AR(m)) \approx \log P(x_t|AR(m), [\hat{\mu}, \hat{\Phi}_1, \dots, \hat{\Phi}_m]) - m. \quad (4.10)$$

A third, and closely related selection criterion, is the HQC (Hannan and Quinn 1979)

$$\log(\hat{\sigma}^2) + \frac{2m \ln(\ln(T))}{T}, \quad (4.11)$$

which makes use of the approximation

$$\log P(x_t|AR(m)) \approx \log P(x_t|AR(m), [\hat{\mu}, \hat{\Phi}_1, \dots, \hat{\Phi}_m]) - m \ln(\ln(T)). \quad (4.12)$$

Therefore, the automatic selection method based on the BIC, AIC, or HQC is to minimize (4.8), (4.9), or (4.11) respectively in reference to $m = 0, \dots, M$ to get \hat{m} , which is the selected lag-order and approximation of p , the true lag-order.

Intuitively, the automatic selection criteria can be thought of as having two parts, a goodness-of-fit measure and a penalty for using higher-order models. The goodness-of-fit measure in the three mentioned criteria is the estimated variance $\hat{\sigma}^2$, while the penalty is different for each of the criteria. For small values of T , the BIC and HQC place less penalty on additional parameters than the AIC, but for large T they actually place a higher penalty on additional parameters. This will have the effect of the AIC tending to select higher lag-orders than the BIC and HQC. In fact, the AIC is not consistent but is biased toward selecting higher lag-orders.

The sequential F test, SEQF, is different from the previous three tests in that it does not use variance as a goodness of fit measure and does not have a specific cost for higher order parameterizations. This automatic selection rule uses the standard t statistics for the parameter estimates to test their significance. To perform this test, start with the maximum possible lag-order M , estimate an AR(M) model, and

compute the t statistic for the M-th parameter estimate. If it is insignificant, drop it, estimate an AR(M-1) model, and check the M-1-th t-statistic. Continue this process until the last parameter estimate is significant. In this study the significance level is set at 5%.

4.4 Effects of Contamination on Selection Criteria

A major source of bias in automatic selection criteria is parameter bias. If there is downward bias in a particular parameter, Φ_i , then the associated lag is less likely to be included. If, however, Φ_i is biased upward, then the associated lag is more likely to be included. Including a lag with a small associated parameter value does not account for much of the variance of the estimated residuals, the goodness-of-fit measure. In this case, the penalty for higher order models tends to dominate. Including a lag with a larger associated parameter value does account for some of the variance of the estimated residuals. Here, the goodness-of-fit measure tends to dominate.

Innovation outliers cause OLS AR parameter estimates to be inefficient but not biased (Martin 1979). The estimate of the constant term, however, is biased (Martin 1979), but this does not greatly affect AR lag-order selection. Therefore, the automatic selection criteria should not be greatly influenced by innovation outliers. The work of Basci and Zaman (1998) supports this claim. To see why there is no bias, consider an AR(1) version of the innovation contamination model from

equation (4.1) without a constant term

$$x_t = \Phi_1 x_{t-1} + \varepsilon_t + w_t \quad (4.13)$$

where ε_t has a mean of 0 and a variance of σ^2 and w_t has a mean of 0 and a variance of σ_w^2 . The OLS estimate of Φ_1 is $\hat{\Phi}_1 = \frac{\sum_{i=1}^{T-1} x_i x_{i+1}}{\sum_{i=1}^{T-1} x_i^2}$. The estimate, $\hat{\Phi}_1$, converges in probability to $\tilde{\Phi}_1 = \rho_1$ where ρ is the first autocorrelation of x , $E[x_t x_{t-1}]/E[x_t^2]$. Substituting in the computed value of ρ from the innovation contamination model results in

$$\tilde{\Phi}_1 = \Phi_1 \frac{\sigma^2 + \sigma_w^2}{\sigma^2 + \sigma_w^2} = \Phi_1. \quad (4.14)$$

Therefore $\hat{\Phi}_1$ is a consistent estimate of Φ_1 ⁴.

Additive contamination is a completely different story. As Martin (1979) illustrates, additive contamination creates AR parameter bias. To see why there is bias with additive contamination, the last experiment is repeated except with model (2). Consider an AR(1) version of the additive contamination model from equation (4.2) without a constant term

$$x_t = \Phi_1 x_{t-1} + \varepsilon_t + (1 - \Phi_1)w_t \quad (4.15)$$

where ε_t has a mean of 0 and a variance of σ^2 and w_t has a mean of 0 and a variance of σ_w^2 . Calculating ρ from $E[x_t x_{t-1}]/E[x_t^2]$ for the additive contamination model results in

$$\tilde{\Phi}_1 = \Phi_1 \frac{\sigma^2}{\sigma^2 + (1 - \Phi_1^2)\sigma_w^2}. \quad (4.16)$$

⁴As with any OLS based estimate of an autoregressive parameter, this estimate is biased for small samples

It is easy to see that $\tilde{\Phi}_1$ has a downward bias for any level of contamination. As $\sigma_w^2 \rightarrow \infty$ $\tilde{\Phi}_1 \rightarrow 0$. If there is no contamination, then the estimate is consistent.

Given this example with an AR(1) it is easy to conclude that additive contamination creates downward bias in the automatic selection criteria as do Le, Raftery, and Martin(1996). However, that conclusion is premature and does not hold for all values of Φ .

To demonstrate this consider estimating the AR(1) model in (4.13) with an AR(2) model. The OLS estimates of Φ_1 and Φ_2 are

$$\begin{bmatrix} \hat{\Phi}_1 \\ \hat{\Phi}_2 \end{bmatrix} = \begin{bmatrix} \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} \\ \sum x_{t-1}x_{t-2} & \sum x_{t-2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_{t-1}x_t \\ \sum x_{t-2}x_t \end{bmatrix} \quad (4.17)$$

These estimates converge in probability to

$$\tilde{\Phi}_1 = \frac{\varrho_1 - \varrho_1\varrho_2}{1 - \varrho_1^2} \quad (4.18)$$

$$\tilde{\Phi}_2 = \frac{\varrho_2 - \varrho_1^2}{1 - \varrho_1^2} \quad (4.19)$$

where ϱ_1 is the first autocorrelation of x and ϱ_2 is the second autocorrelation of x .

Substituting in the values of ϱ_1 and ϱ_2 these equations reduce to

$$\tilde{\Phi}_1 = \Phi_1 \left[\frac{\sigma^2 + \sigma_w^2}{\sigma^2 + \sigma_w^2 + \frac{\sigma_w^2}{\sigma^2}[\sigma^2 + (1 - \Phi^2)\sigma_w^2]} \right] \quad (4.20)$$

$$\tilde{\Phi}_2 = \Phi_1^2 \left[\frac{\sigma_w^2}{\sigma^2 + \sigma_w^2 + \frac{\sigma_w^2}{\sigma^2}[\sigma^2 + (1 - \Phi^2)\sigma_w^2]} \right]. \quad (4.21)$$

Even in this mis-specified case, $\tilde{\Phi}_1$ still displays negative bias. More importantly, notice that for $\sigma_w^2 > 0$, $\tilde{\Phi}_2 > 0$. This upward bias in the second AR parameter may lead to upward bias in automatic selection criteria.

The degree of bias in both parameters, $\tilde{\Phi}_1$ and $\tilde{\Phi}_2$, depends on the true model parameter Φ_1 , the contamination variance, σ_w^2 , and the true variance of the series, σ^2 . Figure 4.1 displays the values of the hypothetical estimated parameters for various true parameter values and contamination variances for $\sigma^2 = 1$.

For any level of contamination, $\tilde{\Phi}_1$ is biased downward: the larger the contamination, the larger the bias. Also, $\tilde{\Phi}_2$ is positive for any positive level of contamination. This bias becomes negligible as σ^2 becomes very large, but the upward bias is more prevalent for higher values of Φ_1 . For a large σ_w^2 relative to σ^2 and Φ_1 close to 1, both $\tilde{\Phi}_1$ and $\tilde{\Phi}_2$ will be close to .5. For a large σ_w^2 relative to σ^2 and Φ_1 relatively small, both $\tilde{\Phi}_1$ and $\tilde{\Phi}_2$ will be close to 0.

4.5 Robust Selection Criteria

There have been a few attempts to make automatic AR selection criteria robust to outliers. Martin (1980) proposed using a bounded-influence likelihood approach that limited the effects of outliers. More recently, Le, Raftery, and Martin (1996) introduced an outlier-robust selection criterion based on a process called smooth filtering (Martin 1981), an outlier-robust Kalman filter. They demonstrate that this criterion performs well when additive outliers are present; they do not test this criterion with innovation outliers.

A simple alternative way to provide some insulation from outliers or simply contamination in the context of model selection is to use robust parameter and

scale estimates instead of the OLS estimates for the above selection criteria. One such robust estimator is the MM estimator introduced by Yohai (1986).⁵ This estimator achieves both a high breakdown point, the percentage of contaminated data the estimator can handle, and high efficiency should the errors conform to a normal distribution. The MM-estimator is a three-part estimator.

In the first step, a high breakdown point estimator is used to compute an initial estimate of the parameter values, $\tilde{\Phi}$. Yohai explains that in the first step the estimator does not have to be efficient. The S-estimator proposed by Rousseeouw and Yohai (1984) and the Least Trimmed Squares estimator purposed by Rousseeouw (1984) provide good initial estimates. However, because of the AR nature of this problem, the residual autocovariance (RA) estimator by Bustos and Yohai (1986) is used. It is faster than and as accurate as the S-estimator in this AR context.

In the second step, an M-estimate of scale, s , is computed from the residuals obtained in the first step. That is, for some function ρ_0 , a scale estimate $s(\varepsilon_1(\tilde{\Phi}), \dots, \varepsilon_n(\tilde{\Phi}))$ is obtained that satisfies

$$\frac{1}{n} \sum_{t=1}^n \rho_0(\varepsilon_t/s(\bullet)) = b \quad (4.22)$$

where $b = E_{\Theta}[\rho(u)]$ for a distribution Θ and $u_t = \varepsilon_t/s$.

In the final step, the MM-estimate Φ_{MM} , is the solution to

$$\min_{\Phi} \sum_{t=1}^n \rho_1(\varepsilon_t, \Phi), \quad (4.23)$$

⁵Pre-made versions of the MM estimator are available in SAS in PROC ROBUSTREG and in S+. My study uses an author-constructed version of Yohai's estimator in OX (Doornik 2001) (see Appendix B).

for some function ρ_1 .

Yohai (1986) suggests choosing ρ_0 and ρ_1 as

$$\rho_B(u_t) = \begin{cases} u_t^2/2 - u_t^4/2 + u_t^6/6 & |u_t| < k_i \\ 1/6 & |u_t| \geq k_i \end{cases} \quad (4.24)$$

for tuning constants k_0 and k_1 , respectively. In the empirical work to follow, $k_0 = 2.9366$, $b = 0.0833$, and $k_1 = 3.44$ which corresponds to 85% efficiency for normally distributed errors. This is set to match the SAS default settings.

Yohai also provides an asymptotic distribution for $\hat{\Phi}$.

$$\hat{\Phi} \sim N \left(\Phi, \sigma^2 \frac{\int_{\Theta} \Psi^2(u_i) d\Theta}{\left[\int_{\Theta} \Psi'(u_i) d\Theta \right]^2} (X'X)^{-1} \right) \quad (4.25)$$

where $\Psi = \rho_1'$ and X is the $T \times p$ vector of right-hand side variables, in the AR case $X_t = [x_{t-1}, x_{t-2}, \dots, x_{t-p}]$. The variance-covariance matrix can be estimated by

$$\frac{s^2}{T} \frac{\sum_i \Psi^2(u_i)}{\left[\sum_i \Psi'(u_i) \right]^2} (X'X)^{-1}. \quad (4.26)$$

The robust selection criteria are created using the above estimator and scale estimate. They are

$$BICMM = \log(s^2) + m \frac{\log(T)}{T} \quad (4.27)$$

$$AICMM = \log(s^2) + \frac{2m}{T} \quad (4.28)$$

$$HQCMM = \log(s^2) + \frac{2m \ln(\ln(T))}{T}. \quad (4.29)$$

The robust sequential F test, SEQFMM, is as described in section 3, but it uses the estimated variance-covariance matrix in equation (4.26).

4.6 Simulations

This study consists of two simulations: 1) a simulation study to determine accuracy of the automatic selection criteria when used with the innovation and additive contamination models and 2) a simulation study to determine the forecasting cost of mis-specifying the true lag order of models with innovation and additive contamination.

For both parts of the simulation study four different model parameterizations are used to formulate the innovation and additive contamination models: A low-persistence AR(1), a high-persistence AR(1), a low-persistence AR(2), and a high-persistence AR(2). Subsequently, each of these models will be matched with high and low variance contamination. Contaminants, w_t , will occur in the data only $\alpha\%$ of the time. When they do occur, they will be drawn from a distribution $G(\bullet)$, as in Lucas (1995). This is done by drawing T numbers from a uniform distribution, and only if the number falls below α , adding the contaminant. The variance of the contamination is simply $\alpha\sigma_{G(\bullet)}^2$. Table 4.1 lists the different parameterizations that are examined. $\sigma^2 = 1$ for all the parameterizations. Each of these parameterizations will be paired with four different contaminant types as indicated in Table 4.2.

For the first part of the simulation study, 500 random series of length 260 are generated using the eight parameterizations. Each of the robust and non-robust selection criteria is then used to estimate lag order⁶. All the criteria search over the

⁶The number 260 was selected because it corresponds to five years of weekly data.

same space $m = 0, 1, \dots, 4$. The distributions of the selected lag order generated from the various selection criteria are reported in Tables 4.3 - 4.10.

In the second part of the simulation study, the same 500 random series of length 260 are used to create out-of-sample forecasts. The forecasts are computed as one-period-ahead forecasts beginning at 234 (90% of the sample) and ending at 260. Parameters are re-estimated with new data for every subsequent period. The results are reported as mean squared forecasting error (MSE) and as relative mean squared forecasting error as compared to the forecasting error of a correctly parameterized model. The results are listed in Tables 4.11 - 4.15.

4.7 Simulation Results

The selection accuracy simulations in Tables 4.3 - 4.10 support the parameter bias theories in Section 4.4. In general, additive outliers in the context of low-persistence parameterizations cause the non-robust selection criteria to under-predict true lag order in all but the smallest contamination case when the contaminating variance is only $9 * 0.05 = .45$. The opposite is true for the high-persistence parameterizations. The non-robust selection criteria tend to over-predict the true lag order in all but the most extreme contamination case when the contamination is from a t distribution with 3 degrees of freedom. Innovation outliers, however, have little effect on the selection criteria. In fact, some selection criteria actually do better when there are

innovation outliers. The robust selection criteria perform better than their non-robust counterparts in the majority of the additive contamination cases.

More specifically, the BIC and HQC perform well in the no-contamination case and with innovation outliers for all the different parameterizations with the exception of model 3. They predict the true lag-order between 92% and 99% of the time for models 1, 2, and 4. Both criteria have a tendency to miss the small AR(2) term in model 3. However, they are greatly affected by moderate and heavy additive contamination. In the heavy-contaminated case with low-persistence models 1 and 3 they correctly assess the true lag order between .6% and 13% of the time and exclusively select AR(0) models. The high-persistence parametrization is more interesting. Both the BIC and HQC greatly over-predict lag-order in all cases. In contrast, they do well in the small-contamination and low-persistence cases. The downward bias of $\hat{\Phi}_1$ is balancing out the upward bias in $\hat{\Phi}_2$.

Their MM-based counterparts, the BICMM and HQCMM, actually perform worse in the no-contamination case for all the models except model 3. For models 1,2, and 4 there is roughly a 15% loss in accuracy when moving to a robust statistic. This is not surprising since the MM estimator was defined to have an efficiency of 85%. They also do slightly worse in all the innovation contamination cases with the exception of model 3. Model 3's second AR parameter is only 0.1 and most of the test statistics have trouble picking it up. The robust techniques do better in finding this parameter. In fact, for innovation contamination, robust techniques actually

do better as the contamination variance gets larger. For example, the HQCMM only selects the correct lag order 33% of the time when there is no contamination, but it selects the correct lag order 69.2% of the time in the extreme-innovation contamination case.

These outlier-robust criteria are superior to the HQC and BIC in the moderate-contamination and heavy-contamination cases for both low-persistence and high-persistence when there are additive outliers. They predict the correct lag order close to 67.6% of the time as compared to only 13% for the HQC and 4.2 % for the BIC in regards to model 1 with extreme-additive contamination. They also do well in model 2 with extreme-additive contamination. In this case the BICMM predicts the correct lag order 92.4% of the time and the HQCMM predicts the correct lag order 88.2% of the time this is opposed to close to 0% of the time for the non-robust versions. Model 4 is more interesting. The HQCMM and BICMM do better in all the additive contamination cases, but they begin to under-select the correct lag order when the contaminating variance becomes large.

The AIC and the sequential F perform poorly in terms of accuracy. They tend to over-predict the true lag order in the innovation and no-contamination cases. They also suffer from the same flaws as the BIC and HQC under the additive contamination case. The over-prediction is actually beneficial in the additive case for the low-persistence cases. For model 1, the AIC correctly selects the true lag order 19.6% of the time as opposed to just 4.2% for the BIC. The AICMM and the

MM based sequential F do better in additive cases. In fact, the MM based sequential F is the best selection criterion in models 3 and 4 with additive contamination. In terms of forecasting this over-prediction may also have benefits.

While one order-selection method may be more accurate in detecting the true lag order, this does not necessarily imply that it will produce a superior forecast when there are outliers, or even in cases without outliers. Under-specifying the lag-order is ordinarily more detrimental because important time series information is lost. Conversely, over-specification is not as costly. The AR parameters on the unneeded lags are small and thus forecasting ability is only slightly decreased. In the high-persistence models 2 and 4, over-selection is actually preferred when there are additive outliers. In this case, parameters on higher-order lagged variables are able to pick up some of the persistence lost from the downward bias in the first AR parameter.

A selection criterion that has a downward bias in its lag-order prediction but is fairly accurate otherwise might perform worse, in forecasting terms, than a less accurate but upwardly biased estimator. The AIC, for instance, has a tendency to over-select lag order, but this actually aids in its forecasting ability when there are additive outliers and there is low-persistence (small AR parameters). Surprisingly, over-selection or even correct lag-order selection is not always preferable. When there are extreme additive outliers, under-selection can decrease mean squared forecasting error. Models 1 and 3 show decreases in mean squared forecasting error.

An AR(0) is selected over an AR(1) when there is extreme additive contamination for both MM and OLS. The MM forecasts are always better when selecting an AR(0) over an AR(1) in the extreme contamination case. Then parameter estimates become so poor that using a simple mean is preferred.

The forecasting costs of mis-specification differ greatly between the various contamination size and type, and model parametrization. In general, parametrization has little effect on the models with low-persistence; this is true for both OLS and MM based forecasts. Mis-specifying a high-persistence model, however, is very costly. In the high-persistence AR(1) case with moderate additive-contamination, the MSE is 163.854% greater for an AR(0) as opposed to the MSE with the estimated AR(1) when using OLS, and 147.110% greater when using MM. This large cost does go down as the contaminating variance increases. In the extreme variance case, under-selection is preferred.

This seems to imply that the most accurate selection criteria are not the best for use in forecasting, especially when outliers are present. The upwardly biased AIC should outperform the more accurate BIC and HQC in terms of forecasting performance simply because it over-predicts more often. Robust selection criteria are more accurate in terms of lag-order selection, but their forecasts are roughly equivalent.⁷

⁷This is not surprising for these short horizon forecasts. As indicated in chapter 3, the only benefit of using robust estimation is in long-run forecast.

4.8 Example with Oil Prices

Oil prices are notorious for being volatile with occasionally large spikes. Therefore, it is not unreasonable to believe that the oil price series exhibit innovation outliers, additive outliers, or even both. Thus, oil prices provide an interesting platform in which to test the implication that outliers affect model specification and therefore forecasts.

This section examines out-of-sample forecasts of real oil price data. The series is the changes in the “Freight on Board Import Price for the United States” from the U.S. Energy Information Agency listed in cents per barrel, recorded monthly from January 1974-June 2002 the prices were adjusted for inflation using the produce price index (PPI). Changes are used here instead of levels because the null hypothesis of a unit root cannot be rejected. See Tables 4.15 and 4.16 for the results of unit root tests using OLS and MM respectively. The forecasts are constructed using the same technique as in Section 6.

In addition, the forecast errors for the alternative models are compared using a simple Morgan-Granger-Newbold test (MGN) (Granger and Newbold 1977). Let $x_t = e_t^j + e_t^i$ and $z_t = e_t^j - e_t^i$ where j represents AR(0) forecasts using OLS and i represents the alternative model. The test statistic is constructed as

$$\frac{\hat{\rho}_{xz}}{\sqrt{\frac{1-\hat{\rho}_{xz}}{T-1}}} \sim t(0, T-1)$$

where $\hat{\rho}_{xz} = \frac{x'z}{\sqrt{(x'x)(z'z)}}$.

Table 4.17 displays the lag-order selection of each of the automatic selection criteria, the root mean squared forecasting error, and the MGN t -statistic based on the chosen model using either OLS or MM-based forecasts. The OLS and MM parameter estimates based on the selected models are reported in Table 4.18.

Notice that in Table 4.18 the MM estimator finds a lower AR(1) parameter but a higher AR(2) parameter in comparison to OLS. Additionally, the OLS estimation finds a significant AR(4) parameter that is not found in the MM estimation. Therefore, it is not surprising that the majority of the OLS-based selection criteria choose an AR(4) model while the MM-based selection criteria choose an AR(2) model. The exception is the BIC that chooses an AR(1) when based on either OLS or MM. This does not necessarily fit the profile of a regression that has additive contamination. If that had been the case, the AR(1) parameter should have been larger and the parameters on higher lags should have been smaller in the MM estimation. This could indicate innovation contamination or some other data problems such as transitory level shifts (see Martin 1979 for the effects of a transitory level shift on the estimation of an AR(1) model.)

The forecasts for the lower lag-order models chosen by the robust selection criteria are slightly better than those from those selected by the OLS-based selection criteria when robust estimation is used to create the forecasts. The forecast based on the AR(1) model selected by the BIC and BICMM is superior to the forecast based on the AR(4) model selected by the other OLS criteria, but inferior to the MM

estimated AR(2) model selected by the MM criteria. However, the MGN t -statistics indicates no significant differences between the forecasts.

The amount of aberrant data in the changes in oil price series makes it difficult to forecast using simple AR models. The out-of-sample forecasts based on lower order AR models and robust estimation do better, but not significantly. Overall, the robust estimation techniques do find some previously unnoticed correlation between current oil price changes and two month lagged oil price changes. Additionally, they indicate no relationship between current oil price changes and four month oil price changes. Unfortunately, this added knowledge may not be very useful to forecasters considering the highly variable nature of oil prices.

