**Title**
Estimation and Forecasting in Time Series Models



**Author**
ZHANG, RU

**Publication Date**
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

# Estimation and Forecasting in Time Series Models

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Ru Zhang

December 2013

Dissertation Committee:

Professor Aman Ullah, Co-Chairperson
Professor Tae-Hwy Lee, Co-Chairperson
Professor Goloria González-Rivera
Professor Marcelle Chauvet
Professor Shujie Ma

The Dissertation of Ru Zhang is approved:

Committee Co-Chairperson

Committee Co-Chairperson

University of California, Riverside

## Acknowledgments

First of all, I would like to thank my two committee chairs, professor Aman Ullah and professor Tae-Hwy Lee, for their inspiration, patience, understanding and encouragement during all these years for my studies. They shaped me in so many levels: made me knowledgable in econometrics, offered me chances to be co-author in diversified research topics, helped me develop teaching skills in both graduate and undergraduate statistics and econometrics courses, all of which prepared me well in the job market. I till remember vividly how professor Ullah instructed me step by step to do research independently starting from my second year's study, and hours of talking with professor Lee through late evenings. I would not have been half way through without all their help.

I also would like to thank my other committee members and professors who helped me through the years. Professor Goloria González-Rivera always give me constructive suggestions in both my research as well as my choice of career path, it was so wonderful talking with her each time. Professor Marcelle Chauvet gave me very helpful suggestions in preparing for job market, encouraging me to pursue my career dream as a female econometrician. I am also thankful for professor Jang-Ting Guo, who always give me the chance to walk into his office talking through my difficulties when I feel frustrated; professor Yong Bao and Shujie Ma who gave me suggestions with my research; and professor Richard Arnott, who offered me a position as his research assistant when I was in financial trouble after my injury from the accident.

I will also thank my parents, who always support me through all these years since I left home for my studies from the age of 18. As the only child of the family, I feel bad not having much chance to be company with them, taking care of them when they need me, but they are always there for me when I need them. They are my motivation to be better and stronger.

Last but not least, I would like to thank all my good friends who were once in Riverside, Zhou Xi who made my life colorful during all these years, Mingming Jiang who always help me out from the most difficulty, Huiling Zhang for being such an amazing roommate, Xiangbo Liu for being a trustful friend, Yingying Sun, Dong Zhou, Wei Lin and Yanpin Su for bringing me lots of cheers. I feel so lucky to have you all as my friends who I can depend on, sharing happiness and sorrows.

I couldn't be thankful enough for all my professors and friends, who made my 4 years and 3 months' staying here in UC Riverside the happiest time in my life so ever. You all made my life here so easy to begin with and so hard to leave by the end. I will always cherish such colorful and joyful memory here, in Riverside, my first home in US. Although I am looking forward to start my new life experience working in a non-academic area, I really hate to say good-bye to the campus I once walked through, the office I once worked in, the street I once lived on, the field I once loved so much, the people who shined my future.

To my parents,

who have always been my motivation to be stronger.

ABSTRACT OF THE DISSERTATION

Estimation and Forecasting in Time Series Models

by

Ru Zhang

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, December 2013
Professor Aman Ullah, Co-Chairperson
Professor Tae-Hwy Lee, Co-Chairperson

This dissertation covers several topics in estimation and forecasting in time series models. Chapter one is about estimation and feasible conditional forecasts properties from the predictive regressions, which extends previous results of OLS estimation bias in the predictive regression model by considering predictive regressions with possible zero intercepts, and also allowing the regressor to follow either a stationary AR(1) process or unit root process. The main thrust of this chapter is to develop an analytical bias reduced estimator and study the mean squared error (MSE) efficiency of the estimator. Then we investigate whether this estimation bias can lead to biased feasible forecasts conditional on the available sample observations, in addition to the expression of the mean squared forecast error (MSFE). The results from this chapter shed lights on the bias reduction estimator of the predictive regressions and its MSE properties in finite samples, as well as the optimal forecasts efficiency. We apply our analytical results to both simulated and financial data with financial return prediction using variables such as dividend yield and short rate. Results show that our bias

reduction works well in estimation even when the data are skewed and having fat tails, and moreover, the bias reduced estimator improves out-of-sample forecasts. All of the results highlight the importance of the bias reduction in estimation and forecasting.

Chapter two explores finite sample bias of the estimators in the first order autoregressive moving average model under a general error distribution. Since the quasi maximum likelihood estimator (QMLE) of parameters in the first order autoregressive moving average model (ARMA(1, 1)) can be biased in finite samples, this chapter discusses bias properties of the QMLE of the ARMA(1,1) model up to order $O(T^{-1})$ by applying the stochastic expansion and the formula and sheds light on the bias correction for the parameter estimation in applied works. The analytical bias expression of the QMLE suggests that the bias is robust to nonnormality and the simulation results show that the bias corrected QML estimators is better even when sample size increased to a moderate size.

Chapter three (joint with Yong Bao) examines estimation bias and feasible conditional forecasts from the first-order moving average model. We develop the second-order analytical bias of the QMLE and investigate whether this estimation bias can lead to biased feasible optimal forecasts conditional on the available sample observations. We find that the feasible multiple-step-ahead forecasts are unbiased under any nonnormal distribution and the one-step-ahead forecast is unbiased under symmetric distributions.

Chapter four (joint with Tae-Hwy Lee and Zhou Xi) discusses using extreme learning machines for out-of-sample prediction. In this chapter, we apply the artificial neural network (ANN) model to out-of-sample prediction of financial return using a set of covariates. The main challenge in ANN model estimation is the multicolinearity between the large numbers of randomly generated hidden layers. We explore several methods to deal with the large dimension regressors, such as general inverse, ridge, pretest and principal components, which are also named extreme learning machines (ELM). We find that although the ELM methods sometimes fit perfectly for in-sample data, it has very poor out-of-sample forecast ability. We then introduce some modifications to the ELM method, which is a two step algorithm, where the first step uses ELM methods with some modifications to get a set of forecasts, and the second step combines the forecasts using principal components weighting scheme. Empirical results show that our method gives best forecast for annually aggregated equity premium among all the alternatives.

Chapter five (joint with Tae-Hwy Lee) considers mallows model averaging in the presence of multicollinearity. A challenge with large dimensional data in regression is the collinearity among covariates. A common solution to this problem is to apply principal component analysis (PCA). Yet one needs to select the number of principal components. Many studies have focused on finding the optimal number of principal components assuming the linear factor model is correctly specified. In this chapter, we do not assume that the data generating process (DGP) is a linear factor model and thus there is no true number of factors. Under this circumstance, we can combine several principal component regressions with different numbers of principal components through the Mallows criteria. Under certain conditions, the model averaging estimator is minimax such that the estimation risk is smaller. We show that the Mallow model averaging estimator can improve the estimation efficiency.

# Contents


**List of Figures** **xiii**

**List of Tables** **xiv**

**1 Estimation and Feasible Conditional Forecasts Properties from the Predictive Regressions** **1**

1.1 Introduction . . . . . 2

1.2 Estimation and Forecasts of No Intercept Model . . . . . 5

1.2.1 Estimator Bias and MSE . . . . . 5

1.2.2 Forecasts Bias and MSFE . . . . . 11

1.3 Estimation and Forecasts of Model With Intercept . . . . . 17

1.3.1 Estimation Bias and MSE . . . . . 17

1.3.2 Forecasts Bias and MSFE . . . . . 22

1.4 Monte Carlo Simulation . . . . . 28

1.5 Predictive Regressions of Financial Return . . . . . 32

1.6 Conclusion . . . . . 36

**Bibliography** **48**

**2 Finite Sample Bias of the Estimators in the First Order Autoregressive Moving Average Model under a General Error Distribution** **50**

2.1 Introduction . . . . . 51

2.2 Main Results . . . . . 51

2.3 Simulation Results . . . . . 58

2.4 Concluding Remarks . . . . . 59

**Bibliography** **71**

**3 Estimation Bias and Feasible Conditional Forecasts from the First-Order Moving Average Model** **72**

3.1 Introduction . . . . . 73

3.2 Main Results . . . . . 74

3.2.1 Estimation Bias . . . . . 76
3.2.2 Forecast Bias . . . . . 78
3.2.3 A Simulation Experiment . . . . . 82
3.3 Concluding Remarks . . . . . 85

**Bibliography** **87**

**4 Using Extreme Learning Machines for Out-of-Sample Prediction** **89**
4.1 Introduction . . . . . 90
4.2 Extreme Learning Machines . . . . . 92
4.2.1 Basic ELM . . . . . 93
4.2.2 Ridge based ELM . . . . . 94
4.2.3 PC based ELM . . . . . 94
4.2.4 Pretest based ELM . . . . . 95
4.3 Forecast Combination using ELM . . . . . 95
4.3.1 Two Step Algorithm . . . . . 96
4.3.2 Weighting Scheme . . . . . 97
4.4 Alternative Models . . . . . 99
4.5 Results . . . . . 102
4.5.1 Data and Variables . . . . . 102
4.5.2 Main Results . . . . . 103
4.5.3 Comparison Criterion . . . . . 106
4.6 Conclusion . . . . . 109

**Bibliography** **117**

**5 Mallows Model Averaging in the Presence of Multicollinearity** **119**
5.1 Introduction . . . . . 120
5.2 OLS estimator . . . . . 121
5.3 Principal Component Estimator . . . . . 123
5.4 Mallows Model Average Estimator with a Fixed $k$ . . . . . 124
5.5 Mallows Model Average Estimator with Many $k$ . . . . . 129
5.6 Conclusion . . . . . 133

**Bibliography** **134**

**A Appendix for Chapter 1** **135**
A.1 Calculation of $E(X'X)^{-1}$ for model without intercept . . . . . 135
A.2 Nagar Expansion of $\hat{\rho}$ for model without intercept . . . . . 136
A.3 Calculation of $E(X'X)^{-1}$ for model with intercept . . . . . 136
A.4 h-step-ahead forecast bias of AR(1) model with intercept . . . . . 137

**B Appendix for Chapter 2** **140**
B.1 Expansion of $\hat{\beta} - \beta$ for the pure model . . . . . 140
B.2 Expansion of $\hat{\beta} - \beta$ for the intercept model . . . . . 145

**C Appendix for Chapter 3** **148**
C.1 Expansion of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ . . . . . . . . . . 148

# List of Figures

4.1 Stochastic Dominance and Distribution of Squared Errors for $Q_t(12)$ . . . . 114
4.2 Stochastic Dominance and Distribution of Squared Errors for $Q_t(3)$ . . . . . 115
4.3 Stochastic Dominance and Distribution of Squared Errors for $Q_t(1)$ . . . . . 116

# List of Tables


1.1 No Intercept Model, T=30, $\rho = 0.2$ . . . . . . . . 38
1.2 No Intercept Model, T=30, $\rho = 0.5$ . . . . . . . . 39
1.3 No Intercept Model, T=30, $\rho = 0.8$ . . . . . . . . 40
1.4 No Intercept Model, T=100, $\rho = 0.5$ . . . . . . . . 41
1.5 Model with Intercept, T=30, $\rho = 0.2$ . . . . . . . . 42
1.6 Model with Intercept, T=30, $\rho = 0.5$ . . . . . . . . 43
1.7 Model with Intercept, T=30, $\rho = 0.8$ . . . . . . . . 44
1.8 Model with Intercept, T=100, $\rho = 0.5$ . . . . . . . . 45
1.9 Unit Root $x_t$ Model with Intercept, T=30 . . . . . . . . 46
1.10 Predictive Regression of Financial Returns . . . . . . . . 47

2.1 Bias Correction for pure ARMA(1,1) model, $\sigma^2 = 1, T = 30$ . . . . . . . . 61
2.2 Bias Correction for pure ARMA(1,1) model, $\sigma^2 = 2, T = 30$ . . . . . . . . 62
2.3 Bias Correction for pure ARMA(1,1) model, $\sigma^2 = 1, T = 100$ . . . . . . . . 63
2.4 Bias Correction for pure ARMA(1,1) model, $\sigma^2 = 2, T = 100$ . . . . . . . . 64
2.5 Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.2$, $\sigma^2 = 1, T = 30$ 65
2.6 Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.5$, $\sigma^2 = 1, T = 30$ 66
2.7 Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.8$, $\sigma^2 = 1, T = 30$ 67
2.8 Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.2$, $\sigma^2 = 1, T = 100$ 68
2.9 Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.5$, $\sigma^2 = 1, T = 100$ 69
2.10 Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.8$, $\sigma^2 = 1, T = 100$ 70

3.1 Finite-Sample Properties of QMLE in MA(1) ($T = 30, 50$) . . . . . . . . 84
3.2 Finite-Sample Properties of QMLE in MA(1) ($T = 100, 200$) . . . . . . . . 85

4.1 In-sample Mean Squared Error . . . . . . . . 111
4.2 Out-of-sample Mean Squared Forecast Error . . . . . . . . 112
4.3 Forecast Comparison . . . . . . . . 113

# Chapter 1

# Estimation and Feasible Conditional Forecasts Properties from the Predictive Regressions

Predictive regression is one of the basic models in financial econometrics. However, the OLS estimator of this model will be biased in finite samples, since the regression disturbance is correlated with regressors innovations. There have been concerns in the financial econometrics literature about the issue of bias in such models. This paper extends previous results by considering predictive regressions with possible zero intercepts, and also allowing the regressor to follow either a stationary AR(1) process or unit root process. The main thrust of this paper is to develop an analytical bias reduced estimator and study the mean squared error (MSE) efficiency of the estimator. Then we investigate whether this estimation bias can lead to biased feasible forecasts conditional on the available sample observations, in addition to the expression of the mean squared forecast error (MSFE). The results from this paper shed lights on the bias reduction estimator of the predictive regressions and its MSE properties in finite samples, as well as the optimal forecasts efficiency. We apply our analytical results to both simulated and financial data with financial return prediction using variables such as dividend yield and short rate. Results show that our bias reduction works well in estimation even when the data are skewed and having fat tails, and moreover, the bias reduced estimator improves out-of-sample forecasts. All of the results highlight the importance of the bias reduction in estimation and forecasting.

## 1.1 Introduction

Predictive regression is one of the basic models in financial econometrics. However, the OLS estimator of this model will be biased in finite samples, since the regression disturbance is correlated with regressor's innovations. There has been concerns in the financial econometrics literature about the issue of estimation bias in such models, for instance, S-tambaugh (1999) gives the OLS estimator bias for the single regressor case, Zhu (2013) gives a method to reduce estimation bias through the jackknife estimator, Amihud and Hurvich (2004) used an augmented regression through adding a proxy for the errors in the AR(1) process to reduce the estimation bias. However, most of current studies assumes the regressor follows a stationary first-order autoregressive (AR(1)) process and considering the bias property of the coefficient of covariate. Such assumption put limitations in the application of the results that in some cases, the regressor may not be stationary, but rather has unit root, in that case, previous studies of bias reduction based on AR(1) process of the regressor will not give appropriate bias correction. And moreover, we find that the bias properties are quite different for predictive regressions model with zero and nonzero intercept, which has not been studied before. In addition, most studies before consider the bias properties for the estimator only, yet the mean squared error efficiency as well as forecast properties are still unclear.

This paper allows the regressor to follow either a stationary AR(1) process or unit root process, and with possible zero intercept in the regressions. The main thrust of this paper is to develop an analytical bias reduced estimator and study the mean squared error (MSE) efficiency of the estimator and feasible optimal forecasts. We find that the estimation bias and MSE of the predictive regression model relates closely to the dynamics of the regressor and the magnitude of bias is different depending on whether the regressor is stationary or unit root, and whether the process has zero or nonzero intercept. As a result, we will study the estimation properties for each model. We derive the analytical bias and mean squared error of the OLS estimator for both nonzero and zero intercept models, where the regressor is allowed to be either a stationary AR(1) or unit root process.

Though the ultimate goal of the predictive regression model is to make prediction or forecasts of the dependent variable, most studies before focus on the marginal effect estimation of the regressor, rather that the forecasts properties of the model. In this paper, we will fill this gap by studying whether estimation bias would lead to feasible optimal forecast bias of the model, as well as the expression of the mean squared forecast error, when the regressor is stationary. Interestingly, we find that although the OLS estimator is biased, the feasible multiple-step-ahead forecasts based on the OLS estimators are unbiased under any non-normal error distribution.

The results from this paper shed lights on the bias reduction estimator of the predictive regressions and the MSE properties of the estimator in small samples, for both stationary AR(1) and unit root regressors, as well as the optimal forecasts efficiency for stationary AR(1) regressor. From simulated data, results show that the bias corrected estimator works better than the OLS estimator and the bias correction will improve estimation even if the error terms are asymmetric and have fat tails. We then apply the results to explain stock return and equity premium using factors such as dividend yield ratio, dividend price ratio, corporate issuing activity as well as short term rates. The results show that if we ignore the estimation bias, the effects of the factors on stock return as well as equity premium will be underestimated. Moreover, feasible forecasts using the bias corrected estimators performs better than the OLS estimators. All of the results highlight the importance of the bias reduction in estimation and forecasting.

The following sections are arranged as follows: section two discusses the estimation and forecasts properties of predictive regression model without intercept, where $x_t$ could be stationary or unit root in estimation, and stationary in forecasting. Section three discusses the estimation and forecasts properties of predictive regression model with intercept, where $x_t$ could be stationary or unit root in estimation, and stationary in forecasting. Section four applies the analytical formula to simulated data and section five applies the formula to financial data. Section six concludes.

## 1.2 Estimation and Forecasts of No Intercept Model

### 1.2.1 Estimator Bias and MSE

Since the estimation as well as forecasts properties are different depending on whether the regressor is an AR(1) or unit root process, as well as whether the model incorporates intercept or not, this paper examines the two models separately: no intercept in both equations and with intercept in both equations. And for each model, we consider the estimation properties allowing the regressor to be either stationary or unit root, and for forecasts properties, we assume the regressor to be stationary. This section consider simple case first were both equations contain no intercept. Next section considers the case where both equations contain a nonzero intercept.

The predictive regression without intercept can be written as a set of two equations:

$$y_t = \beta x_{t-1} + u_t \tag{1.1}$$

$$x_t = \rho x_{t-1} + v_t$$

where $|\rho| < 1$ when $x_t$ is stationary, and $\rho = 1$ when $x_t$ is unit root, and $(u_t, v_t)'$ is joint normally distributed, independent across t, with mean zero and variance covariance matrix $\Sigma$, where

$$\Sigma = \mathrm{cov}((u_t, v_t)'(u_t, v_t)) = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}.$$

Without loss of generality, we assume $x_0 = 0$. Our primary concern in this paper for the estimator properties part is to get the estimation bias as well as its mean squared error (MSE) for $\hat{\beta}$, the coefficient in the predictive regression model. Since with the analytical bias of the coefficient, we are able to make efficient bias reduction in estimation even when the sample size is limited, and such reduced bias in estimation not only reveals the true marginal effect of the regressor on the dependent variable we are interested in, it will also improves the optimal feasible forecasts of the model.

The OLS estimator for $\beta$ is given by $\hat{\beta} = (X'X)^{-1}X'Y$, where $X = (x_0, ..., x_{T-1})'$, $Y = (y_1, ..., y_T)'$, $T$ is the sample size. The OLS estimator for $\rho$ is $\hat{\rho} = (X'X)^{-1}X'X_T$, where $X_T = (x_1, ..., x_T)'$. And the estimation error for $\hat{\beta}$ and $\hat{\rho}$ can then be written as: $\hat{\beta} - \beta = (X'X)^{-1}X'u$, $\hat{\rho} - \rho = (X'X)^{-1}X'v$, where $u = (u_1, ..., u_T)'$, $v = (v_1, ..., v_T)'$. Decompose $u$ as $u = \frac{\sigma_{uv}}{\sigma_v^2}v + \varepsilon$, we can verify that $E(\varepsilon|v) = 0$ followed by the i.i.d. assumption, and this implies $E(\varepsilon|X) = 0$. Then substitute the decomposition of $u$ into the estimation error for $\hat{\beta}$, we have:

$$\begin{aligned} \hat{\beta} - \beta &= (X'X)^{-1}X'u = (X'X)^{-1}X'(\frac{\sigma_{uv}}{\sigma^2 v}v + \varepsilon) \\ &= \frac{\sigma_{uv}}{\sigma_v^2}(X'X)^{-1}X'v + (X'X)^{-1}X'\varepsilon \\ &= \frac{\sigma_{uv}}{\sigma_v^2}(\hat{\rho} - \rho) + (X'X)^{-1}X'\varepsilon \end{aligned} \tag{1.2}$$

Take expectation on both sides, the bias of $\hat{\beta}$ is then given by:

$$B(\hat{\beta}) = E(\hat{\beta} - \beta) = \frac{\sigma_{uv}}{\sigma_v^2}E(\hat{\rho} - \rho) = \frac{\sigma_{uv}}{\sigma_v^2}B(\hat{\rho}) \tag{1.3}$$

where $B(\hat{\rho})$ is the estimation bias for the stationary AR(1) or unit root process in the second equation of the predictive regressions without intercept, depending on whether $|\rho| < 1$ or $\rho = 1$. So the estimation bias from the predictive regressions model is proportional to the estimation bias from the regression of the regressor.

Many studies give the results of the estimation bias for the stationary AR(1) model, such as Kendall (1954), White (1961), Sawa (1978), Tanaka (1984), Nicholls and Pope (1988), Pope (1990), Kiviet and Phillips (1993), among others. Most of these studies assumes error term is normally distributed or a martingale difference process. Bao and Ullah (2007) generalize the result and gives the bias of the OLS estimator $\hat{\rho}$ up to order $O(T^{-1})$ under a general distributed error term, where no distributional assumptions are made other than the same set of moment conditions of $v_t$ made in this paper. Bao (2007a) extends the bias formula up to $O(T^{-2})$ for both AR(1) process with and without intercept under the general distributed error term. Apply this result to equation 1.3, the OLS estimation bias for $\hat{\beta}$ in the predictive regression model without intercept up to order $O(T^{-2})$ is then given by:

$$B(\hat{\beta}) = -\frac{\sigma_{uv}}{\sigma_v^2}\frac{2\rho}{T} + \frac{\sigma_{uv}}{\sigma_v^2}\frac{4\rho}{T^2} + o(T^{-2}) \tag{1.4}$$

given $|\rho| < 1$. Interestingly, if $x_t$ is a stationary AR(1) process, the magnitude of the bias of the OLS estimator $\hat{\beta}$ up to $O(T^{-1})$ depends positively on the value of $\rho$, covariance of $u$ and $v$, and depends negatively on the variance of $v$. And whether the bias is upward or downward depends negatively on the sign of covariance between $u$ and $v$, for a positive $\rho$. While for the estimation bias up to $O(T^{-2})$, the direction between the the bias and $\rho$ is reversed compared to the bias of order $O(T^{-1})$, that the bias is upward or downward depends positively on the sign of covariance between $u$ and $v$, for a positive $\rho$. So that if the error terms $u_t$ and $v_t$ are positively correlated, a more persistent process $x_t$ will lead to smaller bias of order $O(T^{-1})$, yet larger bias of order $O(T^{-2})$, for a stationary process $x_t$.