4.9 Conclusions and Directions for Future Work

The relationship between contamination and automatic selection criteria is complex. The effects of contamination on selection criteria are dependent on the size, frequency, and type of the contamination and the true nature of the model. Table 4.19 provides a summary of the performances of the selection criteria.

Innovation contamination does not greatly affect automatic AR model selection criteria. However, innovation contamination masks some small AR parameters and causes OLS-based criteria to be incorrect when there are small AR parameters. MM-based criteria actually do better with innovation outliers than without them, but still perform worse than OLS in most cases. The one exception is when there are

small AR parameters, then the MM-based criteria are better than the OLS-based criteria. In general, if innovation contamination is suspected, OLS-based techniques should be acceptable unless small AR parameters are expected.

Over-lagging actually helps in the additive contamination cases when the true model exhibits low levels of persistence. However, this same over-selection will be a liability in cases when the true model exhibits high levels of persistence. Therefore, techniques such as the AIC and sequential F will be better suited for low-persistence models, and the HQC and BIC will be better for high-persistence models. Robust estimation techniques offer some insulation from additive contamination; however, they still are somewhat biased.

Selecting the correct lag order can be important in terms of forecasting. It is very costly to select an AR(0) model, when in fact the true model has a higher lag order. As one would suspect, it is more costly to under-select a highly-persistent model than to under-select a lightly-persistent model. Under-selecting is unlikely to occur when one accounts for the directional bias in the selection criteria. High-persistence leads to over-selection, which is not that costly in terms of forecasting high-persistent models. Low persistence leads to under-selection, which is very costly and should be avoided. Fortunately, the cost of under-selecting with a lower-persistence model is not that great.

A simple rule of thumb would be to try both a robust and non-robust technique. If you get very different results, you might have an additive outlier problem and

should probably go with the robust-selected model. The BICMM, for instance, does well in most cases.

Forecasting is not the only case where there could be costs for mis-specifying lag lengths in the presence of outliers. Lucas (1995) shows that unit-root testing is affected by additive outliers. The first step of an Augmented Dicky-Fueller unit-root test is usually to select lag order for the model. As shown in the context of this essay, lag-order selection can be difficult if there are additive outliers. In high-persistence cases – which would be the norm for unit-root testing – with additive outliers, these two things generally happened. There was downward parameter bias in the first AR parameter and upward bias in subsequent parameters. Then because of this, automatic selection criteria generally over-selected the lag order. But does the upward bias in the higher AR terms, make up for the downward bias in the first AR term? This remains to be seen and is a good question for future research.

Figure 4.1: Theoretical Parameter Estimates for Additive Contaminant Model

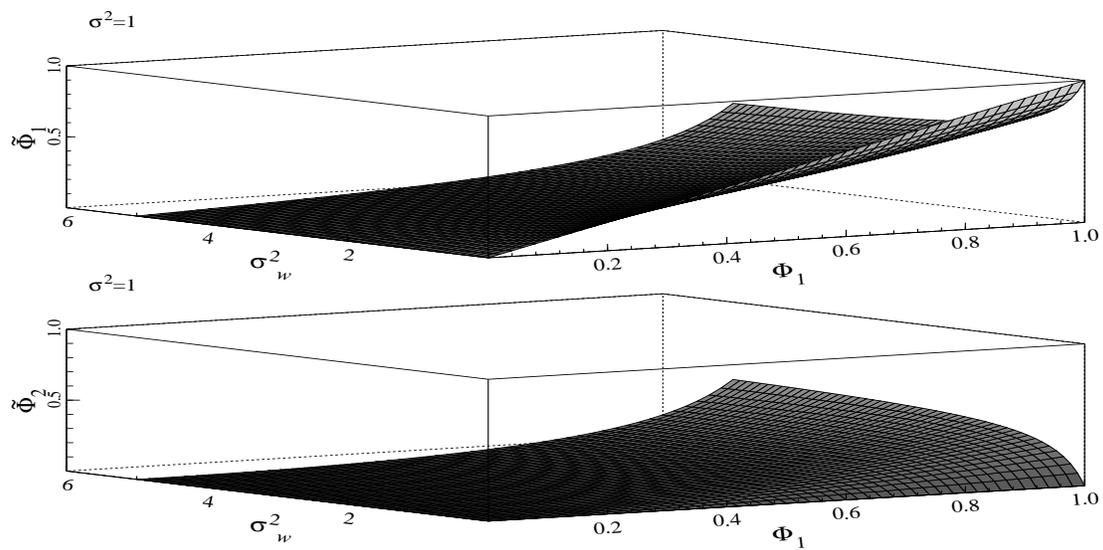


Table 4.1: Parametrization of Innovation Contamination and Additive Contamination Models

Model	μ	$\bar{\Phi}_1$	$\bar{\Phi}_2$
1	2	0.5	-
2	2	0.9	-
3	2	0.3	0.1
4	2	0.5	0.35

Table 4.2: Contamination Types

Parametrization	α	$G(\bullet)$
no contamination	0	-
small contamination	0.05	N(0,9)
moderate contamination	0.10	N(0,9)
heavy contamination	0.10	N(0,100)

Table 4.3: Distributions of Selection Criteria for Model 1: $X_t - 2 - .5X_{t-1} = \varepsilon_t$ with Innovation Contamination

Criterion	Contamination	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
BIC	0%	0.000	0.986	0.014	0.000	0.000
	5% N(0,9)	0.000	0.988	0.012	0.000	0.000
	10% N(0,9)	0.000	0.986	0.012	0.002	0.000
	10% N(0,100)	0.000	0.980	0.018	0.002	0.000
AIC	0%	0.000	0.878	0.092	0.018	0.012
	5% N(0,9)	0.000	0.886	0.084	0.018	0.012
	10% N(0,9)	0.000	0.898	0.074	0.018	0.010
	10% N(0,100)	0.000	0.906	0.046	0.036	0.012
HQC	0%	0.000	0.960	0.034	0.006	0.000
	5% N(0,9)	0.000	0.964	0.030	0.004	0.002
	10% N(0,9)	0.000	0.946	0.050	0.002	0.002
	10% N(0,100)	0.000	0.958	0.030	0.006	0.006
SEQF	0%	0.000	0.858	0.038	0.056	0.048
	5% N(0,9)	0.000	0.840	0.052	0.058	0.050
	10% N(0,9)	0.000	0.872	0.052	0.030	0.046
	10% N(0,100)	0.000	0.872	0.042	0.034	0.052
BICMM	0%	0.000	0.876	0.088	0.024	0.012
	5% N(0,9)	0.000	0.890	0.070	0.026	0.014
	10% N(0,9)	0.000	0.904	0.064	0.020	0.012
	10% N(0,100)	0.000	0.874	0.094	0.020	0.012
AICMM	0%	0.000	0.582	0.176	0.118	0.124
	5% N(0,9)	0.000	0.592	0.164	0.112	0.132
	10% N(0,9)	0.000	0.624	0.158	0.116	0.102
	10% N(0,100)	0.000	0.718	0.178	0.064	0.040
HQCMM	0%	0.002	0.772	0.124	0.068	0.034
	5% N(0,9)	0.000	0.736	0.132	0.082	0.050
	10% N(0,9)	0.000	0.764	0.120	0.072	0.044
	10% N(0,100)	0.000	0.790	0.140	0.050	0.020
SEQFMM	0%	0.000	0.856	0.030	0.062	0.052
	5% N(0,9)	0.000	0.844	0.032	0.052	0.072
	10% N(0,9)	0.000	0.828	0.056	0.050	0.066
	10% N(0,100)	0.000	0.608	0.136	0.120	0.136

Notes: Table entries refer to the percentage of time each AR model was selected. Simulations were based on 500 runs of sample size 260. The BIC, AIC, HQC, and SEQF are estimated using ordinary least squares and the BICMM, AICMM, HQCMM, and SEQFMM are estimated using MM.

Table 4.4: Distributions of Selection Criteria for Model 1: $X_t - 2 - .5X_{t-1} = \varepsilon_t$ with Additive Contamination

Criterion	Contamination	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
BIC	0%	0.000	0.992	0.006	0.002	0.000
	5% N(0,9)	0.002	0.968	0.026	0.004	0.000
	10% N(0,9)	0.018	0.902	0.078	0.002	0.000
	10% N(0,100)	0.948	0.042	0.008	0.002	0.000
AIC	0%	0.000	0.880	0.068	0.032	0.020
	5% N(0,9)	0.000	0.802	0.144	0.034	0.020
	10% N(0,9)	0.006	0.722	0.220	0.042	0.010
	10% N(0,100)	0.704	0.196	0.050	0.034	0.016
HQC	0%	0.000	0.950	0.042	0.006	0.002
	5% N(0,9)	0.002	0.928	0.060	0.008	0.002
	10% N(0,9)	0.004	0.856	0.126	0.010	0.004
	10% N(0,100)	0.836	0.130	0.022	0.010	0.002
SEQF	0%	0.000	0.860	0.050	0.048	0.042
	5% N(0,9)	0.000	0.804	0.098	0.040	0.058
	10% N(0,9)	0.000	0.750	0.162	0.040	0.048
	10% N(0,100)	0.704	0.162	0.056	0.032	0.046
BICMM	0%	0.002	0.866	0.094	0.024	0.014
	5% N(0,9)	0.006	0.874	0.074	0.034	0.012
	10% N(0,9)	0.028	0.802	0.130	0.024	0.016
	10% N(0,100)	0.218	0.676	0.096	0.010	0.000
AICMM	0%	0.004	0.552	0.166	0.142	0.136
	5% N(0,9)	0.002	0.568	0.186	0.130	0.114
	10% N(0,9)	0.006	0.520	0.230	0.132	0.112
	10% N(0,100)	0.180	0.634	0.108	0.052	0.026
HQCMM	0%	0.004	0.750	0.120	0.066	0.060
	5% N(0,9)	0.002	0.742	0.132	0.092	0.032
	10% N(0,9)	0.014	0.668	0.182	0.080	0.056
	10% N(0,100)	0.188	0.676	0.090	0.030	0.016
SEQFMM	0%	0.000	0.878	0.036	0.042	0.044
	5% N(0,9)	0.000	0.814	0.056	0.052	0.078
	10% N(0,9)	0.000	0.752	0.106	0.070	0.072
	10% N(0,100)	0.010	0.582	0.184	0.124	0.100

Notes: Table entries refer to the percentage of time each AR model was selected. Simulations were based on 500 runs of sample size 260. The BIC, AIC, HQC, and SEQF are estimated using ordinary least squares and the BICMM, AICMM, HQCMM, and SEQFMM are estimated using MM.

Table 4.5: Distributions of Selection Criteria for Model 2: $X_t - 2 - .9X_{t-1} = \varepsilon_t$ with Innovation Contamination

Criterion	Contamination	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
BIC	0%	0.000	0.992	0.008	0.000	0.000
	5% N(0,9)	0.000	0.992	0.008	0.000	0.000
	10% N(0,9)	0.000	0.988	0.010	0.002	0.000
	10% N(0,100)	0.000	0.976	0.022	0.002	0.000
AIC	0%	0.000	0.858	0.100	0.036	0.006
	5% N(0,9)	0.000	0.874	0.090	0.022	0.014
	10% N(0,9)	0.000	0.880	0.078	0.030	0.012
	10% N(0,100)	0.000	0.914	0.042	0.034	0.010
HQC	0%	0.000	0.958	0.036	0.006	0.000
	5% N(0,9)	0.000	0.964	0.028	0.004	0.004
	10% N(0,9)	0.000	0.962	0.032	0.006	0.000
	10% N(0,100)	0.000	0.928	0.054	0.008	0.010
SEQF	0%	0.000	0.870	0.048	0.038	0.044
	5% N(0,9)	0.000	0.862	0.052	0.042	0.044
	10% N(0,9)	0.000	0.866	0.040	0.042	0.052
	10% N(0,100)	0.000	0.834	0.054	0.054	0.058
BICMM	0%	0.000	0.874	0.084	0.022	0.020
	5% N(0,9)	0.000	0.892	0.068	0.030	0.010
	10% N(0,9)	0.000	0.882	0.080	0.032	0.006
	10% N(0,100)	0.000	0.912	0.072	0.014	0.002
AICMM	0%	0.000	0.568	0.164	0.138	0.130
	5% N(0,9)	0.000	0.576	0.154	0.138	0.132
	10% N(0,9)	0.000	0.620	0.138	0.122	0.120
	10% N(0,100)	0.000	0.772	0.142	0.048	0.038
HQCMM	0%	0.000	0.726	0.168	0.068	0.038
	5% N(0,9)	0.000	0.754	0.138	0.056	0.052
	10% N(0,9)	0.000	0.768	0.152	0.050	0.030
	10% N(0,100)	0.000	0.852	0.086	0.036	0.026
SEQFMM	0%	0.000	0.832	0.042	0.052	0.074
	5% N(0,9)	0.000	0.840	0.046	0.048	0.066
	10% N(0,9)	0.000	0.806	0.054	0.076	0.064
	10% N(0,100)	0.000	0.632	0.104	0.132	0.132

Notes: Table entries refer to the percentage of time each AR model was selected. Simulations were based on 500 runs of sample size 260. The BIC, AIC, HQC, and SEQF are estimated using ordinary least squares and the BICMM, AICMM, HQCMM, and SEQFMM are estimated using MM.

Table 4.6: Distributions of Selection Criteria for Model 2: $X_t - 2 - .9X_{t-1} = \varepsilon_t$ with Additive Contamination

Criterion	Contamination	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
BIC	0%	0.000	0.984	0.016	0.000	0.000
	5% N(0,9)	0.000	0.298	0.652	0.048	0.002
	10% N(0,9)	0.000	0.042	0.796	0.150	0.012
	10% N(0,100)	0.000	0.002	0.156	0.562	0.280
AIC	0%	0.000	0.892	0.078	0.020	0.010
	5% N(0,9)	0.000	0.118	0.736	0.112	0.034
	10% N(0,9)	0.000	0.010	0.556	0.360	0.074
	10% N(0,100)	0.000	0.000	0.010	0.270	0.720
HQC	0%	0.000	0.956	0.036	0.008	0.000
	5% N(0,9)	0.000	0.184	0.690	0.110	0.016
	10% N(0,9)	0.000	0.026	0.684	0.266	0.024
	10% N(0,100)	0.000	0.004	0.042	0.402	0.552
SEQF	0%	0.000	0.850	0.048	0.056	0.046
	5% N(0,9)	0.000	0.150	0.648	0.150	0.052
	10% N(0,9)	0.000	0.010	0.602	0.304	0.084
	10% N(0,100)	0.000	0.000	0.016	0.366	0.618
BICMM	0%	0.000	0.878	0.082	0.026	0.014
	5% N(0,9)	0.000	0.828	0.126	0.036	0.010
	10% N(0,9)	0.000	0.810	0.136	0.042	0.012
	10% N(0,100)	0.000	0.924	0.046	0.026	0.004
AICMM	0%	0.000	0.584	0.150	0.118	0.148
	5% N(0,9)	0.000	0.626	0.160	0.124	0.090
	10% N(0,9)	0.000	0.630	0.168	0.104	0.098
	10% N(0,100)	0.000	0.840	0.062	0.054	0.044
HQCMM	0%	0.000	0.774	0.124	0.052	0.050
	5% N(0,9)	0.000	0.734	0.142	0.060	0.064
	10% N(0,9)	0.000	0.710	0.176	0.068	0.046
	10% N(0,100)	0.000	0.882	0.062	0.044	0.012
SEQFMM	0%	0.000	0.834	0.036	0.058	0.072
	5% N(0,9)	0.000	0.704	0.134	0.076	0.086
	10% N(0,9)	0.000	0.526	0.306	0.084	0.084
	10% N(0,100)	0.000	0.516	0.240	0.146	0.098

Notes: Table entries refer to the percentage of time each AR model was selected. Simulations were based on 500 runs of sample size 260. The BIC, AIC, HQC, and SEQF are estimated using ordinary least squares and the BICMM, AICMM, HQCMM, and SEQFMM are estimated using MM.

Table 4.7: Distributions of Selection Criteria for Model 3: $X_t - 2 - .3X_{t-1} - .1X_{t-2} = \varepsilon_t$ with Innovation Contamination

Criterion	Contamination	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
BIC	0%	0.006	0.796	0.194	0.004	0.000
	5% N(0,9)	0.002	0.842	0.154	0.002	0.000
	10% N(0,9)	0.000	0.856	0.142	0.002	0.000
	10% N(0,100)	0.002	0.904	0.090	0.002	0.002
AIC	0%	0.002	0.494	0.432	0.060	0.012
	5% N(0,9)	0.000	0.540	0.394	0.046	0.020
	10% N(0,9)	0.000	0.524	0.412	0.042	0.022
	10% N(0,100)	0.006	0.586	0.352	0.038	0.018
HQC	0%	0.002	0.622	0.352	0.022	0.002
	5% N(0,9)	0.000	0.688	0.298	0.008	0.006
	10% N(0,9)	0.002	0.704	0.282	0.008	0.004
	10% N(0,100)	0.004	0.748	0.214	0.024	0.010
SEQF	0%	0.000	0.574	0.312	0.072	0.042
	5% N(0,9)	0.004	0.560	0.318	0.054	0.064
	10% N(0,9)	0.000	0.606	0.310	0.042	0.042
	10% N(0,100)	0.000	0.606	0.310	0.046	0.038
BICMM	0%	0.038	0.600	0.292	0.058	0.012
	5% N(0,9)	0.016	0.550	0.368	0.050	0.016
	10% N(0,9)	0.004	0.526	0.392	0.056	0.022
	10% N(0,100)	0.000	0.208	0.684	0.092	0.016
AICMM	0%	0.020	0.314	0.380	0.162	0.124
	5% N(0,9)	0.006	0.336	0.392	0.152	0.114
	10% N(0,9)	0.000	0.250	0.454	0.152	0.144
	10% N(0,100)	0.000	0.130	0.606	0.176	0.088
HQCMM	0%	0.014	0.474	0.344	0.092	0.076
	5% N(0,9)	0.010	0.412	0.436	0.094	0.048
	10% N(0,9)	0.000	0.344	0.470	0.128	0.058
	10% N(0,100)	0.000	0.124	0.692	0.130	0.054
SEQFMM	0%	0.002	0.646	0.248	0.040	0.064
	5% N(0,9)	0.000	0.580	0.312	0.062	0.046
	10% N(0,9)	0.000	0.492	0.378	0.056	0.074
	10% N(0,100)	0.000	0.062	0.664	0.128	0.146

Notes: Table entries refer to the percentage of time each AR model was selected. Simulations were based on 500 runs of sample size 260. The BIC, AIC, HQC, and SEQF are estimated using ordinary least squares and the BICMM, AICMM, HQCMM, and SEQFMM are estimated using MM.

Table 4.8: Distributions of Selection Criteria for Model 3: $X_t - 2 - .3X_{t-1} - .1X_{t-2} = \varepsilon_t$ with Additive Contamination

Criterion	Contamination	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
BIC	0%	0.006	0.796	0.194	0.004	0.000
	5% N(0,9)	0.106	0.760	0.132	0.002	0.000
	10% N(0,9)	0.302	0.596	0.100	0.002	0.000
	10% N(0,100)	0.952	0.040	0.006	0.002	0.000
AIC	0%	0.002	0.560	0.392	0.034	0.012
	5% N(0,9)	0.022	0.520	0.382	0.060	0.016
	10% N(0,9)	0.098	0.566	0.284	0.032	0.020
	10% N(0,100)	0.814	0.100	0.040	0.036	0.010
HQC	0%	0.000	0.696	0.284	0.016	0.004
	5% N(0,9)	0.054	0.628	0.294	0.022	0.002
	10% N(0,9)	0.156	0.624	0.204	0.014	0.002
	10% N(0,100)	0.918	0.050	0.016	0.012	0.004
SEQF	0%	0.000	0.584	0.320	0.038	0.058
	5% N(0,9)	0.032	0.580	0.278	0.058	0.052
	10% N(0,9)	0.128	0.592	0.198	0.052	0.030
	10% N(0,100)	0.764	0.096	0.062	0.048	0.030
BICMM	0%	0.022	0.574	0.322	0.062	0.020
	5% N(0,9)	0.074	0.540	0.308	0.052	0.026
	10% N(0,9)	0.144	0.510	0.262	0.054	0.030
	10% N(0,100)	0.588	0.354	0.046	0.010	0.002
AICMM	0%	0.008	0.322	0.376	0.142	0.152
	5% N(0,9)	0.046	0.306	0.370	0.130	0.148
	10% N(0,9)	0.068	0.342	0.326	0.144	0.120
	10% N(0,100)	0.480	0.330	0.096	0.054	0.040
HQCMM	0%	0.024	0.442	0.370	0.100	0.064
	5% N(0,9)	0.024	0.464	0.328	0.116	0.068
	10% N(0,9)	0.108	0.410	0.302	0.112	0.068
	10% N(0,100)	0.562	0.338	0.058	0.024	0.018
SEQFMM	0%	0.000	0.652	0.262	0.036	0.050
	5% N(0,9)	0.012	0.586	0.272	0.074	0.056
	10% N(0,9)	0.034	0.554	0.258	0.086	0.068
	10% N(0,100)	0.072	0.392	0.252	0.126	0.158

Notes: Table entries refer to the percentage of time each AR model was selected. Simulations were based on 500 runs of sample size 260. The BIC, AIC, HQC, and SEQF are estimated using ordinary least squares and the BICMM, AICMM, HQCMM, and SEQFMM are estimated using MM.

Table 4.9: Distributions of Selection Criteria for Model 4: $X_t - 2 - .5X_{t-1} - .35X_{t-2} = \varepsilon_t$ with Innovation Contamination

Criterion	Contamination	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
BIC	0%	0.000	0.000	0.986	0.014	0.000
	5% N(0,9)	0.000	0.000	0.994	0.006	0.000
	10% N(0,9)	0.000	0.000	0.992	0.004	0.004
	10% N(0,100)	0.000	0.002	0.982	0.016	0.000
AIC	0%	0.000	0.000	0.898	0.076	0.026
	5% N(0,9)	0.000	0.000	0.902	0.070	0.028
	10% N(0,9)	0.000	0.000	0.912	0.070	0.018
	10% N(0,100)	0.000	0.000	0.908	0.066	0.026
HQC	0%	0.000	0.000	0.962	0.032	0.006
	5% N(0,9)	0.000	0.000	0.960	0.034	0.006
	10% N(0,9)	0.000	0.000	0.960	0.030	0.010
	10% N(0,100)	0.000	0.002	0.960	0.028	0.010
SEQF	0%	0.000	0.000	0.904	0.044	0.052
	5% N(0,9)	0.000	0.000	0.908	0.040	0.052
	10% N(0,9)	0.000	0.000	0.922	0.044	0.034
	10% N(0,100)	0.000	0.000	0.926	0.042	0.032
BICMM	0%	0.000	0.022	0.882	0.072	0.024
	5% N(0,9)	0.000	0.022	0.876	0.072	0.030
	10% N(0,9)	0.000	0.008	0.916	0.058	0.018
	10% N(0,100)	0.000	0.000	0.872	0.108	0.020
AICMM	0%	0.000	0.018	0.606	0.212	0.164
	5% N(0,9)	0.000	0.012	0.616	0.220	0.152
	10% N(0,9)	0.000	0.002	0.628	0.210	0.160
	10% N(0,100)	0.000	0.000	0.712	0.196	0.092
HQCMM	0%	0.000	0.032	0.734	0.134	0.100
	5% N(0,9)	0.000	0.020	0.750	0.136	0.094
	10% N(0,9)	0.000	0.002	0.748	0.148	0.102
	10% N(0,100)	0.000	0.000	0.812	0.128	0.060
SEQFMM	0%	0.000	0.000	0.882	0.060	0.058
	5% N(0,9)	0.000	0.000	0.856	0.076	0.068
	10% N(0,9)	0.000	0.000	0.836	0.080	0.084
	10% N(0,100)	0.000	0.000	0.746	0.104	0.150

Notes: Table entries refer to the percentage of time each AR model was selected. Simulations were based on 500 runs of sample size 260. The BIC, AIC, HQC, and SEQF are estimated using ordinary least squares and the BICMM, AICMM, HQCMM, and SEQFMM are estimated using MM.

Table 4.10: Distributions of Selection Criteria for Model 4: $X_t - 2 - .5X_{t-1} - .35X_{t-2} = \varepsilon_t$ with Additive Contamination

Criterion	Contamination	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
BIC	0%	0.000	0.000	0.982	0.018	0.000
	5% N(0,9)	0.000	0.000	0.790	0.204	0.006
	10% N(0,9)	0.000	0.000	0.500	0.436	0.064
	10% N(0,100)	0.012	0.068	0.252	0.390	0.278
AIC	0%	0.000	0.000	0.872	0.100	0.028
	5% N(0,9)	0.000	0.000	0.486	0.394	0.120
	10% N(0,9)	0.000	0.000	0.194	0.530	0.276
	10% N(0,100)	0.006	0.008	0.060	0.260	0.666
HQC	0%	0.000	0.000	0.966	0.034	0.000
	5% N(0,9)	0.000	0.000	0.574	0.344	0.082
	10% N(0,9)	0.000	0.000	0.306	0.530	0.164
	10% N(0,100)	0.004	0.026	0.112	0.360	0.498
SEQF	0%	0.000	0.000	0.892	0.062	0.046
	5% N(0,9)	0.000	0.000	0.508	0.368	0.124
	10% N(0,9)	0.000	0.000	0.280	0.514	0.206
	10% N(0,100)	0.000	0.008	0.082	0.338	0.572
BICMM	0%	0.000	0.032	0.862	0.084	0.022
	5% N(0,9)	0.000	0.058	0.746	0.148	0.048
	10% N(0,9)	0.000	0.070	0.622	0.228	0.080
	10% N(0,100)	0.002	0.508	0.376	0.082	0.032
AICMM	0%	0.000	0.014	0.606	0.202	0.178
	5% N(0,9)	0.000	0.042	0.584	0.214	0.160
	10% N(0,9)	0.000	0.042	0.378	0.302	0.278
	10% N(0,100)	0.002	0.448	0.314	0.142	0.094
HQCMM	0%	0.000	0.056	0.748	0.118	0.078
	5% N(0,9)	0.000	0.038	0.628	0.194	0.140
	10% N(0,9)	0.000	0.050	0.508	0.298	0.144
	10% N(0,100)	0.002	0.452	0.348	0.130	0.068
SEQFMM	0%	0.000	0.000	0.916	0.042	0.042
	5% N(0,9)	0.000	0.000	0.786	0.136	0.078
	10% N(0,9)	0.000	0.000	0.570	0.294	0.136
	10% N(0,100)	0.000	0.022	0.400	0.286	0.292

Notes: Table entries refer to the percentage of time each AR model was selected. Simulations were based on 500 runs of sample size 260. The BIC, AIC, HQC, and SEQF are estimated using ordinary least squares and the BICMM, AICMM, HQCMM, and SEQFMM are estimated using MM.

Table 4.11: Forecasting Costs of Misspecification for Model 1: $X_t - 2 - .5X_{t-1} = \varepsilon_t$
Under Various Contaminations

Contamination		OLS				MM		
Type	Size	<i>Selected Model</i>	MSE	<i>Quartile Range</i>	<i>%Relative Loss</i>	MSE	<i>Quartile Range</i>	<i>%Relative Loss</i>
None	0%	AR(0)	1.349	0.272	33.992	1.329	0.333	31.946
		AR(1)	1.007	0.203	0.000	1.016	0.197	0.933
		AR(2)	1.012	0.200	0.506	1.025	0.199	1.827
		AR(3)	1.017	0.194	1.033	1.036	0.191	2.860
		AR(4)	1.022	0.191	1.460	1.043	0.196	3.605
	5% N(0,9)	AR(0)	1.85	0.407	29.544	1.955	0.452	36.880
		AR(1)	1.428	0.267	0.000	1.449	0.291	1.456
		AR(2)	1.432	0.276	0.280	1.454	0.291	1.757
		AR(3)	1.438	0.287	0.672	1.465	0.299	2.576
		AR(4)	1.444	0.278	1.085	1.476	0.295	3.332
IC	10% N(0,9)	AR(0)	2.524	0.582	33.107	2.496	0.658	32.978
		AR(1)	1.896	0.467	0.000	1.877	0.473	0.000
		AR(2)	1.9	0.473	0.227	1.887	0.467	0.554
		AR(3)	1.906	0.467	0.517	1.89	0.464	0.677
		AR(4)	1.912	0.484	0.854	1.904	0.483	1.412
	10% N(0,100)	AR(0)	15.62	4.859	32.722	14.89	5.881	32.571
		AR(1)	11.77	4.049	0.000	11.23	4.667	0.000
		AR(2)	11.84	3.994	0.561	11.26	4.633	0.214
		AR(3)	11.9	4.008	1.113	11.28	4.640	0.401
		AR(4)	11.96	4.103	1.631	11.3	4.640	0.570
	5% N(0,9)	AR(0)	1.773	0.451	15.587	1.716	0.399	14.269
		AR(1)	1.534	0.381	0.000	1.502	0.259	0.000
		AR(2)	1.538	0.370	0.248	1.501	0.255	-0.033
		AR(3)	1.542	0.364	0.495	1.511	0.253	0.613
		AR(4)	1.548	0.388	0.939	1.52	0.266	1.179
AC	10% N(0,9)	AR(0)	2.254	0.496	9.175	2.258	0.594	6.540
		AR(1)	2.064	0.428	0.000	2.119	0.566	0.000
		AR(2)	2.065	0.417	0.034	2.121	0.585	0.071
		AR(3)	2.07	0.448	0.257	2.127	0.587	0.349
		AR(4)	2.075	0.472	0.518	2.136	0.581	0.769
	10% N(0,100)	AR(0)	11.33	4.134	-0.115	11.02	3.632	-8.877
		AR(1)	11.35	4.135	0.000	12.09	4.050	0.000
		AR(2)	11.37	4.231	0.212	11.82	3.990	-2.258
		AR(3)	11.4	4.226	0.458	11.82	3.987	-2.250
		AR(4)	11.43	4.209	0.767	11.82	3.962	-2.192

Note: Simulations are based on 500 runs of sample size 260. Table entries refer to the mean squared forecasting error. The forecasts are computed as one-period-ahead forecasts beginning at 234 (90% of the sample) and ending at 260. IC and AC represent innovation contamination and additive contamination.

Table 4.12: Forecasting Costs of Misspecification for Model 2: $X_t - 2 - .9X_{t-1} = \varepsilon_t$
Under Various Contaminations

Contamination		OLS				MM		
Type	Size	<i>Selected Model</i>	MSE	<i>Quartile Range</i>	<i>%Relative Loss</i>	MSE	<i>Quartile Range</i>	<i>%Relative Loss</i>
None	0%	AR(0)	5.468	1.845	449.187	5.708	1.613	473.324
		AR(1)	0.996	0.203	0.000	1.01	0.195	1.468
		AR(2)	1	0.209	0.484	1.017	0.202	2.162
		AR(3)	1.004	0.216	0.826	1.023	0.200	2.764
		AR(4)	1.007	0.216	1.177	1.032	0.207	3.638
IC	5% N(0,9)	AR(0)	8.179	2.396	457.550	8.252	2.112	462.513
		AR(1)	1.467	0.278	0.000	1.471	0.290	0.307
		AR(2)	1.474	0.277	0.484	1.478	0.289	0.736
		AR(3)	1.478	0.285	0.750	1.488	0.295	1.418
		AR(4)	1.485	0.300	1.227	1.497	0.310	2.079
	10% N(0,9)	AR(0)	10.81	3.333	464.736	10.14	3.419	422.862
		AR(1)	1.915	0.509	0.000	1.94	0.429	0.000
		AR(2)	1.926	0.505	0.575	1.944	0.440	0.201
		AR(3)	1.93	0.527	0.815	1.956	0.431	0.804
		AR(4)	1.939	0.512	1.243	1.963	0.441	1.191
10% N(0,100)	AR(0)	58.39	18.951	429.164	63.88	18.799	474.847	
	AR(1)	11.03	4.581	0.000	11.11	4.096	0.000	
	AR(2)	11.08	4.658	0.408	11.13	4.104	0.117	
	AR(3)	11.12	4.634	0.770	11.14	4.078	0.252	
	AR(4)	11.16	4.572	1.124	11.16	4.080	0.459	
AC	5% N(0,9)	AR(0)	6.203	2.081	250.976	5.912	1.666	239.374
		AR(1)	1.767	0.352	0.000	1.742	0.355	0.000
		AR(2)	1.73	0.376	-2.111	1.731	0.352	-0.614
		AR(3)	1.735	0.372	-1.845	1.738	0.354	-0.258
		AR(4)	1.747	0.380	-1.149	1.743	0.358	0.069
	10% N(0,9)	AR(0)	6.566	1.856	163.854	6.216	1.825	147.110
		AR(1)	2.489	0.625	0.000	2.516	0.711	0.000
		AR(2)	2.349	0.575	-5.614	2.433	0.677	-3.280
		AR(3)	2.339	0.592	-6.024	2.425	0.686	-3.621
		AR(4)	2.348	0.592	-5.662	2.428	0.699	-3.470
10% N(0,100)	AR(0)	16.52	5.348	7.909	15.16	4.784	-18.281	
	AR(1)	15.31	5.655	0.000	18.56	6.921	0.000	
	AR(2)	14.4	5.136	-5.963	18.12	6.681	-2.344	
	AR(3)	14	5.017	-8.595	17.99	6.495	-3.029	
	AR(4)	13.89	4.973	-9.287	17.94	6.521	-3.293	

Note: Simulations are based on 500 runs of sample size 260. Table entries refer to the mean squared forecasting error. The forecasts are computed as one-period-ahead forecasts beginning at 234 (90% of the sample) and ending at 260. IC and AC represent innovation contamination and additive contamination.

Table 4.13: Forecasting Costs of Misspecification for Model 3: $X_t - 2 - .3X_{t-1} - .1X_{t-2} = \varepsilon_t$ Under Various Contaminations

Contamination		OLS				MM		
Type	Size	<i>Selected Model</i>	MSE	<i>Quartile Range</i>	<i>%Relative Loss</i>	MSE	<i>Quartile Range</i>	<i>%Relative Loss</i>
None	0%	AR(0)	1.158	0.237	14.194	1.134	0.249	11.896
		AR(1)	1.021	0.223	0.690	1.015	0.214	0.158
		AR(2)	1.014	0.207	0.000	1.011	0.204	-0.266
		AR(3)	1.019	0.201	0.513	1.018	0.206	0.375
		AR(4)	1.024	0.190	0.967	1.026	0.206	1.154
IC	5% N(0,9)	AR(0)	1.618	0.373	11.601	1.707	0.384	17.739
		AR(1)	1.457	0.331	0.455	1.519	0.315	4.773
		AR(2)	1.45	0.296	0.000	1.511	0.291	4.235
		AR(3)	1.458	0.302	0.579	1.518	0.294	4.690
		AR(4)	1.465	0.318	1.035	1.527	0.306	5.345
	10% N(0,9)	AR(0)	2.253	0.637	13.036	2.062	0.493	12.065
		AR(1)	2.013	0.545	1.019	1.854	0.390	0.783
		AR(2)	1.993	0.549	0.000	1.84	0.373	0.000
		AR(3)	2	0.521	0.361	1.844	0.376	0.196
		AR(4)	2.009	0.526	0.798	1.849	0.386	0.511
10% N(0,100)	AR(0)	12.46	4.177	12.588	13.22	5.175	14.218	
	AR(1)	11.14	3.813	0.642	11.71	4.540	1.183	
	AR(2)	11.07	3.754	0.000	11.58	4.482	0.000	
	AR(3)	11.12	3.737	0.461	11.59	4.549	0.147	
	AR(4)	11.14	3.751	0.669	11.62	4.486	0.389	
AC	5% N(0,9)	AR(0)	1.614	0.398	6.381	1.622	0.347	5.909
		AR(1)	1.519	0.354	0.105	1.547	0.294	1.005
		AR(2)	1.517	0.335	0.000	1.532	0.296	0.000
		AR(3)	1.524	0.347	0.461	1.538	0.318	0.437
		AR(4)	1.532	0.332	1.009	1.549	0.311	1.123
	10% N(0,9)	AR(0)	2.035	0.428	3.163	1.945	0.494	2.471
		AR(1)	1.98	0.460	0.345	1.903	0.478	0.300
		AR(2)	1.973	0.446	0.000	1.898	0.452	0.000
		AR(3)	1.982	0.458	0.451	1.902	0.460	0.216
		AR(4)	1.989	0.463	0.796	1.909	0.467	0.590
10% N(0,100)	AR(0)	10.87	3.924	-0.749	10.87	3.624	-2.345	
	AR(1)	10.92	3.945	-0.283	11.22	3.913	0.845	
	AR(2)	10.95	3.919	0.000	11.13	3.758	0.000	
	AR(3)	10.99	3.906	0.320	11.09	3.850	-0.332	
	AR(4)	11.01	3.927	0.521	11.09	3.835	-0.323	

Note: Simulations are based on 500 runs of sample size 260. Table entries refer to the mean squared forecasting error. The forecasts are computed as one-period-ahead forecasts beginning at 234 (90% of the sample) and ending at 260. IC and AC represent innovation contamination and additive contamination.

Table 4.14: Forecasting Costs of Misspecification for Model 4: $X_t - 2 - .5X_{t-1} - .35X_{t-2} = \varepsilon_t$ Under Various Contaminations

Contamination		OLS				MM		
Type	Size	<i>Selected Model</i>	MSE	<i>Quartile Range</i>	<i>%Relative Loss</i>	MSE	<i>Quartile Range</i>	<i>%Relative Loss</i>
None	0%	AR(0)	3.067	0.853	200.000	2.88	0.681	181.700
		AR(1)	1.168	0.245	14.212	1.153	0.253	12.813
		AR(2)	1.022	0.209	0.000	1.011	0.209	-1.076
		AR(3)	1.027	0.206	0.469	1.02	0.211	-0.264
		AR(4)	1.032	0.206	0.890	1.025	0.194	0.245
IC	5% N(0,9)	AR(0)	4.2	1.106	176.764	4.165	1.074	174.445
		AR(1)	1.729	0.325	13.911	1.693	0.374	11.591
		AR(2)	1.518	0.273	0.000	1.474	0.252	-2.847
		AR(3)	1.525	0.275	0.474	1.482	0.273	-2.320
		AR(4)	1.534	0.276	1.100	1.488	0.268	-1.957
	10% N(0,9)	AR(0)	5.295	1.517	174.387	5.35	1.373	189.611
		AR(1)	2.194	0.482	13.669	2.149	0.439	16.360
		AR(2)	1.93	0.441	0.000	1.847	0.421	0.000
		AR(3)	1.936	0.439	0.290	1.854	0.434	0.363
		AR(4)	1.945	0.437	0.762	1.863	0.438	0.855
10% N(0,100)	AR(0)	31.98	10.653	171.676	30.78	10.458	183.688	
	AR(1)	13.32	4.988	13.151	12.69	4.672	16.920	
	AR(2)	11.77	4.478	0.000	10.85	4.157	0.000	
	AR(3)	11.81	4.473	0.331	10.87	4.135	0.212	
	AR(4)	11.86	4.470	0.739	10.87	4.056	0.157	
AC	5% N(0,9)	AR(0)	3.513	1.001	114.020	1.622	0.347	5.909
		AR(1)	1.894	0.399	15.428	1.547	0.294	1.005
		AR(2)	1.641	0.328	0.000	1.532	0.296	0.000
		AR(3)	1.632	0.333	-0.585	1.538	0.318	0.437
		AR(4)	1.635	0.335	-0.384	1.549	0.311	1.123
	10% N(0,9)	AR(0)	4.019	1.186	78.306	3.709	1.083	68.389
		AR(1)	2.587	0.698	14.782	2.575	0.582	16.879
		AR(2)	2.254	0.560	0.000	2.203	0.499	0.000
		AR(3)	2.211	0.547	-1.912	2.174	0.479	-1.285
		AR(4)	2.207	0.575	-2.072	2.171	0.494	-1.462
10% N(0,100)	AR(0)	13.55	5.024	1.712	13.16	4.673	-17.046	
	AR(1)	13.57	5.258	1.922	18.23	6.844	14.890	
	AR(2)	13.32	5.175	0.000	15.86	6.093	0.000	
	AR(3)	13.13	4.946	-1.412	15.64	6.100	-1.406	
	AR(4)	13	4.942	-2.358	15.29	5.981	-3.631	

Note: Simulations are based on 500 runs of sample size 260. Table entries refer to the mean squared forecasting error. The forecasts are computed as one-period-ahead forecasts beginning at 234 (90% of the sample) and ending at 260. IC and AC represent innovation contamination and additive contamination.

Table 4.15: Oil Price Unit Root Test

Estimated autoregression using OLS

$$op_t = \frac{0.17681}{(0.15340)} + \frac{0.99039}{(0.0075456)} op_{t-1}$$

Dicky-Fuller Statistic t-test

$$(\hat{\rho} - 1)/\sigma = (.99039 - 1)/0.0075456 = -1.2736$$

5% Critical Value for Dicky-Fuller t-test

$$-3.51$$

Therefore, the null of a unit root cannot be rejected.

Notes: Real Oil Price refers to the “FOB Import Price for the U.S.” divided by the producer price index (PPI) listed in cents per barrel from January 1974-June 2002.

Table 4.16: Oil Price Unit Root Test

Estimated autoregression using MM

$$op_t = \frac{0.19227}{(0.11066)} + \frac{0.98963}{(0.0054435)} op_{t-1}$$

Dicky-Fuller Statistic t-test

$$(\hat{\rho} - 1)/\sigma = (.98963 - 1)/0.0054435 = -1.905$$

5% Critical Value for Dicky-Fuller t-test

$$-3.52$$

Therefore, the null of a unit root cannot be rejected.

Notes: Real Oil Price refers to the “FOB Import Price for the U.S.” divided by the producer price index (PPI) listed in cents per barrel from January 1974-June 2002. Standard errors are computed according to Yohai (1986). Under assumption of no contamination, the critical values for the MM-based DF test are nearly identical to those of the OLS-based DF test for a large number of observations, T (Lucas 1995). This is due to the high efficiency of the MM estimator under normal errors.

Table 4.17: Forecasting Changes in Oil Prices with Alternative Selection Criteria

Criterion	<i>Selected LagOrder</i>	Root Mean Squared Forecasting Error	
		OLS	MM
BIC	1	1.0012 (-)	0.99200 (0.19652)
AIC	4	1.0323 (-0.60078)	1.0194 (-0.32784)
HQC	4	1.0323 (-0.60078)	1.0194 (-0.32784)
seqf	4	1.0323 (-0.60078)	1.0194 (-0.32784)
BICMM	1	1.0012 (-)	0.99200 (0.19652)
AICMM	2	1.0398 (-0.75765)	0.98831 (0.75024)
HQCMM	2	1.0398 (-0.75765)	.98831 (0.75024)
seqfMM	2	1.0398 (-0.75765)	.98831 (0.75024)

Notes: Real Oil Price refers to the “FOB Import Price for the U.S.” divided by the producer price index (PPI) listed in cents per barrel from January 1974-June 2002. Numbers in parentheses are t-statistics for a MGN test for comparing forecast accuracy between the listed forecast and the OLS AR(1) forecast.

Table 4.18: Autoregressive Parameter Estimates for Changes in Real Oil Prices

Method	AR_{model}	$\hat{\mu}$	$\hat{\Phi}_1$	$\hat{\Phi}_2$	$\hat{\Phi}_3$	$\hat{\Phi}_4$
OLS	1	-0.00704 (0.0476)	0.52659 (0.0465)	-	-	-
	2	-0.007508 (0.0474)	0.57921 (0.0545)	-0.0999 (0.0544)	-	-
	4	-0.01347 (0.04658)	0.53836 (0.05321)	-0.04040 (0.06061)	0.00222 (0.06063)	-0.16507 (0.05389)
MM	1	-0.05171 (0.03697)	0.40147 (0.03611)	-	-	-
	2	-0.06798 (0.03617)	0.41074 (0.04156)	-0.23774 (0.04154)	-	-
	4	-0.09008 (0.03660)	0.31020 (0.04181)	-0.22997 (0.04763)	-0.032017 (0.04764)	0.04549 (0.04234)

Notes: An asterisk indicates significant at the 5% level. Standard errors for the MM estimator are as described in section 5. Real Oil Price refers to the “FOB Import Price for the U.S.” divided by the producer price index (PPI) listed in cents per barrel from January 1974-June 2002.

Table 4.19: Performance Summary of the Various Selection Criteria

	BIC	AIC	HQC	SEQF
OLS	<p>Best criterion in NC cases except for model 3.</p> <p>Best criterion in IC cases except for model 3.</p> <p>Performs poorly when there is AC.</p>	<p>Better than other OLS criteria in models 1 and 3.</p> <p>Worse than other OLS criteria in models 2 and 4.</p> <p>Performs poorly when there is AC.</p>	<p>Similar to BIC in NC and IC cases.</p> <p>Slightly better than BIC in heavy AC cases.</p> <p>Performs poorly when there is AC.</p>	<p>Worse than other OLS criteria in all cases.</p> <p>Performs poorly when there is AC.</p>
MM	<p>Slightly less accurate than OLS BIC in most IC and no contamination cases.</p> <p>Performs much better than OLS BIC in model 3 with IC and in all models with AC.</p> <p>Best MM criterion in all cases except for model 3 and model 4 with AC.</p> <p>Accuracy improves with higher IC.</p>	<p>Best criterion in model 3 when there is AC.</p> <p>Performs poorly in all other cases.</p> <p>Accuracy improves with higher IC.</p>	<p>Best criterion in model 3 when there is IC.</p> <p>Less accurate than BICMM in all other cases.</p> <p>Accuracy improves with higher IC.</p>	<p>Best criteria in model 4 when there is AC.</p> <p>Worse than other MM criteria in all other cases.</p>

Notes: NC, IC, and AC refer to no contamination, innovation contamination, and additive contamination respectively. Model numbers refer to those designated in Table 4.2. All comparisons are in terms of selection accuracy.