When $x_t$ is a unit root process, that is, when $\rho = 1$, Abadir (1993, 1995) gives the exact bias of $\hat{\rho}$ under normally distributed error term $v_t$ through rewriting the bias as:

$E(\hat{\rho}-\rho) = \frac{\sqrt{2}}{T}E(\frac{T}{\sqrt{2}}(\hat{\rho}-\rho)) = \frac{\sqrt{2}}{T}E(\hat{\rho}^*)$. Then the paper gives the bias approximation using results from Evans and Savin (1981), that $E(\hat{\rho}-\rho) \simeq -1.78143(\rho/T)e^{-2.6138/T}$. Substitute to the bias of $\hat{\beta}$ in 1.3, we have, when $x_t$ is a unit root process and $(u_t, v_t)$ are joint normally distributed, the estimation bias of $\hat{\beta}$ of the predictive regression model is:

$$B(\hat{\beta}) = -1.78143\frac{\sigma_{uv}}{\sigma_v^2}\frac{\rho}{T}e^{-2.6138/T} + o(T^{-1}) \tag{1.5}$$

Besides the estimation bias, another interesting property of the estimator in prediction regressions is the estimator mean squared error. We will examine both conditional MSE and unconditional MSE in the following, where the condition is upon the information set $\mathcal{I}_T$, given observed $X$ and up to time $T$. Notice that the conditional bias of $\hat{\beta}$ given observed $X$ is always equal to zero, since $B(\hat{\beta}|X) = E((\hat{\beta}-\beta)|X) = E((X'X)^{-1}X'u|X) = 0$. Then the conditional mean squared error for $\hat{\beta}$ is equal to its conditional variance, given observed $X$, substitute the decomposition of $\beta$ in 1.2, we have:

$$\begin{aligned} M(\hat{\beta}|X) &= V(\hat{\beta}|X) = V((\hat{\beta}-\beta)|X) = V\left[(\frac{\sigma_{uv}}{\sigma_v^2}(\hat{\rho}-\rho) + (X'X)^{-1}X'\varepsilon)|X\right] \\ &= (X'X)^{-1}(\frac{\sigma_{uv}^2}{\sigma_v^2} + \sigma_\varepsilon^2) = (X'X)^{-1}\sigma_u^2 = \frac{\sigma_u^2}{\sum_{t=0}^{T-1} x_t^2} \end{aligned} \tag{1.6}$$

since by the decomposition of $u$ where $u = \frac{\sigma_{uv}}{\sigma_v^2}v + \varepsilon$, we can verify that $\sigma_u^2 = \frac{\sigma_{uv}^2}{\sigma_v^2} + \sigma_\varepsilon^2$.

Then consider the unconditional MSE of $\hat{\beta}$. From equation 1.2, the unconditional variance of $\hat{\beta}$ can be written as:

$$\begin{aligned} V(\hat{\beta}) &= V(\hat{\beta}-\beta) = V\left[\frac{\sigma_{uv}}{\sigma_v^2}(\hat{\rho}-\rho) + (X'X)^{-1}X'\varepsilon\right] \\ &= \frac{\sigma_{uv}^2}{\sigma_v^4}V(\hat{\rho}-\rho) + \sigma_\varepsilon^2 E(X'X)^{-1} \end{aligned}$$

Then combine $V(\hat{\beta})$ and $B(\hat{\beta})$, the MSE of $\hat{\beta}$ is then given by:

$$\begin{aligned} M(\hat{\beta}) &= B(\hat{\beta})^2 + V(\hat{\beta}) = \frac{\sigma_{uv}^2}{\sigma_v^4}(E(\hat{\rho}-\rho))^2 + \frac{\sigma_{uv}^2}{\sigma_v^4}V(\hat{\rho}-\rho) + \sigma_\varepsilon^2 E(X'X)^{-1} \\ &= \frac{\sigma_{uv}^2}{\sigma_v^4}M(\hat{\rho}) + \sigma_\varepsilon^2 E(X'X)^{-1} \end{aligned}$$

where $M(\hat{\rho})$ is the MSE from the stationary AR(1) or unit root process of $x_t$ without intercept.

To get $E(X'X)^{-1}$, we can use the Nagar-type expansion (Nagar, 1959). First consider the case when $x_t$ is stationary AR(1). Denote $D = X'X = \sum_{t=0}^{T-1} x_t^2$, then $E(X'X)^{-1} = E(\frac{1}{D})$, and note that $E(D) = O(T)$, $\frac{D-ED}{ED} = O(T^{-1/2})$, then we have the following Nagar-type expansion of $\frac{1}{D}$:

$$\begin{aligned} \frac{1}{D} &= \frac{1}{D+ED-ED} = \frac{1}{ED}(1+\frac{D-ED}{ED})^{-1} \\ &= \frac{1}{ED}(1-\frac{D-ED}{ED}+\frac{(D-ED)^2}{(ED)^2}) + o_P(T^{-2}) \\ &= a_{-1} + a_{-3/2} + a_{-2} + o_P(T^{-2}) \end{aligned}$$

where $a_{-1} = \frac{1}{ED}$, $a_{-3/2} = -\frac{D-ED}{(ED)^2}$, $a_{-2} = \frac{(D-ED)^2}{(ED)^3}$. Notice that $E(a_{-3/2}) = 0$, and $a_{-i} = O(T^{-})$, for $i = 1, 2, 3$, then we have:

$$E(X'X)^{-1} = E(a_{-1}) + E(a_{-2}) + o(T^{-2})$$

Then followed by some calculation, we can verify the following when $x_t$ is stationary AR(1):

$$\begin{aligned} E(a_{-1}) &= \frac{1}{ED} = \frac{1-\rho^2}{T\sigma_v^2} \\ E(a_{-2}) &= \frac{E(D-ED)^2}{(ED)^3} = \frac{2(1-\rho^2)}{T^2\sigma_v^2} \end{aligned}$$

The proof is in the appendix. Then up to order $O(T^{-1})$, we have $E(X'X)^{-1} = \frac{1-\rho^2}{T\sigma_v^2} + \frac{2(1-\rho^2)}{T^2\sigma_v^2}$.

The MSE for the AR(1) process, $M(\hat{\rho})$, has been discussed in some studies. White (1961), Shenton and Johnson (1965), Phillips (1977) has results for normally distributed error terms. Bao and Ullah (2007) gives method to calculate it under general distributed error terms. Bao (2007a) gives the analytical formula of MSE for $\hat{\rho}$ of AR(1) process both with and without intercept, up to order $O(T^{-2})$, under general distributed error terms. Applying this result and combine with the expression of $E(X'X)^{-1}$, the MSE for $\hat{\beta}$ form the predictive regressions when $x_t$ is stationary is:

$$M(\hat{\beta}) = \frac{(1-\rho^2)}{T}\frac{\sigma_u^2}{\sigma_v^2} + \frac{1}{T^2}\left[\frac{\sigma_{uv}^2}{\sigma_v^4}(14\rho^2-1) + 2\frac{\sigma_\varepsilon^2}{\sigma_v^2}(1-\rho^2)\right] + o(T^{-2}) \tag{1.7}$$

provided $|\rho| < 1$. Notice that again, we apply the relation that $\sigma_u^2 = \frac{\sigma_{uv}^2}{\sigma_v^2} + \sigma_\varepsilon^2$.

It's interesting to notice that up to order $O(T^{-1})$, $M(\hat{\beta}) = M(\hat{\rho})\lambda$, where $\lambda = \frac{\sigma_u^2}{\sigma_v^2}$. That is, the MSE of $\hat{\beta}$ in the predictive regressions is proportional to the MSE of $\hat{\rho}$ in the AR(1) process, where their ratio is equal to the ratio of the variance of error terms in the two equations. This property is similar to the estimation bias, which is also proportional to the estimation bias of the AR(1) model.

When $x_t$ is a unit root process, Abadir (1993, 1995) gives the exact variance of $\hat{\rho}$ through rewriting it as: $E(\hat{\rho}-\rho)^2 = \frac{2}{T^2}E(\hat{\rho}^* - E(\hat{\rho}^*))^2$. Then apply similar method as for the bias of $\hat{\rho}$ for the unit root model, the MSE of $\hat{\rho}$ can be written as: $M(\hat{\rho}) \simeq \frac{1}{T^2}(3.1735e^{-5.2276/T} + 10.1124e^{-5.4462/T+14.519/T^2})$, assuming $v_t$ is normally distributed.

To get $E(X'X)^{-1}$ for the unit root process, note that if $x_t$ has unit root, $E(D) = O(T^2)$, $\frac{D-ED}{ED} = O(T^{-1/2})$, so the Nagar-type expansion for $D = X'X$ in this case is written as:

$$\begin{aligned}\frac{1}{D} &= \frac{1}{ED} - \frac{D-ED}{(ED)^2} + \frac{(D-ED)^2}{(ED)^3} + o_P(T^{-3}) \\ &= a_{-2} + a_{-5/2} + a_{-3} + o_P(T^{-3})\end{aligned}$$

Following similar method as stationary case (see appendix), we can check that up to order $O(T^{-2})$, for unit root $x_t$, we have $E(X'X)^{-1} = (ED)^{-1} = \frac{2}{T^2\sigma_v^2} + o(T^{-2})$. So when $x_t$ has unit root, the MSE of $\hat{\beta}$ from the predictive regressions has the form:

$$M(\hat{\beta}) = \frac{\sigma_{uv}^2}{\sigma_v^4}(\frac{1}{T^2}(3.1735e^{-5.2276/T} + 10.1124e^{-5.4462/T+14.519/T^2})) + \frac{2\sigma_\varepsilon^2}{T^2\sigma_v^2} + o(T^{-2}) \quad (1.8)$$

### 1.2.2 Forecasts Bias and MSFE

For the forecasts properties of the predictive regressions both with and without intercept, we will focus on two of them in this study: the $h$-step-ahead feasible conditional forecasts bias given information set up to time $T$, and the $h$-step-ahead feasible conditional mean squared forecast error (MSFE), for all $h \geq 1$. And when discussing the forecasts properties, we will focus on the case when $x_t$ is a stationary AR(1) process, since the analytical forecasts properties for the unit root process is limited in literature, especially when the error terms is not necessarily normal.

We first consider the simple case of one-step-ahead forecast. Since from model 1.1 we have: $y_{T+1} = \beta x_T + u_{T+1}$, then the one-step-ahead conditional forecast of $y_{t+1}$ given observations of $x_t$ up to time $T$ is given by: $\hat{y}_{T+1|T} = E(y_{T+1}|\mathcal{I}_T) = \hat{\beta}x_T$, where $\mathcal{I}_T$ denotes the information set at time $T$. Then the one-step-ahead forecast error is the difference between the true realization of future $y_{t+1}$ and its conditional forecasts made at

$T$: $e_{T+1|T} = y_{T+1} - \hat{y}_{T+1|T} = (\beta - \hat{\beta})x_T + u_{T+1}$, and the forecast bias is the expectation of the forecast error, denoted by $E(e_{T+1|T})$. Substitute the decomposition of $\hat{\beta}$ in 1.2, the one-step-ahead forecast bias can be rewritten as:

$$\begin{aligned} E(e_{T+1|T}) &= E((\beta - \hat{\beta})x_T) = E\Big[\big(\frac{\sigma_{uv}}{\sigma_v^2}(\rho - \hat{\rho}) - (X'X)^{-1}X'\varepsilon\big)x_T\Big] \\ &= \frac{\sigma_{uv}}{\sigma_v^2}E\big[(\rho - \hat{\rho})x_T\big] = \frac{\sigma_{uv}}{\sigma_v^2}E(e^{AR}_{T+1|T}) \end{aligned}$$

where $E(e^{AR}_{T+1|T})$ is the one-step-ahead forecast bias for the stationary AR(1) or unit root process without intercept, depending the value of $\rho$. When $|\rho| < 1$, that is, when $x_t$ is stationary, Bao (2007b) studied the $h$-step-ahead forecast bias as well as its mean squared forecast error for AR(1) model without intercept for any $h \geq 1$, where the $h$-step-ahead forecast bias is equal to zero up to order $O(T^{-1})$ for any $h \geq 1$. Substitute this result to the above forecast bias equation, we can conclude that the one-step-ahead forecast bias for the predictive regressions without intercept is unbiased up to order $O(T^{-1})$, that is,

$$E(e_{T+1|T}) = 0 + o_P(\frac{1}{T}) \tag{1.9}$$

In general, consider the $h$-step-ahead forecast for $h > 1$ when $|\rho| < 1$. Since $y_{T+h} = \beta x_{T+h-1} + u_{T+h}$, the conditional forecast for $y_{T+h}$ at time $T$ is equal to its conditional mean: $\hat{y}_{T+h|T} = \hat{\beta}\hat{x}_{T+h-1|T}$, where $\hat{x}_{T+h-1|T} = E(x_{T+h-1}|\mathcal{I}_t)$ is the $h-1$ step ahead forecast for $x_{t+h-1}$ from the AR(1) process at time $T$. Then the $h$-step-ahead forecast error is given by $e_{T+h|T} = y_{T+h} - \hat{y}_{T+h|T} = \beta x_{T+h-1} - \hat{\beta}\hat{x}_{T+h-1|T} + u_{T+h}$. Since $x_{T+h-1}$ can be written recursively as $x_{T+h-1} = \rho^{h-1}x_T + \sum_{t=1}^{h-1} v_{T+t}\rho^{h-t-1}$, so $\hat{x}_{T+h-1|T} = E(x_{T+h-1}|\mathcal{I}_t) = \hat{\rho}^{h-1}x_T$. Notice that the forecast error can be rewritten as: $e_{T+h|T} = (\beta - \hat{\beta})x_{T+h-1} + \hat{\beta}(x_{T+h-1} - \hat{x}_{T+h-1}) + u_{T+h}$.

Then the forecast bias is given by taking the expectation of the forecast error:

$$E(e_{T+h|T}) = E(y_{T+h} - \hat{y}_{T+h|T}) = E[(\beta - \hat{\beta})x_{T+h-1}] + E[\hat{\beta}(x_{T+h-1} - \hat{x}_{T+h-1})]$$

Substitute the decomposition of $\hat{\beta}$ in 1.2 to the above two terms, we have:

$$\begin{aligned}
E[(\beta - \hat{\beta})x_{T+h-1}] &= E\left[\Big(\frac{\sigma_{uv}}{\sigma_v^2}(\rho - \hat{\rho}) - (X'X)^{-1}X'\varepsilon\Big)x_{T+h-1}\right] \\
&= E\left[\frac{\sigma_{uv}}{\sigma_v^2}(\rho - \hat{\rho})x_{T+h-1}\right] \\
E[\hat{\beta}(x_{T+h-1} - \hat{x}_{T+h-1})] &= E\left[\Big(\beta + \frac{\sigma_{uv}}{\sigma_v^2}(\hat{\rho} - \rho) + (X'X)^{-1}X'\varepsilon\Big)(x_{T+h-1} - \hat{x}_{T+h-1})\right] \\
&= \beta E(x_{T+h-1} - \hat{x}_{T+h-1}) + E\left[\frac{\sigma_{uv}}{\sigma_v^2}(\hat{\rho} - \rho)x_{T+h-1}\right] \\
&\quad - E\left[\frac{\sigma_{uv}}{\sigma_v^2}(\hat{\rho} - \rho)\hat{x}_{T+h-1}\right]
\end{aligned}$$

Notice that $E(x_{T+h-1} - \hat{x}_{T+h-1})$ is just the $h-1$-step-ahead forecast bias for the AR(1) process with no intercept, which is equal to zero by Bao (2007b) up to order $O(T^{-1})$. Moreover, since $\hat{x}_{T+h-1} = \hat{\rho}^{h-1}x_T$, so $E[(\hat{\rho} - \rho)\hat{x}_{T+h-1}] = E[(\hat{\rho} - \rho)\hat{\rho}^{h-1}x_T] = E(\hat{\rho}^h x_T) - \rho E(\hat{\rho}^{h-1}x_T)$. Collect all the terms, the h-step-ahead forecast bias can be simplified to:

$$E(e_{T+h|T}) = -\frac{\sigma_{uv}}{\sigma_v^2}(E(\hat{\rho}^h x_T) - \rho E(\hat{\rho}^{h-1}x_T))$$

To get the expectation involving $\hat{\rho}^h$, we use similar method as in Bao (2007b), by applying the Nagar-type stochastic expansion of $\hat{\rho}^h$ as:

$$\hat{\rho}^h = \rho^h + a_{-1/2}^{(h)} + a_{-1}^{(h)} + a_{-3/2}^{(h)} + o_P(T^{-3/2}) \tag{1.10}$$

where

$$a_{-1/2}^{(h)} = h\rho^{h-1}a_{-1/2}$$

$$a_{-1}^{(h)} = h\rho^{h-1}a_{-1} + \frac{h(h-1)}{2}\rho^{h-2}a_{-1/2}^2$$
$$a_{-3/2}^{(h)} = h\rho^{h-1}a_{-3/2} + h(h-1)\rho^{h-2}a_{-1/2}a_{-1} + \frac{h(h-1)(h-2)}{6}\rho^{h-3}a_{-1/2}^3$$

and $\rho \neq 0$, $h > 0$. Also $a_{-1/2}$, $a_{-1}$, $a_{-3/2}$ are functions of the parameter $\rho$, and are of order $O(T^{-1/2})$, $O(T^{-1})$ and $O(T^{-3/2})$, respectively. The exact forms of $a_{-1/2}$, $a_{-1}$, $a_{-3/2}$ are in the appendix. Substitute the above expansion into $E(\hat{\rho}^h x_T)$, we have:

$$\begin{aligned} E(\hat{\rho}^h x_T) &= \rho^h x_T + h\rho^{h-1}E(a_{-1/2}x_T) + h\rho^{h-1}E(a_{-1}x_T) + \frac{h(h-1)}{2}\rho^{h-2}E(a_{-1/2}^2 x_T) \\ &\quad + o_P(T^{-1}) \end{aligned}$$

After some calculation applying expectations of quadratic forms, it can be verified that $E(a_{-1/2}x_T)$, $E(a_{-1}x_T)$ and $E(a_{-1/2}^2 x_T)$ are all of order $o(T^{-1})$, so $E(\hat{\rho}^h x_T) = \rho^h x_T$ up to order $O(T^{-1})$. Similarly, we also have $E(\hat{\rho}^{h-1}x_T) = \rho^{h-1}x_T$ up to order $O(T^{-1})$. Substitute the results to the forecast bias above, we find that the $h$-step-ahead forecast bias for the predictive regressions is also unbiased up to order $O(T^{-1})$ for $h > 1$.

So to sum up, for stationary AR(1) process of $x_t$ without intercept, the $h$-step-ahead forecast from the predictive regressions is unbiased up to order $O(T^{-1})$ for all $h \geq 1$, that is:

$$E(e_{T+h|T}) = 0 + o_P(\frac{1}{T}) \tag{1.11}$$

Besides forecast bias, another property of conditional feasible forecasts that worth studying is the mean squared forecast error. We first examine the MSFE for the one-step-ahead feasible forecast. Given the one-step-ahead forecast error $e_{T+1|T} = y_{T+1} - \hat{y}_{T+1|T} = (\beta - \hat{\beta})x_T + u_{T+1}$, the MSFE is just equal to the expectation of the squared forecast error, denoted by $E(e^2_{T+1|T})$. Substitute the decomposition of $\hat{\beta}$ in 1.2, we have,

$$\begin{aligned}
E(e^2_{T+1|T}) &= E((\beta-\hat{\beta})^2 x_T^2) + \sigma_u^2 = E(\big(\frac{\sigma_{uv}}{\sigma_v^2}(\rho-\hat{\rho}) - (X'X)^{-1}X'\varepsilon\big)^2 x_T^2) + \sigma_u^2 \\
&= \frac{\sigma_{uv}^2}{\sigma_v^4}E(\rho-\hat{\rho})^2 x_T^2 + E((X'X)^{-1}X'\varepsilon)^2 x_T^2 + \sigma_u^2 \\
&= \frac{\sigma_{uv}^2}{\sigma_v^4}(\rho^2 E(x_T^2) - 2\rho E(\hat{\rho}x_T^2) + E(\hat{\rho}^2 x_T^2)) + E((X'X)^{-1}X'\varepsilon)^2 x_T^2 + \sigma_u^2
\end{aligned}$$

then substitute the Nagar decomposition of $\hat{\rho}$ and $\hat{\rho}^2$ as in 1.10 for $h = 1, 2$, we have, up to $O(T^{-1})$,

$$\begin{aligned}
E(\hat{\rho}x_T^2) &= E(\rho + a_{-1/2} + a_{-1})x_T^2 = E(\rho x_T^2) + E(a_{-1/2}x_T^2) + E(a_{-1}x_T^2) \\
E(\hat{\rho}^2 x_T^2) &= E(\rho^2 + a^{(2)}_{-1/2} + a^{(2)}_{-1})x_T^2 = E(\rho^2 + 2\rho a_{-1/2} + 2\rho a_{-1} + a^2_{-1/2})x_T^2 \\
&= \rho^2 E(x_T^2) + 2\rho E(a_{-1/2}x_T^2) + 2\rho E(a_{-1}x_T^2) + E(a^2_{-1/2}x_T^2)
\end{aligned}$$

substitute back, the MSFE is then given by:

$$E(e^2_{T+1|T}) = \frac{\sigma_{uv}^2}{\sigma_v^4}E(a^2_{-1/2}x_T^2) + E((X'X)^{-1}X'\varepsilon)^2 x_T^2 + \sigma_u^2$$

We can verify that $E(a^2_{-1/2}x_T^2) = \frac{\sigma_v^2}{T}$, applying the vector form for each term and use the expectation of quadratic forms. To get $E((X'X)^{-1}X'\varepsilon)^2 x_T^2$, we can apply the Nagar-type expansion again, since $E((X'X)^{-1}X'\varepsilon)^2 x_T^2 = \sigma_\varepsilon^2 E(\frac{(\iota_T' X)^2}{X'X}) = E(\frac{N}{D})$, where $N = (\iota_T' X)^2$, $D = X'X$, and $\iota_T$ is a $T$ dimension vector with the last element equal to 1 and others zero. Then we have:

$$\frac{N}{D} = \frac{N}{ED}(1 + (D-ED)ED^{-1})^{-1} = a_{-1} + a_{-3/2} + o_P(T^{-3/2})$$

where $a_{-1} = \frac{N}{ED}$, $a_{-3/2} = -\frac{N}{ED}\frac{D-ED}{ED}$. Applying similar method as before, we can verify that up to order $O(T^{-1})$, $E((X'X)^{-1}X'\varepsilon)^2 x_T^2 = \frac{EN}{ED} = \frac{\sigma_\varepsilon^2(1-\rho^2)}{T\sigma_v^2}\frac{\sigma_v^2}{1-\rho^2} = \frac{\sigma_\varepsilon^2}{T}$. Substitute to the above equation, the one-step-ahead MSFE from the predictive regressions is given by:

$$E(e^2_{T+1|T}) = \frac{\sigma^2_{uv}}{\sigma^2_v T} + \frac{\sigma^2_\varepsilon}{T} + \sigma^2_u + o(T^{-1}) = \sigma^2_u(1+\frac{1}{T}) + o(T^{-1}) \tag{1.12}$$

It's interesting to notice that, the one-step-ahead MSFE depends only upon the variance of the error term in the first equation $u_t$ up to order $O(T^{-1})$, for any distribution of $x_t$, as long as $x_t$ is stationary.