Chapter 5

International Real Business Cycle Models and Decisions before Shocks

5.1 Introduction

International Real Business Cycle (IRBC) models do reasonably well at matching some of the moments of the international time-series data. They do well at matching domestic volatility of and correlations among consumption, output, investment, and labor. However, they perform more poorly at matching international correlations. In particular, there are two areas in which these simple stochastic models consistently fail: the “consumption/output anomaly” and the “terms of trade anomaly”. The consumption/output anomaly involves the international correlation of output and consumption. Typically, theoretical models predict a higher correlation of consumption across countries than output across countries. The actual data shows that the opposite is true; higher correlation between outputs than consumptions exists. The terms of trade anomaly simply states that in the data there is excess variability

in the terms of trade that cannot be explained by the theoretical models.

Much of the work in the IRBC literature involves adding complexity to the benchmark model in order to help account for the two anomalies. For instance, Stockman and Tesar (1995) add nontraded goods, Baxter and Crucini (1995) incorporate incomplete capital markets, and more recently Ambler, Cardia and Zimmermann (2002) add multiple sectors and intermediate goods. Most of these augmentations help explain one or both of the anomalies, but at the cost of complexity. Often, while fixing one thing another breaks down, as is the case with the dynamic general equilibrium models. One of the goals of this paper is to provide a simple step to help solve the consumption/output anomaly while keeping changes to the benchmark IRBC model to a minimum.

A potential improvement for IRBC models comes from a single country model by Boldrin, Christiano, and Fisher (2001). Their paper on asset pricing, habit persistence, and business cycles incorporates a simple decision-before-shock (DBS) methodology in their closed multi-sector business cycle model. Capital goods and employment are determined prior to the realization of a contemporaneous shock. This introduces some additional friction in the model, in that aggregate employment and capital investment become difficult to adjust quickly in response to a shock. In this chapter, I apply the DBS framework to the benchmark IRBC model of Backus, Kehoe, and Kydland (1994) (BKK), with dramatic results. The predicted cross-country correlation of consumption drops drastically, putting it much more in line

with the actual data. At the same time, the moments that already matched do not change significantly. The model in this paper can be thought of as a single-sector adaptation of Boldrin et al.'s model to an international framework focusing on matching cross-country correlations rather than asset prices.

The basic story is one of risk-sharing. In standard IRBC models, if one country experiences a negative productivity shock, agents in that country are able to shift their investments to another country that has higher productivity and thus a higher return. If decisions must be made before a shock is realized, there is less flexibility to share risk across countries. It is not a lack of desire to share risk, but simply the inability to do so fully. The only flaw the DBS approach generates is that it greatly reduces first-order autocorrelation of consumption, contrary to the actual data. Incorporating habit persistence, as in Boldrin et al. (2001), easily rectifies this flaw.

Another potential problem for these simulated general-equilibrium models is their sensitivity to calibration, particularly, the calibration of technology innovations – productivity is typically calculated using Solow residuals. Kehoe and Perri (2000), Baxter and Crucinni (1995), and Zimmerman (1994) all demonstrate that changes in persistence levels and cross-country correlations of technology innovations can alter the results of IRBC models. Because of this sensitivity, it is imperative to correctly estimate the structure of these productivity movements. A potential source of problems for properly estimating this structure is outliers. Balke and

Fomby (1994) and Zaman, Rousseeuw, and Orhan (2001) all show that outliers are present and influence macroeconomic data. Autoregressive estimates, such as the vector autoregression (VAR) used to formulate the technology innovations over time in IRBC models, are influenced by outliers (Martin 1979). To correct for the influence of potential outliers, I use an outlier-robust alternative to ordinary least squares known as the MM-estimator (Yohai 1986) to estimate the structure of productivity movements over time.

In total, this essay adds to the literature in two ways. First, it provides a simple way to reduce the consumption/output anomaly. Second, it provides more accurate, robust estimates of the structure of technology innovations over time and examines how those robust estimates alter the simulation results of an IRBC model.

The layout of the essay is as follows. Section 2 reviews the previous literature on IRBC models and the relevant closed Real Business Cycle (RBC) models. It also provides a brief overview of outliers in macroeconomic data. Section 3 presents the model, calibration, and solution methods used for analysis. Section 4 compares the moments generated by this model and those generated by a benchmark IRBC model to the data. Additional modifications to the model, including habit persistence, and a sensitivity analysis are conducted in Section 5. Section 6 draws conclusions.

5.2 Previous Literature

5.2.1 International Real Business Cycles

The story of the basic IRBC model goes like this. In a complete market and open economy agents are able to smooth lifetime consumption by sharing exposure to technology shocks. A positive technology shock in *one* country leads to increases in consumption for partner countries and vice versa. If the countries only trade two goods that are perfect substitutes, then we expect consumption correlations of one in our models. If technology shocks are persistent, as they typically are in RBC models, agents want to shift investment to the most productive location. Therefore, we typically see theoretical cross-country output correlations that are close to zero or negative.

Essentially, a country has two investment options in this simple open economy as opposed to one in a closed economy. It can invest in its own production or it can indirectly invest in the other country's production by way of trade.

To better see how trade channel works, consider the simple world budget constraints

$$y_t = c_t + i_t + nx_t \tag{5.1}$$

$$y_t^* = c_t^* + i_t^* + nx_t^* \tag{5.2}$$

$$nx_t = - nx_t^* \tag{5.3}$$

where y is output, c is consumption, i is investment, and nx_t is net exports. An

asterisk represents the foreign country.

Suppose that the home country experiences a positive technology shock, thus increasing y_t . Because shocks are persistent but not permanent, y_{t+1} also increases, but to a lesser degree than y_t . Agents in the home country realize this and increase c_t and i_t . Their combined increase is more than the increase in y_t . Therefore, nx_t decreases and the home country has a trade deficit. The foreign country experiences an increase in nx_t^* , and cuts i_t^* by more than it increases nx_t^* , allowing it to increase consumption. It has no problem lowering its own investment because it knows the home country will still experience the positive shock next period. The foreign country provides capital and consumption goods for the home country by lowering its investment because it knows there will be a higher return if the home country increases its own investment.

There have been two basic strategies to solve the consumption/output anomaly through risk-sharing. The first is to limit the *ability* of agents to share risk. The second is to limit the *desire* of agents to share risk.

Baxter and Crucini (1995) take the first approach. They argue that the typical IRBC model does not adequately portray international financial markets. They introduce a model that restricts asset trading to noncontingent real debt. They find that restricting asset trading might be important in solving the IRBC anomalies when the underlying stochastic process is highly persistent, with little international

spillover. Kehoe and Perri (2000) add the ability to default on loans as a factor limiting risk sharing. This reduces the cross-country consumption correlation, but their model fails to generate a cross-country output correlation higher than the consumption correlation.

Nadnednick (2000) finds that by incorporating durable goods, the IRBC model can help explain the consumption/output anomaly. Stockman and Tesar find that nontradable goods decrease the cross-country consumption correlation, but similar to Kehoe and Perri, the cross-country consumption correlation is still larger than the cross-country output correlation. Guo and Sturzenegger (1998) find that adding increasing returns subjects the economy to sunspot shocks. Their belief-driven model helps to account for the consumption/output anomaly.

The model in this paper resembles Baxter and Crucini's (1995) by attempting to limit the ability of agents to share risk. Unlike their model, however, this model maintains complete asset markets. Agents still have all the tools necessary to share risk, but it is their inability to respond quickly to a shock that limits their ability to share risk.

5.2.2 Outliers in Macroeconomic Data

A modified outlier identification procedure from Tsay (1988), Balke and Fomby (1994) finds evidence of outliers in post-World War II data for the United States. They state that "Large shocks appear to be present in all the series we examined." In particular they find a few outliers in GNP growth, employment growth, and

investment all of which are commonly used in the computation of Solow residuals, the typical measure of productivity in IRBC models.

Zaman, Rousseeuw, and Orhan (2001) revisit De Long and Summer's 1991 paper studying the growth of 61 countries from 1960-1985, using robust estimation techniques. They find evidence of outliers in the international data. In addition, when using a robust estimation technique, they find that non-equipment investment has a significant influence on GDP that was missed when using ordinary least squares. This paper provides evidence that the presence of outliers in macroeconomic time series is not simply found in U.S. data. Most other countries seem to have the same presence of "large shocks" as the U.S.

Outliers are a widely studied phenomenon in statistics (see Barnett and Lewis (1994) for a overview of outliers in statistics), and outliers are found to consistently bias parameter estimates. This is particularly true for autoregressive time-series models (Martin 1979). To combat this problem, many authors have developed outlier-robust statistical techniques.

Huber (1964) introduced the class of M-estimators where parameter estimates are based on minimizing an objective function other than the sum of squares that attach less weight to outlying observations. The S-estimator developed by Rousseeuw and Yohai (1984) is even more resilient to extreme observations but lacks efficiency if the disturbances follow a normal distribution. The MM-estimator (Yohai 1986) has more recently been introduced to be both highly efficient when the disturbances

follow a normal distribution and highly robust to outlying observations. The MM-estimator, because it is highly robust and highly efficient, will be used to estimate the vector autoregression of the Solow residuals as part of a sensitivity analysis.

5.3 General Model

The model in this section is essentially the benchmark IRBC from BKK (1994), but it incorporates decisions made before the shocks. Representative agents from two countries, a home country and a foreign country, attempt to maximize the present value of their lifetime utility, based on consumption and leisure. Production, and thus income, is determined by combining capital and labor inputs with some persistent technology shock. As is standard in the literature, the model assumes that the countries are symmetrical. Variables for the foreign country are denoted with an asterisk.

Consider the consumer maximization problem

$$\max E_0 \sum_{t=0}^{\infty} \beta^t U(c_t, L_t) \quad (5.4)$$

where preferences are represented by a power utility function $U(c, L) = \frac{1}{1-\psi} (c_t^\sigma L_t^{1-\sigma})^{1-\psi}$.

Here $c_t > 0$ represents consumption at time t and $L_t > 0$ denotes leisure at time t .

Each economy produces a unique good in quantity y_t from a Cobb-Douglas production function

$$y_t = e^{z_t} k_t^\alpha N_t^{1-\alpha} \quad (5.5)$$

where k_t denotes capital stock at time t , N_t represents labor at time t , and z_t are the stochastic technology innovations.¹ The domestic and foreign innovations, (z_t, z_t^*) , follow the VAR

$$\begin{pmatrix} z_t \\ z_t^* \end{pmatrix} = \begin{pmatrix} \rho_1 & \rho_2 \\ \rho_2 & \rho_1 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ z_{t-1}^* \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \varepsilon_t^* \end{pmatrix}. \quad (5.6)$$

The diagonal term ρ_1 accounts for the persistence of technology innovations. The closer ρ_1 is to one, the longer it takes the model economy to converge back to a steady-state. The off-diagonal term ρ_2 represents a spillover effect – the direct transmission of a technology shock between countries. The shocks to technology, $(\varepsilon_t, \varepsilon_t^*)$, are serially uncorrelated and follow a multivariate-normal distribution with variance-covariance matrix Ω . The off-diagonal terms of Ω are nonzero allowing for correlation of contemporaneous shocks between countries.

Output can either be used domestically or exported:

$$y_t = y_{ht} + ex_t. \quad (5.7)$$

Exports at time t are represented by ex_t and y_{ht} represents domestic production that remains in the home country.

Specialization in this economy is incorporated in the model through the use of an Armington aggregator (1969) as in BKK (1994). Essentially each country produces its own unique good, but this domestic good needs to be combined with the foreign good to produce usable products for consumption or investment. Mathematically,

¹The simulated moments are reported as percentages and therefore there is no need to calibrate the size of the economy.

this aggregator is represented by

$$H(y_{ht}, im_t) = [(\lambda)y_{ht}^{-\mu} + (1 - \lambda)im_t^{-\mu}]^{-\frac{1}{\mu}}. \quad (5.8)$$

Here, im_t denotes the amount of the foreign good imported into the domestic country and H_t is the amount of the final good that can be used for domestic consumption or investment. Note that H_t and not y_{ht} is final output of the multi-purpose good, so the budget constraint is

$$H(y_{ht}, im_t) = c_t + i_t, \quad (5.9)$$

where i_t is investment at time t .

The next period's capital stock is the depreciated current capital stock plus any new investment. Adjustment costs are implicit due to the decision-before-shock approach discussed below and are thus not explicitly defined in this equation of motion,

$$k_{t+1} = (1 - \delta)k_t + i_t. \quad (5.10)$$

Total hours available to agents are fixed and normalized to one. Agents can choose between labor and leisure from their total hours available:

$$1 = L_t + N_t. \quad (5.11)$$

Finally, there is an accounting identity indicating that whatever one country exports, the other country imports. In other words, net world trade is zero:

$$im = ex^*. \quad (5.12)$$

Additionally, there are foreign analogs to all these equations.

5.3.1 Decision Before Shock

So far, the model is similar to the benchmark IRBC model from BKK (1994). To incorporate the frictions in the model I assume that i_t , N_t , and ex_t must be chosen prior to the realization of z_t , similarly to Boldrin et al. (2001). Here im_t is chosen prior to realization of the shock, because the foreign country must set ex_t^* which is equal to im_t in equation (5.12). This leaves agents with only one control variable, c_t , to absorb the contemporaneous shock – c_t is automatically chosen because there are no degrees of freedom.

We can think of the decision process as a two-stage problem. Initially, agents find their controls based on the expected technology level. During the period, there is an unanticipated shock to technology in the economy. Agents now must address this shock, but investment, labor supply, and exports are difficult to adjust quickly. This leaves agents with consumption as their only route to absorb the technology shock. Technology is persistent and agents are able to adjust to the persistent component of the shock in the next period, but it is too late to help them in the current period.

For example, suppose there is a small tribe of hunters with a fixed number of members.² Each hunter can do one of three things, he can hunt (supply consumption goods), he can make weapons (invest), or he can go work for another tribe for the day (export) (conversely the tribe could import workers for the day.) At the beginning of the day the chief makes a decision about what the members of the tribe will do

²Note that this example is not perfectly analogous to the full-fledged model. It is simply provided to give intuition for decisions-before-shocks.

to maximize the tribe's well-being over time. Hunters are out hunting, exported workers are working at the other tribe, and weapons-makers are at home making weapons, so it is not logistically feasible for members of the tribe to switch jobs easily. Now suppose the hunters unexpectedly run into a large group of deer (a technology shock). They are able to bring home much more food than expected. When they come home there is nothing to do with the food but to eat it (consumption) – it is assumed that consumption goods are not durable and cannot be transformed into investment goods.

Had the chief known that there would be this many deer, he would have had more people making weapons and sent out fewer hunters – fewer hunters can still provide an increase in consumption. It is also likely that the tribe would have tried to hire more workers from a neighboring tribe to help reap the rewards of the sudden increase in deer. Altogether, this would have resulted in a higher capacity to hunt in the future (which would be how the benchmark model works) rather than to have one sudden influx of food. The chief does know that deer like to stay in one place for awhile (technology is persistent) and the next day he makes decisions about making weapons, hunting, and importing or exporting workers in order to maximize his tribe's life-time well-being.

This specification is not without support. A similar way to look at this model is that it incorporates time to build, labor contracts, and time to ship constraints but without any discounting.³ Time-to-build and time-to-ship augmentations to the

³If we were to specify a model with time to build, labor contracts, and time to ship constraints,

benchmark model have already been incorporated by BKK (1994). By themselves they do not significantly reduce the consumption/output problems as seen in BKK (1994). They simply do not provide enough frictions. In addition, the low elasticity of capital supply inherent in this type of model with constraints to investment also fits the empirical evidence in Goolsbee(1998) as indicated by Boldrin et all (2001). Frictions in labor movements are also very plausible. BKK (1992) incorporates a distributed lag on leisure that in effect reduces labor supply. Phelan and Trejos (2000) show that small search-and-matching costs may slow down labor movements.

5.3.2 Calibration and Solution Method

Table 5.1 lists all the parameter values used for the benchmark model and the base DBS model. Where possible, the parameter values are those used by BKK (1994) . With complete markets, the second welfare theorem holds so that the social planner's solution is the same as the competitive solution. Operationally, I solve the social planner's problem using linear quadratic programming, as in BKK (1994). All non-linear constraints are substituted into the objective function. Then a second-order Taylor approximation about the steady state is formed. Therefore we have a quadratic objective function and linear constraints. Solution methods for linear-quadratic problems are well known. I use a Bellman iteration scheme. Appendix A describes the complete derivation and solution procedure. All simulations are based

then investment, imports, and labor would all be determined in the previous period and would be subject to a discount factor, β . This is not the case with the specification in this paper. Investment, imports, and labor are all determined during the present period and are not discounted.

on 100 runs of 100 quarters.

Following BKK (1994) net exports and the terms of trade are derived. From equation (5.9), $c + i = H(y_{ht}, im_t)$ where H is homogeneous of degree 1. So in equilibrium $c + i = \frac{\partial H(y_{ht}, im_t)}{\partial y_{ht}} y_{ht} + \frac{\partial H(y_{ht}, im_t)}{\partial im_t} im_t$. Therefore $\frac{\partial H(y_{ht}, im_t)}{\partial y_{ht}}$ is the price of domestic goods and $\frac{\partial H(y_{ht}, im_t)}{\partial im_t}$ is the price of imports. Thus the ratio of net exports to output, nx_t is defined as

$$nx_t = \frac{(ex_t - p_t im_t^*)}{y_t}. \quad (5.13)$$

p_t is the terms of trade and is defined as

$$p_t = \frac{\partial H(y_{ht}, im_t) / \partial im_t}{\partial H(y_{ht}, im_t) / \partial y_{ht}} = \frac{1 - \lambda}{\lambda} \left(\frac{y_{ht}}{im_t} \right)^{\mu+1}. \quad (5.14)$$

5.4 Results

Table 5.2 and Table 5.3 report the results. The U.S. data and the cross-country average data are from Ambler et al. (2002). The data are quarterly from 1960:1 to 1991:2 and from the O.E.C.D with missing data filled in from national sources. The U.S. cross-country correlations are between the U.S. and a European aggregate composed of Austria, Finland, France, Germany, Italy, Sweden and the U.K. The cross-country average is an average of 19 industrialized countries. The most striking result of this simple augmentation is the decrease in the consumption correlation across countries. In my DBS model, cross-country consumption correlation is much more in line with that of the data. As opposed to finding a correlation of

approximately 0.8, it is closer to 0.2. Any contemporaneous shock is now absorbed by consumption rather than all the state variables. If there is a positive shock in the home country, consumers there consume more, but foreign consumers do not necessarily see an increase in their consumption. This, however, depends on the spillover parameter. If there is a higher spillover, then foreign consumers would see some increase in consumption.

Figures 5.1 - 5.4 display the responses of the countries to a 1% increase in technology for the home country in the context of the benchmark and DBS models. The difference between the two models can be seen in the first period. In the DBS model for the home country, investment, employment, imports, and exports all remain fixed for the first period while they respond quickly in the benchmark case. Consumption increases greatly during the first period because it is the only variable that can adjust quickly. After the shock has been realized, the impulse response curves for both countries are virtually identical. The same is true for the foreign country.

Consumption variance has also increased considerably in this model. This matches the data better than the benchmark model for the aggregate of European countries, but it is higher than what is found in the U.S. In the DBS model, agents lack the flexibility to share risk and smooth consumption. Therefore, we see higher consumption variance. This is an improvement over many of the other models that attempt to explain the consumption/output problem. For instance, Guo and Sturzenegger

(1998) correct for the consumption/output anomaly, but they have low consumption variance.

Most of the other correlations and volatility measures remain similar to those found in the benchmark model. Some of the problems inherent in the benchmark model still persist. The terms of trade are much less volatile in theory than in the data. We also see negative cross-country correlation of investment and labor supply. Finally, cross-country output correlation is still close to zero, although now it is slightly positive. So, we cannot yet state that the model solves the consumption/output anomaly, only that it helps explain it. In the next section, we see that small modifications to the structure of the technology innovations can eliminate these final two problems.

The only problems that arise in this model as opposed to the benchmark model involve persistence. Consumption is highly persistent in both the data and the benchmark model, but not in the DBS model. Adjusting to new shocks each period has the effect of eliminating the ability of agents to smooth consumption over time. If we reconstruct the model such that agents strongly desire to smooth consumption across time, we should be able to correct for this problem. One way to go about this incorporates external habit persistence as discussed in the following.

5.5 Modifications and Sensitivity

This section examines modifications to the standard DBS model that alleviate some of its flaws. Sensitivity to various parameter choices are also investigated.

5.5.1 Habit Persistence Model

In an attempt to correct for the persistence problem found in the DBS model, I incorporate habit persistence preferences as in Constantinides (1990) (habits are across time and not across countries.) The model is identical to the base DBS model with the exception of the utility function. Preferences are now defined as $U(c, L) = \frac{1}{1-\psi}((c_t - bc_{t-1})^\sigma L_t^{1-\sigma})^{1-\psi}$, where c_{t-1} is the previous period's consumption level and b is the habit persistence parameter. As in Boldrin, Christiano, and Fisher (2001), a more complex form of habit persistence was avoided. With this specification, it is possible to have negative values for $c_t - bc_{t-1}$. However, this did not occur in my simulations.

The solution method is similar to the one used for the base case. The only difference is that now last period's control variables must be included as state variables. I set $b = .7$ in my simulations, as do Boldrin, Christiano, and Fisher (2001). An alternative, and perhaps superior (as there is no other rational basis for selecting b 's value,) method might determine the value of b that gives the model the correct consumption persistence levels.

As predicted, the results in Table 5.3 show that consumption persistence is much

higher in the habit formation model as compared to the DBS model. The other moments are little changed with this addition to the model. One item to note, however, is that the variance of consumption relative to output is actually higher in the DBS model with habit persistence than without it.

Figures 5.5 and 5.6 display the responses of the countries to a 1% increase in technology for the home country in the context of the DBS model with habit formation. The most notable difference between these curves is the smoothness of the consumption path. The consumption path is much smoother over time when habit formation is incorporated in the model.⁴ Also, note the muted response of investment compared to the standard DBS model. Agents sacrifice some future gains in consumption that they could realize if they invested, because they have a strong desire to smooth their consumption path over time. In addition, the response of imports is slightly higher than in the standard DBS model. The home country attempts to import more to meet its habit. The foreign country also loses some future consumption to meet its habit.

5.5.2 Solow Residual Sensitivity

In typical IRBC models, BKK's VAR specification of technology is used. They find that technological innovations are highly persistent with some international spillover effects. Unlike more refined measurements of the residuals for the United

⁴Here smooth consumption refers to the smoothness of the consumption path over time. It does not refer to the consumption having a small variance.

States, they construct the international Solow residuals from output and employment statistics only. Capital is not included in their analysis because of the lack of data. They argue that short-run capital should be orthogonal to output and should not matter much. Furthermore, Baxter and Crucinni (1995) found that there might be cointegrating relationships between certain countries' Solow residuals and that error-correction models should be used. The hypothesis that the residuals follow a unit-root process with no spillovers cannot be rejected. All of this brings into question the reliability of the point estimates. As seen in the simulations, even a 10% change in the point estimates can lead to rather dramatic results. I analyze several cases of the structure of technology to test for robustness.

In particular, I examine three cases in addition to the standard BKK specification. The first is a high-persistence, no-spillover parametrization. In terms of the specification of technology innovations in (5.3), I let $\rho_1 = .99$ and $\rho_2 = 0$. The second is a high-spillover parametrization of $\rho_1 = .85$ and $\rho_2 = .15$. Finally, there is a low-persistence, moderate-spillover parametrization of $\rho_1 = .8$ and $\rho_2 = .1$.

The results are reported in Tables 5.2 and 5.3. We can see that the moments are sensitive to the specification of the structure of the shocks. The cross-country correlations of labor, investment, and output change considerably depending on the technology specification. With a high-persistence technology, the results differ little from the base case. A higher spillover increases the output correlations noticeably, but investment and labor are still negatively related. The most surprising result is

that of the low-persistence technology. Here, the cross-country output correlation exceeds the cross-country consumption correlation and investment and labor are positively correlated across countries. This is consistent with the data, unlike the other specifications.

Recall that investment does not increase the capital stock until the next period. For today's shock to affect investment and thus future capital stocks, there has to be motivation to move investment to a place that experienced a current productivity shock. This motivation comes from the persistence of the shock. A positive shock in the home country today entices the agent to invest in the home country, because they know the shock is persistent. Much of that positive shock will still be present next period to make the investment worthwhile. In the benchmark model with the benchmark shock specification, this leads to investment being negatively correlated between countries. Agents shift investment to whichever location is more productive. Reducing the persistence decreases the ability of agents to shift investment to the more productive country. They simply do not know which place will be more productive. This is compounded by the DBS shock specification.

Unfortunately, the data indicates that the Solow residuals – the measure of technology innovations – are highly persistent. Therefore, the low-persistence specification may not be justifiable even though it provides the closest fit to the data. However, it is troublesome that a decrease of roughly 10% in persistence can change the results so dramatically. Measurement error alone could account for such a small

change in persistence.

5.5.3 MM-based Technology Specification

As indicated, another potential problem with the VAR technology estimation is bias from outliers. In this section, I address this problem using the MM-estimator (Yohai 1986).

The MM-estimator is a three-step estimator providing both high resistance to outliers and high efficiency under Gaussian errors. In the first step, an initial highly outlier-robust but not necessarily efficient estimator is computed. Next, a robust M-estimate of scale, \hat{s} , is determined. (A scale estimate is analogous to a standard deviation of the residuals). This robust scale estimate is then used in the determination of the MM-estimate, Θ_{MM} , that is equal to

$$\min_{\Theta} \sum_{i=1}^T \eta \left(\frac{y_i - x_i' \Theta}{\hat{s}} \right) \quad (5.15)$$

where y is the dependent variable, x is the independent variable, T is the number of observations, and η is a function that meets a certain set of assumptions as described by Yohai (1986). For a more technical description, see Yohai (1986) or see the description of the MM-estimator in Chapters 3 and 4.

The estimated VAR is

$$\begin{pmatrix} z_t \\ z_t^* \end{pmatrix} = \begin{pmatrix} \rho_1 & \rho_2 \\ \rho_2^* & \rho_1^* \end{pmatrix} \begin{pmatrix} z_{t-1} \\ z_{t-1}^* \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \varepsilon_t^* \end{pmatrix}. \quad (5.16)$$

It is estimated with both OLS and MM for 18 countries. Data are from Christian Zimmermann's "Technology Innovations and the Volatility of Output: An International Perspective" (1994). The data are quarterly from 1960:1 to 1991:2 and are from the O.E.C.D with missing data filled in from national sources. The results are reported in Table 5.4.

With a few exceptions, the MM-estimator typically finds higher levels of persistence and lower levels of spillover. The estimates that were unusually small with OLS, such as those for Sweden and Switzerland, are now much more in line with the estimates from the rest of the countries. Two major exceptions to this are the United States and Germany lower levels of persistence are found.

There does not appear to be any general trend in the parameters for the rest of the world. None of these parameters differ substantially between OLS and MM. This is not surprising because outliers are not as influential in high-level aggregation.

Overall, the VAR estimates using the MM-estimator tend to lend more support to the high-persistence, no-spillover parametrization examined in the previous section. This is consistent with the results of Baxter and Crucini (1995) and further discredits the use of the low-persistence technology, which would make the model fit the actual data better.

5.5.4 Elasticity of Substitution

Some uncertainty exists about the value of the elasticity of substitution between foreign and domestic goods, $\frac{1}{1+\mu}$. Therefore it is typical to see some sensitivity

analysis to this parameter in the IRBC literature. Following BKK (1994), two alternative elasticities of substitution are examined: (1) a low elasticity, $\frac{1}{1+\mu} = .5$ and (2) a high elasticity $\frac{1}{1+\mu} = 10$. Each is matched with the standard DBS with Habit model and the DBS with Habit model with highly persistent technology. The results are reported in Tables 5.2 and 5.3.

The changes in the elasticity of substitution only marginally change the simulation results in the standard case. As the elasticity rises, the standard deviations of investment and net exports also rise. As the goods become closer substitutes, it becomes easier for two countries to risk-share.

The low-elasticity and high-persistence case is more interesting. It matches the cross-country correlations better than all the other models with the exception of the low-persistence model. However, it does poorly at matching the domestic moments.

Altogether, altering the elasticity of substitution provides only marginal increases in the ability of this IRBC model to replicate the actual data.

5.6 Conclusions

This paper has examined a DBS structure, habit persistence, and Solow residual sensitivity in an IRBC framework. All of these modifications affect the ability of

the stochastic model to adequately replicate economic data. A DBS model with low-persistence shock structure and habit persistence does remarkably well at matching most of the moments of the data. This model resolves the classic consumption/output anomaly. It also correctly shows positive correlations between foreign and domestic labor and investment. Furthermore, it does not create any new problems in the simulated results.

The DBS method alone provides a simple way to decrease the consumption correlation between countries by reducing their ability to share risk. Habit formation only improves the results. Sensitivity to the Solow residuals is apparent in this model. Small changes in its structural specification can dramatically change certain cross-country correlations. Using the MM-estimator in the VAR specification for technology gives more indication of high persistence and low spillovers.

Figure 5.1: Response of the Home Country to a 1% Increase in the Home Country's Technology: Benchmark Model

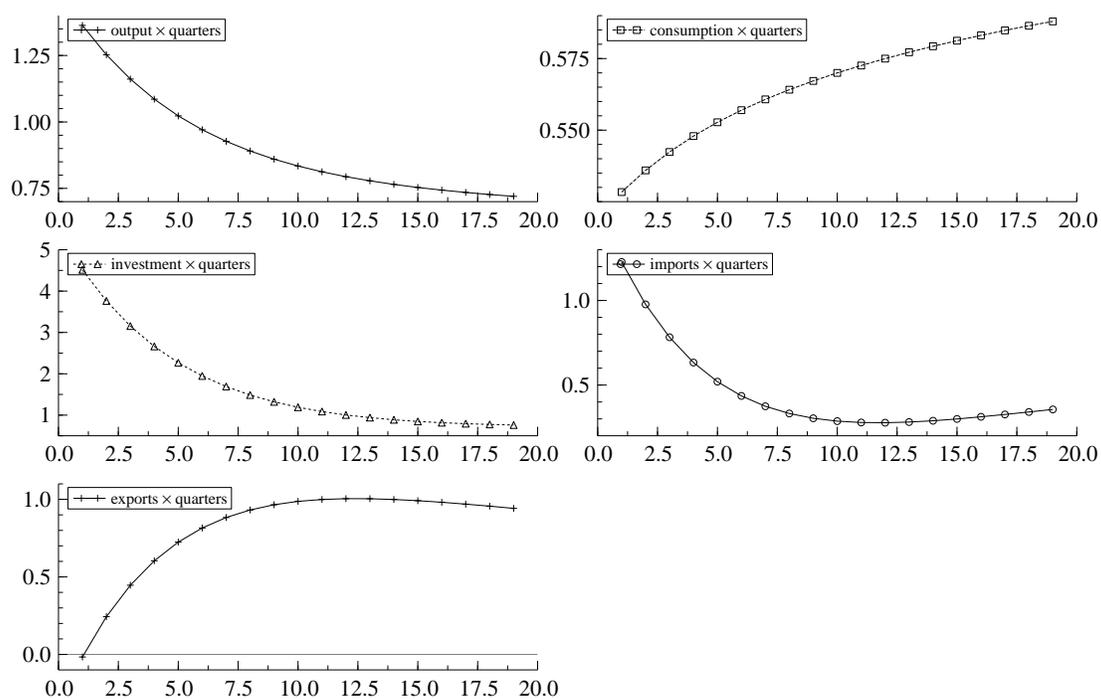


Figure 5.2: Response of the Foreign Country to a 1% Increase in the Home Country's Technology: Benchmark Model

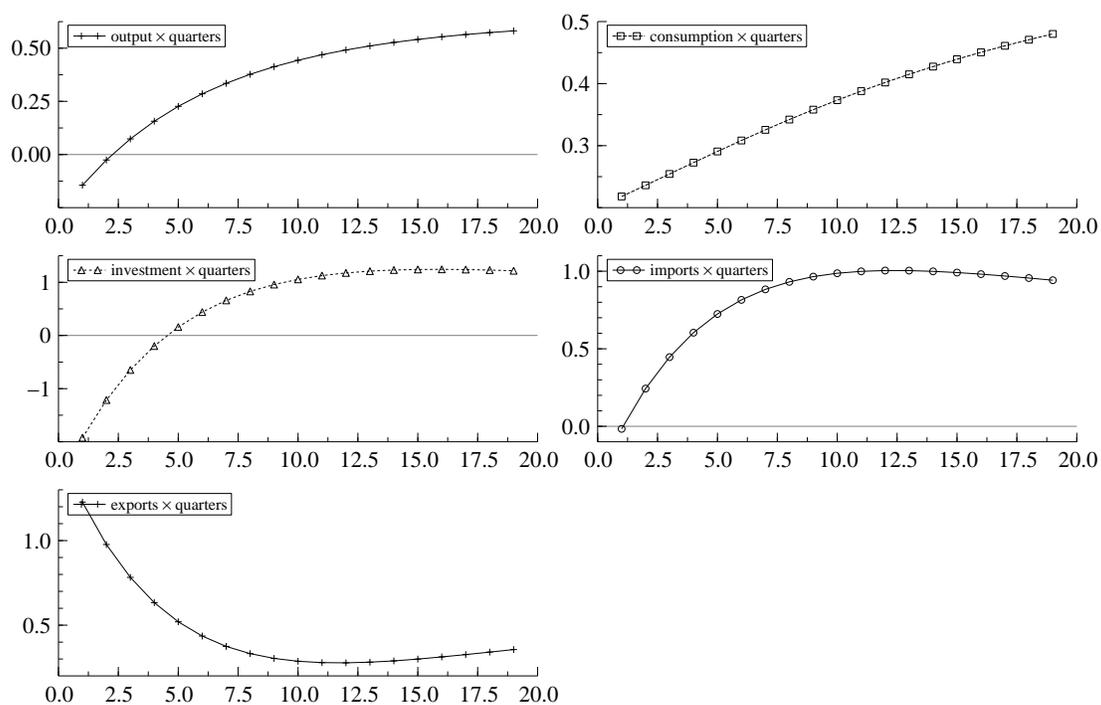


Figure 5.3: Response of the Home Country to a 1% Increase in the Home Country's Technology: DBS Model

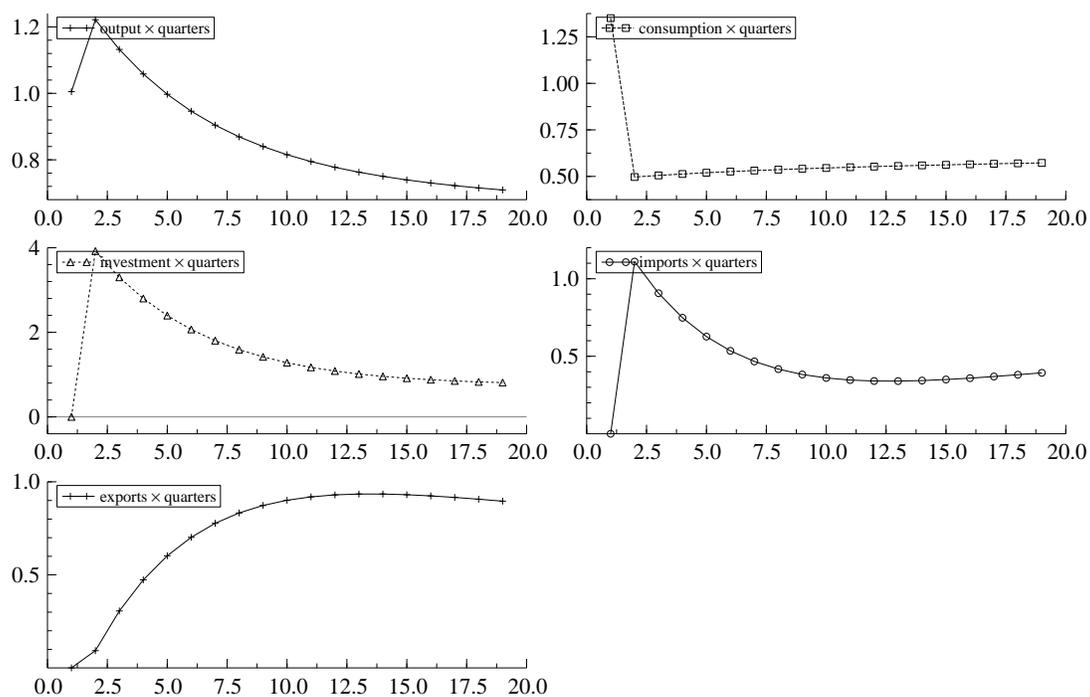


Figure 5.4: Response of the Foreign Country to a 1% Increase in the Home Country's Technology: DBS Model

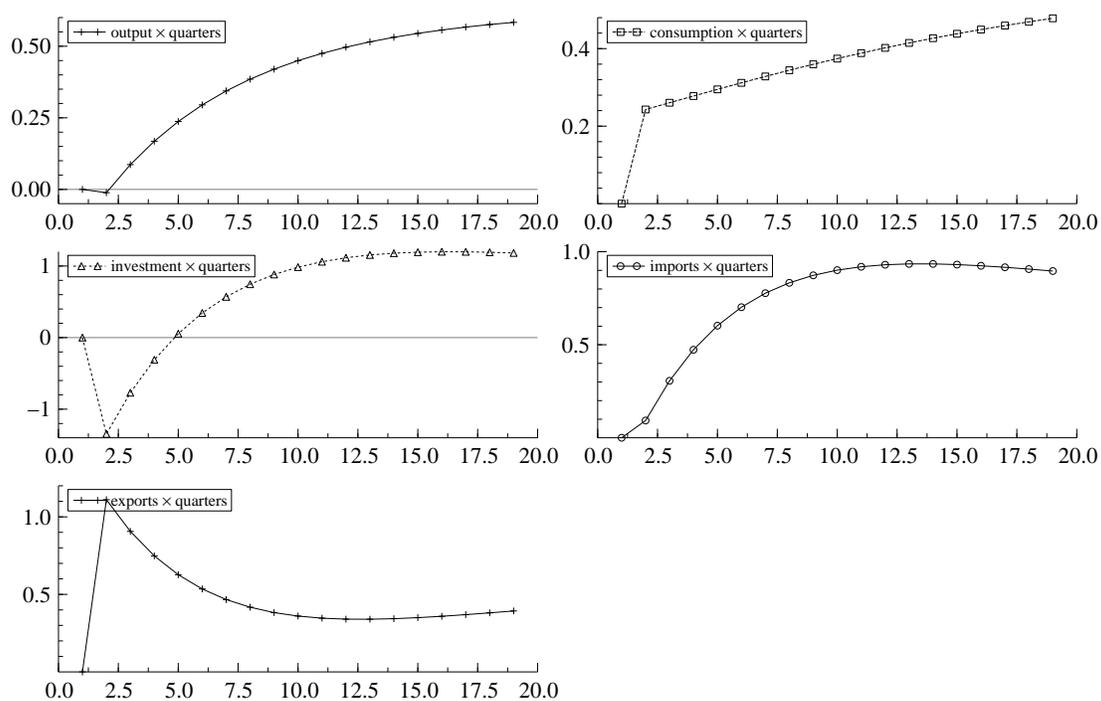


Figure 5.5: Response of the Home Country to a 1% Increase in the Home Country's Technology: DBS Model with Habit Formation

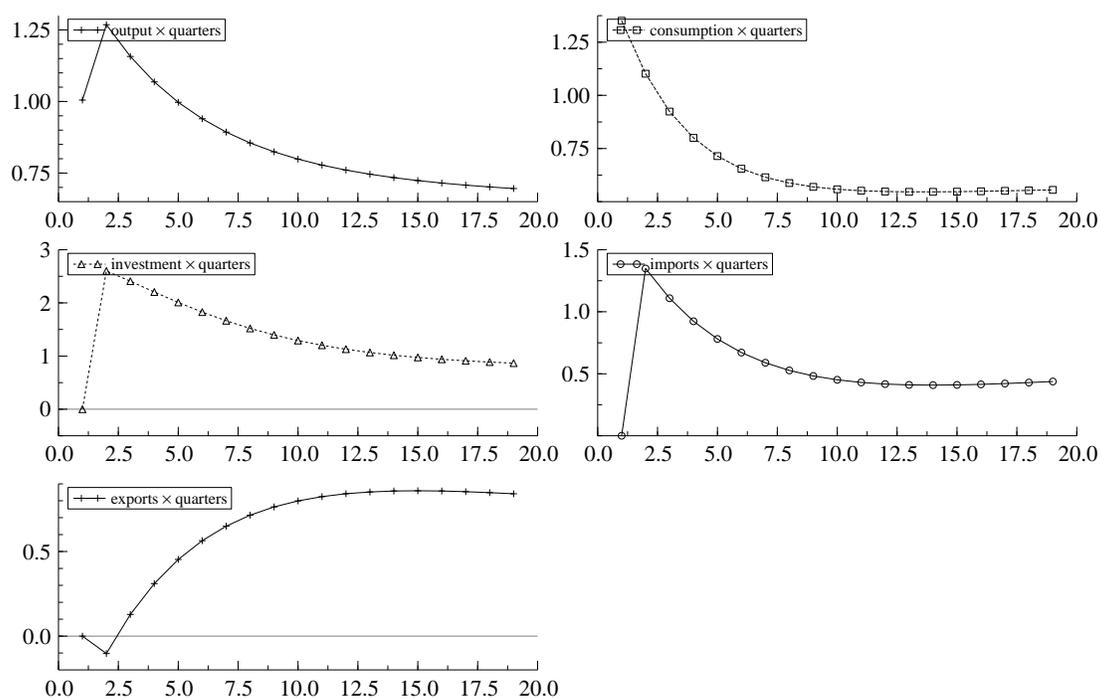


Figure 5.6: Response of the Foreign Country to a 1% Increase in the Home Country's Technology: DBS Model with Habit Formation

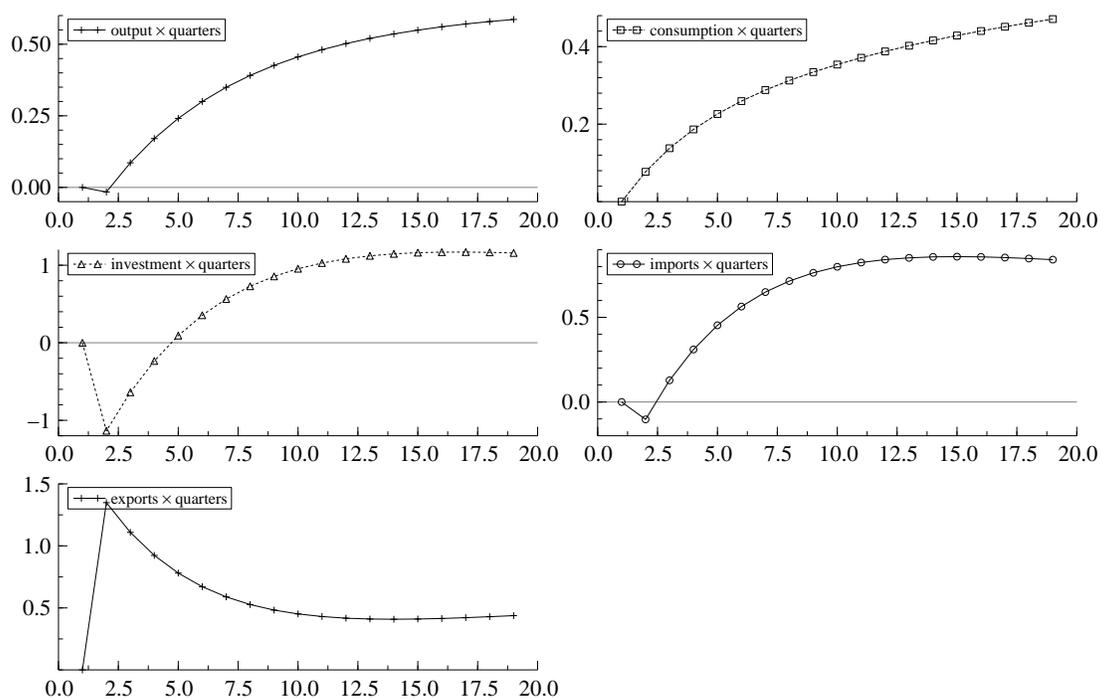


Table 5.1: Calibration Parameters

variable	value	description
β	.99	Time discount rate
α	.36	Capital's share of output
δ	.025	Capital's quarterly depreciation rate
$1/\psi$.5	Intertemporal elasticity of substitution
σ	.34	Leisure/Consumption parameter
λ	a	Import share parameter
$1/(1 + \mu)$	1.5	Elasticity of substitution between foreign and domestic goods
ρ	$\begin{pmatrix} .908 & .088 \\ .088 & .908 \end{pmatrix}$	Persistence and spillover
Ω	$10^{-5} * \begin{pmatrix} .726 & .187 \\ .187 & .726 \end{pmatrix}$	Variance-covariance matrix
<p>a set so that the import share of output = .15 Parameters are those found in Backus, Kehoe, and Kydland (1994).</p>		

Table 5.2: Output Variation and Correlations with Output

	St. Dev. (%)	St. Dev. Rel. to Output				Corr. with Output		
	<i>y</i>	<i>c</i>	<i>i</i>	<i>n</i>	<i>nx</i>	<i>c</i>	<i>i</i>	<i>nx</i>
U.S. data ^a	1.92	0.75	3.27	0.61	0.27	0.82	0.94	-0.37
Cross-country average ^a	1.62	0.30	2.97	0.71	0.65	0.69	0.80	-0.32
Benchmark model	1.45	0.42	3.76	0.42	0.22	0.87	0.94	-0.68
DBS model	1.30	0.83	3.51	0.40	0.21	0.70	0.78	-0.66
Habit	1.44	0.30	4.26	0.41	0.22	0.64	0.95	-0.67
Habit and DBS	1.32	0.98	2.52	0.44	0.24	0.92	0.76	-0.68
Habit and DBS Sensitivity Analysis								
High Spillover	1.23	1.02	2.28	0.38	0.22	0.93	0.72	-0.65
High Persistence	1.31	0.91	2.58	0.40	0.14	0.79	0.76	-0.54
Low Persistence	1.36	0.96	2.56	0.54	0.21	0.88	0.78	-0.61
Low Elasticity $\frac{1}{1+\mu} = .5$	1.31	1.00	2.39	0.41	0.31	0.93	0.74	-0.67
High Elasticity $\frac{1}{1+\mu} = 10$	1.37	0.95	3.85	0.49	0.59	0.89	0.73	-0.52
Low Elast + High Persist	1.20	1.06	1.63	0.30	0.34	0.92	0.69	-0.58
High Elast + High Persist	1.36	0.96	2.56	0.54	0.21	0.88	0.78	-0.61

^a From Ambler et al. (2002)

The U.S. data and the cross-country average data are from Ambler et al. (2002). The data are quarterly from 1960:1 to 1991:2 and from the O.E.C.D with missing data filled in from national sources. The U.S. cross-country correlations are between the U.S. and a European aggregate composed of Austria, Finland, France, Germany, Italy, Sweden and the U.K. The cross-country average is an average of 19 industrialized countries.