In general, consider the h-step-ahead conditional MSFE for $h > 1$, which can be written as: $E(e^2_{T+h|T}) = \beta^2 E x^2_{T+h-1} + \sigma^2_u + E\hat{\beta}^2 \hat{x}^2_{T+h-1|T} - 2\beta E\hat{\beta} x_{T+h-1}\hat{x}_{T+h-1}$. Substitute the decomposition of $\hat{\beta}$ in 1.2, we get:

$$\begin{aligned} E(e^2_{T+h|T}) &= \beta^2(Ex^2_{T+h-1} + E\hat{x}^2_{T+h-1|T} - 2Ex_{T+h-1}\hat{x}_{T+h-1|T}) + \sigma^2_u \\ &\quad + \frac{\sigma^2_{uv}}{\sigma^4_v} E(\hat{\rho}-\rho)^2 \hat{x}^2_{T+h-1|T} + E((X'X)^{-1}X'\varepsilon)^2 \hat{x}^2_{T+h-1|T} \\ &\quad + 2\beta\frac{\sigma_{uv}}{\sigma^2_v} E(\hat{\rho}-\rho)(\hat{x}^2_{T+h-1|T} - x_{T+h-1}\hat{x}_{T+h-1|T}) \end{aligned}$$

Then substitute the recursive form $x_{T+h-1} = \rho^{h-1}x_T + \sum_{t=1}^{h-1} v_{T+t}\rho^{h-t-1}$, $\hat{x}_{T+h-1|T} = \hat{\rho}^{h-1}x_T$, after simplification, we have:

$$\begin{aligned} E(e^2_{T+h|T}) &= \sigma^2_u + \beta^2\sigma^2_v \frac{1-\rho^{2(h-1)}}{1-\rho^2} + \frac{\sigma^2_\varepsilon}{T}\rho^{2(h-1)} + 2(h-1)\rho^{2h-3}\frac{\sigma^2_\varepsilon(1-\rho^2)}{\sigma^2_v T}(Ea_{-1/2}x^2_T \\ &\quad + Ea_{-1}x^2_T) + Ea^2_{-1/2}x^2_T\left[\frac{\sigma^2_{uv}}{\sigma^4_v}\rho^{2(h-1)} + 2\beta\frac{\sigma_{uv}}{\sigma^2_v}(h-1)\rho^{2h-3} \right. \\ &\quad \left. + \frac{\sigma^2_\varepsilon(1-\rho^2)}{\sigma^2_v T}(h-1)(2h-3)\rho^{2(h-2)} + \beta^2(h-1)^2\rho^{2(h-2)}\right] \end{aligned}$$

Since we have verified that $Ea^2_{-1/2}x^2_T = \frac{\sigma^2_v}{T}$ from $h=1$, in addition, $Ea_{-1/2}x^2_T$, $Ea_{-1}x^2_T$ are also of order $O(T^{-1})$, substitute to the above equation, we have the h-step-ahead MSFE for the predictive regressions when $h > 1$ up to order $O(T^{-1})$ as:

$$E(e^2_{T+h|T}) = \sigma^2_u(1+\frac{\rho^{2(h-1)}}{T}) + \beta^2\sigma^2_v\frac{1-\rho^{2(h-1)}}{1-\rho^2} + \beta^2(h-1)^2\rho^{2(h-2)}\frac{\sigma^2_v}{T}$$

$$+2\beta\frac{\sigma_{uv}}{T}(h-1)\rho^{2h-3}+o(T^{-1}) \tag{1.13}$$

Notice that when $h = 1$, the above equation simplifies to the one-step-ahead MSFE in equation 1.12, so the above formula gives the analytical MSFE for all $h \geq 1$. Notice that when $h > 1$, the MSFE in $h$-step-ahead forecast up to order $O(T^{-1})$ depends on the parameter of $\beta$, $\rho$, the forecast horizon $h$ as well as the variance of the error terms in both equations and their covariance. However, if we only approximate the MSFE up to order 1, the expression is simplified to $E(e^2_{T+h|T}) = \sigma_u^2 + \beta^2\sigma_v^2\frac{1-\rho^{2(h-1)}}{1-\rho^2}$, where the covariance between $u_t$ and $v_t$ does not play a role.

## 1.3 Estimation and Forecasts of Model With Intercept

### 1.3.1 Estimation Bias and MSE

Then consider a more general case where the predictive regressions have nonzero intercept in both equations. We will first study the estimation properties for the stationary AR(1) process of $x_t$ and then consider the case where $x_t$ is a unit root process. The predictive regressions model with intercept has the form:

$$y_t = \alpha + \beta x_{t-1} + u_t$$

$$x_t = \theta + \rho x_{t-1} + v_t$$

where $|\rho| < 1$ or $\rho = 1$. Similarly, we assume $(u_t, v_t)'$ joint normally distributed, independent across t, with mean zero and variance-covariance matrix $\Sigma$, which has the same form as the case without intercept. Denote $b_1 = (\alpha, \beta)'$, $b_2 = (\theta, \rho)'$, then the OLS estimator for $b_1$ is $\hat{b}_1 = (X'X)^{-1}X'Y$, where $X = [\iota, X_{T-1}]$, $\iota$ is a vector of ones, $X_{T-1} = (x_0, ..., x_{T-1})'$, $Y =$

$(y_1, ..., y_T)'$. And the OLS estimator for $b_2$ is $\hat{b}_2 = (X'X)^{-1}X'X_T$, where $X_T = (x_1, ..., x_T)'$. Then the estimation error for $\hat{b}_1$ and $\hat{b}_2$ can be written as: $\hat{b}_1 - b_1 = (X'X)^{-1}X'u$, $\hat{b}_2 - b_2 = (X'X)^{-1}X'v$, where $u = (u_1, ..., u_T)'$, $v = (v_1, ..., v_T)'$. Use the same decomposition of $u$, where $u = \frac{\sigma_{uv}}{\sigma_v^2}v + \varepsilon$, which also implies $E(\varepsilon|X) = 0$. Then the estimation error for $\hat{b}_1$ can be rewritten as:

$$\begin{aligned}
\hat{b}_1 - b_1 &= (X'X)^{-1}X'u = (X'X)^{-1}X'(\frac{\sigma_{uv}}{\sigma^2 v}v + \varepsilon) \\
&= \frac{\sigma_{uv}}{\sigma_v^2}(X'X)^{-1}X'v + (X'X)^{-1}X'\varepsilon \\
&= \frac{\sigma_{uv}}{\sigma_v^2}(\hat{b}_2 - b_2) + (X'X)^{-1}X'\varepsilon \qquad (1.14)
\end{aligned}$$

Take expectation on both sides, the estimation bias of $\hat{b}_1$ is then given by:

$$B(\hat{b}_1) = E(\hat{b}_1 - b_1) = \frac{\sigma_{uv}}{\sigma_v^2}E(\hat{b}_2 - b_2) = \frac{\sigma_{uv}}{\sigma_v^2}B(\hat{b}_2) \qquad (1.15)$$

where $B(\hat{b}_2)$ is the estimation bias in the AR(1) or unit root process.

When $x_t$ is stationary, many studies give the form of the estimation bias of the AR(1) process with intercept assuming the error term to be normally distributed and up to order $O(T^{-1})$. Bao and Ullah (2007) gives the bias of $\hat{\rho}$ up to order $O(T^{-1})$, for the stationary AR(1) model with intercept under a general distributed error term and Bao (2007) extend the bias formula up to order $O(T^{-2})$. Applying this result to the above expression, the estimation bias in the predictive regressions for $\hat{\beta}$ up to order $O(T^{-2})$ can be written as:

$$\begin{aligned}
B(\hat{\beta}) &= E(\hat{\beta} - \beta) = \frac{\sigma_{uv}}{\sigma_v^2}E(\hat{\rho} - \rho) \\
&= -\frac{\sigma_{uv}}{\sigma_v^2}\frac{1 + 3\rho}{T} + \frac{\sigma_{uv}}{\sigma_v^2}\frac{1}{T^2}\frac{3\rho - 9\rho^2 - 1}{1 - \rho} + o(T^{-2}) \qquad (1.16)
\end{aligned}$$

provided $|\rho| < 1$. From the formula we can see that the bias of the coefficient estimation depends on the parameter of $\rho$, as well as the variance of the error term in the second equation $v_t$, and the covariance between the two equations. And up to order $O(T^{-1})$, given positively correlated $u_t$ and $v_t$, a more persistent series of $x_t$ will have larger estimation bias for stationary series.

Notice that although in the predictive regression model, our primary parameter of interest is the coefficient of the covariate, $\beta$. However, it's not hard to get the estimation properties of the intercept form the coefficient estimation bias of $\hat{\beta}$ by the decomposition of 1.14, where we have

$$B(\hat{\alpha}) = E(\hat{\alpha} - \alpha) = \frac{\sigma_{uv}}{\sigma_v^2} E(\hat{\theta} - \theta) = -\frac{\sigma_{uv}}{\sigma_v^2} \frac{\theta}{1-\rho} E(\hat{\rho} - \rho) = -\frac{\theta}{1-\rho} B(\hat{\beta}) \qquad (1.17)$$

the third equation holds since $\hat{\theta}$ and $\hat{\rho}$ are both the OLS estimator of the AR(1) process.

On the other hand, if $x_t$ follows a unit root process with intercept, that is, $x_t = \theta + x_{t-1} + v_t$, Kiviet and Phillips (2005) gives the formula for the bias of $\hat{\rho}$ up to $O(T^{-3})$ under the assumption that $v_t$ is normally distributed. As a more general case, Bao, Ullah, Zhang (2013) generalize the results assuming a generally distributed error term, where the formula of the bias of $\hat{\rho}$ incorporates the skewness of the error term $v_t$. Apply their results, we have the estimation bias of $\hat{\beta}$ when $x_t$ is a unit root process is given by:

$$B(\hat{\beta}) = -6\frac{\sigma_{uv}}{T^2\theta^2} + 18\frac{\sigma_{uv}}{T^3\theta^2} - \frac{84}{5}\frac{\sigma_{uv}\sigma_v^2}{T^3\theta^4} + o(T^{-3}) \qquad (1.18)$$

Notice the estimation bias of $\hat{\beta}$ when $x_t$ is unit root process depend on the parameter of the intercept in the unit root process, $\theta$, as well as the covariance of the two equations up to order $O(T^{-2})$, and the estimation bias up to order $O(T^{-3})$ also depends on the variance of the unit root model.

To get the conditional MSE of the predictive regressions, observe that from the OLS estimation, the conditional bias for $\hat{b}_1$ is zero, similar as the no intercept case, that is, $E(\hat{b}_1|X) = 0$, then the conditional variance of $\hat{b}_1$ is:

$$\begin{aligned} V(\hat{b}_1|X) &= V((\hat{b}_1 - b_1)|X) = V\left[(\frac{\sigma_{uv}}{\sigma_v^2}(\hat{b}_2 - b_2) + (X'X)^{-1}X'\varepsilon)|X\right] \\ &= \frac{\sigma_{uv}^2}{\sigma_v^4}V((\hat{b}_2 - b_2)|X) + (X'X)^{-1}\sigma_\varepsilon^2 = (X'X)^{-1}(\frac{\sigma_{uv}^2}{\sigma_v^2} + \sigma_\varepsilon^2) = (X'X)^{-1}\sigma_u^2 \end{aligned}$$

Then the conditional MSE of $\hat{b}_1$ is equal to the conditional variance, and for $\hat{\alpha}$ and $\hat{\beta}$ respectively, the formula are given by:

$$\begin{aligned} M(\hat{\alpha}|X) = V(\hat{\alpha}|X) &= \frac{\sigma_u^2 \sum_{t=1}^{T-1} x_t^2}{(T-1)\sum_{t=1}^{T-1} x_t^2 - (\sum_{t=1}^{T-1} x_t)^2} \\ M(\hat{\beta}|X) = V(\hat{\beta}|X) &= \frac{\sigma_u^2 (T-1)}{(T-1)\sum_{t=1}^{T-1} x_t^2 - (\sum_{t=1}^{T-1} x_t)^2} \end{aligned}$$

On the other hand, the unconditional variance of $\hat{b}_1$ can be written as:

$$\begin{aligned} V(\hat{b}_1) &= V(\hat{b}_1 - b_1) = V\left[\frac{\sigma_{uv}}{\sigma_v^2}(\hat{b}_2 - b_2) + (X'X)^{-1}X'\varepsilon\right] \\ &= \frac{\sigma_{uv}^2}{\sigma_v^4}V(\hat{b}_2 - b_2) + \sigma_\varepsilon^2 E(X'X)^{-1} \end{aligned}$$

And the unconditional MSE for $\hat{b}_1$ is then given by $M(\hat{b}_1) = B(\hat{b}_1)^2 + V(\hat{b}_1)$, and the MSE for $\hat{\beta}$ is followed by:

$$\begin{aligned} M(\hat{\beta}) &= B(\hat{\beta})^2 + V(\hat{\beta}) = \frac{\sigma_{uv}^2}{\sigma_v^4}E(\hat{\rho} - \rho)^2 + \frac{\sigma_{uv}^2}{\sigma_v^4}V(\hat{\rho} - \rho) + \sigma_\varepsilon^2(E(X'X))^{-1}_{(2,2)} \\ &= \frac{\sigma_{uv}^2}{\sigma_v^4}M(\hat{\rho}) + \sigma_\varepsilon^2(E(X'X))^{-1}_{(2,2)} \end{aligned}$$

where $M(\hat{\rho})$ is the parameter estimation MSE for the AR(1) or unit root process with intercept.

First consider the case when $x_t$ is stationary. To get $E(X'X)^{-1}$, we again use the Nagar-type expansion similar with the no intercept case. Denote $D = X'X$, then $E(X'X)^{-1} = E(D^{-1})$, where $E(D) = O(T)$, $(ED)^{-1}(D - ED) = O(T^{-1/2})$, then we have the following expansion for the stationary model with intercept:

$$\begin{aligned} D^{-1} &= (D - ED + ED)^{-1} = (ED)^{-1}(I + (ED)^{-1}(D - ED))^{-1} \\ &= (ED)^{-1}(I - (ED)^{-1}(D - ED) + (ED)^{-1}(D - ED)(D - ED)'(ED')^{-1}) + o(T^{-2}) \\ &= a_{-1} + a_{-3/2} + a_{-2} + o(T^{-2}) \end{aligned}$$

where $a_{-1} = (ED)^{-1}$, $a_{-3/2} = -(ED)^{-1}(ED)^{-1}(D - ED)$, $a_{-2} = (ED)^{-1}(ED)^{-1}(D - ED)(D - ED)'(ED')^{-1}$, and notice $E(a_{-3/2}) = 0$, so we have: $E(X'X)^{-1} = ED^{-1} = E(a_{-1}) + E(a_{-2}) + o(T^{-2})$. To get the MSE of $\hat{\beta}$, we only need the $(2, 2)$th element of $E(X'X)^{-1}$, denoted by $E(X'X)^{-1}_{(2,2)}$, that means, we only need to get $E(a_{-1(2,2)})$ and $E(a_{-2(2,2)})$. And apply similar method as the no intercept case, we can verify that:

$$\begin{aligned} E(a_{-1(2,2)}) &= (E(X'X))^{-1}_{(2,2)} = \frac{1-\rho^2}{T\sigma_v^2} \\ E(a_{-2(2,2)}) &= \frac{2(1-\rho^2)}{T^2\sigma_v^2} - \frac{2\rho\theta^2}{(1-\rho)T^2\sigma_v^2} \end{aligned}$$

the detailed proof is in the appendix.

As noted before, several researches considered the MSE for parameter estimation of the stationary AR(1) model, yet most under the assumption of normally distributed error terms. Bao (2007a) gave the exact formula of MSE of $\hat{\rho}$ up to order of $O(T^{-2})$ under a generally distributed error term. Apply this result, and combine with the expression for $(E(X'X))^{-1}_{(2,2)}$, the MSE for $\hat{\beta}$ for the predictive regressions with intercept and stationary

AR(1) $x_t$ is given by:

$$M(\hat{\beta}) = \frac{1-\rho^2}{T}\frac{\sigma_u^2}{\sigma_v^2} + \frac{\sigma_{uv}^2}{\sigma_v^4}\frac{1}{T^2}(23\rho^2 + 10\rho) + \frac{\sigma_\varepsilon^2}{T^2}\left[\frac{2(1-\rho^2)}{T^2\sigma_v^2} - \frac{2\rho\theta^2}{(1-\rho)T^2\sigma_v^2}\right] + o(T^{-2}) \quad (1.19)$$

Note that up to order $O(T^{-1})$, the MSE of $\hat{\beta}$ for the intercept model is exactly the same as the simple model without intercept in the previous part. Yet for MSE up to order $O(T^{-2})$, the intercept model is very different from the model without intercept.

When $x_t$ is a unit root process with intercept, first we can verify using similar method as the stationary case that $E(X'X)^{-1} = \frac{1}{ED} + o(T^{-3}) = \frac{12}{\theta^2 T^3} + o(T^{-3})$, where $D = X'X$. For the MSE of $\hat{\rho}$, Kiviet and Phillips (2005) give the formula under normally distributed error terms for unit root $x_t$ with intercept. As a more general case, Bao, Ullah and Zhang (2013) extends the formula to generally distributed error terms. Apply their results, we have the MSE for $\hat{\beta}$ of unit root $x_t$ with intercept as:

$$M(\hat{\beta}) = \frac{12\sigma_u^2}{\theta^2 T^3} + o(T^3) \quad (1.20)$$

Notice that the MSE of $\hat{\beta}$ with unit root $x_t$ depends on the parameter of intercept in the unit root process, $\theta$, the variance the first equation, that a larger intercept for the unit root process as well as smaller variance of the dependent variable will lead to smaller MSE of order $O(T^{-3})$.

### 1.3.2 Forecasts Bias and MSFE

Consider the one-step-ahead forecast for the predictive regressions with intercept when $x_t$ is stationary, since $y_{T+1} = \alpha + \beta x_T + u_{T+1}$, then the one-step-ahead forecast of $y_{t+1}$ given information set at time $T$ can be written as: $\hat{y}_{T+1|T} = E(y_{T+1}|\mathcal{I}_T) = \hat{\alpha} + \hat{\beta}x_T$, where

$\mathcal{I}_T$ denotes the information set at time $T$. The one-step-ahead forecast error is then equal to $e_{T+1|T} = y_{T+1} - \hat{y}_{T+1|T} = (\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_T + u_{T+1}$. And the one-step-ahead forecast bias is just the expectation of the forecast error, $E(e_{T+1|T})$. Substitute the decomposition of $\hat{\beta}$ by (1.14), we have:

$$
\begin{aligned}
E(e_{T+1|T}) &= E(\alpha - \hat{\alpha}) + E((\beta - \hat{\beta})x_T) \\
&= E(\alpha - \hat{\alpha}) + E\Big[\big(\frac{\sigma_{uv}}{\sigma_v^2}(\rho - \hat{\rho}) - (X'X)^{-1}X'\varepsilon_{(2)}\big)x_T\Big] \\
&= \frac{\sigma_{uv}}{\sigma_v^2}E(\theta - \hat{\theta}) + \frac{\sigma_{uv}}{\sigma_v^2}E\big[(\rho - \hat{\rho})x_T\big] \\
&= \frac{\sigma_{uv}}{\sigma_v^2}E(\theta - \hat{\theta} + (\rho - \hat{\rho})x_T) = \frac{\sigma_{uv}}{\sigma_v^2}E[e_{T+1|T}^{AR}]
\end{aligned}
$$

where $E(e_{T+1|T}^{AR})$ denotes the one-step-ahead forecast bias for the AR(1) process of $x_t$ with intercept.

For the forecast bias of stationary AR(1) model, Bao (2007b) studied for the case where the model does not contain intercept, where the results shows that the feasible forecast is unbiased up to order $O(T^{-1})$ for any forecast horizon $h$. In the appendix, we show that for the AR(1) process with intercept, similar results follow, that the $h$-step-ahead feasible forecast is unbiased up to order $O(T^{-1})$, for all $h \geq 1$. In this part, we need the one-step-ahead forecast bias, $E(e_{T+1|T}^{AR})$. Substitute to the above equation, we have, the one-step-ahead forecast bias for the predictive regressions with intercept is unbiased up to order $O(T^{-1})$, that is:

$$E(e_{T+1|T}) = 0 + o_P(\frac{1}{T}) \tag{1.21}$$

In general, consider the $h$-step-ahead forecast for $h > 1$, since $y_{T+h} = \alpha + \beta x_{T+h-1} + u_{T+h}$, the forecast for $y_{T+h}$ at time $T$ is equal to $\hat{y}_{T+h|T} = \hat{\alpha} + \hat{\beta}\hat{x}_{T+h-1|T}$, where $\hat{x}_{T+h-1|T}$ is the $h-1$ step ahead forecast from the AR(1) model with intercept. And the h-step-ahead forecast error for $y_{T+h}$ is equal to $e_{T+h|T} = y_{T+h} - \hat{y}_{T+h|T} = (\alpha - \hat{\alpha}) + \beta x_{T+h-1} - \hat{\beta}\hat{x}_{T+h-1|T} + u_{T+h}$. Notice that $x_{T+h-1}$ can be written recursively as $x_{T+h-1} = \rho^{h-1}x_T + \sum_{t=1}^{h-1}\theta\rho^{t-1} + \sum_{t=1}^{h-1}v_{T+t}\rho^{h-t-1}$, so $\hat{x}_{T+h-1|T} = E(x_{T+h-1}|\mathcal{I}_t) = \hat{\rho}^{h-1}x_T + \sum_{t=1}^{h-1}\hat{\theta}\hat{\rho}^{t-1}$. And the forecast error can also be written as: $e_{T+h|T} = (\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_{T+h-1} + \hat{\beta}(x_{T+h-1} - \hat{x}_{T+h-1}) + u_{T+h}$.