Table 5.3: Cross-Country Correlations and Country-Specific Persistence Levels

	Cross Correlations				Persistence	
	<i>y</i>	<i>c</i>	<i>i</i>	<i>n</i>	<i>y</i>	<i>c</i>
U.S. data ^a	0.66	0.51	0.53	0.33	0.84 ^b	0.87 ^c
Cross-country average ^a	0.24	0.14	0.16	0.21	0.73 ^c	0.71 ^c
Benchmark model	-0.03	0.77	-0.57	-0.62	0.66	0.65
DBS model	0.06	0.21	-0.42	-0.47	0.74	0.20
Habit	-0.04	0.64	-0.67	-0.41	0.67	0.86
Habit and DBS	0.06	0.19	-0.39	-0.48	0.74	0.51
Habit and DBS Sensitivity Analysis						
High Spillover	0.14	0.18	-0.39	-0.46	0.71	0.52
High Persistence	0.01	0.25	-0.30	-0.20	0.75	0.48
Low Persistence	0.31	0.19	0.13	0.38	0.73	0.49
Low Elasticity $\frac{1}{1+\mu} = .5$	0.09	0.17	-0.43	-0.24	0.74	0.56
High Elasticity $\frac{1}{1+\mu} = 10$	-0.02	0.17	-0.80	-0.53	0.76	0.56
Low Elast + High Persist	0.25	0.18	0.43	0.92	0.72	0.54
High Elast + High Persist	-0.32	0.30	-0.87	-0.71	0.80	0.56

^a From Ambler et al. (2002)

^b From Baxter and Crucini (1995)

^c Author computed

The U.S. data and the cross-country average data are from Ambler et al. (2002). The data are quarterly from 1960:1 to 1991:2 and from the O.E.C.D with missing data filled in from national sources. The U.S. cross-country correlations are between the U.S. and a European aggregate composed of Austria, Finland, France, Germany, Italy, Sweden and the U.K. The cross-country average is an average of 19 industrialized countries.

Table 5.4: Comparison of OLS and MM VAR estimates for Solow Residuals

Country	OLS				MM			
	Home Country		Rest of World		Home Country		Rest of World	
	ρ_1	ρ_2	ρ_1^*	ρ_2^*	ρ_1	ρ_2	ρ_1^*	ρ_2^*
Australia	0.759	0.244	-0.019	1.011	0.767	0.241	0.000	0.993
Austria	0.850	0.172	-0.028	1.023	0.916	0.091	-0.004	0.995
Canada	0.912	0.087	0.004	0.987	0.980	0.013	-0.013	1.006
Denmark	0.477	0.353	0.056	0.956	0.577	0.287	-0.042	1.023
Finland	0.879	0.157	-0.004	0.997	0.914	0.107	0.006	0.982
France	0.845	0.164	-0.022	1.019	0.892	0.111	-0.019	1.011
Germany	0.832	0.185	-0.029	1.026	0.779	0.231	-0.011	1.003
Greece	0.872	0.145	-0.005	0.996	0.929	0.076	-0.009	1.002
Italy	0.876	0.151	-0.017	1.016	0.955	0.049	0.036	0.944
Japan	0.932	0.119	0.001	0.989	0.940	0.108	0.009	0.969
Netherlands	0.799	0.187	-0.027	1.021	0.866	0.120	-0.023	1.012
Norway	0.936	0.059	-0.004	0.996	0.925	0.054	-0.005	0.996
South Africa	0.825	0.105	-0.037	1.016	0.935	0.023	-0.003	0.991
Spain	0.991	0.005	0.024	0.943	1.030	-0.091	0.019	0.949
Sweden	0.747	0.181	-0.025	1.007	0.902	0.068	0.059	0.943
Switzerland	0.770	0.200	-0.052	1.037	0.977	0.018	-0.007	0.997
United Kingdom	0.823	0.151	0.026	0.968	0.888	0.101	0.021	0.973
United States	0.980	0.005	0.076	0.954	0.924	0.031	0.069	0.958

Note: An asterisk denotes foreign countries as discussed in the text. The data for the Solow residuals are from Zimmermann (1994). The data are quarterly from 1960:1 to 1991:2 and are from the O.E.C.D with missing data filled in from national sources. The results are reported in Table 5.4. The MM-estimator is from Yohai (1996).

Appendix A

Technical Appendix for Chapter 5

Linear quadratic programming is used to generate an approximate solution of the problem near a deterministic steady state. This method has been used by BKK (1994) in the context of IRBC models. For a more general description of linear quadratic programming see Díaz-Giménez (1999) or Judd (2000).

A.1 The Model

Consider the IRBC model with habit persistence:

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \frac{1}{1-\psi} ((c_t - bc_{t-1})^\sigma L_t^{1-\sigma})^{1-\psi} \quad (\text{A.1})$$

subject to

$$y_t = F(z_t, k_t, N_t) \quad (\text{A.2})$$

$$F(z_t, k_t, N_t) = e^{z_t} k_t^\alpha N_t^{1-\alpha} \quad (\text{A.3})$$

$$\begin{pmatrix} z_t \\ z_t^* \end{pmatrix} = \begin{pmatrix} \rho_1 & \rho_2 \\ \rho_2 & \rho_1 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ z_{t-1}^* \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \varepsilon_t^* \end{pmatrix} \quad (\text{A.4})$$

$$y_t = y_{ht} + ex_t \quad (\text{A.5})$$

$$H(y_{ht}, im_t) = [(\lambda)y_{ht}^{-\mu} + (1 - \lambda)im_t^{-\mu}]^{\frac{-1}{\mu}} \quad (\text{A.6})$$

$$H(y_{ht}, im_t) = c_t + i_t \quad (\text{A.7})$$

$$k_{t+1} = (1 - \delta)k_t + i_t \quad (\text{A.8})$$

$$1 = L_t + N_t \quad (\text{A.9})$$

$$ex = im^*. \quad (\text{A.10})$$

Note that benchmark model is just a special case of this model with $b = 0$.

A.2 Deterministic Steady State

The first stage in approximating a solution to the this model is to obtain the deterministic steady state. To do this, set $(z_t, z_t^*) = (0, 0)$ and construct the Lagrangian

$$\begin{aligned}
L = & \max E_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{1-\psi} \left((c_t - bc_{t-1})^\sigma L_t^{1-\sigma} \right)^{1-\psi} \right. & (A.11) \\
& + \frac{1}{1-\psi} \left((c_t^* - bc_{t-1}^*)^\sigma L_t^{*1-\sigma} \right)^{1-\psi} \\
& + \theta_t (F(k_t, N_t) - y_{ht} - ex_t) + \theta_t^* (F(k_t^*, N_t^*) - y_{ht} - ex_t^*) \\
& + \tau_t (H(y_{ht}, ex_t^*) - c_t - i_t) + \tau_t^* (H(y_{ht}^*, ex_t) - c_t^* - i_t^*) \\
& + \omega_t (k_{t+1} - (1-\delta)k_t - i_t) + \omega_t^* (k_{t+1}^* - (1-\delta)k_t^* - i_t^*) \\
& \left. + \pi_t (L_t - (1 - N_t)) + \pi_t^* (L_t^* - (1 - N_t^*)) \right].
\end{aligned}$$

Because the countries are symmetrical, we only need to compute the steady state for one country. The steady state for the other country will be identical. The first order conditions for the home country are as follows:

$$\begin{aligned}
c_t : \quad & \sigma \left((c_t - bc_{t-1})^\sigma L_t^{1-\sigma} \right)^{-\psi} \left(\frac{c_t - bc_{t-1}}{L_t} \right)^{\sigma-1} \\
& - \beta \sigma b \left((c_{t+1} - bc_t)^\sigma L_t^{1-\sigma} \right)^{-\psi} \left(\frac{c_{t+1} - bc_t}{L_{t+1}} \right)^{\sigma-1} - \tau_t = 0
\end{aligned} \tag{A.12}$$

$$i_t : \quad -\tau_t - \omega_t = 0 \tag{A.13}$$

$$N_t : \quad \theta_t D_2 F + \pi_t = 0 \tag{A.14}$$

$$L_t : \quad (1 - \sigma) \left((c_t - bc_{t-1})^\sigma L_t^{1-\sigma} \right)^{-\psi} \left(\frac{c_t - bc_{t-1}}{L_t} \right)^\sigma + \pi_t = 0 \tag{A.15}$$

$$y_{ht} : \quad \tau_t D_1 H + \theta_t = 0 \tag{A.16}$$

$$ex : \quad \tau_t^* D_2 H - \theta_t = 0 \tag{A.17}$$

$$k_{t+1} : \quad \beta \theta_{t+1} D_1 F - \beta \omega_{t+1} (1 - \delta) + \omega_t = 0 \tag{A.18}$$

$$\theta_t : \quad F(k_t, N_t) = y_{ht} + ex_t \tag{A.19}$$

$$\tau_t : \quad H(y_{ht}, ex_t^*) = c_t + i_t \tag{A.20}$$

$$\omega_t : \quad k_{t+1} = (1 - \delta)k_t + i_t \tag{A.21}$$

$$\pi_t : \quad 1 = L_t + N_t \tag{A.22}$$

where D_i represents a derivative with respect to element i . To find the steady state, drop the time subscript and solve for (c, y_h, L, k, N, i, ex) .

Additional information is provided for the steady state relationship. The ratio $\frac{im}{y}$, import share of income, is set equal to .20 and relative prices are set equal to one, $\frac{D_1 H}{D_2 H} = 1$.

With $\frac{im}{y} = .20$, D_1H is uniquely determined as $\lambda[\lambda + (1 - \lambda).25^{-\mu}]^{\frac{-1-\mu}{\mu}}$ and

$$H = \Xi F \quad (\text{A.23})$$

where $\Xi = [\lambda.8^{-\mu} + (1 - \lambda).2^{-\mu}]^{-\frac{1}{\mu}}$. So D_1H is simply a constant and H is a function of k and N .

Combine equations A.13, A.16, and A.18 to get k as a function of N

$$k = \left(\frac{r + \delta}{D_1H} \right)^{\frac{1}{\alpha-1}} N \quad (\text{A.24})$$

where $r = \frac{1}{\beta} - 1$ which is simply the steady state interest rate.

Also combine equations A.12, A.14, A.15, A.16, and A.22 to get c as a function of k and N

$$c = \frac{D_1H D_2F * (1 - N)}{\zeta} \quad (\text{A.25})$$

where $\zeta = \left(\frac{1-\sigma}{\sigma} \right) \left(\frac{1-b}{1-\beta b} \right)$.

From the constraint in equation A.21, i can be defined as a function of k :

$$i = \delta k. \quad (\text{A.26})$$

Substituting equations A.23 - A.26 into equation A.20 and solving for N gives us the steady state relation for N ,

$$\bar{N} = \frac{(1 - \alpha)D_1H(r + \delta)}{\Xi\zeta + D_1H(r + \delta) - \alpha D_1H(r + (\zeta + 1)\delta)}. \quad (\text{A.27})$$

Substituting this result back into equation A.24 determines \bar{k} that then is used in equations A.25 and A.26 to find \bar{c} and \bar{i} . \bar{y} and thus $\bar{e}x$ are also determined by \bar{k} and \bar{N} .

A.3 Linear Quadratic Form

The second step in the solution problem is to express the problem in linear quadratic form

$$\begin{aligned} \max \quad & x'Rx + u'Qu + u'Hx \\ \text{s.t.} \quad & Ax + Bu + C\varepsilon, \end{aligned}$$

where x are the state variables and u are the control variables.

A.3.1 Quadratic Approximation

To obtain this form, first substitute all the non-linear constraints into the objective function, in this case c . Next take a second order Taylor approximation of the objective function about the steady state, $SS = (\bar{x}, \bar{u})$. In this particular problem

$$x = (1, z_t, z_t^*, z_{t-1}, z_{t-1}^*, k_t, k_t^*, k_{t-1}, k_{t-1}^*, \quad (\text{A.28})$$

$$i_{t-1}, i_{t-1}^*, N_{t-1}, N_{t-1}^*, ex_{t-1}, ex_{t-1}^*, \varepsilon_{t-1}, \varepsilon_{t-1}^*)$$

$$u = (i_t, i_t^*, N_t, N_t^*, ex_t, ex_t^*). \quad (\text{A.29})$$

Notice that that lagged-control variables must be included in the state vector because of habit formation. Also, notice that c is not listed as a control variable. This is because c was substituted out of the problem. The import variables, (im_t, im_t^*) , could be used in place of the export variables, (ex_t, ex_t^*) , because of equation A.10.

The approximation can be found using the Hessian and Jacobian matrices. Let

$$\Phi = \begin{bmatrix} \bar{\Gamma} - SS + J + \frac{1}{2}SSHessSS' & \frac{1}{2}(J - HessSS')' \\ \frac{1}{2}(J - HessSS') & \frac{1}{2}Hess \end{bmatrix} \quad (\text{A.30})$$

where $\bar{\Gamma}$ is the objective function evaluated at the steady state, J is the Jacobian of the objective function evaluated at the steady state, and $Hess$ is the Hessian of the objective function evaluated at the steady state.

R is the submatrix of Φ associated with the state variables, H is the submatrix of Φ associated with the cross terms of the state and control variables, and Q is the submatrix of Φ associated with the control variables.

A.3.2 Equation of Motion

The next step is to determine the linear portion of the approximation, $Ax + Bu + C\varepsilon$.

In this problem, A is a 17x17 matrix, B is a 17x6 matrix, and C is a 17x3 matrix.

More precisely

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho_{11} & \rho_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho_{21} & \rho_{22} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 1 - \delta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 - \delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & \cdot & & & \\ & & & & & \cdot & & & \\ & & & & & \cdot & & & \end{bmatrix} \quad (\text{A.31})$$

$$B = \begin{bmatrix} & & & \cdot & & \\ & & & \cdot & & \\ & & & \cdot & & \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (\text{A.32})$$

$$C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ & \cdot \\ & \cdot \\ & \cdot \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (\text{A.33})$$

A.4 Solution

The optimal value of our objective function, Γ , can be found by iterating on

$$\Gamma_i = R + \beta A' \Gamma_{i-1} A - (\beta A' \Gamma_{i-1} B + H')(Q + \beta B' \Gamma_{i-1} B)^{-1} (\beta B' \Gamma_{i-1} A + H) \quad (\text{A.34})$$

until the sequence Γ_i converges to Γ . Our optimal policy function, F , can then be found as

$$F = (Q + \beta B' \Gamma B)^{-1} (\beta B' \Gamma A + H). \quad (\text{A.35})$$

Therefore, our optimal control vector, \tilde{u}_t , at any time t is $\tilde{u}_t = -F x_t$. The equation of motion becomes

$$x_{t+1} = (A - BF)x_t + C\varepsilon_t. \quad (\text{A.36})$$

This is the equation that drives the theoretic economy. It is essentially a VAR model. To simulate an economy, all that is needed is to specify ε and then iterate the VAR forward in time. The optimal controls, \tilde{u}_t , can be found from equation A.38.

Decisions-before-shocks are implemented slightly differently. The optimal controls, \tilde{u}_t , are determined prior to the realization of the shock, ε .

$$\hat{x}_{t+1} = E[(A - BF)x_t + C\varepsilon_t] = (A - BF)x_t \quad (\text{A.37})$$

$$\tilde{u}_{t+1} = -F x_{t+1} \quad (\text{A.38})$$

$$x_{t+1} = \hat{x}_{t+1} + C\varepsilon_t \quad (\text{A.39})$$

where \hat{x}_{t+1} is the expected state of the economy at time period $t + 1$. The optimal controls are thus determined without the realization of the shock to the economy.

Consumption is implicitly determined from output, trade, and investment. Trade and investment are fixed before the shock is realized. As output changes from the shock, consumption must also change. After the shock realization, the agents have

no degrees of freedom to adjust any of their explicit controls. Thus, the only way to balance the equations and not have an excess or shortage of output is to allow consumption to change.

The decision rule, F , is still optimal with respect to the state of the economy prior to the shock, because of the certainty equivalence associated with linear quadratic approximations. For this reason, notice that the value function iteration in (A.34) does not and need not include the stochastic shocks.

Appendix B

An Outlier Robust Estimation

Package for OX

This appendix will list and describe several robust estimation procedures for use in the Ox programming language (Doornik 2001). All procedures are based on algorithms from the estimator's respective creator.

B.1 Program `robust.ox`

M-estimator (Huber 1964) based on Huber's Ψ function ($k=1.654$)

HM(const y, const x, const beta, const scale)

y in: $T \times 1$ vector

x in: $T \times k$ matrix

beta in: $k \times 1$ matrix initial estimate usually OLS

scale in: address of variable

out: estimate of scale

Returns a $k \times 1$ matrix of parameter estimates

M-estimator (Huber 1964) based on Tukey's bisquare function ($k=1.56$)

TM(const y, const x, const beta, const scale)

y in: $T \times 1$ vector

x in: $T \times k$ matrix

beta in: $k \times 1$ matrix initial estimate usually OLS

scale in: address of variable

out: estimate of scale

Returns a $k \times 1$ matrix of parameter estimates

MM-estimator (Yohai 1986) ($k_0=1.56$, $k_1=4.68$) .5 breakdown 95% efficiency

myMM(const y, const x, const beta, const scale)

y in: $T \times 1$ vector

x in: $T \times k$ matrix

beta in: $k \times 1$ matrix initial robust estimate

scale in: address of variable

out: estimate of scale

Returns a $k \times 1$ matrix of parameter estimates

S-estimator (Rousseeuw and Yohai 1984)($k=1.547$)

myS(const y, const x, const beta, const scale)

y in: $T \times 1$ vector

x in: $T \times k$ matrix

beta in: $k \times 1$ matrix initial robust estimate

scale in: address of variable

out: estimate of scale

Returns a $k \times 1$ matrix of parameter estimates

τ -estimator (Yohai and Zamar 1988) $k_0=1.56$. $k_1=6.08$) .5 breakdown 95% efficiency

mytau(const y, const x, const beta, const scale)

y in: $T \times 1$ vector

x in: $T \times k$ matrix

beta in: $k \times 1$ matrix initial robust estimate

scale in: address of variable

out: estimate of scale

Returns a $k \times 1$ matrix of parameter estimates

Least Median of Squares (Rousseeuw 1984)

myLMS(const y, const x, const scale)

y in: $T \times 1$ vector

x in: $T \times k$ matrix

scale in: address of variable

out: estimate of scale

Returns a $k \times 1$ matrix of parameter estimates

RA-estimator (Bustos and Yohai 1986)

myMM(const y, const p, const scale)

y in: $T \times 1$ vector

p in: scalar number of lags to use

scale in: address of variable

out: estimate of scale

Returns a $(p+1) \times 1$ matrix of parameter estimates

B.2 Program rscale.ox

Q-scale estimate (Rousseeuw and Croux 1993)

$Q_n(x)$

x in: $T \times 1$ vector

Return a scalar scale estimate

MAD-scale estimate

$MAD(x)$

x in: $T \times 1$ vector

Return a scalar scale estimate

Robust Variance Covariance Matrix (Yohai 1986)

$VC0(\text{const } y, \text{const } x, \text{const } \beta, \text{const } \text{scale})$

y in: $T \times 1$ vector

x in: $T \times k$ matrix

beta in: $k \times 1$ matrix parameter estimate

scale in: address of variable

Returns a $k \times k$ matrix containing the robust variance covariance matrix

Robust Autocovariance (Yanyuan and Genton 2000)

`racov(const x, const h)`

x in: T×1 vector

h in: scalar lag number

Return a scalar containing the robust autocovariance of **x** at lag **h**

B.3 robust.ox code

```
#include <oxstd.h>
#include <oxprob.h>
#include <arma.h>

pMedian(const py) {
    decl t, res1, sortedr, medy, oddeven;
    t = rows(py);
    //compute the median of py
    sortedr = sortc(py);
    oddeven = imod(t,2);
    if(oddeven==1)
    {
        medy = sortedr[(t+1)/2-1]; //index starts at 0
    }
    //need to subtract 1
    else
    { //interpolate if even #
        medy = (sortedr[(t/2)-1]+sortedr[(t/2)+1-1])/2;
    }
    return medy;
} pMAD(const py) //mad(y)=mediani(yi-medjyj) {
    decl t, res1, sortedr, medy, oddeven, madest;
    t = rows(py);
    //compute the median of py
    medy=pMedian(py);
    //compute the median of adjusted series
    res1=((py-medy).^2).^0.5;
    sortedr = sortc(res1);
    oddeven = imod(t,2);
    if(oddeven==1)
    {
        madest = 1.4826*sortedr[(t+1)/2-1]; //index starts at 0
    }
    //need to subtract 1
    else
    {
        //interpolate if even #
        madest = 1.4286*(sortedr[(t/2)-1]+sortedr[(t/2)+1-1])/2;
    }
}
```

```

    return madest;
}
//This function computes a LMS estimate from a subsample
itLMS(const py, const px) {
    decl n, t, subsamp, mB, mA, Theta, i;
    //determine size of x and y
    n = columns(px);
    t = rows(px);
    //randomly generate a subsample of x and y
    subsamp = ransubsample(n, t);
    mB = zeros(n, 1);
    mA = zeros(n, n);
    for(i=0; i < n; ++i)
    {
        mB[i] = py[subsamp[i]];
        mA[i][] = px[subsamp[i]][];
    }
    //solve the system to obtain an estimate
// if (determinant(mA)!=0) //possibility of singular matrices
// {

        Theta = invertgen(mA,0)*mB; //compute the estimate if possible
        return Theta;
// }

}

//
//This function computes the med square error
mdserror(const py, const px, const ptheta) {
    decl t, half, res1, ressq, sortedr, medsqer,oddeven;
    t = rows(py);
    res1 = py - px*ptheta;
    ressq = res1.*res1;

    sortedr = sortc(ressq);
    medsqer=pMedian(sortedr);
    return medsqer;
}
//
//This function computes the number of reps to do
numreps(const px) {