So the h-step-ahead forecast bias is given by taking the expectation of the forecast error:

$$\begin{aligned} E(e_{T+h|T}) &= E(y_{T+h} - \hat{y}_{T+h|T}) = E(\alpha - \hat{\alpha}) + E[(\beta - \hat{\beta})x_{T+h-1}] \\ &\quad + E[\hat{\beta}(x_{T+h-1} - \hat{x}_{T+h-1})] \end{aligned} \tag{1.22}$$

Substitute the decomposition of $\hat{\beta}$ in 1.14 to the second and third term, we have:

$$\begin{aligned} E[(\beta - \hat{\beta})x_{T+h-1}] &= E\left[\Big(\frac{\sigma_{uv}}{\sigma_v^2}(\rho - \hat{\rho}) - (X'X)^{-1}X'\varepsilon_{(2)}\Big)x_{T+h-1}\right] \\ &= E\left[\frac{\sigma_{uv}}{\sigma_v^2}(\rho - \hat{\rho})x_{T+h-1}\right] \\ E[\hat{\beta}(x_{T+h-1} - \hat{x}_{T+h-1})] &= E\left[\Big(\beta + \frac{\sigma_{uv}}{\sigma_v^2}(\hat{\rho} - \rho) + (X'X)^{-1}X'\varepsilon_{(2)}\Big)(x_{T+h-1} - \hat{x}_{T+h-1})\right] \\ &= \beta E(x_{T+h-1} - \hat{x}_{T+h-1}) + \frac{\sigma_{uv}}{\sigma_v^2}E\Big[(\hat{\rho} - \rho)x_{T+h-1}\Big] \\ &\quad -\frac{\sigma_{uv}}{\sigma_v^2}E\Big[(\hat{\rho} - \rho)\hat{x}_{T+h-1}\Big] \end{aligned}$$

Notice that $E(x_{T+h-1} - \hat{x}_{T+h-1})$ is just the $h-1$ step ahead forecast bias for the AR(1) model with intercept, substitute in to 1.22, the h-step-ahead forecast bias for $h > 1$ is then given by:

$$E(e_{T+h|T}) = E(\alpha - \hat{\alpha}) + \beta E(x_{T+h-1} - \hat{x}_{T+h-1}) - \frac{\sigma_{uv}}{\sigma_v^2}E[(\hat{\rho} - \rho)(\hat{\rho}^{h-1}x_T + \hat{\theta}\sum_{t=1}^{h-1}\hat{\rho}^{t-1})]$$

Then apply the expansion of $\hat{\rho}^h$ similar as the no intercept case as in 1.10, except the expression for $a_{-i/2}$, $i = 1, 2, 3$ are different, see the appendix for the expression of the model with intercept. Also, since we have shown that the h-step-ahead forecast bias for AR(1) model with intercept is unbiased up to order $O(T^{-1})$, for any $h \geq 1$, then we have up to order $O(T^{-1})$:

$$\begin{aligned} E(e_{T+h|T}) &= -\frac{\sigma_{uv}}{\sigma_v^2}[E(\hat{\theta}-\theta) + \sum_{t=1}^{h-1}\theta\rho^{t-1}Ea_{-1} + (\sum_{t=1}^{h-2}h\theta\rho^{t-1} - \frac{\theta}{1-\rho}\sum_{t=1}^{h-1}\rho^{t-1})Ea_{-1/2}^2] \\ &\quad + o(T^{-1}) \end{aligned}$$

where $a_{-1}$ and $a_{-1/2}^2$ are the terms in the Nagar expansion for the OLS estimator $\hat{\rho}$ in the AR(1) process with intercept, also see appendix. Check $Ea_{-1} = \frac{\theta}{1-\rho}$ and $Ea_{-1/2}^2 = \frac{1-\rho^2}{T}$, substitute to the above equation, we can see that the $h$-step-ahead forecast for the predictive regressions with intercept for $h > 1$ is also unbiased up to order $O(T^{-1})$. So we can conclude that the h-step-ahead forecast is unbiased up to order $O(T^{-1})$ for all forecast horizon $h$, that is,

$$E(e_{T+h|T}) = 0 + o_P(T^{-1}) \tag{1.23}$$

for all $h \geq 1$.

Then consider the MSFE for the predictive regressions with intercept and stationary AR(1) process of $x_t$. Since the one-step-ahead forecast error for $y_{T+1}$ is $e_{T+1|T}$, the one-step-ahead MSFE is then $E(e_{T+1|T}^2)$, where

$$\begin{aligned} E(e_{T+1|T}^2) &= E(\alpha-\hat{\alpha})^2 + E((\beta-\hat{\beta})^2 x_T^2) + 2E((\alpha-\hat{\alpha})(\beta-\hat{\beta})x_T) + \sigma_u^2 \\ &= \frac{\sigma_{uv}^2}{\sigma_v^4}E(\theta-\hat{\theta})^2 + E\big(\big(\frac{\sigma_{uv}}{\sigma_v^2}(\rho-\hat{\rho}) - (X'X)^{-1}X'\varepsilon\big)^2 x_T^2\big) + \sigma_u^2 \end{aligned}$$

$$
\begin{aligned}
& +2\frac{\sigma_{uv}^2}{\sigma_v^4}E(\theta-\hat{\theta})(\rho-\hat{\rho})x_T \\
= \; & \frac{\sigma_{uv}^2}{\sigma_v^4}[E(\theta-\hat{\theta})^2+E(\rho-\hat{\rho})^2x_T^2+2E(\theta-\hat{\theta})(\rho-\hat{\rho})x_T] \\
& +E((X'X)^{-1}X'\varepsilon)^2x_T^2+\sigma_u^2
\end{aligned}
$$

For the first part, substitute $\theta-\hat{\theta}=-\frac{\theta}{1-\rho}(\rho-\hat{\rho})$ and we have:

$$
\begin{aligned}
& E(\theta-\hat{\theta})^2+E(\rho-\hat{\rho})^2x_T^2+2E(\theta-\hat{\theta})(\rho-\hat{\rho})x_T \\
= \; & \frac{\theta^2}{(1-\rho)^2}E(\hat{\rho}-\rho)^2+(\rho^2E(x_T^2)-2\rho E(\hat{\rho}x_T^2)+E(\hat{\rho}^2x_T^2))-2\frac{\theta}{1-\rho}E(\hat{\rho}-\rho)^2x_T)
\end{aligned}
$$

to calculate, substitute for the decomposition of $\hat{\rho}$ and $\hat{\rho}^2$ as in 1.10 except that the terms of $a_{-i/2}$ are for the intercept model case, we have, up to $O(T^{-1})$,

$$
\begin{aligned}
E(\hat{\rho}x_T^2) &= E(\rho+a_{-1/2}+a_{-1})x_T^2=E(\rho x_T^2)+E(a_{-1/2}x_T^2)+E(a_{-1}x_T^2) \\
E(\hat{\rho}^2x_T^2) &= E(\rho^2+a_{-1/2}^{(2)}+a_{-1}^{(2)})x_T^2=E(\rho^2+2\rho a_{-1/2}+2\rho a_{-1}+a_{-1/2}^2)x_T^2 \\
&= \rho^2E(x_T^2)+2\rho E(a_{-1/2}x_T^2)+2\rho E(a_{-1}x_T^2)+E(a_{-1/2}^2x_T^2)
\end{aligned}
$$

and similar formula holds for $E(\hat{\rho}x_T)$ and $E(\hat{\rho}^2x_T)$, when substitute $x_T$ to $x_T^2$ respectively. Then we have,

$$
\begin{aligned}
& E(\theta-\hat{\theta})^2+E(\rho-\hat{\rho})^2x_T^2+2E(\theta-\hat{\theta})(\rho-\hat{\rho})x_T \\
& =\frac{\theta^2}{(1-\rho)^2}E(\hat{\rho}-\rho)^2+Ea_{-1/2}^2x_T^2-2\frac{\theta}{1-\rho}Ea_{-1/2}^2x_T
\end{aligned}
$$

and we can verify that $E(a_{-1/2}^2x_T^2)=\frac{\sigma_v^2}{T}$, $Ea_{-1/2}^2x_T=o(T^{-1})$.

To get $E((X'X)^{-1}X'\varepsilon)^2x_T^2$, we apply the Negar-type expansion again similar as before, where $E((X'X)^{-1}X'\varepsilon)^2x_T^2=\sigma_\varepsilon^2E(\frac{(\iota_T'X)^2}{X'X})=E(\frac{N}{D})$, and up to order $O(T^{-1})$, we have, $E((X'X)^{-1}X'\varepsilon)^2x_T^2=\sigma_\varepsilon^2\frac{EN}{ED}=\frac{\sigma_\varepsilon^2}{T}(1+\frac{(1+\rho)\theta^2}{\sigma_v^2(1-\rho)})$.

Substitute all the items, then the one-step-ahead MSFE of the predictive regressions up to order $T^{-1}$ is given by:

$$E(e_{T+1|T}^2) = \sigma_u^2[1 + \frac{1}{T} + \frac{\theta^2(1+\rho)}{T\sigma_v^2(1-\rho)}] + o(T^{-1}) \tag{1.24}$$

We can see that this result is more general that it nests the case for the one-step-ahead MSFE for the no intercept model where $E(e_{T+1|T}^2) = \sigma_u^2(1+\frac{1}{T}) + o(T^{-1})$ by substituting $\theta = 0$ to the above formula. Moreover, a lager magnitude of the coefficient in the AR(1) process will lead to a larger value of MSFE.

Then consider the h-step-ahead forecast MSFE for $h > 1$. For the predictive regressions with intercept, the h-step MSFE can be written as: $E(e_{T+h|T}^2) = E(\alpha - \hat{\alpha})^2 + \beta^2 E x_{T+h-1}^2 + \sigma_u^2 + E\hat{\beta}^2\hat{x}_{T+h-1|T}^2 - 2\beta E\hat{\beta}x_{T+h-1}\hat{x}_{T+h-1} + 2\beta E(\alpha-\hat{\alpha})x_{T+h-1} - 2E(\alpha - \hat{\alpha})\hat{\beta}\hat{x}_{T+h-1|T}$. Substitute the decomposition of $\hat{\beta}$ in 1.14 and the relation $E(\theta - \hat{\theta}) = -\frac{\theta}{1-\rho}E(\rho - \hat{\rho})$, we get:

$$\begin{aligned}
E(e_{T+h|T}^2) &= \beta^2(Ex_{T+h-1}^2 + E\hat{x}_{T+h-1|T}^2 - 2Ex_{T+h-1}\hat{x}_{T+h-1|T}) + \sigma_u^2 \\
&\quad + E((X'X)^{-1}X'\varepsilon)^2\hat{x}_{T+h-1|T}^2 + \frac{\sigma_{uv}^2}{\sigma_v^4}E(\hat{\rho}-\rho)^2\hat{x}_{T+h-1|T}^2 + \frac{\sigma_{uv}^2}{\sigma_v^4}E(\theta-\hat{\theta})^2 \\
&\quad - 2\frac{\sigma_{uv}}{\sigma_v^2}E(\theta-\hat{\theta})\big(\beta + \frac{\sigma_{uv}}{\sigma_v^2}(\hat{\rho}-\rho) + (X'X)^{-1}X'\varepsilon_{(2)}\big)\hat{x}_{T+h-1|T} \\
&\quad + 2\beta\frac{\sigma_{uv}}{\sigma_v^2}E(\theta-\hat{\theta})x_{T+h-1} + 2\beta\frac{\sigma_{uv}}{\sigma_v^2}E(\hat{\rho}-\rho)(\hat{x}_{T+h-1|T}^2 - x_{T+h-1}\hat{x}_{T+h-1|T}) \\
&= \beta^2 E(x_{T+h-1} - \hat{x}_{T+h-1|T})^2 + \sigma_u^2 + \frac{\sigma_{uv}^2}{\sigma_v^4}E(\hat{\rho}-\rho)^2\hat{x}_{T+h-1|T}^2 \\
&\quad + 2\beta\frac{\sigma_{uv}}{\sigma_v^2}E(\hat{\rho}-\rho)(\hat{x}_{T+h-1|T}^2 - x_{T+h-1}\hat{x}_{T+h-1|T}) + \frac{\sigma_{uv}^2}{\sigma_v^4}\frac{\theta^2}{(1-\rho)2}E(\hat{\rho}-\rho)^2 \\
&\quad + 2\beta\frac{\sigma_{uv}}{\sigma_v^2}\frac{\theta}{1-\rho}E(\hat{\rho}-\rho)(x_{T+h-1} - \hat{x}_{T+h-1|T}) + E((X'X)^{-1}X'\varepsilon)^2\hat{x}_{T+h-1|T}^2 \\
&\quad - 2\frac{\sigma_{uv}^2}{\sigma_v^4}\frac{\theta}{1-\rho}E(\hat{\rho}-\rho)^2\hat{x}_{T+h-1|T}
\end{aligned}$$

Then substitute $x_{T+h-1} = \rho^{h-1}x_T + \sum_{t=1}^{h-1} v_{T+t}\rho^{h-t-1}$, $\hat{x}_{T+h-1|T} = \hat{\rho}^{h-1}x_T$, after simplification, we have:

$$\begin{aligned} E(e^2_{T+h|T}) &= \sigma_u^2 + \beta^2\sigma_v^2 \frac{1-\rho^{2(h-1)}}{1-\rho^2} + 2(h-1)\rho^{2h-3}\frac{\sigma_\varepsilon^2(1-\rho^2)}{\sigma_v^2 T}(Ea_{-1/2}x_T^2 + Ea_{-1}x_T^2) \\ &+ Ea^2_{-1/2}x_T^2\left[\frac{\sigma^2_{uv}}{\sigma_v^4}\rho^{2(h-1)} + 2\beta\frac{\sigma_{uv}}{\sigma_v^2}(h-1)\rho^{2h-3} + \beta^2(h-1)^2\rho^{2(h-2)}\right. \\ &\left. + \frac{\sigma_\varepsilon^2(1-\rho^2)}{\sigma_v^2 T}(h-1)(2h-3)\rho^{2(h-2)}\right] + \sigma_u^2\frac{\theta^2(1+\rho)}{T\sigma_v^2(1-\rho)} + \frac{\sigma_\varepsilon^2}{T}\rho^{2(h-1)} \end{aligned}$$

Similarly we can check $Ea^2_{-1/2}x_T^2 = \frac{\sigma_v^2}{T}$, and $Ea_{-1/2}x_T^2$, $Ea_{-1}x_T^2$ are of order $O(T^{-1})$, substitute to the above equation, we have the h-step-ahead MSFE for predictive regressions with intercept up to order $O(T^{-1})$:

$$\begin{aligned} E(e^2_{T+h|T}) &= \sigma_u^2(1 + \frac{\rho^{2(h-1)}}{T} + \frac{\theta^2(1+\rho)}{T\sigma_v^2(1-\rho)}) + \beta^2\sigma_v^2\frac{1-\rho^{2(h-1)}}{1-\rho^2} + \beta^2(h-1)^2\rho^{2(h-2)}\frac{\sigma_v^2}{T} \\ &+ 2\beta\frac{\sigma_{uv}}{T}(h-1)\rho^{2h-3} + o(T^{-1}) \end{aligned} \tag{1.25}$$

Notice that this is the general formula of the MSFE for any forecast horizon and for model with and without intercept. If $h = 1$, the above equation simplifies to the one-step-ahead MSFE in equation 1.24 for the predictive regression model with intercept. And if we substitute $\theta = 0$, then the formula simplifies to the MSFE for the model without intercept in 1.13.

## 1.4 Monte Carlo Simulation

To see how the above analytical results of the estimation bias and MSE, forecast bias and MSFE works in estimation bias correction as well as for inference, in this part, we use Monte Carlo simulated data to verify our results, for both the model with and without intercept and for both stationary AR(1) and unit root process of $x_t$. Although we assumed the error terms to follow a joint normal distribution, we would like to see how the bias reduction works under nonnormally distributed error terms for the robustness of the bias correction as well as the forecast properties, since in empirical studies, most financial data are asymmetric as well as fat tailed. We allow the error terms $(u_t, v_t)$ have a joint distribution of bivariate normal, bivariate student-t, and bivariate log Normal, respectively. And we will verify the above results applying different values of parameters. Table 1 through 9 reports some of the simulated results. To save space, we only report results for certain parameter values: for stationary AR(1) case, $\rho = 0.2, 0.5, 0.8$, $\beta = 0.3, 0.6, 0.9$, correlation coefficient between $u_t$ and $v_t$ is set to 0.5 for no intercept model, and in addition, $\alpha = 0.5$, $\theta = 0.4$ for the stationary intercept model and $\alpha = 0.5$, $\theta = 2$ for the unit root model with intercept. For the unit root case, we only report the model with intercept case for simplicity. We have a small sample size of 30 and moderate sample size of 100. For the case of $T = 100$, we only report the results when $\rho = 0.5$, the results for the other values of $\rho$ follows similarly. The simulation is over 10,000 repeated samples.

For the parameter estimation part, in each table, we first report the OLS estimators, denoted by $\hat{\rho}$, $\hat{\beta}$,$\hat{\theta}$, $\hat{\alpha}$. And then we report the bias corrected estimators, denoted by $\tilde{\rho}$, $\tilde{\beta}$, $\tilde{\theta}$ , $\tilde{\alpha}$, where the bias corrected estimators are calculated using the OLS estimator minus the estimation bias given by the formula, for example, $\tilde{\beta} = \hat{\beta} - B(\hat{\beta})$. And depending on the bias formula we use, the bias correction can be up to order either $O(T^{-1})$ or $O(T^{-2})$. And since the bias formula given before is a function of the model parameters, we can substitute the real value of the parameters in simulated data, yet we can only substitute the estimated values in real data, so we also try the feasible bias corrected estimator here. The feasible bias corrected estimators are denoted by $\ddot{\rho}$, $\ddot{\beta}$, $\ddot{\theta}$, $\ddot{\alpha}$, where the feasible bias corrected estimators are calculated by substituting the OLS estimators to the bias correction formula, $B(\hat{\beta})$, for example. Since for the predictive regression model, the estimation of the coefficient of $\beta$ is our primary interest, we will have another improved bias corrected estimator for it, which is calculated by substituting the feasible bias corrected estimator of $\rho$, $\ddot{\rho}$, to the bias formula of $\hat{\beta}$, since if the feasible bias corrected estimator of $\rho$ works better than the OLS estimator $\hat{\rho}$, we should expect that the improved bias correction of $\beta$ works better than the feasible bias corrected estimator of $\beta$. We will denote the improved bias corrected estimator for $\beta$ as $\check{\beta}$. Similarly, depending on whether we substitute the true parameter value, the OLS estimator, or the feasible bias corrected estimator, we have different results for the MSE of $\beta$, denoted by $M(\tilde{\beta})$, $M(\ddot{\beta})$ and $M(\check{\beta})$ respectively. Also, we compute the simulated MSE of the OLS estimator of $\beta$, denoted by $M(\hat{\beta})$.

To check whether the forecast is unbiased for any forecast step, as well as how the MSFE works, we check the 1-step, 2-step and 5-step forecasts. We cut the data into equal size of in-sample and pseudo out-of sample observations, and use the rolling window method to get the forecasts with the fixed window size equal to the in-sample size. For each case, we report the forecast bias, the MSFE based on the data, denoted by MSFE($h = 1, 2, 5$), the MSFE calculated applying the formula above applying the true parameter value or its OLS estimator, as well as the MSFE for the forecasts based on two simple models: martingale difference model and historical mean model as a comparison, where for the martingale difference model, the optimal forecast is always equal to zero for any forecast horizon.

Table 1 through 3 reports the estimation and forecast results for model without intercept, when sample size is equal to 30, for different values of $\rho$. Table 4 gives one case for sample size equal to 100 when $\rho = 0.5$. Table 5 through 8 gives the corresponding results for model with intercept, and table 9 shows the case for unit root model with intercept when sample size is equal to 30. From panel a of tables 1 through 8, we can see that given any values of $\rho$ and for both models, with or without intercept, the bias corrected estimators are always better than the OLS estimators in the way that after bias correction, all the estimators are closer to their corresponding true values. Moreover, the feasible bias corrected estimator as well as the improved bias corrected estimator performs similar to the bias corrected estimator where true parameter values are used, indicating that the parameter estimation uncertainty does not affect the efficiency of bias reduction. In addition, even for distribution with fat tails such t-distribution or with both asymmetry and fat tails such as log Normal distribution, bias correction can reduce estimation bias most of the time, indicating that the bias reduction is robust to the distribution of error terms. Compare the MSE calculated from formula with simulated value, we can see that substituting the bias corrected estimator to the analytical formula will give a even better approximation to the simulated MSE than using the true value of the estimator, and the formula gives a good approximation even when sample size is small (T=30). Panel b of table 1 though 8 gives the forecasts results for different parameter values corresponding to each model. First we can observe that the forecast bias is small for all cases, verified the results that forecasts for all horizons are unbiased. For the MSFE, we can conclude from the tables that the values calculated using formula is close to the simulated value most of the time, indicating that the formula gives a good approximation. However, due to model estimation uncertainty, sometime the predictive regressions model may not produce better forecasts compared to simple models such as historical mean model or martingale difference model in terms of having a lower MSFE, especially when the forecast horizon is longer.

When $X$ is unit root process, we have similar results that the bias corrected estimator works better than the OLS estimator where the bias is significantly reduced after correction. Notice that for the unit root model, the value of intercept for the unit root process is set as $\theta = 2$, since when the ratio of $\theta/\sigma_v^2$ is small, in particular, smaller than 1, the OLS estimation will suffer from series bias problem when sample size is small, as noted in Kivet and Phillips (2005), while when the ratio is larger, the problem will be unnoticeable. For the forecasts from unit root process, we compare the MSFE using OLS estimators with the MSFE assuming a known unit root (i.e. $\rho = 1$), where there is no parameter estimation uncertainty in $\rho$. Since when $h = 1$, the forecasts is not a function of $\rho$, so we can only compare the two for $h = 2, 5$. From the result, we can see that when there is no parameter estimation uncertainty in $\rho$, the MSFE is reduced compared with the case where the OLS estimation of $\rho$ is substituted. Notice that when $X$ is a unit root process, historical mean model and martingale difference model no long work, since forecasts based on these two models ignores the trend and thus will have large bias.

## 1.5 Predictive Regressions of Financial Return

We apply the previous analytical estimation and forecasts properties of the predictive regressions to modeling financial return in this part. In particular, we will apply the estimation bias formula to do bias correction on the parameter estimation as well as applying the bias reduced estimator in the feasible conditional forecasts. We will consider the effects of several factors such as dividend price ratio, dividend yield ratio, net equity expansion and T-bill rate on stock return as well as equity premium, using the data in Goyal and Welch (2008). In addition to their original monthly data, we calculate the annualized monthly stock return as in Campbell and Thompson (2008).

Denote $P_t$ as the S&P500 index at month $t$. The monthly simple one-month return from month $t$ to month $t+1$ is defined as $R_t(1) \equiv P_{t+1}/P_t - 1$, and equity premium is defined as $Q_t(1) \equiv R_t(1) - r_t^f$ with $r_t^f$ being the risk-free interest rate. Following Campbell, Lo and MacKinlay (1997, p. 10), we define the aggregated $k$-period return from month $t$ to month $t+k$ as

$$
\begin{aligned}
R_t(k) &\equiv \frac{P_{t+k}}{P_t} - 1 \\
&= \left(\frac{P_{t+k}}{P_{t+k-1}}\right) \times \cdots \times \left(\frac{P_{t+1}}{P_t}\right) - 1 \\
&= (1 + R_{t+k-1}(1)) \times \cdots \times (1 + R_t(1)) - 1 \qquad (1.26)
\end{aligned}
$$

and following Campbell and Thompson (2008) we define the aggregated $k$-period equity premium as

$$
\begin{aligned}
Q_t(k) &\equiv (1 + R_{t+k-1}(1) - r_{t+k-1}^f) \times \cdots \times (1 + R_t(1) - r_t^f) - 1 \\
&= (Q_{t+k-1}(1) + 1) \times \cdots \times (Q_t(1) + 1) - 1 \\
&= \left[\prod_{j=1}^{k} (Q_{t+k-j}(1) + 1)\right] - 1. \qquad (1.27)
\end{aligned}
$$

To model financial return using certain factors we are interested in, we let $y_t = R_t(k)$ or $Q_t(k)$, and consider $k = 1, 3, 12$ denoting the monthly data, quarterly aggregated monthly data, and the annually aggregated monthly data respectively. We consider four factors separately for our predictive regressions to examine their effects on financial returns respectively: dividend price ratio (d/p), dividend yield (d/y), net equity expansion (ntis) and treasure bill rates (tbl), where the ntis is a variable reflects the issuing activities of corporations. For a detailed definition and calculation of each of the predictors as well as the return, please refer to Goyal and Welch (2008).

We consider the data from January 1952 to December 1989 in particular, since as noted in Zhu (2013), the interest rate is hard to interpret before the 1951 Treasury Accord and there was significant structural break around the 1990s. We divide the data equally into $R$ in-sample observations and $P$ pseudo out-of-sample observations to evaluate predictability with bias correction. The models are estimated using rolling windows of the fixed size $R$. For annualized or quarterly aggregated monthly data, to avoid using future information, we only use data up to month $t-11$ or $t-2$ for estimation. For forecast comparison, we compare the forecasts from the predictive regression model with two simple models: the historical mean forecast and the martingale difference forecast where the optimal forecasts for $y_t$ is zero at all forecast horizon.

Table 10 reports the results for estimation and forecasts from data. Since the predictability of the covariates on the month or even quarterly aggregated return does not vary much between models, we only report the case for $k = 12$ here. Panel a of table 10 summarizes the 95 % confidence interval of the OLS estimation of the parameters in the predictive regressions using different regressors for both annualized stock return and equity premium. According to the table, it is obvious that when using d/p or d/y, for both equity premium and stock return, the predictive regressions has stationary AR(1) process for $x_t$ and nonzero intercept for both equations. However, when using ntis or tbl, the estimation of $\rho$ for both two stock return and equity premium is very close to one, though the confidence interval itself does not contain 1. Also notice that for both of these two variables, the estimated intercepts are close to zero as well for the AR(1) process. Since based on current results, the bias correction result for unit root process of $x_t$ with zero intercept is approximated under a different frame work in Abadir (1993, 1995) through calculations under normality assumptions rather than expansions. And moreover, for unit root model with intercept, when the true intercept is small relative to variance of the error term, the bias correction maynot work well, according to Kiviet and Phillips (2005) and our simulation results. And based on the data observations, the error terms of the covariates is far away from normal distribution since both the skewness and excess kurtosis is significantly not zero. Based on these facts, we apply the bias correction formula for stationary AR(1) process of $x_t$ of model with intercept when $x$ is ntis or tbl as well.

Panel b of table 10 reports the parameter estimation and bias correction results as well as the corresponding MSFE of the feasible conditional forecasts. We report the OLS estimator for the four parameters in the predictive regressions as well as the feasible bias corrected estimator using the bias correction formula up to order $O(T^{-2})$ for stationary model with intercept in section three. The notations are similar as the simulation part. From the results, we can see that after bias correction, the effects of all the four variables, d/p, d/y, ntis and tbl, increased in absolute value compared to the OLS estimator, indicating that the OLS estimation underestimates the effects of these variables on both stock return and equity premium. And after bias correction, we will expect a larger effect of each variable on the return. To see how the bias correction to the coefficients works in forecasting, we will compare the MSFE of the 1, 2, and 5 step feasible conditional forecasts using OLS estimators and the bias corrected estimators. In table 10 panel b, MSFE (h=1,2,5) denotes the corresponding mean squared forecast error using OLS estimators, and MSFE-C (h=1,2,5) denotes the one using the bias corrected estimators. We multiply the numbers by 100 in order to compare more decimal numbers for the result. Results show that for d/p, d/y, and ntis, for almost all the forecast horizon, applying bias corrected estimators gives lower MSFE compared with using OLS estimators, while for tbl, the OLS estimators gives lower MSFE. In the table we also report the MSFE according the formula, which is smaller than the realized value from data, since the formula only approximates up to order $O(T^{-2})$. In the bottom part, we also give the MSFE from the two simple models, historical mean and martingale difference. Results show that for equity premium, martingale difference gives the best forecasts compared to other models, and for stock returns, historical mean gives the best forecasts. This result is consistent to most empirical studies for financial return that, the stock return and equity premium is hard to predict using covariates, especially applying linear models.

## 1.6 Conclusion

This paper focuses on the analytical result of the estimation and forecasting properties of the predictive regressions, which has been used widely in modeling financial return. We allow the regressor to follow either a stationary AR(1) process or unit root process, as well as zero or nonzero intercept in both equations. The main thrust of this paper is to develop an analytical bias reduced estimator and study its mean squared error (MSE) efficiency as well as properties of feasible optimal forecasts.

The results from this paper shed lights on the bias reduction estimator of the predictive regressions in small samples, for both stationary AR(1) and unit root regressors, as well as the optimal forecasts efficiency for stationary AR(1) regressor. We applied our analytical results to both simulation and empirical applications. Simulation results show that the bias correction can effectively reduce estimation bias and the MSE and MSFE formula gives good approximation. Moreover, by allowing the error terms to have different distributions, including asymmetric as well as fat tailed, results show that the bias reduction process works well under distributions other than normal as well. We model stock return and equity premium using four factors such as dividend yield ratio, dividend price ratio, corporate issuing activity as well as short term rates. The results show that if we ignore the estimation bias, the effects of all the factors on stock return as well as equity premium will be underestimated. Such results not only give inaccurate conclusion of the magnitude to which those factors affect return, moreover, the optimal forecasts using biased OLS estimators will be inefficient as well. On the other hand, feasible forecasts using the bias reduced estimators applying correction performs better than applying the OLS estimators directly. All of the above results highlight the importance of the bias reduction in estimation and forecasting.

Table 1.1: No Intercept Model, T=30, $\rho = 0.2$

(a) Estimator Bias and MSE

| $\rho$=0.2 corr$(u,v)$=0.5 | bivN | $\beta$=0.3 bivT | bivlogN | bivN | $\beta$=0.6 bivT | bivlogN | bivN | $\beta$=0.9 bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\rho}$ | 0.1877 | 0.1881 | 0.2120 | 0.1874 | 0.1886 | 0.2119 | 0.1872 | 0.1876 | 0.2111 |
| $\tilde{\rho}$, $O(T^{-1})$ | 0.2010 | 0.2014 | 0.2253 | 0.2008 | 0.2019 | 0.2252 | 0.2005 | 0.2009 | 0.2244 |
| $\ddot{\rho}$, $O(T^{-1})$ | 0.2002 | 0.2006 | 0.2261 | 0.1999 | 0.2012 | 0.2260 | 0.1996 | 0.2001 | 0.2251 |
| $\tilde{\rho}$, $O(T^{-2})$ | 0.2001 | 0.1978 | 0.2244 | 0.1999 | 0.1984 | 0.2243 | 0.1996 | 0.1973 | 0.2235 |
| $\ddot{\rho}$, $O(T^{-2})$ | 0.1989 | 0.1974 | 0.2074 | 0.1987 | 0.1979 | 0.2072 | 0.1984 | 0.1968 | 0.2064 |
| $\hat{\beta}$ | 0.2942 | 0.2949 | 0.3023 | 0.5931 | 0.5941 | 0.6013 | 0.8929 | 0.8932 | 0.9010 |
| $\tilde{\beta}$, $O(T^{-1})$ | 0.3009 | 0.3016 | 0.3089 | 0.5998 | 0.6008 | 0.6079 | 0.8995 | 0.8998 | 0.9077 |
| $\ddot{\beta}$, $O(T^{-1})$ | 0.3004 | 0.3012 | 0.3093 | 0.5993 | 0.6004 | 0.6082 | 0.8991 | 0.8994 | 0.9080 |
| $\check{\beta}$, $O(T^{-1})$ | 0.3009 | 0.3016 | 0.3098 | 0.5997 | 0.6008 | 0.6087 | 0.8995 | 0.8999 | 0.9085 |
| $\tilde{\beta}$, $O(T^{-2})$ | 0.3004 | 0.2998 | 0.3085 | 0.5993 | 0.5990 | 0.6075 | 0.8991 | 0.8981 | 0.9073 |
| $\ddot{\beta}$, $O(T^{-2})$ | 0.2998 | 0.2996 | 0.3015 | 0.5987 | 0.5988 | 0.6004 | 0.8985 | 0.8978 | 0.9002 |
| $\check{\beta}$, $O(T^{-2})$ | 0.3002 | 0.3000 | 0.3017 | 0.5991 | 0.5991 | 0.6007 | 0.8989 | 0.8982 | 0.9005 |
| $M(\hat{\beta})$, | 0.0351 | 0.0353 | 0.0522 | 0.0348 | 0.0350 | 0.0508 | 0.0350 | 0.0350 | 0.0522 |
| $M(\tilde{\beta})$, $O(T^{-1})$ | 0.0320 | 0.0320 | 0.0320 | 0.0320 | 0.0320 | 0.0320 | 0.0320 | 0.0320 | 0.0320 |
| $M(\ddot{\beta})$, $O(T^{-1})$ | 0.0327 | 0.0340 | 0.0523 | 0.0327 | 0.0341 | 0.0520 | 0.0327 | 0.0340 | 0.0522 |
| $M(\check{\beta})$, $O(T^{-1})$ | 0.0324 | 0.0337 | 0.0516 | 0.0324 | 0.0338 | 0.0513 | 0.0324 | 0.0337 | 0.0515 |
| $M(\tilde{\beta})$, $O(T^{-2})$ | 0.0334 | 0.0399 | 0.0389 | 0.0334 | 0.0399 | 0.0389 | 0.0334 | 0.0399 | 0.0390 |
| $M(\ddot{\beta})$, $O(T^{-2})$ | 0.0342 | 0.0358 | 0.0567 | 0.0342 | 0.0358 | 0.0564 | 0.0342 | 0.0357 | 0.0567 |
| $M(\check{\beta})$, $O(T^{-2})$ | 0.0339 | 0.0354 | 0.0559 | 0.0339 | 0.0355 | 0.0556 | 0.0339 | 0.0354 | 0.0559 |

(b) Forecast Bias and MSFE

| $\rho$=0.2 corr$(u,v)$=0.5 | bivN | $\beta$=0.3 bivT | bivlogN | bivN | $\beta$=0.6 bivT | bivlogN | bivN | $\beta$=0.9 bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $E(e_{T+1\|T})$ | 0.0001 | 0.0014 | 0.0201 | 0.0008 | -0.0008 | 0.0200 | -0.0001 | -0.0009 | 0.0209 |
| $E(e_{T+2\|T})$ | 0.0002 | 0.0005 | -0.0063 | -0.0006 | -0.0010 | 0.0086 | -0.0004 | -0.0012 | 0.0309 |
| $E(e_{T+5\|T})$ | 0.0004 | 0.0001 | -2.9353 | -0.0002 | -0.0047 | -5.2951 | -0.0011 | -0.0018 | -1.5595 |
| MSFE$(h=1)$ | 1.0686 | 1.7591 | 4.6902 | 1.0717 | 1.7612 | 4.6648 | 1.0715 | 1.7568 | 4.6671 |
| MSFE$(h=2)$ | 1.1004 | 1.8090 | 4.8773 | 1.3951 | 2.2921 | 6.2905 | 1.8739 | 3.0836 | 8.5972 |
| MSFE$(h=5)$ | 1.1937 | 1.9550 | 5.5204 | 1.5080 | 2.4697 | 7.1095 | 2.0197 | 3.3145 | 9.7292 |
| $E(e^2_{T+1\|T})$ | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 |
| $E(e^2_{T+2\|T})$ | 1.1027 | 1.1027 | 1.1027 | 1.3947 | 1.3947 | 1.3947 | 1.8787 | 1.8787 | 1.8787 |
| $E(e^2_{T+5\|T})$ | 1.0938 | 1.0938 | 1.0938 | 1.3750 | 1.3750 | 1.3750 | 1.8438 | 1.8438 | 1.8438 |
| $E(\tilde{e}^2_{T+1\|T})$ | 1.0000 | 1.6689 | 4.6622 | 1.0020 | 1.6673 | 4.6338 | 1.0013 | 1.6700 | 4.6532 |
| $E(\tilde{e}^2_{T+2\|T})$ | 1.0949 | 1.8271 | 5.0998 | 1.3705 | 2.2778 | 6.3694 | 1.8193 | 3.0401 | 8.5497 |
| $E(\tilde{e}^2_{T+5\|T})$ | 1.0998 | 1.8380 | 5.7892 | 1.3910 | 2.3125 | 7.4048 | 1.8639 | 3.1134 | 9.6352 |
| MSFE-HM$(h=1)$ | 1.1617 | 1.9399 | 5.4570 | 1.4636 | 2.4321 | 6.8202 | 1.9582 | 3.2670 | 9.1908 |
| MSFE-HM$(h=2)$ | 1.1050 | 1.8438 | 5.1651 | 1.4129 | 2.3488 | 6.5399 | 1.9032 | 3.1790 | 8.8722 |
| MSFE-HM$(h=5)$ | 0.8724 | 1.4546 | 4.0670 | 1.1189 | 1.8591 | 5.1599 | 1.5110 | 2.5254 | 7.0268 |
| MSFE-MD$(h=1)$ | 1.0910 | 1.8216 | 5.1241 | 1.3766 | 2.2888 | 6.4092 | 1.8445 | 3.0743 | 8.6451 |
| MSFE-MD$(h=2)$ | 1.0179 | 1.6983 | 4.7783 | 1.2848 | 2.1369 | 5.9818 | 1.7204 | 2.8690 | 8.0673 |
| MSFE-MD$(h=5)$ | 0.7995 | 1.3334 | 3.7495 | 1.0098 | 1.6773 | 4.6877 | 1.3524 | 2.2563 | 6.3411 |

Table 1.2: No Intercept Model, T=30, $\rho = 0.5$

(a) Estimator Bias and MSE

| $\rho$=0.5<br>corr$(u, v)$=0.5 | bivN | $\beta$=0.3<br>bivT | bivlogN | bivN | $\beta$=0.6<br>bivT | bivlogN | bivN | $\beta$=0.9<br>bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\rho}$ | 0.4691 | 0.4695 | 0.4963 | 0.4683 | 0.4722 | 0.4969 | 0.4693 | 0.4707 | 0.4956 |
| $\tilde{\rho}$, $O(T^{-1})$ | 0.5025 | 0.5028 | 0.5296 | 0.5016 | 0.5055 | 0.5303 | 0.5026 | 0.5040 | 0.5289 |
| $\ddot{\rho}$, $O(T^{-1})$ | 0.5004 | 0.5008 | 0.5294 | 0.4995 | 0.5037 | 0.5301 | 0.5006 | 0.5021 | 0.5286 |
| $\tilde{\rho}$, $O(T^{-2})$ | 0.5002 | 0.4940 | 0.5274 | 0.4994 | 0.4966 | 0.5280 | 0.5004 | 0.4951 | 0.5267 |
| $\ddot{\rho}$, $O(T^{-2})$ | 0.4978 | 0.4948 | 0.4988 | 0.4969 | 0.4976 | 0.4993 | 0.4980 | 0.4960 | 0.4980 |
| $\hat{\beta}$ | 0.2832 | 0.2847 | 0.2944 | 0.5841 | 0.5855 | 0.5946 | 0.8846 | 0.8853 | 0.8944 |
| $\tilde{\beta}$, $O(T^{-1})$ | 0.2999 | 0.3013 | 0.3111 | 0.6008 | 0.6021 | 0.6112 | 0.9013 | 0.9020 | 0.9110 |
| $\ddot{\beta}$, $O(T^{-1})$ | 0.2988 | 0.3003 | 0.3104 | 0.5998 | 0.6012 | 0.6105 | 0.9003 | 0.9010 | 0.9103 |
| $\check{\beta}$, $O(T^{-1})$ | 0.2999 | 0.3013 | 0.3114 | 0.6008 | 0.6022 | 0.6116 | 0.9013 | 0.9020 | 0.9114 |
| $\tilde{\beta}$, $O(T^{-2})$ | 0.2988 | 0.2969 | 0.3100 | 0.5997 | 0.5977 | 0.6101 | 0.9002 | 0.8975 | 0.9099 |
| $\ddot{\beta}$, $O(T^{-2})$ | 0.2976 | 0.2973 | 0.2977 | 0.5985 | 0.5981 | 0.5978 | 0.8990 | 0.8980 | 0.8976 |
| $\check{\beta}$, $O(T^{-2})$ | 0.2985 | 0.2982 | 0.2981 | 0.5994 | 0.5990 | 0.5982 | 0.9000 | 0.8989 | 0.8980 |
| $M(\hat{\beta})$, | 0.0291 | 0.0296 | 0.0417 | 0.0285 | 0.0289 | 0.0416 | 0.0287 | 0.0291 | 0.0409 |
| $M(\tilde{\beta})$, $O(T^{-1})$ | 0.0250 | 0.0250 | 0.0250 | 0.0250 | 0.0250 | 0.0250 | 0.0250 | 0.0250 | 0.0250 |
| $M(\ddot{\beta})$, $O(T^{-1})$ | 0.0265 | 0.0275 | 0.0402 | 0.0265 | 0.0274 | 0.0408 | 0.0264 | 0.0274 | 0.0405 |
| $M(\check{\beta})$, $O(T^{-1})$ | 0.0253 | 0.0263 | 0.0378 | 0.0253 | 0.0262 | 0.0383 | 0.0253 | 0.0262 | 0.0381 |
| $M(\tilde{\beta})$, $O(T^{-2})$ | 0.0269 | 0.0320 | 0.0312 | 0.0269 | 0.0320 | 0.0313 | 0.0269 | 0.0320 | 0.0313 |
| $M(\ddot{\beta})$, $O(T^{-2})$ | 0.0284 | 0.0296 | 0.0436 | 0.0284 | 0.0295 | 0.0443 | 0.0284 | 0.0295 | 0.0439 |
| $M(\check{\beta})$, $O(T^{-2})$ | 0.0273 | 0.0284 | 0.0412 | 0.0273 | 0.0283 | 0.0418 | 0.0273 | 0.0284 | 0.0415 |

(b) Forecast Bias and MSFE

| $\rho$=0.5<br>corr$(u, v)$=0.5 | bivN | $\beta$=0.3<br>bivT | bivlogN | bivN | $\beta$=0.6<br>bivT | bivlogN | bivN | $\beta$=0.9<br>bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $E(e_{T+1\|T})$ | 0.0011 | -0.0003 | 0.0270 | -0.0020 | -0.0003 | 0.0274 | 0.0021 | -0.0008 | 0.0287 |
| $E(e_{T+2\|T})$ | 0.0005 | 0.0001 | 0.0197 | -0.0030 | 0.0002 | 0.0400 | 0.0034 | -0.0015 | 0.0671 |
| $E(e_{T+5\|T})$ | 0.0007 | 0.0024 | -0.4267 | -0.0039 | -0.0004 | -0.7742 | 0.0054 | 0.0008 | -0.7880 |
| MSFE$(h = 1)$ | 1.0743 | 1.7619 | 4.6045 | 1.0731 | 1.7634 | 4.5959 | 1.0736 | 1.7577 | 4.6250 |
| MSFE$(h = 2)$ | 1.1215 | 1.8391 | 4.8475 | 1.4135 | 2.3378 | 6.2751 | 1.9030 | 3.1478 | 8.5704 |
| MSFE$(h = 5)$ | 1.2388 | 2.0383 | 5.6313 | 1.6445 | 2.7429 | 7.7386 | 2.3256 | 3.8663 | 11.0489 |
| $E(e^2_{T+1\|T})$ | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 |
| $E(e^2_{T+2\|T})$ | 1.1227 | 1.1227 | 1.1227 | 1.4207 | 1.4207 | 1.4207 | 1.9107 | 1.9107 | 1.9107 |
| $E(e^2_{T+5\|T})$ | 1.1219 | 1.1219 | 1.1219 | 1.4856 | 1.4856 | 1.4856 | 2.0914 | 2.0914 | 2.0914 |
| $E(\tilde{e}^2_{T+1\|T})$ | 1.0062 | 1.6783 | 4.6514 | 1.0053 | 1.6783 | 4.6924 | 1.0056 | 1.6761 | 4.6911 |
| $E(\tilde{e}^2_{T+2\|T})$ | 1.0974 | 1.8310 | 5.0852 | 1.3629 | 2.2841 | 6.4133 | 1.8133 | 3.0388 | 8.5376 |
| $E(\tilde{e}^2_{T+5\|T})$ | 1.1397 | 1.9033 | 5.5004 | 1.5086 | 2.5371 | 7.7570 | 2.1270 | 3.5647 | 10.9474 |
| MSFE-HM$(h = 1)$ | 1.1875 | 1.9788 | 5.5370 | 1.5569 | 2.6044 | 7.2956 | 2.1839 | 3.6395 | 10.1825 |
| MSFE-HM$(h = 2)$ | 1.1342 | 1.8892 | 5.2768 | 1.5195 | 2.5439 | 7.0934 | 2.1617 | 3.6017 | 10.0250 |
| MSFE-HM$(h = 5)$ | 0.9082 | 1.5146 | 4.2066 | 1.2393 | 2.0805 | 5.7586 | 1.7848 | 2.9662 | 8.2028 |
| MSFE-MD$(h = 1)$ | 1.1205 | 1.8655 | 5.2163 | 1.4774 | 2.4736 | 6.9162 | 2.0797 | 3.4711 | 9.6762 |
| MSFE-MD$(h = 2)$ | 1.0458 | 1.7405 | 4.8781 | 1.3781 | 2.3089 | 6.4678 | 1.9410 | 3.2403 | 9.0397 |
| MSFE-MD$(h = 5)$ | 0.8210 | 1.3683 | 3.8246 | 1.0832 | 1.8176 | 5.0766 | 1.5261 | 2.5412 | 7.0786 |