```

```

decl P, T, poss, i;

    //compute possible combinations
P = columns(px);
T = rows(px);
    poss = T;
for (i = 1; i<(P); ++i)
{
    poss = poss*(T-i)/(P-i+1);
}
    //poss holds the # of possible choices for LMS
if (poss>10000)
{
    poss = 10000;
}
return poss;
}

//This will compute the scale estimate for LMS
Scale(const py, const px, const BMS) {
decl
    prelim, mredsqa, weight, res1, test, P, T, wres, pho;

P = columns(px);
T = rows(px);
weight = zeros(T,1);
mredsqa = mdserror(py, px, BMS);
//compute intial scale estimate
prelim = 1.4826*(1+(5/(T-P)))*sqrt(mredsqa);
    //determine the weights
res1 = (py - px*BMS);
test = res1/prelim;
test = (test.*test).^0.5;
weight = test.<=2.5 .*1 .: 0;
wres = weight.*res1;
    //compute scale estimate
pho = sqrt((wres'*wres)/(weight'*weight - P));
return pho;
}

```

```

//This function will use the exact algorithm
//to adjust the constant term
AdjConst(const py, const px, const BMS) {
    decl T, i, nconst, newest, meder, wocon, woones, contemp, test;

    test = 100000; //used for first case
    T = rows(px);
    if(columns(px>1))
    {
        //remove the constant estimate and row of ones
        wocon = BMS[1:][];
        woones = px[][1:];

        //get the constants to be tried
        contemp = py - woones*wocon;

// determine the constant with the smallest med squared error
for(i=0; i<T; ++i)
    {
        newest = contemp[i]|wocon;
        meder = mdserror(py, px, newest);
        if(meder<test)
        {
            nconst = contemp[i];
            test = meder;
        }
    }
// add the new constant estimate back to the other param estimates
newest = nconst|wocon;
return newest;
}
else
{
return BMS;
}
}

```

```

//This procedure computes the LMS estimate
myLMS(const py, const px, const pscale) {
    decl i, poss, mederr, Btemp, medtemp, BLMS;

```

```

    mederr = 100000;
    poss = numreps(px);
// print("\n", "reps ", poss, "\n");
    for(i=0; i < poss; ++i)
    {
        Btemp = itLMS(py, px);
        medtemp = mdserror(py, px, Btemp);
        if (medtemp < mederr)
        {
            mederr = medtemp;
            BLMS = Btemp;
        }
    }
    BLMS = AdjConst(py, px, BLMS);
    pscale[0] = Scale(py, px, BLMS);
// print("Least Median of Squares\n ");
// print("%r", xname, BLMS, "\n");
// print("scale ", pscale, "\n");
    return BLMS;
}
////////////////////////////////////
///end LMS estimate////////////////////////////////////

////////////////////////////////////
///begin MM estimate////////////////////////////////////
///.5 breakdown .95 efficiency
//// k0=1.56
//// k1=4.68
//////////
///// .25 breakdown .85 efficiency
//// k0=2.9366
//// k1=3.440

//This function will compute the pho function with k0 = 1.56
pho0(const phi) {
    decl test, result, u;
// u = phi/1.56;
u=phi/2.9366;
    test = (u.^2)/2 - (u.^4)/2 + (u.^6)/6;
    result = ((u.^2).^(.5)).<=1 .?test .: 1/6;
    return result;
}

```

```

}

pho1(const phi) {
    decl test, result, u;
//    u = phi/4.68;
    u = phi/3.44;
    test = (u.^2)/2 - (u.^4)/2 + (u.^6)/6;
    result = ((u.^2).^(.5)).<=1 .?test .: 1/6;
    return result;
}

//This function will compute the bisquare psi-function
//with k1 = 4.68
gamma1(const phi) {
    decl test, result, u;
//    u = phi/4.68;
    u=phi/3.44;
    test = u.*(1-u.^2).^2;
    result = ((u.^2).^(.5)).<=1 .?test .: 0;
    return result;
}

//This function will compute the M-Scale for T0 with b = .0833
Mscale(const py, const px, const Bo, const sc0) //sc0 is the
initial scale estimate {
    decl T, sntemp, test, b, res1, i, phosum, snsq, sn;
    T = rows(py);
    sn = sc0;
    b = 0.0833;
    res1 = (py - px*Bo);
    test = 1;
    i = 0;
    while (test>.005 && i < 500)
    {
        phosum = sumc(pho0(res1/sn));
        snsq = ((sn^2)/(T*b))*phosum;
        sntemp = snsq^(.5);
        test = ((sn - sntemp)^2)^(.5);
        sn = sntemp;
        ++i;
    }
//    print("\n", "Mscale", sn, "\n");
}

```

```

// print("did the fixed point equation work\n");
// print("these numbers should be the same\n");
// print("pho\n", phosum/T, "b\n      ", b, "\n");
return sn;
}

Subject(const py, const px, const theta, const sn) {
    decl resids, S;

    resids = py - px*theta;
    S = sumc(pho1(resids/sn));
    return S;
}

//This procedure will generate 1 iteration for the MM-estimate
MMiter(const py, const px, const beta, const sn) {
    decl weight, dwght, resids, g, M, est0, est1, dest, msc, T, n,
    S0, S1;

    n = 0;
    est0 = beta;
    T = rows(px);
    dwght = zeros(T,T);
    msc = sn;
    resids = py - px*beta;
    weight = gamma1(resids/msc)./(resids/msc);
    weight = resids .== 0 .?1 .:weight;
    g = ((1/msc)*sumc(gamma1(resids/msc).*px))';
    dwght = setdiagonal(dwght, weight);
    M = px'*dwght*px;
    M = (1/msc^2)*M;
    dest = invert(M)*g;
    S0 = Subject(py, px, est0, msc);
    S1 = S0 + 1;
    while (S1>S0 && n<100)
    {
        est1 = est0 + dest/(2^n);
        S1 = Subject(py, px, est1, msc);
        ++n;
    }
}

```

```

    return est1;
}

myMM(const py, const px, const BLMS, const sc) {
    decl LMSscale, Msn, BMM1, test, BMM0, n;

    n = 0;
    test = 1;
// BLMS = myLMS(py, yname, px, xname);
    LMSscale = Scale(py, px, BLMS);
    Msn = Mscale(py, px, BLMS, LMSscale);
    sc[0] = Msn;
    BMM0 = BLMS;
    while (test>.005 && n <500)
    {
        BMM1 = MMiter(py, px, BMM0, Msn);
        test = (BMM1 - BMM0)'(BMM1 - BMM0);
        ++n;
        BMM0 = BMM1;
    }
    return BMM1;
}
//////////end MM

//////////
//////////
//Tukey M
//////////
//////////

TMiter(const py, const px, const beta, const sn) {
    decl weight, dwght, resids, g, M, est0, est1, dest, msc, T, n,
    S0, S1;

    n = 0;
    est0 = beta;
    T = rows(px);
    dwght = zeros(T,T);
    msc = sn;
    resids = py - px*beta;

```

```

weight = gamma1(resids/msc)./(resids/msc);
weight = resids ./= 0 .?1 .:weight;
g = ((1/msc)*sumc(gamma1(resids/msc).*px))';
dwght = setdiagonal(dwght, weight);
M = px'*dwght*px;
M = (1/msc^2)*M;
dest = invert(M)*g;
S0 = Subject(py, px, est0, msc);
S1 = S0 + 1;
while (S1>S0 && n<1000)
{
est1 = est0 + dest/(2^n);
S1 = Subject(py, px, est1, msc);
++n;
}

return est1;

} TM(const py,const px, const BLMS, const sc) {
decl mad, BMM1, test, BMM0, n;

n = 0;
test = 1;
// BLMS = myLMS(py, yname, px, xname);
mad=pMAD(py);
sc[0] = mad;
BMM0 = BLMS;
while (test>.001 && n <1000)
{
BMM1 = TMiter(py, px, BMM0, mad);
test = (BMM1 - BMM0)'(BMM1 - BMM0);
++n;
BMM0 = BMM1;
}
return BMM1;
}

////////////////////////////////////
////////////////////////////////////
//Huber M
////////////////////////////////////
////////////////////////////////////

```

```

phoH(const phi) {
    decl good, bad, result, u,k;
    k=1.654;
    u = phi;
    good = (u.^2)/2;
    bad=k*fabs(u)-k^2/2;
    result = fabs(u).<=k .?good .:bad ;
    return result;
}

Hgamma(const phi) { decl maxi, mini,k,u;
    k=1.654;
    u=phi;
    maxi=maxc((-k~u)')';
    mini=minc((k~maxi)')';
    return mini;
}

Hobject(const py, const px, const theta, const sn) {
    decl resids, S;

    resids = py - px*theta;
    S = sumc(phoH(resids/sn));
    return S;
}

HMiter(const py, const px, const beta, const sn) {
    decl weight, dwght, resids, g, M, est0, est1, dest, msc, T, n,
    S0, S1;

    n = 0;
    est0 = beta;
    T = rows(px);
    dwght = zeros(T,T);
    msc = sn;
    resids = py - px*beta;
    weight = Hgamma(resids/msc)./(resids/msc);
    weight = resids .== 0 .?1 .:weight;
    g = ((1/msc)*sumc(Hgamma(resids/msc).*px))';
    dwght = setdiagonal(dwght, weight);
}

```

```

M = px'*dwght*px;
M = (1/msc^2)*M;
dest = invert(M)*g;
S0 = Hobject(py, px, est0, msc);
S1 = S0 + 1;
while (S1>S0 && n<1000)
{
est1 = est0 + dest/(2^n);
S1 = Hobject(py, px, est1, msc);
++n;
}

return est1;

} HM(const py,const px, const BLMS, const sc) {
  decl mad, BMM1, test, BMM0, n;

  n = 0;
  test = 1;
// BLMS = myLMS(py, yname, px, xname);
  mad=pMAD(py);
  sc[0] = mad;
  BMM0 = BLMS;
  while (test>.001 && n <1000)
  {
  BMM1 = HMIter(py, px, BMM0, mad);
  test = (BMM1 - BMM0)'(BMM1 - BMM0);
  ++n;
  BMM0 = BMM1;
  }
  return BMM1;
}

////////////////////////////////////
////////////////////////////////////
////////////////////////////////////begin tau
//compute p1 with c1 = 1.56
tphi1(const phi) {

  decl test, result, u, c;
  c = 1.56;
  u = phi;
  test = ((u.^2)/2).*(1 - (u.^2/c^2) + (u.^4/(3*c^4)));

```

```

    result = ((u.^2).^(.5)).<=c .*test .: c^2/6;
    return result;
}
//This function will compute the bisquare psi-function
tgamma1(const phi) {
    decl test, result, u, c;
    c = 1.56;
    u = phi;
    test = u.*(1-(u.^2/c^2)).^2;
    result = ((u.^2).^(.5)).<=c .*test .: 0;
    return result;
} tpho2(const phi) {

    decl test, result, u, c;
    c = 6.08;
    u = phi;
    test = ((u.^2)/2).*(1 - (u/c).^2 + ((u/c).^4)/3);
    result = ((u.^2).^(.5)).<=c .*test .: c^2/6;
    return result;

}
//This function will compute the bisquare psi-function
tgamma2(const phi) {
    decl test, result, u, c;
    c = 6.08;
    u = phi;
    test = u.*(1-(u/c).^2).^2;
    result = ((u.^2).^(.5)).<=c .*test .: 0;
    return result;
}

//This function will compute the t-Scale for T0 with b = .203
tscale(const py, const px, const Bo, const sc0) //sc0 is the
initial scale estimate {
    decl T, sntemp, test, b, res1, i, phosum, snsq, sn;
    T = rows(py);
    sn = sc0;
    b = 0.203;
    res1 = (py - px*Bo);
    test = 1;

```

```

    i = 0;
    while (test>.001 && i < 10000)
    {
    phosum = sumc(tpho1(res1/sn));
    snsq = ((sn^2)/(T*b))*phosum;
    sntemp = snsq^(.5);
    test = ((sn - sntemp)^2)^(.5);
    sn = sntemp;
    ++i;
    }
    return sn;
}
//This procedure will compute the tau scale estimate
tau(const py, const px, const theta, const sn) {
    decl resid, S;

    resid = py - px*theta;
    S = (sn^2/(rows(px)))*sumc(tpho2(resid/sn));
    S = S^(.5);
    return S;
}
//This procedure will generate 1 iteration for the MM-estimate
tauiter(const py, const px, const beta, const sn) {
    decl Wn, weight, dwght, resid, g, M, est0, est1, dest, msc, T, n,
    S0, S1, tsc;

    n = 0;
    est0 = beta;
    T = rows(px);
    dwght = zeros(T,T);
    msc = sn;
    resid = py - px*beta;
    Wn = sumc(2*tpho2(resid/msc)- tgamma2(resid/msc).*(resid/msc))/
        (sumc(tgamma1(resid/msc).*(resid/msc)));

    weight = (Wn*tgamma1(resid/msc)+tgamma2(resid/msc))/(resid/msc);
    weight = resid .== 0 .?1 .:weight;
    g = ((-2/T)*sumc(weight.*resid.*px))';
    dwght = setdiagonal(dwght, weight);
    M = px'*dwght*px;
    M = (2/T)*M;

```

```

dest = -1*invert(M)*g;
S0 = tau(py, px, est0, msc);
S1 = S0 + 1;

while (S1>S0 && n<1000)
{
est1 = est0 + dest/(2^n);
tsc = tscale(py, px, est1, msc);
msc = tsc;
S1 = tau(py, px, est1, msc);
++n;
}

return est1;
}

mytau(const py,const px, const BLMS, const sc) {
decl LMSscale, ttemp, tsn, Btau1, test, Btau0, n;

n = 0;
test = 1;
LMSscale = Scale(py, px, BLMS);
tsn = tscale(py, px, BLMS, LMSscale);

Btau0 = BLMS;
test = 10000;
while (test > .0005 && n <1000) //conergence criteria
{
Btau1 = tauiter(py, px, Btau0, tsn);
ttemp = tscale(py, px, Btau1, tsn);
++n;
tsn = ttemp;
test = (Btau1 - Btau0)'(Btau1 - Btau0);
Btau0 = Btau1;

print(n, test, tsn);
}
sc[0] = tsn;
return Btau1;
}

```

```

////////////////////////////////////
Scale0(const py, const px, const BMS) {
  decl
    prelim, mredsqa, weight, res1, test, P, T, wres, pho;

    P = columns(px);
    T = rows(px);
    weight = zeros(T,1);
    mredsqa = mdserror(py, px, BMS);
    //compute initial scale estimate
    prelim = 1.4826*(1+(5/(T-P)))*sqrt(mredsqa);
    //determine the weights
    res1 = (py - px*BMS);
    test = res1/prelim;
    test = (test.*test).^0.5;
    weight = test.<=2.5 .*1 .: 0;
    wres = weight.*res1;
    //compute scale estimate
    pho = sqrt((wres'*wres)/(weight'*weight - P));
    return pho;
} srho(const phi) {

  decl test, result, u, c;
  u = phi;
  c = 1.547;
  test = (u.^2)/2 - (u.^4)/(2*c^2) + (u.^6)/(6*c^4);
  result = abs(u).<=c .*test .: c^2/6;
  return result;
} sgam(const phi) {
  decl test, result, u,c;
  u = phi;
  c = 1.547;
  test = u.*(1-(u/c).^2).^2;
  result = abs(u).<=c .*test .: 0;
  return result;
}
//S- estimator
//This procedure will generate 1 iteration for the S-estimate
//sc0 is the initial scale estimate
Sscale(const py, const px, const Bo, const sc0)

```

```

{
  decl T, sntemp, test, b, res1, i, phosum, snsq, sn, check;
  T = rows(py);
  sn = sc0[0];
  b = 0.2;
  res1 = (py - px*Bo);
  test = 1;
  i = 0;
  while (test>.0001 && i < 10000)
  {
    phosum = sumc(srho(res1/sn));
    snsq = ((sn^2)/(T*b))*phosum;
    sntemp = snsq^(.5);
    test = abs(sn-sntemp);
    sn = sntemp;
    ++i;
  }
  return sn;
}

```

```

Siter(const py, const px, const beta, const sn) {
  decl Wn, weight, dwght, resids, g, M, est0, est1, dest, msc, T, n,
  S0, S1;
  n = 0;
  est0 = beta;
  T = rows(px);
  dwght = zeros(T,T);
  msc = sn;
  resids = py - px*beta;
  Wn = sumc(2*srho(resids/msc)- sgam(resids/msc).*(resids/msc))/
    (sumc(sgam(resids/msc).*(resids/msc)));

  weight = (Wn*sgam(resids/msc)+sgam(resids/msc))/(resids/msc);
  weight = resids .== 0 .?1 .:weight;
  g = ((-2/T)*sumc(weight.*resids.*px))';
  dwght = setdiagonal(dwght, weight);
  M = px'*dwght*px;
  M = (2/T)*M;
  dest = -1*invert(M)*g;
  S0 = Sscale(py, px, est0, msc);
  S1 = S0 + 1;
}

```

```

while (S1>S0 && n<1000)
{
est1 = est0 + dest/(2^n);
S1 = Sscale(py, px, est1, msc);
++n;
}

return est1;

}

myS(const py,const px, const B0, const sc) {
decl B0scale, stemp, ssn, Bs1, test, Bs0, n;

n = 0;
test = 1;
B0scale = Scale(py, px, B0);
ssn = Sscale(py, px, B0, B0scale);

Bs0 = B0;
test = 10000;
while (test > .0001 && n <1000)
{
Bs1 = Siter(py, px, Bs0, ssn);
stemp = Sscale(py, px, Bs1, ssn);
++n;
ssn = stemp;
test = (Bs1 - Bs0)'(Bs1 - Bs0);
Bs0 = Bs1;

// print(n, test, ssn);
}
sc[0] = ssn;
return Bs1;

}
////////////////////////////////////

gamma(const phi) {
decl test, result, u;
u = phi/4.65;

```

```

    test = u.*(1-(u.^2)).^2;
    result = ((u.^2).^(.5)).<=1 .?test .: 0;
    return result;
} gammaH(const phi) {
    decl result,u, neg;

    u=phi;
    result=u.<=1.65 .?u .: 1.65;
    result=u.>=-1.65 .?result .: -1.65;

// result=u.<=2.65 .?u .: 2.65;
// result=u.>=-2.65 .?result .: -2.65;
    return result;
}

//This Procedure computes Yohai's (1988) RA estimate
RA(const py, const p, const sc) {
    decl scale, z, r, beta, beta2,X, n, resids, test, t, temp,i,con;

    t=0;
    test=10000;
    n=rows(py);

    X=zeros(n,p+1);
    X[][0]=constant(1,n,1);
    for (i=1; i<=p; ++i)
    {
        X[][i]=lag0(py,i);
    }
// beta=myLMS(py,X,&scale);
// beta2=beta;
    olsc(py,X,&beta);
// print(beta);
// scale=((py-X*beta)'(py-X*beta));
    z=py;
while(test>.005 && t<20) {
    temp=beta;
    resids=z-X*beta;
// print(resids);
    scale=pMAD(resids);
// scale=((py-X*beta)'(py-X*beta));
// print(gammaH(resids/scale));

```

```

// print(scale);
r=gammaH(resids/scale)*scale;
// r=resids;
// print(scale, "\n");
// print(r);
con=constant(beta[0],n,1);
for(i=0; i<p; ++i)
{
    con[i]=py[i];
}
if(p==0)
{
    z=armagen(con,r,<>,0,0);
}
else
{
    z=armagen(con, r, beta[1:]', p, 0);
}
// print(beta');
// print(py~X*beta~r~z);
for (i=1; i<=p; ++i)
{
    X[][i]=lag0(z,i);
}
olsc(z,X,&beta);
test=(temp-beta)'(temp-beta);
// print(beta);
}
// sc[0]=scale;
sc[0]=resids'*resids/(n-i-1);
// if (beta[1]>=1)
// {beta= beta2; print("*");}
return beta;
}

```

B.4 rscale.ox code

```
#include <oxstd.h> #include <oxprob.h>

himed(const py) {
    decl t, res1, sortedr, medy, oddeven;
    t = rows(py);
    //compute the median of py
    sortedr = sortc(py);
    oddeven = imod(t,2);
    if(oddeven==1)
    {
        medy = sortedr[(t+1)/2-1]; //index starts at 0
    } //need to subtract 1
    else
    {
        medy = sortedr[(t/2)+1-1];
    }
    return medy;
}

lomed(const py) {
    decl t, res1, sortedr, medy, oddeven;
    t = rows(py);
    //compute the median of py
    sortedr = sortc(py);
    oddeven = imod(t,2);
    if(oddeven==1)
    {
        medy = sortedr[(t+1)/2-1]; //index starts at 0
    } //need to subtract 1
    else
    {
        medy = sortedr[(t/2)-1]; //interpolate if even #
    }
    return medy;
}

whimed(const x, const weight, const length) {
    decl iw, a, n, acand, iw cand, wtotal, wrest, wleft, wmid, wright,
    nn, i, trial, pwhimed, flag, kcand;
```

```

a=x[:length-1];
iw=weight[:length-1];
flag=0;
n=rows(a);
acand=zeros(n,1);
iwcand=zeros(n,1);
nn=n;
wtotal=0;
for(i=1; i<=nn; ++i)
{
    wtotal = wtotal +iw[i-1];
} //end for
wrest=0;

//while statement?
while (flag==0)
{
    trial = a[nn/2];
    wleft=0;
    wmid=0;
    wright=0;
    for(i=1; i<=nn; ++i)
    {
        if (a[i-1]<trial)
        {
            wleft=int(wleft+iw[i-1]);
        }
        else
        {
            if(a[i-1]>trial)
            {
                wright=int(wright+iw[i-1]);
            }
            else
            {
                wmid=int(wmid+iw[i-1]);
            }
        }
    }
}
if((2*wrest+2*wleft)>wtotal)
{