Table 1.3: No Intercept Model, T=30, $\rho = 0.8$

(a) Estimator Bias and MSE

| $\rho$=0.8 corr$(u,v)$=0.5 | bivN | $\beta$=0.3 bivT | bivlogN | bivN | $\beta$=0.6 bivT | bivlogN | bivN | $\beta$=0.9 bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\rho}$ | 0.7516 | 0.7522 | 0.7731 | 0.7506 | 0.7527 | 0.7735 | 0.7503 | 0.7524 | 0.7731 |
| $\tilde{\rho}$, $O(T^{-1})$ | 0.8050 | 0.8055 | 0.8265 | 0.8039 | 0.8061 | 0.8269 | 0.8036 | 0.8058 | 0.8264 |
| $\ddot{\rho}$, $O(T^{-1})$ | 0.8017 | 0.8023 | 0.8247 | 0.8006 | 0.8029 | 0.8251 | 0.8003 | 0.8026 | 0.8247 |
| $\tilde{\rho}$, $O(T^{-2})$ | 0.8014 | 0.7913 | 0.8229 | 0.8003 | 0.7918 | 0.8233 | 0.8001 | 0.7915 | 0.8229 |
| $\ddot{\rho}$, $O(T^{-2})$ | 0.7978 | 0.7933 | 0.7805 | 0.7966 | 0.7939 | 0.7810 | 0.7964 | 0.7935 | 0.7806 |
| $\hat{\beta}$ | 0.2754 | 0.2760 | 0.2838 | 0.5747 | 0.5766 | 0.5842 | 0.8746 | 0.8769 | 0.8839 |
| $\tilde{\beta}$, $O(T^{-1})$ | 0.3021 | 0.3027 | 0.3105 | 0.6014 | 0.6033 | 0.6109 | 0.9013 | 0.9036 | 0.9105 |
| $\ddot{\beta}$, $O(T^{-1})$ | 0.3005 | 0.3011 | 0.3084 | 0.5997 | 0.6016 | 0.6088 | 0.8996 | 0.9020 | 0.9085 |
| $\check{\beta}$, $O(T^{-1})$ | 0.3022 | 0.3027 | 0.3100 | 0.6014 | 0.6033 | 0.6104 | 0.9013 | 0.9037 | 0.9101 |
| $\tilde{\beta}$, $O(T^{-2})$ | 0.3003 | 0.2956 | 0.3087 | 0.5996 | 0.5961 | 0.6091 | 0.8995 | 0.8965 | 0.9088 |
| $\ddot{\beta}$, $O(T^{-2})$ | 0.2985 | 0.2966 | 0.2901 | 0.5978 | 0.5972 | 0.5906 | 0.8976 | 0.8975 | 0.8902 |
| $\check{\beta}$, $O(T^{-2})$ | 0.3001 | 0.2980 | 0.2907 | 0.5993 | 0.5986 | 0.5912 | 0.8992 | 0.8989 | 0.8908 |
| $M(\hat{\beta})$, | 0.0179 | 0.0182 | 0.0236 | 0.0178 | 0.0182 | 0.0234 | 0.0178 | 0.0182 | 0.0240 |
| $M(\tilde{\beta})$, $O(T^{-1})$ | 0.0120 | 0.0120 | 0.0120 | 0.0120 | 0.0120 | 0.0120 | 0.0120 | 0.0120 | 0.0120 |
| $M(\ddot{\beta})$, $O(T^{-1})$ | 0.0148 | 0.0154 | 0.0204 | 0.0149 | 0.0153 | 0.0202 | 0.0148 | 0.0154 | 0.0205 |
| $M(\check{\beta})$, $O(T^{-1})$ | 0.0120 | 0.0125 | 0.0153 | 0.0121 | 0.0124 | 0.0150 | 0.0120 | 0.0125 | 0.0153 |
| $M(\tilde{\beta})$, $O(T^{-2})$ | 0.0148 | 0.0172 | 0.0169 | 0.0148 | 0.0172 | 0.0169 | 0.0148 | 0.0172 | 0.0169 |
| $M(\ddot{\beta})$, $O(T^{-2})$ | 0.0177 | 0.0184 | 0.0243 | 0.0177 | 0.0184 | 0.0240 | 0.0177 | 0.0184 | 0.0243 |
| $M(\check{\beta})$, $O(T^{-2})$ | 0.0151 | 0.0157 | 0.0194 | 0.0151 | 0.0157 | 0.0192 | 0.0151 | 0.0157 | 0.0195 |

(b) Forecast Bias and MSFE

| $\rho$=0.8 corr$(u,v)$=0.5 | bivN | $\beta$=0.3 bivT | bivlogN | bivN | $\beta$=0.6 bivT | bivlogN | bivN | $\beta$=0.9 bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $E(e_{T+1\|T})$ | 0.0001 | 0.0010 | 0.0418 | -0.0007 | -0.0007 | 0.0382 | -0.0005 | -0.0007 | 0.0413 |
| $E(e_{T+2\|T})$ | 0.0001 | 0.0012 | 0.0563 | -0.0007 | -0.0012 | 0.0860 | -0.0004 | -0.0012 | 0.1269 |
| $E(e_{T+5\|T})$ | 0.0000 | 0.0008 | 0.1913 | -0.0016 | -0.0060 | 6.0454 | 0.0003 | -0.0006 | 0.0303 |
| MSFE$(h=1)$ | 1.0789 | 1.7709 | 4.5374 | 1.0826 | 1.7681 | 4.4930 | 1.0787 | 1.7725 | 4.5314 |
| MSFE$(h=2)$ | 1.1545 | 1.9027 | 4.8789 | 1.4656 | 2.4108 | 6.2803 | 1.9616 | 3.2412 | 8.5900 |
| MSFE$(h=5)$ | 1.4166 | 2.3737 | 6.2379 | 2.2324 | 3.7691 | 9.9887 | 3.5814 | 6.0379 | 16.1641 |
| $E(e^2_{T+1\|T})$ | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 | 1.0667 |
| $E(e^2_{T+2\|T})$ | 1.1547 | 1.1547 | 1.1547 | 1.4587 | 1.4587 | 1.4587 | 1.9547 | 1.9547 | 1.9547 |
| $E(e^2_{T+5\|T})$ | 1.2612 | 1.2612 | 1.2612 | 1.9776 | 1.9776 | 1.9776 | 3.1605 | 3.1605 | 3.1605 |
| $E(\tilde{e}^2_{T+1\|T})$ | 1.0146 | 1.6936 | 4.7301 | 1.0172 | 1.6868 | 4.6291 | 1.0134 | 1.6937 | 4.7097 |
| $E(\tilde{e}^2_{T+2\|T})$ | 1.1026 | 1.8439 | 5.1631 | 1.3705 | 2.2841 | 6.3375 | 1.8139 | 3.0429 | 8.6012 |
| $E(\tilde{e}^2_{T+5\|T})$ | 1.2482 | 2.0969 | 5.9210 | 1.8906 | 3.1729 | 9.1377 | 2.9441 | 4.9990 | 14.6364 |
| MSFE-HM$(h=1)$ | 1.2614 | 2.1072 | 5.9590 | 1.9068 | 3.1815 | 8.8974 | 2.9989 | 5.0183 | 14.1902 |
| MSFE-HM$(h=2)$ | 1.2179 | 2.0363 | 5.7363 | 1.9091 | 3.1864 | 8.8626 | 3.0643 | 5.1317 | 14.4321 |
| MSFE-HM$(h=5)$ | 1.0203 | 1.7039 | 4.7775 | 1.6930 | 2.8345 | 7.8431 | 2.8076 | 4.7077 | 13.1128 |
| MSFE-MD$(h=1)$ | 1.2493 | 2.0842 | 5.8751 | 1.9971 | 3.3277 | 9.2555 | 3.2389 | 5.4237 | 15.2592 |
| MSFE-MD$(h=2)$ | 1.1649 | 1.9451 | 5.4835 | 1.8643 | 3.1046 | 8.6307 | 3.0235 | 5.0630 | 14.2600 |
| MSFE-MD$(h=5)$ | 0.9148 | 1.5268 | 4.3121 | 1.4627 | 2.4374 | 6.7927 | 2.3754 | 3.9734 | 11.2087 |

Table 1.4: No Intercept Model, T=100, $\rho = 0.5$

(a) Estimator Bias and MSE

| $\rho$=0.5<br>corr$(u, v)$=0.5 | bivN | $\beta$=0.3<br>bivT | bivlogN | bivN | $\beta$=0.6<br>bivT | bivlogN | bivN | $\beta$=0.9<br>bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\rho}$ | 0.4900 | 0.4901 | 0.4960 | 0.4901 | 0.4906 | 0.4958 | 0.4900 | 0.4906 | 0.4955 |
| $\tilde{\rho}$, $O(T^{-1})$ | 0.5000 | 0.5001 | 0.5060 | 0.5001 | 0.5006 | 0.5058 | 0.5000 | 0.5006 | 0.5055 |
| $\ddot{\rho}$, $O(T^{-1})$ | 0.4998 | 0.4999 | 0.5059 | 0.4999 | 0.5005 | 0.5057 | 0.4998 | 0.5004 | 0.5055 |
| $\tilde{\rho}$, $O(T^{-2})$ | 0.4998 | 0.4993 | 0.5058 | 0.4999 | 0.4998 | 0.5056 | 0.4998 | 0.4998 | 0.5053 |
| $\ddot{\rho}$, $O(T^{-2})$ | 0.4996 | 0.4993 | 0.4998 | 0.4997 | 0.4998 | 0.4996 | 0.4996 | 0.4998 | 0.4993 |
| $\hat{\beta}$ | 0.2953 | 0.2954 | 0.2973 | 0.5949 | 0.5955 | 0.5975 | 0.8949 | 0.8959 | 0.8975 |
| $\tilde{\beta}$, $O(T^{-1})$ | 0.3003 | 0.3004 | 0.3023 | 0.5999 | 0.6005 | 0.6025 | 0.8999 | 0.9009 | 0.9025 |
| $\ddot{\beta}$, $O(T^{-1})$ | 0.3002 | 0.3003 | 0.3016 | 0.5998 | 0.6004 | 0.6018 | 0.8998 | 0.9008 | 0.9017 |
| $\check{\beta}$, $O(T^{-1})$ | 0.3003 | 0.3004 | 0.3016 | 0.5999 | 0.6005 | 0.6018 | 0.8999 | 0.9009 | 0.9018 |
| $\tilde{\beta}$, $O(T^{-2})$ | 0.3002 | 0.3000 | 0.3022 | 0.5998 | 0.6001 | 0.6024 | 0.8998 | 0.9005 | 0.9024 |
| $\ddot{\beta}$, $O(T^{-2})$ | 0.3001 | 0.3000 | 0.2993 | 0.5997 | 0.6000 | 0.5994 | 0.8997 | 0.9005 | 0.8994 |
| $\check{\beta}$, $O(T^{-2})$ | 0.3002 | 0.3001 | 0.2993 | 0.5998 | 0.6001 | 0.5995 | 0.8998 | 0.9006 | 0.8995 |
| $M(\hat{\beta})$, | 0.0078 | 0.0078 | 0.0094 | 0.0078 | 0.0078 | 0.0094 | 0.0079 | 0.0079 | 0.0094 |
| $M(\tilde{\beta})$, $O(T^{-1})$ | 0.0075 | 0.0075 | 0.0075 | 0.0075 | 0.0075 | 0.0075 | 0.0075 | 0.0075 | 0.0075 |
| $M(\ddot{\beta})$, $O(T^{-1})$ | 0.0076 | 0.0078 | 0.0098 | 0.0076 | 0.0078 | 0.0098 | 0.0076 | 0.0078 | 0.0098 |
| $M(\check{\beta})$, $O(T^{-1})$ | 0.0075 | 0.0077 | 0.0097 | 0.0075 | 0.0077 | 0.0097 | 0.0075 | 0.0077 | 0.0097 |
| $M(\tilde{\beta})$, $O(T^{-2})$ | 0.0077 | 0.0082 | 0.0081 | 0.0077 | 0.0082 | 0.0081 | 0.0077 | 0.0082 | 0.0081 |
| $M(\ddot{\beta})$, $O(T^{-2})$ | 0.0078 | 0.0080 | 0.0104 | 0.0078 | 0.0080 | 0.0104 | 0.0078 | 0.0080 | 0.0104 |
| $M(\check{\beta})$, $O(T^{-2})$ | 0.0077 | 0.0079 | 0.0102 | 0.0077 | 0.0079 | 0.0102 | 0.0077 | 0.0079 | 0.0102 |

(b) Forecast Bias and MSFE

| $\rho$=0.5<br>corr$(u, v)$=0.5 | bivN | $\beta$=0.3<br>bivT | bivlogN | bivN | $\beta$=0.6<br>bivT | bivlogN | bivN | $\beta$=0.9<br>bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $E(e_{T+1\|T})$ | -0.0004 | 0.0005 | 0.0044 | 0.0003 | -0.0005 | 0.0030 | -0.0002 | -0.0004 | 0.0044 |
| $E(e_{T+2\|T})$ | -0.0007 | 0.0006 | 0.0051 | 0.0002 | -0.0014 | 0.0065 | 0.0001 | -0.0007 | 0.0120 |
| $E(e_{T+5\|T})$ | -0.0010 | 0.0007 | -0.0108 | 0.0001 | -0.0022 | -0.0286 | 0.0005 | -0.0005 | -0.0078 |
| MSFE$(h = 1)$ | 1.0199 | 1.6765 | 4.4635 | 1.0192 | 1.6806 | 4.4471 | 1.0201 | 1.6804 | 4.4459 |
| MSFE$(h = 2)$ | 1.0983 | 1.8067 | 4.8297 | 1.3746 | 2.2705 | 6.1130 | 1.8352 | 3.0380 | 8.2061 |
| MSFE$(h = 5)$ | 1.1441 | 1.8821 | 5.0502 | 1.5100 | 2.4951 | 6.7592 | 2.1214 | 3.5145 | 9.5435 |
| $E(e^2_{T+1\|T})$ | 1.0200 | 1.0200 | 1.0200 | 1.0200 | 1.0200 | 1.0200 | 1.0200 | 1.0200 | 1.0200 |
| $E(e^2_{T+2\|T})$ | 1.0998 | 1.0998 | 1.0998 | 1.3782 | 1.3782 | 1.3782 | 1.8402 | 1.8402 | 1.8402 |
| $E(e^2_{T+5\|T})$ | 1.1202 | 1.1202 | 1.1202 | 1.4804 | 1.4804 | 1.4804 | 2.0805 | 2.0805 | 2.0805 |
| $E(\tilde{e}^2_{T+1\|T})$ | 1.0007 | 1.6610 | 4.6527 | 1.0006 | 1.6663 | 4.6576 | 1.0013 | 1.6658 | 4.6213 |
| $E(\tilde{e}^2_{T+2\|T})$ | 1.0887 | 1.8079 | 5.0681 | 1.3557 | 2.2596 | 6.3362 | 1.8039 | 3.0104 | 8.3985 |
| $E(\tilde{e}^2_{T+5\|T})$ | 1.1206 | 1.8605 | 5.2138 | 1.4783 | 2.4630 | 6.8919 | 2.0765 | 3.4644 | 9.6304 |
| MSFE-HM$(h = 1)$ | 1.1417 | 1.8973 | 5.3312 | 1.5066 | 2.5163 | 7.0515 | 2.1165 | 3.5328 | 9.8437 |
| MSFE-HM$(h = 2)$ | 1.1271 | 1.8734 | 5.2570 | 1.4976 | 2.5010 | 6.9983 | 2.1128 | 3.5272 | 9.8087 |
| MSFE-HM$(h = 5)$ | 1.0651 | 1.7704 | 4.9581 | 1.4233 | 2.3774 | 6.6480 | 2.0154 | 3.3657 | 9.3499 |
| MSFE-MD$(h = 1)$ | 1.1200 | 1.8606 | 5.2291 | 1.4786 | 2.4699 | 6.9205 | 2.0785 | 3.4707 | 9.6563 |
| MSFE-MD$(h = 2)$ | 1.0976 | 1.8236 | 5.1249 | 1.4490 | 2.4203 | 6.7846 | 2.0369 | 3.4017 | 9.4648 |
| MSFE-MD$(h = 5)$ | 1.0306 | 1.7125 | 4.8080 | 1.3604 | 2.2730 | 6.3781 | 1.9120 | 3.1948 | 8.8966 |

Table 1.5: Model with Intercept, T=30, $\rho = 0.2$

(a) Estimator Bias and MSE

| $\rho$=0.2, $\alpha = 0.5$ | | $\beta$=0.3 | | | $\beta$=0.6 | | | $\beta$=0.9 | |
|---|---|---|---|---|---|---|---|---|---|
| corr$(u, v)$=0.5, $\theta$=0.4 | bivN | bivT | bivlogN | bivN | bivT | bivlogN | bivN | bivT | bivlogN |
| $\hat{\rho}$ | 0.1466 | 0.1477 | 0.1485 | 0.1463 | 0.1474 | 0.1480 | 0.1462 | 0.1466 | 0.1489 |
| $\tilde{\rho}$, $O(T^{-2})$ | 0.2010 | 0.1994 | 0.2029 | 0.2007 | 0.1991 | 0.2023 | 0.2006 | 0.1984 | 0.2032 |
| $\ddot{\rho}$, $O(T^{-2})$ | 0.1960 | 0.1967 | 0.1953 | 0.1957 | 0.1964 | 0.1947 | 0.1956 | 0.1956 | 0.1956 |
| $\hat{\beta}$ | 0.2734 | 0.2736 | 0.2752 | 0.5726 | 0.5735 | 0.5743 | 0.8730 | 0.8732 | 0.8735 |
| $\tilde{\beta}$, $O(T^{-2})$ | 0.3006 | 0.2994 | 0.3024 | 0.5998 | 0.5994 | 0.6015 | 0.9002 | 0.8990 | 0.9007 |
| $\ddot{\beta}$, $O(T^{-2})$ | 0.2982 | 0.2981 | 0.2976 | 0.5973 | 0.5980 | 0.5967 | 0.8977 | 0.8976 | 0.8960 |
| $\check{\beta}$, $O(T^{-2})$ | 0.3007 | 0.3005 | 0.2996 | 0.5999 | 0.6004 | 0.5987 | 0.9002 | 0.9001 | 0.8980 |
| $\hat{\theta}$ | 0.4246 | 0.4252 | 0.4236 | 0.4261 | 0.4258 | 0.4254 | 0.4256 | 0.4248 | 0.4240 |
| $\tilde{\theta}$, $O(T^{-2})$ | 0.3974 | 0.3993 | 0.3964 | 0.3989 | 0.4000 | 0.3982 | 0.3984 | 0.3990 | 0.3968 |
| $\ddot{\theta}$, $O(T^{-2})$ | 0.3999 | 0.4006 | 0.3648 | 0.4012 | 0.4013 | 0.3976 | 0.4008 | 0.4003 | 0.3642 |
| $\hat{\alpha}$ | 0.5123 | 0.5125 | 0.5087 | 0.5123 | 0.5135 | 0.5112 | 0.5127 | 0.5125 | 0.5122 |
| $\tilde{\alpha}$, $O(T^{-2})$ | 0.4987 | 0.4996 | 0.4951 | 0.4987 | 0.5006 | 0.4976 | 0.4991 | 0.4996 | 0.4986 |
| $\ddot{\alpha}$, $O(T^{-2})$ | 0.5003 | 0.4928 | 0.3476 | 0.5002 | 0.4939 | 0.4351 | 0.5007 | 0.4928 | 0.3816 |
| $M(\hat{\beta})$, | 0.0367 | 0.0376 | 0.0571 | 0.0369 | 0.0372 | 0.0580 | 0.0367 | 0.0378 | 0.0572 |
| $M(\tilde{\beta})$, $O(T^{-1})$ | 0.0320 | 0.0320 | 0.0320 | 0.0320 | 0.0320 | 0.0320 | 0.0320 | 0.0320 | 0.0320 |
| $M(\ddot{\beta})$, $O(T^{-1})$ | 0.0332 | 0.0346 | 0.0560 | 0.0333 | 0.0347 | 0.0568 | 0.0333 | 0.0347 | 0.0565 |
| $M(\check{\beta})$, $O(T^{-1})$ | 0.0324 | 0.0338 | 0.0547 | 0.0325 | 0.0338 | 0.0555 | 0.0325 | 0.0338 | 0.0552 |
| $M(\tilde{\beta})$, $O(T^{-2})$ | 0.0328 | 0.0312 | 0.0328 | 0.0328 | 0.0312 | 0.0328 | 0.0328 | 0.0312 | 0.0328 |
| $M(\ddot{\beta})$, $O(T^{-2})$ | 0.0341 | 0.0350 | 0.0543 | 0.0341 | 0.0351 | 0.0551 | 0.0342 | 0.0351 | 0.0548 |
| $M(\check{\beta})$, $O(T^{-2})$ | 0.0336 | 0.0345 | 0.0532 | 0.0336 | 0.0345 | 0.0540 | 0.0337 | 0.0346 | 0.0537 |

(b) Forecast Bias and MSFE

| $\rho = 0.2$, $\alpha = 0.5$ | | $\beta = 0.3$ | | | $\beta$=0.6 | | | $\beta$=0.9 | |
|---|---|---|---|---|---|---|---|---|---|
| corr$(u, v)$=0.5, $\theta$=0.4 | bivN | bivT | bivlogN | bivN | bivT | bivlogN | bivN | bivT | bivlogN |
| $E(e_{T+1\|T})$ | 0.0000 | 0.0014 | 0.0006 | -0.0003 | -0.0002 | 0.0003 | 0.0007 | 0.0010 | 0.0038 |
| $E(e_{T+2\|T})$ | -0.0005 | 0.0015 | -0.0346 | 0.0006 | -0.0002 | -0.0372 | 0.0023 | 0.0006 | -0.0348 |
| $E(e_{T+5\|T})$ | -0.0009 | 0.0014 | -0.0489 | 0.0019 | -0.0010 | -0.0782 | 0.0035 | 0.0000 | -0.1033 |
| MSFE$(h = 1)$ | 1.1472 | 1.8826 | 4.9622 | 1.1495 | 1.8814 | 5.0089 | 1.1507 | 1.8781 | 5.0042 |
| MSFE$(h = 2)$ | 1.1902 | 1.9500 | 5.2428 | 1.5193 | 2.4924 | 6.9226 | 2.0567 | 3.3634 | 9.3568 |
| MSFE$(h = 5)$ | 1.3033 | 2.1293 | 6.4199 | 1.6695 | 2.7335 | 8.6371 | 2.2609 | 3.6981 | 11.8418 |
| $E(e^2_{T+1\|T})$ | 1.0827 | 1.0827 | 1.0827 | 1.0827 | 1.0827 | 1.0827 | 1.0827 | 1.0827 | 1.0827 |
| $E(e^2_{T+2\|T})$ | 1.1187 | 1.1187 | 1.1187 | 1.4107 | 1.4107 | 1.4107 | 1.8947 | 1.8947 | 1.8947 |
| $E(e^2_{T+5\|T})$ | 1.1098 | 1.1098 | 1.1098 | 1.3910 | 1.3910 | 1.3910 | 1.8598 | 1.8598 | 1.8598 |
| $E(\tilde{e}^2_{T+1\|T})$ | 1.0110 | 1.6993 | 4.8650 | 1.0114 | 1.7002 | 4.9930 | 1.0130 | 1.6965 | 4.9328 |
| $E(\tilde{e}^2_{T+2\|T})$ | 1.0882 | 1.8266 | 5.2496 | 1.3313 | 2.2332 | 6.5778 | 1.7575 | 2.9334 | 8.5196 |
| $E(\tilde{e}^2_{T+5\|T})$ | 1.0882 | 1.8296 | 6.1243 | 1.3375 | 2.2429 | 8.2741 | 1.7722 | 2.9581 | 13.6071 |
| MSFE-HM$(h = 1)$ | 1.1631 | 1.9445 | 5.3464 | 1.4611 | 2.4384 | 6.8654 | 1.9592 | 3.2570 | 9.1082 |
| MSFE-HM$(h = 2)$ | 1.1065 | 1.8494 | 5.0684 | 1.4105 | 2.3530 | 6.5809 | 1.9059 | 3.1672 | 8.8054 |
| MSFE-HM$(h = 5)$ | 0.8737 | 1.4612 | 3.9854 | 1.1181 | 1.8634 | 5.1956 | 1.5127 | 2.5127 | 6.9817 |
| MSFE-MD$(h = 1)$ | 1.0933 | 1.8266 | 5.0190 | 1.3750 | 2.2918 | 6.4500 | 1.8452 | 3.0662 | 8.5734 |
| MSFE-MD$(h = 2)$ | 1.0202 | 1.7045 | 4.6860 | 1.2833 | 2.1386 | 6.0187 | 1.7216 | 2.8618 | 8.0111 |
| MSFE-MD$(h = 5)$ | 0.8017 | 1.3404 | 3.6678 | 1.0089 | 1.6792 | 4.7185 | 1.3528 | 2.2491 | 6.3034 |

Table 1.6: Model with Intercept, T=30, $\rho = 0.5$

(a) Estimator Bias and MSE

| $\rho$=0.5, $\alpha = 0.5$ | | $\beta$=0.3 | | | $\beta$=0.6 | | | $\beta$=0.9 | |
|---|---|---|---|---|---|---|---|---|---|
| corr$(u, v)$=0.5, $\theta$=0.4 | bivN | bivT | bivlogN | bivN | bivT | bivlogN | bivN | bivT | bivlogN |
| $\hat{\rho}$ | 0.4158 | 0.4171 | 0.4232 | 0.4149 | 0.4168 | 0.4229 | 0.4154 | 0.4168 | 0.4228 |
| $\tilde{\rho}$, $O(T^{-2})$ | 0.5030 | 0.4976 | 0.5104 | 0.5021 | 0.4974 | 0.5102 | 0.5027 | 0.4973 | 0.5100 |
| $\ddot{\rho}$, $O(T^{-2})$ | 0.4945 | 0.4944 | 0.4934 | 0.4935 | 0.4940 | 0.4931 | 0.4941 | 0.4940 | 0.4928 |
| $\hat{\beta}$ | 0.2576 | 0.2589 | 0.2616 | 0.5577 | 0.5582 | 0.5625 | 0.8574 | 0.8580 | 0.8614 |
| $\tilde{\beta}$, $O(T^{-2})$ | 0.3012 | 0.2992 | 0.3052 | 0.6014 | 0.5985 | 0.6061 | 0.9010 | 0.8982 | 0.9050 |
| $\ddot{\beta}$, $O(T^{-2})$ | 0.2969 | 0.2975 | 0.2958 | 0.5970 | 0.5969 | 0.5963 | 0.8966 | 0.8966 | 0.8953 |
| $\check{\beta}$, $O(T^{-2})$ | 0.3019 | 0.3024 | 0.2993 | 0.6022 | 0.6022 | 0.6002 | 0.9017 | 0.9014 | 0.8990 |
| $\hat{\theta}$ | 0.4623 | 0.4602 | 0.4550 | 0.4618 | 0.4612 | 0.4577 | 0.4630 | 0.4605 | 0.4549 |
| $\tilde{\theta}$, $O(T^{-2})$ | 0.3925 | 0.3957 | 0.3852 | 0.3920 | 0.3967 | 0.3879 | 0.3932 | 0.3960 | 0.3851 |
| $\ddot{\theta}$, $O(T^{-2})$ | 0.3992 | 0.3977 | 0.2696 | 0.3988 | 0.3999 | 0.1837 | 0.3999 | 0.3992 | 0.3313 |
| $\hat{\alpha}$ | 0.5317 | 0.5311 | 0.5230 | 0.5305 | 0.5305 | 0.5232 | 0.5305 | 0.5293 | 0.5227 |
| $\tilde{\alpha}$, $O(T^{-2})$ | 0.4968 | 0.4989 | 0.4881 | 0.4956 | 0.4983 | 0.4883 | 0.4956 | 0.4971 | 0.4879 |
| $\ddot{\alpha}$, $O(T^{-2})$ | 0.5012 | 0.4812 | -0.0709 | 0.5000 | 0.4820 | 0.0465 | 0.5000 | 0.4811 | -0.0908 |
| $M(\hat{\beta})$, | 0.0329 | 0.0334 | 0.0502 | 0.0326 | 0.0334 | 0.0494 | 0.0328 | 0.0335 | 0.0485 |
| $M(\tilde{\beta})$, $O(T^{-1})$ | 0.0250 | 0.0250 | 0.0250 | 0.0250 | 0.0250 | 0.0250 | 0.0250 | 0.0250 | 0.0250 |
| $M(\ddot{\beta})$, $O(T^{-1})$ | 0.0283 | 0.0293 | 0.0480 | 0.0283 | 0.0294 | 0.0474 | 0.0283 | 0.0294 | 0.0473 |
| $M(\check{\beta})$, $O(T^{-1})$ | 0.0257 | 0.0266 | 0.0436 | 0.0257 | 0.0267 | 0.0431 | 0.0257 | 0.0267 | 0.0430 |
| $M(\tilde{\beta})$, $O(T^{-2})$ | 0.0280 | 0.0267 | 0.0280 | 0.0280 | 0.0267 | 0.0280 | 0.0280 | 0.0267 | 0.0280 |
| $M(\ddot{\beta})$, $O(T^{-2})$ | 0.0310 | 0.0316 | 0.0481 | 0.0309 | 0.0318 | 0.0475 | 0.0309 | 0.0318 | 0.0473 |
| $M(\check{\beta})$, $O(T^{-2})$ | 0.0291 | 0.0298 | 0.0449 | 0.0291 | 0.0299 | 0.0443 | 0.0291 | 0.0299 | 0.0442 |

(b) Forecast Bias and MSFE

| $\rho$=0.5, $\alpha = 0.5$ | | $\beta$=0.3 | | | $\beta$=0.6 | | | $\beta$=0.9 | |
|---|---|---|---|---|---|---|---|---|---|
| corr$(u, v)$=0.5, $\theta$=0.4 | bivN | bivT | bivlogN | bivN | bivT | bivlogN | bivN | bivT | bivlogN |
| $E(e_{T+1\|T})$ | 0.0003 | 0.0014 | 0.0006 | -0.0005 | 0.0005 | 0.0069 | -0.0004 | -0.0012 | 0.0019 |
| $E(e_{T+2\|T})$ | 0.0000 | 0.0016 | -0.0350 | 0.0000 | 0.0000 | -0.0338 | -0.0004 | -0.0013 | -0.0394 |
| $E(e_{T+5\|T})$ | -0.0007 | 0.0028 | -0.0715 | -0.0007 | -0.0016 | -0.1087 | -0.0004 | -0.0034 | -0.1476 |
| MSFE$(h = 1)$ | 1.1575 | 1.8980 | 4.9829 | 1.1583 | 1.8965 | 5.0790 | 1.1579 | 1.9052 | 5.0050 |
| MSFE$(h = 2)$ | 1.2120 | 1.9830 | 5.3490 | 1.5498 | 2.5416 | 7.1504 | 2.0922 | 3.4573 | 9.6511 |
| MSFE$(h = 5)$ | 1.3723 | 2.2557 | 7.5395 | 1.8823 | 3.1145 | 11.4793 | 2.7044 | 4.5038 | 16.4595 |
| $E(e^2_{T+1\|T})$ | 1.0987 | 1.0987 | 1.0987 | 1.0987 | 1.0987 | 1.0987 | 1.0987 | 1.0987 | 1.0987 |
| $E(e^2_{T+2\|T})$ | 1.1547 | 1.1547 | 1.1547 | 1.4527 | 1.4527 | 1.4527 | 1.9427 | 1.9427 | 1.9427 |
| $E(e^2_{T+5\|T})$ | 1.1539 | 1.1539 | 1.1539 | 1.5176 | 1.5176 | 1.5176 | 2.1234 | 2.1234 | 2.1234 |
| $E(\tilde{e}^2_{T+1\|T})$ | 1.0333 | 1.7457 | 5.3429 | 1.0348 | 1.7440 | 5.3944 | 1.0340 | 1.7562 | 5.2889 |
| $E(\tilde{e}^2_{T+2\|T})$ | 1.1016 | 1.8594 | 5.7434 | 1.3326 | 2.2449 | 6.9462 | 1.7363 | 2.9402 | 8.7805 |
| $E(\tilde{e}^2_{T+5\|T})$ | 1.1218 | 1.8969 | 13.3469 | 1.4078 | 2.3807 | 14.7301 | 1.9055 | 3.2317 | 15.7844 |
| MSFE-HM$(h = 1)$ | 1.1864 | 1.9791 | 5.5844 | 1.5565 | 2.5994 | 7.3796 | 2.1773 | 3.6536 | 10.1836 |
| MSFE-HM$(h = 2)$ | 1.1337 | 1.8911 | 5.3251 | 1.5204 | 2.5364 | 7.1618 | 2.1552 | 3.6155 | 10.0276 |
| MSFE-HM$(h = 5)$ | 0.9085 | 1.5140 | 4.2639 | 1.2402 | 2.0708 | 5.8251 | 1.7796 | 2.9826 | 8.2403 |
| MSFE-MD$(h = 1)$ | 1.1188 | 1.8677 | 5.2672 | 1.4793 | 2.4656 | 6.9978 | 2.0740 | 3.4821 | 9.7019 |
| MSFE-MD$(h = 2)$ | 1.0443 | 1.7436 | 4.9297 | 1.3813 | 2.2994 | 6.5304 | 1.9362 | 3.2479 | 9.0688 |
| MSFE-MD$(h = 5)$ | 0.8206 | 1.3685 | 3.8838 | 1.0854 | 1.8060 | 5.1408 | 1.5231 | 2.5497 | 7.1407 |

## Table 1.7: Model with Intercept, T=30, $\rho = 0.8$

(a) Estimator Bias and MSE

| $\rho$=0.8, $\alpha = 0.5$ | | $\beta$=0.3 | | | $\beta$=0.6 | | | $\beta$=0.9 | |
|---|---|---|---|---|---|---|---|---|---|
| corr$(u,v)$=0.5, $\theta$=0.4 | bivN | bivT | bivlogN | bivN | bivT | bivlogN | bivN | bivT | bivlogN |
| $\hat{\rho}$ | 0.6859 | 0.6837 | 0.6889 | 0.6855 | 0.6828 | 0.6887 | 0.6868 | 0.6832 | 0.6892 |
| $\tilde{\rho}$, $O(T^{-2})$ | 0.8235 | 0.8106 | 0.8265 | 0.8231 | 0.8097 | 0.8262 | 0.8243 | 0.8101 | 0.8268 |
| $\ddot{\rho}$, $O(T^{-2})$ | 0.8034 | 0.7993 | 0.7842 | 0.8061 | 0.7989 | 0.7885 | 0.8071 | 0.7997 | 0.7954 |
| $\hat{\beta}$ | 0.2431 | 0.2416 | 0.2454 | 0.5423 | 0.5419 | 0.5447 | 0.8430 | 0.8413 | 0.8455 |
| $\tilde{\beta}$, $O(T^{-2})$ | 0.3119 | 0.3050 | 0.3142 | 0.6111 | 0.6053 | 0.6135 | 0.9118 | 0.9048 | 0.9142 |
| $\ddot{\beta}$, $O(T^{-2})$ | 0.3018 | 0.2993 | 0.2895 | 0.6025 | 0.6000 | 0.5928 | 0.9033 | 0.8997 | 0.8960 |
| $\check{\beta}$, $O(T^{-2})$ | 0.3348 | 0.2930 | 0.2980 | 0.5976 | 0.6191 | 0.6241 | 0.9326 | 0.8643 | 0.8917 |
| $\hat{\theta}$ | 0.5859 | 0.5905 | 0.5802 | 0.5882 | 0.5893 | 0.5827 | 0.5845 | 0.5897 | 0.5804 |
| $\tilde{\theta}$, $O(T^{-2})$ | 0.3108 | 0.3367 | 0.3051 | 0.3131 | 0.3355 | 0.3076 | 0.3094 | 0.3360 | 0.3053 |
| $\hat{\alpha}$ | 0.5926 | 0.5955 | 0.5752 | 0.5950 | 0.5930 | 0.5759 | 0.5919 | 0.5952 | 0.5737 |
| $\tilde{\alpha}$, $O(T^{-2})$ | 0.4550 | 0.4686 | 0.4376 | 0.4575 | 0.4661 | 0.4383 | 0.4543 | 0.4684 | 0.4361 |
| $M(\hat{\beta})$, | 0.0227 | 0.0248 | 0.0380 | 0.0230 | 0.0245 | 0.0383 | 0.0230 | 0.0248 | 0.0374 |
| $M(\tilde{\beta})$, $O(T^{-1})$ | 0.0120 | 0.0120 | 0.0120 | 0.0120 | 0.0120 | 0.0120 | 0.0120 | 0.0120 | 0.0120 |
| $M(\ddot{\beta})$, $O(T^{-1})$ | 0.0183 | 0.0192 | 0.0311 | 0.0183 | 0.0192 | 0.0311 | 0.0183 | 0.0192 | 0.0310 |
| $M(\check{\beta})$, $O(T^{-1})$ | 0.0129 | 0.0136 | 0.0220 | 0.0129 | 0.0136 | 0.0220 | 0.0128 | 0.0136 | 0.0219 |
| $M(\tilde{\beta})$, $O(T^{-2})$ | 0.0183 | 0.0177 | 0.0183 | 0.0183 | 0.0177 | 0.0183 | 0.0183 | 0.0177 | 0.0183 |
| $M(\ddot{\beta})$, $O(T^{-2})$ | 0.0240 | 0.0248 | 0.0361 | 0.0239 | 0.0248 | 0.0360 | 0.0239 | 0.0248 | 0.0360 |
| $M(\check{\beta})$, $O(T^{-2})$ | 0.0199 | 0.0208 | 0.0298 | 0.0199 | 0.0208 | 0.0297 | 0.0199 | 0.0208 | 0.0297 |

(b) Forecast Bias and MSFE

| $\rho$=0.8, $\alpha = 0.5$ | | $\beta$=0.3 | | | $\beta$=0.6 | | | $\beta$=0.9 | |
|---|---|---|---|---|---|---|---|---|---|
| corr$(u,v)$=0.5, $\theta$=0.4 | bivN | bivT | bivlogN | bivN | bivT | bivlogN | bivN | bivT | bivlogN |
| $E(e_{T+1\|T})$ | 0.0052 | 0.0026 | 0.0026 | 0.0029 | 0.0033 | 0.0033 | 0.0055 | 0.0050 | 0.0050 |
| $E(e_{T+2\|T})$ | 0.0069 | 0.0041 | -0.0273 | 0.0065 | 0.0065 | -0.0288 | 0.0139 | 0.0143 | -0.0343 |
| $E(e_{T+5\|T})$ | 0.0087 | 0.0055 | -0.0822 | 0.0097 | 0.0081 | -0.1345 | 0.0240 | 0.0271 | -0.2024 |
| MSFE($h = 1$) | 1.1859 | 1.9417 | 5.0671 | 1.1867 | 1.9419 | 5.0745 | 1.1846 | 1.9469 | 5.0750 |
| MSFE($h = 2$) | 1.2635 | 2.0672 | 5.5859 | 1.6202 | 2.6655 | 7.3297 | 2.1926 | 3.6166 | 10.0543 |
| MSFE($h = 5$) | 1.6078 | 2.6822 | 9.9191 | 2.6313 | 4.4078 | 16.3384 | 4.2753 | 7.1736 | 27.6880 |
| $E(e^2_{T+1\|T})$ | 1.1627 | 1.1627 | 1.1627 | 1.1627 | 1.1627 | 1.1627 | 1.1627 | 1.1627 | 1.1627 |
| $E(e^2_{T+2\|T})$ | 1.2507 | 1.2507 | 1.2507 | 1.5547 | 1.5547 | 1.5547 | 2.0507 | 2.0507 | 2.0507 |
| $E(e^2_{T+5\|T})$ | 1.3572 | 1.3572 | 1.3572 | 2.0736 | 2.0736 | 2.0736 | 3.2565 | 3.2565 | 3.2565 |
| $E(\tilde{e}^2_{T+1\|T})$ | 1.1850 | 2.0464 | 7.3057 | 1.1859 | 2.0455 | 7.3577 | 1.1864 | 2.0581 | 7.5011 |
| $E(\tilde{e}^2_{T+2\|T})$ | 1.2393 | 2.1401 | 7.6813 | 1.4494 | 2.4948 | 8.7840 | 1.8288 | 3.1428 | 10.7590 |
| $E(\tilde{e}^2_{T+5\|T})$ | 1.3046 | 2.2632 | 12.6103 | 1.7056 | 2.9492 | 14.2016 | 2.3975 | 4.1207 | 16.3624 |
| MSFE-HM($h = 1$) | 1.2678 | 2.1033 | 5.8885 | 1.9317 | 3.2028 | 9.0109 | 3.0545 | 5.0586 | 14.2095 |
| MSFE-HM($h = 2$) | 1.2248 | 2.0316 | 5.6820 | 1.9351 | 3.2090 | 8.9831 | 3.1234 | 5.1698 | 14.4830 |
| MSFE-HM($h = 5$) | 1.0264 | 1.7069 | 4.7497 | 1.7260 | 2.8574 | 7.9312 | 2.8731 | 4.7435 | 13.1630 |
| MSFE-MD($h = 1$) | 1.2488 | 2.0791 | 5.7851 | 2.0019 | 3.3336 | 9.3383 | 3.2575 | 5.4082 | 15.2708 |
| MSFE-MD($h = 2$) | 1.1652 | 1.9398 | 5.4117 | 1.8678 | 3.1097 | 8.7200 | 3.0396 | 5.0467 | 14.2778 |
| MSFE-MD($h = 5$) | 0.9150 | 1.5267 | 4.2692 | 1.4684 | 2.4443 | 6.8593 | 2.3851 | 3.9596 | 11.1759 |