```

```

kcand=0;
for(i=1; i<=nn; ++i)
{
    if(a[i-1]<trial)
    {
        kcand=kcand+1;
        acand[kcand-1]=a[i-1];
        iw cand[kcand-1]=iw[i-1];
    }//end if
} //end for
nn=kcand;
}
else
{
    if((2*wrest+2*wleft+2*wmid)>wtotal)
    {
        pwhimed=trial;
        flag=1;
    }
    else
    {
        flag=0;
        kcand=0;
        for(i=1; i<=nn; ++i)
        {
            if(a[i-1]>trial)
            {
                kcand=kcand+1;
                acand[kcand-1]=a[i-1];
                iw cand[kcand-1]=iw[i-1];
            } //end if
        } //end for
        nn=kcand;
        wrest=wrest+wleft+wmid;
    } //end if
} //end if
if (flag==0)
{
    for(i=1; i<=nn; ++i)
    {
        a[i-1]=acand[i-1];
        iw[i-1]=iw cand[i-1];
    }
}

```

```

    }//end for
  }//end if
} //end while
return pwhimed;
} //end proc

Sn(const x) {
  decl n, medA, medB, rightA, rightB, tryA, tryB, diff,
  Amin, Amax, even, half, y, a2, length, i, nB, nA, leftA,
  cn, leftB, sn, sa2;

  n=rows(x);
  a2=zeros(n,1);
  y=zeros(n,1);
  y = sortc(x);
  a2[0]= y[n/2]-y[0];

  for(i=2; i<=int((n+1)/2); ++i)
  {
//   print("loop1 i ", i, "\n");
    nA=i-1;
    nB=n-i;
    diff=nB-nA;
    leftA=1;
    leftB=1;
    rightA=nB;
    rightB=nB;
    Amin=int(diff/2+1);
    Amax=int(diff/2+nA);

    while(leftA<rightA)
    {
//   print("leftA ", leftA, "rightA ", rightA, "\n");
    length=rightA-leftA+1;
    even = 1-imod(length,2);
    half=int((length-1)/2);
    tryA=leftA+half;
    tryB=leftB+half;
    if(tryA<Amin)
    {

```



```

{
//  print("loop2 i ", i, "\n");
    nA=n-i;
    nB=i-1;
    diff=nB-nA;
    leftA=1;
    leftB=1;
    rightA=nB;
    rightB=nB;
    Amin=int(diff/2+1);
    Amax=int(diff/2+nA);

while(leftA<rightA)
{

    length=rightA-leftA+1;
    even = 1-imod(length,2);
    half=int((length-1)/2);
    tryA=leftA+half;
    tryB=leftB+half;
    if(tryA<Amin)
    {
        rightB=tryB;
        leftA=tryA+even;
    }
    else
    {
        if(tryA>Amax)
        {
            rightA=tryA;
            leftB=tryB+even;
        }
        else
        {
            medA=y[i+tryA-Amin]-y[i-1];
            medB=y[i-1]-y[i-tryB-1];
            if(medA>=medB)
            {
                rightA=tryA;
                leftB=tryB+even;
            }
            else

```

```

        {
            rightB=tryB;
            leftA=tryA+even;
        }
    }
}
if(leftA>Amax)
{
    a2[i-1]=y[i-1]-y[i-leftB-1];
}
else
{
    medA=y[i+leftA-Amin]-y[i-1];
    medB=y[i-1]-y[i-leftB-1];
    a2[i-1]=min(medA,medB);
}
} //end while loop
} //end 2nd do loop
a2[n-1]=y[n-1]-y[(n+1)/2-1];
cn=1;
if(n<9)
{
    if(n==2){cn=.743;}
    if(n==3){cn=1.851;}
    if(n==4){cn=.954;}
    if(n==5){cn=1.351;}
    if(n==6){cn=.993;}
    if(n==7){cn=1.198;}
    if(n==8){cn=1.005;}
    if(n==9){cn=1.131;}
}
else
{
    if(imod(n,2)==1)
    {
        cn=n/(n-0.9);
    }
}
}

sn=cn*1.1926*lomed(a2);
return sn;
} //end procedure

```

```

naiveS(const py) {
  decl n, i, j, a2, a1, s, cn;

  n=rows(py);
  a1=zeros(n,1);
  a2=zeros(n,1);
  for(i=1;i<=n; ++i)
  {
    for(j=1; j<=n; ++j)
    {
      a1[j-1]=((py[i-1]-py[j-1])^2)^.5;
    }
    a2[i-1]=himed(a1);
  }
  if(n<=9)
  {
    if(n==2){cn=.743;}
    if(n==3){cn=1.851;}
    if(n==4){cn=.954;}
    if(n==5){cn=1.351;}
    if(n==6){cn=.993;}
    if(n==7){cn=1.198;}
    if(n==8){cn=1.005;}
    if(n==9){cn=1.131;}
  }
  else
  {
    if(imod(n,2)==1)
    {
      cn=n/(n-0.9);
    }
  }
  s=cn*1.1926*lomed(a2);
  return s;
}

Qn(const px) {
  decl pQn, i, n, x, y, work, left, right, weight, Q, P,
  h, k, knew, jhelp, nL, nR, sumQ, sumP, found, j,
  jj, trial, dn, twork;

```

```

x=px;
n=rows(px);
y=zeros(n,1);
work=zeros(n,1);
left=zeros(n,1);
right=zeros(n,1);
weight=zeros(n,1);
Q=zeros(n,1);
P=zeros(n,1);
found=0;
y=sortc(x);
h=int(n/2+1);
k=int(h*(h-1)/2);
for(i=1; i<=n; ++i)
{
    left[i-1]=n-i+2;
    right[i-1]=n;
} //end for
jhhelp=int(n*(n+1)/2);
knew=k+jhhelp;
nL=jhhelp;
nR=n*n;
found=0;

//while statement

while(((nR-nL)>n)&&(found!=1))
{
    j=1;
    for (i=2; i<=n; ++i)
    {
        if(left[i-1]<=right[i-1])
        {
            weight[j-1]=right[i-1]-left[i-1]+1;
            jhhelp=left[i-1]+int(weight[j-1]/2);
            work[j-1]=y[i-1]-y[n-jhhelp];
            j=j+1;
        } //end if
    } //end for
    trial=whimed(work, weight, j-1);
    j=0;
    for(i=n; i>=1; --i)

```

```

{
  while((j<n)&&((y[i-1]-y[n-j-1])<trial))
  {
    j=j+1;
  } //end while
  P[i-1]=j;
} //end for
j=n+1;
for(i=1;i<=n;++i)
{
  while(((y[i-1]-y[n-j+1])>trial)&&(j>=0))
  {
    j=j-1;
  } //end while
  Q[i-1]=j;
} //end for
sumP=sumc(P);
sumQ=sumc(Q)-n;
if(knew<=sumP)
{
  for(i=1; i<=n; ++i)
  {
    right[i-1]=P[i-1];
  } //end for
  nR=sumP;
} //end if
else
{
  if(knew>sumQ)
  {
    for(i=1; i<=n; ++i)
    {
      left[i-1]=Q[i-1];
    } //end for
    nL=sumQ;
  }
  else
  {
    pQn=trial;
    found=1;
  } //end else
} //end else

```

```

} //end while

if(found==0)
{
    j=1;
    for(i=2; i<=n; ++i)
    {
        if((left[i-1]<=right[i-1]))
        {
            for(jj=int(left[i-1]); jj<=int(right[i-1]); ++jj)
            {
                if(j<=n){ work[j-1]=y[i-1]-y[n-jj];}
                j=j+1;
            } //end for
        } //end if
    } //end for
    twork=work[:j-2];
    twork=sortc(twork);
    pQn=twork[knew-nL-1];
} //end if
if(n<=9)
{
    if(n==2){dn=.399;}
    if(n==3){dn=.994;}
    if(n==4){dn=.512;}
    if(n==5){dn=.844;}
    if(n==6){dn=.611;}
    if(n==7){dn=.857;}
    if(n==8){dn=.669;}
    if(n==9){dn=.872;}
}
else
{
    if(imod(n,2)==1)
    {
        dn=n/(n+1.4);
    }
    else
    {
        dn=n/(n+3.8);
    } //end else
} //end else

```

```

    pQn=dn*2.2219*pQn;
    return pQn;
} //end proc

//robust autocovariance function
racov(const x, const h)
{
    decl u, v, n, gamma;

    n=rows(x);
    u=x[:n-h-1];
    v=x[h:];
    gamma=1/4*(Qn(u+v)^2-Qn(u-v)^2);
    return gamma;
}

Median(const py) {
    decl t, res1, sortedr, medy, oddeven;
    t = rows(py);
    //compute the median of py
    sortedr = sortc(py);
    oddeven = imod(t,2);
    if(oddeven==1)
    {
        medy = sortedr[(t+1)/2-1]; //index starts at 0
    } //need to subtract 1
    else
    { //interpolate if even #
        medy = (sortedr[(t/2)-1]+sortedr[(t/2)+1-1])/2;
    }
    return medy;
}

MAD(const py) //mad(y)=mediani(yi-medjyj) {
    decl t, res1, sortedr, medy, oddeven, madest;
    t = rows(py);
    //compute the median of py
    medy=whimed(py,constant(1,t,1),t);
//medy=Median(py);
    //compute the median of adjusted series
    res1=((py-medy).^2).^0.5;
    sortedr = sortc(res1);

```

```

oddeven = imod(t,2);
if(oddeven==1)
{
    madest = 1.4826*sortedr[(t+1)/2-1]; //index starts at 0
}
//need to subtract 1
else
{
    //interpolate if even #
    madest = 1.4286*(sortedr[(t/2)-1]+sortedr[(t/2)+1-1])/2;
}

return madest;
}

```

```

rho(const phi) {
    decl test, result, u,k;
    u = phi;
    k=4.68;
    test = 3*(phi/k).^2-3*(phi/k).^4+(phi/k).^6;
    result = ((u.^2).^(.5)).<=k .*test .: 1;
    return result;
}

```

```

gam(const phi) {
    decl test, result, u, n,k;
    n = rows(phi);
    u = phi;
    k=4.68;
    test = 6/k*(phi/k-2*(phi/k).^3+(phi/k).^5);
    result = ((u.^2).^(.5)).<=k .*test .: 0;
    return result;
}

```

//This function will compute the bisquare psi-function derivative
//with k1 = 4.68

```

gamprime(const phi) {
    decl test, result, u, n,k;
    n = rows(phi);
    u = phi;
    k=4.68;

```

```

        test = 6/k^2*(1-6*(phi/k).^2+5*(phi/k).^4);
        result = ((u.^2).^(.5)).<=4.68 .?test .: 0;
// result = (sumc(result)/n)^2;
        return result;
}

VC0(const y, const x, const bmm,const sc) { decl
K,p,n,phi,phiprime,res,varhip,Ehip,V;

    p=columns(x);
    n=rows(x);
    res=(y-x*bmm);
    V=sc^2*(invert((x)')*(x))*meanc(gam(res/sc).^2)
    /(meanc(gamprime(res/sc))).^2);
    return V;
}

VC1(const y, const x, const bmm,const sc) { decl
K,p,n,phi,phiprime,res,varhip,Ehip,V;

    p=columns(x);
    n=rows(x);
    res=(y-x*bmm)/sc;
    phiprime=gamprime(res);
    phi=gam(res);
    varhip=varc(phiprime);
    Ehip=meanc(phiprime);
    K=1+p/n*(varhip/Ehip^2);
    V=sc^2*(K^2*(1/(n-p))*sumsqrc(phi)/Ehip^2*invert(x'x));
    return V;
}

VC2(const y, const x, const bmm,const sc) { decl
K,p,n,phi,phiprime,res,varhip,Ehip,V,W;

    p=columns(x);
    n=rows(x);
    res=(y-x*bmm)/sc;
    phiprime=gamprime(res);
    phi=gam(res);
    varhip=varc(phiprime);
    Ehip=meanc(phiprime);

```

```

W=x'*diag(hiprime)*x;
K=1+p/n*(varhip/Ehip^2);
V=sc^2*(K*(1/(n-p))*sumsqrc(phi)/Ehip*invert(W));
return V;
}

```

```

VC3(const y, const x, const bmm,const sc) { decl
K,p,n,phi,hiprime,res,varhip,Ehip,V,W;

```

```

p=columns(x);
n=rows(x);
res=(y-bmm*x)/sc;
hiprime=gamprime(res);
phi=gam(res);
varhip=varc(hiprime);
Ehip=meanc(hiprime);
W=x'*diag(hiprime)*x;
K=1+p/n*(varhip/Ehip^2);
V=sc^2*(K^(-1)*(1/(n-p))*sumsqrc(phi)*invert(W)*(x'x)*invert(W));
return V;
}

```

Bibliography

- Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, *in* B.Petro & Csàki, eds, '2nd International Symposium on Information Theory', Akadèmiai Kiadó, Budapest, pp. 267–281.
- Allen, D.M. (1974), 'The relationship between variable selection and data augmentation and a method for prediction', *Technometrics* **16**, 125–127.
- Ambler, Steve, Emanuela Cardia & Christian Zimmermann (2002), 'International transmission of the business cycle in a multi-sector model', *European Economic Review* **46**, 273–300.
- Armington, P. (1969), 'A theory of demand for products distinguished by place of production', *International Monetary Fund Staff Papers* **27**(27).
- Backus, David, Patrick Kehoe & Finn Kydland (1992), 'International real business cycles', *Journal of Political Economy* **100**(41), 745–775.
- Backus, David, Patrick Kehoe & Finn Kydland (1994), 'J-curve', *The American Economic Review* **84**(1), 84–103.
- Başçi, Sidika & Asad Zaman (1998), 'Effects of skewness and kurtosis on model selection criteria', *Economics Letters* **59**, 17–22.
- Balke, N. S. & T. B. Fomby (1994), 'Large shocks, small shocks, and economic fluctuations: Outliers in macroeconomic time series', *Journal of Applied Econometrics* **9**, 181–200.
- Balke, Nathan (1993), 'Detecting level shifts in time series', *Journal of Business and Economic Statistics* **11**(1), 81–92.
- Barnett, V. & T. Lewis (1994), *Outliers in Statistical Data*, John Wiley and Sons, Chichester.
- Baxter, Marianne & Mario Crucini (1995), 'Business cycles and the asset structure of foreign trade', *International Economic Review* **36**(4), 821–854.

- Blattberg, R. & N. Gonedes (1974), 'A comparison of student and stable distributions as statistical models of stock prices', *Journal of Business* **47**, 244–289.
- Boldrin, M., L. Christiano & J. Fisher (2001), 'Habit persistence, asset returns, and the business cycle', *The American Economic Review* **91**(1), 149–166.
- Bollerslev, T. (1987), 'A conditional heteroskedastic time series model for speculative preices and rates of return', *Review of Economics and Statistics* **69**, 542–547.
- Boothe, P. & D. Glassman (1987), 'The statistical distribution of exchange rates: Empirical evidence and economic implications', *Journal of International Economics* **22**, 297–319.
- Box, G. E. P. & G. C. Tiao (1968), 'A bayesian approach to some outlier problems', *Biometrika* **55**, 119–129.
- Box, G. & G. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, John Wiley, New York.
- Bustos, Oscar H. & Victor J. Yohai (1986), 'Robust estimates for arma models', *Journal of the American Statistical Association* **81**, 155–168.
- Chen, Chung & Lon-Mu Liu (1993), 'Joint estimation of model parameters and outlier effects in time series', *Journal of the American Statistical Association* **88**, 284–297.
- Constantinides, G. (1990), 'Habit formation: a resolution of the equity premium puzzle', *Journal of Political Economy* **98**, 519–542.
- DeLong, J. B. & L. H. Summers (1991), 'Equipment, investment, and economic growth', *Quarterly Journal of Economics* **106**, 445–501.
- Denby, L. & R. D. Martin (1986), 'Robust estimation of the first-order autoregressive parameter', *Journal of the American Statistical Association* **74**, 140–146.
- Díaz-Giménez, J. (1999), Linear quadratic approximations: An introduction, in R. Marimon & A. Scott, eds, 'Computational Methods for the Study of Dynamic Economics', Oxford University Press.
- Diebold, F. & R. S. Mariano (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics* **13**, 253–263.
- Fama, E. (1965), 'The behavior of stock market prices', *Journal of Business* **50**, 987–1007.
- Franses, P. & N. Haldrup (1994), 'The effect of additive outliers on test for unit roots and cointegration', *Journal of Business and Economic Statistics* **12**, 471–478.

- Goolsbee, A. (1998), 'Investment tax incentives, prices, and the supply of capital goods', *Quarterly Journal of Economics* **113**, 121–148.
- Guo, Jang-Ting & Federico Sturzenegger (1998), 'Crazy explanations of international business cycles', *International Economic Review* **39**(1), 111–133.
- Guttman, I. (1973), 'Care and handling of univariate or multivariate outliers in detecting spuriousity—a bayesian approach', *Technometrics* **15**, 723–738.
- Hall, J. A., B. W. Brorsen & S. H. Irwin (1989), 'The distribution of futures prices: A test of the stable paretian and mixture of normals hypotheses', *Journal of Financial and Quantitative Analysis* **24**, 105–116.
- Hamilton, J. D. (1996), 'The daily market for federal funds', *Journal of Political Economy* **104**, 26–56.
- Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw & W. A. Stahel (1986), *Robust Statistics: The Approach Based on Influence Functions*, John Wiley, New York.
- Hannan, E.J. & B.G. Quinn (1979), 'The determination of the order of an autoregression', *Journal of the Royal Statistical Society* **B41**, 190–195.
- Huber, P. J. (1964), 'Robust estimation of a location parameter', *Annals of Mathematical Statistics* **35**, 73–101.
- Huber, P. J. (1977), *Robust Statistical Procedures*, SIAM, Philadelphia.
- Huber, P. J. (1979), 'Robust statistics: A review', *Annals of Mathematical Statistics* **43**, 1041–1067.
- Hurvich, C.M. & C.L. Tsai (1989), 'Regression and time series model selection in small samples', *Biometrika* **76**(2), 297–207.
- Judd, K. L. (1999), *Numerical Methods in Economics*, The MIT Press.
- Kanter, M. & W. Steiger (1974), 'Regression and autoregression with infinite variance', *Advances in Applied Probability* **6**, 768–783.
- Kehoe, Patrick & Fabrizio Perri (2000), 'International business cycles with endogenous incomplete markets', *NBER Working Paper* (7870).
- Kleiner, B., R. D. Martin & D. J. Thomson (1979), 'Robust estimation of power spectra', *Journal of the Royal Statistical Society, B* **41**, 313–351.
- Kon, S. (1984), 'Models of stock returns: A comparison', *Journal of Finance* **39**, 153–173.

- Le, Nhu D., Adrian E. Raftery & R. Douglas Martin (1996), ‘Robust bayesian model selection for autoregressive processes with additive outliers’, *Journal of the American Statistical Association* **91**(433), 123–131.
- Lucas, Andre (1995), ‘An outlier robust unit root test with an application to the extended nelson-plosser data’, *Journal of Econometrics* **66**, 153–173.
- Mandelbrot, B. (1963), ‘New methods in statistical economics’, *Journal of Political Economy* **71**, 421–440.
- Mann, H. B. & A. Wald (1943), ‘On the statistical treatment of linear stochastic difference equations’, *Econometrica* **11**, 173–220.
- Martin, R. D. (1979), Robust estimation for time series, *in* R. L.Launer & G. N.Wilkinson, eds, ‘Robustness in Statistics’, Academic Press, New York.
- Martin, R. Douglas & Judith Zeh (1977), ‘Determining the character of time series outliers’, *American Statistical Association. Business and Economics Statistics Section* pp. 818–823.
- McCulloch, J. H. (1998), Linear regression with stable disturbances, *in* R. J.Adler, R. E.Feldman & M. S.Taqqu, eds, ‘A Practical Guide to Heavy Tails’, Birkhauser, Boston.
- Nadenichek, Jon (1999), ‘Consumer durable goods in an international real business cycle framework’, *The Quaterly Review of Economics and Finance* **39**, 233–247.
- Newey, W. K. & K. D. West (1987), ‘A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix’, *Econometrica* **55**, 703–708.
- Phelan, C. & A. Trejos (2000), ‘The aggregate effects of sectoral reallocations’, *Journal of Monetary Economics* **45**, 249–268.
- Resnick, S. I. (1998), Why non-linearities can ruin the heavy-tailed modeler’s day, *in* R. J.Adler, R. E.Feldman & M. S.Taqqu, eds, ‘A Practical Guide to Heavy Tails’, Birkhauser, Boston.
- Rousseeuw, P. J. & Croux (1993), ‘Alternatives to the median absolute deviation’, *Annals of Statistics* .
- Rousseeuw, P. J. & V. J. Yohai (1984), Robust regression by means of s-estimators, *in* W. H.Franke & R. D.Martin, eds, ‘Robust and Nonlinear Time Series’, Springer-Verlag, New York.
- Rousseeuw, Peter (1984), ‘Least median of squares regression’, *Journal of the American Statistical Association* **79**, 871–880.

- Schwarz, G. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**, 461–464.
- Stockman, Alan & Linda Tesar (1995), 'Tastes and technology in a two-country model of the business cycle:explaining international comovements', *The American Economic Review* **85**(1), 168–185.
- Tsay, R. S. (1988), 'Outliers, level shifts, and variance changes in time series', *Journal of Forecasting* **7**, 1–20.
- VanDijk, D., P. H. Franses & A. Lucas (1999a), 'Testing for arch in the presence of additive outliers', *Journal of Applied Econometrics* **14**, 539–562.
- VanDijk, D., P. H. Franses & A. Lucas (1999b), 'Testing for smooth transition nonlinearity in the presence of outliers', *Journal of Business and Economic Statistics* **17**, 217–235.
- Yanyuan, M. & M. Genton (2000), 'Highly robust version of the autocovariance function', *Journal of Time Series Analysis* **21**(6).
- Yohai, V. J. (1986), 'High breakdown-point estimates of regression by means of the minimization of an efficient scale', *Journal of the American Statistical Association* **83**, 406–413.
- Yohai, V. J. & R. H. Zamar (1988a), 'High breakdown-point estimates of regression by means of the minimization of an efficient scale', *Journal of the American Statistical Association* **83**, 406–413.
- Yohai, V.J & R.H. Zamar (1988b), 'High breakdown-point estimates of regression by means of the minimization of an efficient scale', *Journal of the American Statistical Association* **83**, 406–413.
- Zaman, A., P. J. Rousseeuw & M. Orhan (2001), 'Econometric applications of high-breakdown robust regression techniques', *Econometrics Letters* **71**, 1–8.
- Zimmermann, Christian (1994), Technology innovations and the volatility of output: An international perspective. Research Center on Employment and Economic Fluctuations.