Table 1.8: Model with Intercept, T=100, $\rho = 0.5$

(a) Estimator Bias and MSE

| $\rho$=0.5, $\alpha = 0.5$ corr$(u,v)$=0.5, $\theta$=0.4 | $\beta$=0.3 bivN | bivT | bivlogN | $\beta$=0.6 bivN | bivT | bivlogN | $\beta$=0.9 bivN | bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\rho}$ | 0.4745 | 0.4749 | 0.4767 | 0.4749 | 0.4756 | 0.4769 | 0.4747 | 0.4749 | 0.4768 |
| $\tilde{\rho}$, $O(T^{-2})$ | 0.4999 | 0.4996 | 0.5021 | 0.5003 | 0.5004 | 0.5023 | 0.5000 | 0.4997 | 0.5021 |
| $\ddot{\rho}$, $O(T^{-2})$ | 0.4991 | 0.4992 | 0.4990 | 0.4995 | 0.4999 | 0.4993 | 0.4993 | 0.4993 | 0.4991 |
| $\hat{\beta}$ | 0.2870 | 0.2875 | 0.2899 | 0.5872 | 0.5877 | 0.5904 | 0.8876 | 0.8873 | 0.8901 |
| $\tilde{\beta}$, $O(T^{-2})$ | 0.2997 | 0.2998 | 0.3026 | 0.5999 | 0.6001 | 0.6030 | 0.9002 | 0.8997 | 0.9027 |
| $\ddot{\beta}$, $O(T^{-2})$ | 0.2993 | 0.2996 | 0.2996 | 0.5995 | 0.5999 | 0.6000 | 0.8999 | 0.8995 | 0.8997 |
| $\check{\beta}$, $O(T^{-2})$ | 0.2997 | 0.3000 | 0.2999 | 0.5999 | 0.6003 | 0.6003 | 0.9003 | 0.8998 | 0.9000 |
| $\hat{\theta}$ | 0.4202 | 0.4195 | 0.4186 | 0.4196 | 0.4193 | 0.4181 | 0.4199 | 0.4197 | 0.4179 |
| $\tilde{\theta}$, $O(T^{-2})$ | 0.3999 | 0.3997 | 0.3983 | 0.3993 | 0.3995 | 0.3978 | 0.3996 | 0.3999 | 0.3977 |
| $\ddot{\theta}$, $O(T^{-2})$ | 0.4005 | 0.4000 | 0.4010 | 0.3999 | 0.4001 | 0.4005 | 0.4003 | 0.4002 | 0.4004 |
| $\hat{\alpha}$ | 0.5105 | 0.5102 | 0.5086 | 0.5099 | 0.5104 | 0.5066 | 0.5103 | 0.5095 | 0.5079 |
| $\tilde{\alpha}$, $O(T^{-2})$ | 0.5004 | 0.5003 | 0.4985 | 0.4998 | 0.5005 | 0.4965 | 0.5002 | 0.4996 | 0.4977 |
| $\ddot{\alpha}$, $O(T^{-2})$ | 0.5008 | 0.4942 | 0.4730 | 0.5002 | 0.4994 | 0.4708 | 0.5006 | 0.4935 | 0.4722 |
| $M(\hat{\beta})$, | 0.0082 | 0.0082 | 0.0098 | 0.0082 | 0.0082 | 0.0098 | 0.0081 | 0.0083 | 0.0099 |
| $M(\tilde{\beta})$, $O(T^{-1})$ | 0.0075 | 0.0075 | 0.0075 | 0.0075 | 0.0075 | 0.0075 | 0.0075 | 0.0075 | 0.0075 |
| $M(\ddot{\beta})$, $O(T^{-1})$ | 0.0078 | 0.0079 | 0.0102 | 0.0078 | 0.0079 | 0.0101 | 0.0078 | 0.0079 | 0.0102 |
| $M(\check{\beta})$, $O(T^{-1})$ | 0.0075 | 0.0077 | 0.0099 | 0.0075 | 0.0077 | 0.0098 | 0.0075 | 0.0077 | 0.0098 |
| $M(\tilde{\beta})$, $O(T^{-2})$ | 0.0078 | 0.0077 | 0.0078 | 0.0078 | 0.0077 | 0.0078 | 0.0078 | 0.0077 | 0.0078 |
| $M(\ddot{\beta})$, $O(T^{-2})$ | 0.0080 | 0.0081 | 0.0099 | 0.0080 | 0.0081 | 0.0098 | 0.0080 | 0.0081 | 0.0099 |
| $M(\check{\beta})$, $O(T^{-2})$ | 0.0078 | 0.0079 | 0.0096 | 0.0078 | 0.0079 | 0.0095 | 0.0078 | 0.0079 | 0.0096 |

(b) Forecast Bias and MSFE

| $\rho$=0.5 corr$(u,v)$=0.5 | $\beta$=0.3 bivN | bivT | bivlogN | $\beta$=0.6 bivN | bivT | bivlogN | $\beta$=0.9 bivN | bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $E(e_{T+1\|T})$ | 0.0002 | -0.0003 | -0.0003 | 0.0006 | -0.0002 | -0.0016 | -0.0005 | -0.0008 | 0.0010 |
| $E(e_{T+2\|T})$ | 0.0003 | -0.0005 | -0.0032 | 0.0006 | 0.0002 | -0.0052 | -0.0004 | -0.0012 | -0.0045 |
| $E(e_{T+5\|T})$ | 0.0003 | -0.0006 | -0.0076 | 0.0006 | 0.0002 | -0.0124 | 0.0002 | -0.0018 | -0.0156 |
| MSFE$(h=1)$ | 1.0406 | 1.7124 | 4.4740 | 1.0403 | 1.7127 | 4.4538 | 1.0402 | 1.7139 | 4.4903 |
| MSFE$(h=2)$ | 1.1248 | 1.8513 | 4.8608 | 1.4135 | 2.3331 | 6.1820 | 1.8919 | 3.1207 | 8.3526 |
| MSFE$(h=5)$ | 1.1804 | 1.9427 | 5.1281 | 1.5761 | 2.6057 | 6.9576 | 2.2319 | 3.6838 | 9.9206 |
| $E(e^2_{T+1\|T})$ | 1.0296 | 1.0296 | 1.0296 | 1.0296 | 1.0296 | 1.0296 | 1.0296 | 1.0296 | 1.0296 |
| $E(e^2_{T+2\|T})$ | 1.1094 | 1.1094 | 1.1094 | 1.3878 | 1.3878 | 1.3878 | 1.8498 | 1.8498 | 1.8498 |
| $E(e^2_{T+5\|T})$ | 1.1298 | 1.1298 | 1.1298 | 1.4900 | 1.4900 | 1.4900 | 2.0901 | 2.0901 | 2.0901 |
| $E(\tilde{e}^2_{T+1\|T})$ | 1.0098 | 1.6832 | 4.7520 | 1.0095 | 1.6834 | 4.7003 | 1.0096 | 1.6847 | 4.7531 |
| $E(\tilde{e}^2_{T+2\|T})$ | 1.0890 | 1.8154 | 5.1377 | 1.3462 | 2.2470 | 6.3230 | 1.7847 | 2.9761 | 8.4325 |
| $E(\tilde{e}^2_{T+5\|T})$ | 1.1140 | 1.8564 | 5.2527 | 1.4466 | 2.4141 | 6.7851 | 2.0108 | 3.3520 | 9.4705 |
| MSFE-HM$(h=1)$ | 1.1411 | 1.9028 | 5.3753 | 1.5067 | 2.5150 | 7.0354 | 2.1179 | 3.5324 | 9.9319 |
| MSFE-HM$(h=2)$ | 1.1266 | 1.8781 | 5.3029 | 1.4973 | 2.5005 | 6.9805 | 2.1145 | 3.5264 | 9.8962 |
| MSFE-HM$(h=5)$ | 1.0644 | 1.7739 | 4.9995 | 1.4230 | 2.3773 | 6.6221 | 2.0174 | 3.3624 | 9.4171 |
| MSFE-MD$(h=1)$ | 1.1194 | 1.8664 | 5.2719 | 1.4794 | 2.4680 | 6.9048 | 2.0805 | 3.4687 | 9.7462 |
| MSFE-MD$(h=2)$ | 1.0971 | 1.8287 | 5.1699 | 1.4495 | 2.4190 | 6.7671 | 2.0390 | 3.3992 | 9.5526 |
| MSFE-MD$(h=5)$ | 1.0301 | 1.7163 | 4.8489 | 1.3608 | 2.2718 | 6.3516 | 1.9142 | 3.1896 | 8.9617 |

Table 1.9: Unit Root $x_t$ Model with Intercept, T=30

(a) Estimator Bias and MSE

| $\rho$=1, $\alpha=0.5$ corr$(u,v)$=0.5, $\theta$=2 | $\beta$=0.3 bivN | bivT | bivlogN | $\beta$=0.6 bivN | bivT | bivlogN | $\beta$=0.9 bivN | bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\rho}$ | 0.9985 | 0.9974 | 0.9945 | 0.9985 | 0.9975 | 0.9945 | 0.9984 | 0.9974 | 0.9945 |
| $\tilde{\rho}$, $O(T^{-2})$ | 1.0002 | 0.9991 | 0.9962 | 1.0001 | 0.9991 | 0.9962 | 1.0001 | 0.9990 | 0.9962 |
| $\ddot{\rho}$, $O(T^{-2})$ | 1.0001 | 1.0002 | 1.0009 | 1.0001 | 1.0003 | 1.0009 | 1.0001 | 1.0002 | 1.0009 |
| $\tilde{\rho}$, $O(T^{-3})$ | 1.0000 | 0.9989 | 0.9960 | 1.0000 | 0.9990 | 0.9961 | 1.0000 | 0.9989 | 0.9960 |
| $\ddot{\rho}$, $O(T^{-3})$ | 1.0000 | 1.0001 | 0.9609 | 1.0000 | 1.0002 | 0.9583 | 1.0000 | 1.0001 | 0.9590 |
| $\hat{\beta}$ | 0.2992 | 0.2986 | 0.2976 | 0.5992 | 0.5987 | 0.5977 | 0.8992 | 0.8986 | 0.8977 |
| $\tilde{\beta}$, $O(T^{-2})$ | 0.3000 | 0.2995 | 0.2984 | 0.6000 | 0.5996 | 0.5985 | 0.9000 | 0.8995 | 0.8986 |
| $\ddot{\beta}$, $O(T^{-2})$ | 0.3000 | 0.3001 | 0.3004 | 0.6000 | 0.6001 | 0.6004 | 0.9000 | 0.9001 | 0.9005 |
| $\tilde{\beta}$, $O(T^{-3})$ | 0.2999 | 0.2994 | 0.2984 | 0.6000 | 0.5995 | 0.5985 | 0.9000 | 0.8994 | 0.8985 |
| $\ddot{\beta}$, $O(T^{-3})$ | 0.2999 | 0.3000 | 0.2998 | 0.6000 | 0.6001 | 0.5998 | 0.9000 | 0.9000 | 0.8999 |
| $\hat{\theta}$ | 2.0602 | 2.1010 | 2.2467 | 2.0599 | 2.0989 | 2.2452 | 2.0617 | 2.1020 | 2.2457 |
| $\hat{\alpha}$ | 0.5314 | 0.5521 | 0.6002 | 0.5318 | 0.5502 | 0.5970 | 0.5308 | 0.5515 | 0.5951 |
| $M(\hat{\beta})$, | 0.0001 | 0.0002 | 0.0006 | 0.0001 | 0.0002 | 0.0006 | 0.0001 | 0.0002 | 0.0006 |
| $M(\tilde{\beta})$, $O(T^{-3})$ | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| $M(\ddot{\beta})$, $O(T^{-3})$ | 0.0001 | 0.0002 | 0.0008 | 0.0001 | 0.0002 | 0.0008 | 0.0001 | 0.0002 | 0.0008 |
| $M(\tilde{\beta})$, $O(T^{-4})$ | 0.0001 | 0.0001 | 0.0004 | 0.0001 | 0.0001 | 0.0004 | 0.0001 | 0.0001 | 0.0004 |
| $M(\ddot{\beta})$, $O(T^{-4})$ | 0.0001 | 0.0002 | 0.0007 | 0.0001 | 0.0002 | 0.0007 | 0.0001 | 0.0002 | 0.0007 |

(b) Forecast Bias and MSFE

| $\rho$=1, $\alpha=0.5$ corr$(u,v)$=0.5, $\theta$=2 | $\beta$=0.3 bivN | bivT | bivlogN | $\beta$=0.6 bivN | bivT | bivlogN | $\beta$=0.9 bivN | bivT | bivlogN |
|---|---|---|---|---|---|---|---|---|---|
| $E(e_{T+1\|T})$ | 0.0272 | 0.0446 | 0.0676 | 0.0273 | 0.0453 | 0.0660 | 0.0258 | 0.0465 | 0.0680 |
| $E(e_{T+2\|T})$ | 0.0418 | 0.0686 | 0.0998 | 0.0593 | 0.0985 | 0.1491 | 0.0717 | 0.1267 | 0.1873 |
| $E(e_{T+5\|T})$ | 0.0740 | 0.1173 | 0.0650 | 0.1387 | 0.2312 | 0.2786 | 0.1797 | 0.3143 | 0.2085 |
| MSFE$(h=1)$ | 1.2900 | 2.0870 | 5.1400 | 1.2890 | 2.0959 | 5.1424 | 1.2865 | 2.0960 | 5.1365 |
| MSFE$(h=2)$ | 1.5579 | 2.5295 | 6.3565 | 2.0000 | 3.2673 | 8.3106 | 2.6719 | 4.3845 | 11.213 |
| MSFE$(h=2)$, $(\rho=1)$ | 1.3989 | 2.2780 | 5.8709 | 1.7636 | 2.8969 | 7.9638 | 2.4464 | 4.0393 | 11.614 |
| MSFE$(h=5)$ | 3.3091 | 5.3992 | 15.195 | 6.8503 | 11.264 | 32.052 | 12.414 | 20.325 | 57.901 |
| MSFE$(h=5)$, $(\rho=1)$ | 2.4791 | 4.1136 | 12.496 | 5.8698 | 9.8102 | 31.613 | 12.036 | 19.966 | 65.188 |
| MSFE-HM$(h=1)$ | 24.440 | 25.360 | 29.696 | 94.854 | 96.665 | 105.18 | 212.39 | 215.69 | 231.64 |
| MSFE-HM$(h=2)$ | 28.604 | 29.519 | 33.835 | 111.70 | 113.61 | 122.54 | 250.39 | 253.97 | 271.18 |
| MSFE-HM$(h=5)$ | 39.311 | 40.146 | 44.116 | 155.06 | 157.08 | 166.40 | 348.22 | 352.15 | 371.23 |
| MSFE-MD$(h=1)$ | 27.334 | 29.307 | 38.133 | 106.31 | 112.05 | 138.68 | 238.17 | 250.77 | 307.24 |
| MSFE-MD$(h=2)$ | 27.176 | 29.050 | 37.426 | 105.88 | 111.34 | 136.69 | 237.29 | 249.31 | 303.15 |
| MSFE-MD$(h=5)$ | 26.337 | 27.873 | 34.743 | 103.12 | 107.63 | 128.71 | 231.31 | 241.33 | 286.38 |

Table 1.10: Predictive Regression of Financial Returns

(a) Confidence Interval of Parameters

| | | d/p | | d/y | | ntis | | tbl | |
|---|---|---|---|---|---|---|---|---|---|
| $Q_t(12)$ | $\alpha$ | 0.4390 | 0.7748 | 0.4349 | 0.7699 | 0.0313 | 0.0735 | 0.0455 | 0.1033 |
| | $\beta$ | 0.1304 | 0.2331 | 0.1294 | 0.2320 | -3.0444 | -1.2685 | -1.5503 | -0.6323 |
| $R_t(12)$ | $\alpha$ | 0.5963 | 0.9263 | 0.5896 | 0.9191 | 0.1005 | 0.1420 | 0.0526 | 0.1118 |
| | $\beta$ | 0.1613 | 0.2623 | 0.1596 | 0.2605 | -3.7264 | -1.9808 | -0.6702 | 0.2706 |
| $X_t$ | $\theta$ | -1.0749 | -0.8221 | -1.0779 | -0.8227 | -0.0001 | 0.0010 | 0.0001 | 0.0021 |
| | $\rho$ | 0.6725 | 0.7498 | 0.6710 | 0.7492 | 0.9494 | 0.9949 | 0.9676 | 0.9984 |

(b) Estimation Correction and Forecast MSFE

| | $Q_t(12)$ | | | | $R_t(12)$ | | | |
|---|---|---|---|---|---|---|---|---|
| | d/p | d/y | ntis | tbl | d/p | d/y | ntis | tbl |
| $\hat{\beta}$ | 0.1818 | 0.1807 | -2.1564 | -1.0913 | 0.2118 | 0.2100 | -2.8536 | -0.1998 |
| $\ddot{\beta}$ | 0.1831 | 0.1820 | -2.1937 | -1.1513 | 0.2135 | 0.2117 | -2.8993 | -0.2533 |
| $\hat{\rho}$ | 0.7112 | 0.7101 | 0.9722 | 0.9830 | 0.7112 | 0.7101 | 0.9722 | 0.9830 |
| $\ddot{\rho}$ | 0.7164 | 0.7153 | 0.9817 | 0.9934 | 0.7164 | 0.7153 | 0.9817 | 0.9934 |
| $\hat{\alpha}$ | 0.6069 | 0.6024 | 0.0524 | 0.0744 | 0.7613 | 0.7544 | 0.1212 | 0.0822 |
| $\ddot{\alpha}$ | 0.6069 | 0.6025 | 0.0524 | 0.0744 | 0.7614 | 0.7544 | 0.1212 | 0.0822 |
| $\hat{\theta}$ | -0.9485 | -0.9503 | 0.0005 | 0.0011 | -0.9485 | -0.9503 | 0.0005 | 0.0011 |
| $\ddot{\theta}$ | -0.9313 | -0.9334 | 0.0003 | 0.0004 | -0.9313 | -0.9334 | 0.0003 | 0.0004 |
| M($\ddot{\beta}$) | 0.0018 | 0.0018 | 0.1237 | 0.0309 | 0.0017 | 0.0017 | 0.1208 | 0.0311 |
| MSFE(h=1) | 3.0895 | 3.0859 | 2.4113 | 2.8729 | 2.9551 | 2.9496 | 2.6267 | 3.1585 |
| MSFE(h=2) | 3.0093 | 3.0265 | 2.4390 | 2.9082 | 2.8850 | 2.9009 | 2.6675 | 3.1640 |
| MSFE(h=5) | 2.9037 | 2.9499 | 2.6426 | 2.9853 | 2.8011 | 2.8497 | 2.9148 | 3.1794 |
| MSFE-C(h=1) | 2.9910 | 2.9418 | 2.3922 | 3.1083 | 2.8737 | 2.8346 | 2.6155 | 3.3872 |
| MSFE-C(h=2) | 2.9397 | 2.9103 | 2.4286 | 3.0739 | 2.8296 | 2.8102 | 2.6698 | 3.4544 |
| MSFE-C(h=5) | 2.8923 | 2.8947 | 2.6624 | 3.3988 | 2.7956 | 2.8134 | 2.9629 | 4.2822 |
| $E(e^2_{T+1\|T})$ | 1.8168 | 1.8198 | 1.9677 | 2.1073 | 1.7558 | 1.7648 | 1.9787 | 2.2112 |
| $E(e^2_{T+2\|T})$ | 1.8447 | 1.8473 | 1.9909 | 2.1122 | 1.7892 | 1.7972 | 2.0062 | 2.2156 |
| $E(e^2_{T+5\|T})$ | 1.9172 | 1.9187 | 2.0512 | 2.1265 | 1.8771 | 1.8828 | 2.0782 | 2.2279 |
| | HM | MD | | | HM | MD | | |
| MSFE(h=1) | 2.7539 | 2.4620 | | | 2.9907 | 3.1588 | | |
| MSFE(h=2) | 2.7575 | 2.4610 | | | 2.9926 | 3.1550 | | |
| MSFE(h=5) | 2.7601 | 2.4598 | | | 2.9907 | 3.1493 | | |

# Bibliography


[1] Abadir, K. M. "OLS bias in a nonstationary autoregression". *Econometric Theory*, 9(1):81–93, 1993.

[2] Abadir, K. M. "Unbiased estimation as a solution to testing for random walks". *Economics Letters*, 47:263–268, 1995.

[3] Amihud Y. and C. M. Hurvich. "Predictive regressions: A reduced-bias estimation method". *Journal of Financial and Quantitative Analysis*, 39(4):813–841, 2004.

[4] Bao, Y. "The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution". *Ecomometric Theory*, 23:1013–1021, 2007a.

[5] Bao, Y. "Finite-sample properties of forecasts from the stationary first-order autoregressive model under a general error distribution". *Ecomometric Theory*, 23:767–773, 2007b.

[6] Bao, Y., A. Ullah and R. Zhang. "Finite sample properties for estimtors in unit root models". *working paper*, 2013.

[7] Bao, Y. and A. Ullah. "The second-order bias and mean squared error of estimators in time-series models". *Journal of Econometrics*, 140:650–669, 2007.

[8] Bao, Y. and A. Ullah. "Expectation of quadratic forms in normal and nonnormal variables with applications". *Journal of Statistical Planning and Inference*, 140:1193–1205, 2010.

[9] A.W. Lo Campbell, J.Y. and A.C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.

[10] Campbell, J.Y. and S. Thompson. "Predicting the equity premium out of sample: Can anything beat the historical average?". *Review of Financial Studies*, 21(4):1511–1531, 2008.

[11] Evans, G. B. A. and N. E. Savin. "Testing for unit roots: 1". *Econometrica*, 49:753–779, 1981.

[12] Goyal, A. and I. Welch. “A comprehensive look at the empirical performance of equity premium prediction”. *The Review of Financial Studies*, 21(4):1455–1508, 2008.

[13] Kendall, M.G. “Note on bias in estimation of autocorrelation”. *Biometrika*, 41:403–404, 1954.

[14] Kiviet, J. F. and G. D. A. Phillips. “Moment approximation for least-squares estimators in dynamic regression models with a unit root”. *Econometrics Journal*, 8:115–142, 2005.

[15] Kiviet, J.F. and G.D.A. Phillips. “Alternative bias approximations in regressions with a lagged dependent variable”. *Econometric Theory*, 9:62–80, 1993.

[16] Nagar, A.L. “The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations”. *Econometrica*, 27:575–595, 1959.

[17] Nicholls, D.F. and A.L. Pope. “Bias in the estimation of multivariate autoregressions”. *Australian Journal of Statistics*, 30A:296–309, 1988.

[18] Phillips, P.C.B. “Approximations to some finite sample distributions associated with a first-order stochastic difference equations”. *Econometrica*, 45:463–485, 1977.

[19] Pope, A.L. “Biases of estimators in multivariate non-Gaussian autoregressions”. *Journal of Time Series Analysis*, 11:249–258, 1990.

[20] Sawa, T. “The exact moments of the least squares estimator for the autoregressive model”. *Journal of Econometrics*, 8:159–172, 1978.

[21] Shenton, L.R. and W.L. Johnson. “Moments of a serial correlation coefficient”. *Journal of the Royal Statistical Society, Series B*, 27:308–320, 1965.

[22] Stambaugh, R. F. “Predictive regressions”. *Journal of Financial Econometrics*, 54:375–421, 1999.

[23] Tanaka, K. “An asymptotic expansion associated with the maximum likelihood estimators in ARMA models”. *Journal of the Royal Statistical Society, Series B*, 46(1):58–67, 1984.

[24] A. Ullah. *Finite Sample Ecomometrics*. Oxford Universit Press, New York, 2004.

[25] White, J.S. “Asymptotic expansions for the mean and variance of the serial correlation coefficient”. *Biometrika*, 48:85–94, 1961.

[26] Zhu, M. “Jackknife for bias reduction in predictive regressions”. *Journal of Financial Econometrics*, 11(1):193–220, 2013.

## Chapter 2

# Finite Sample Bias of the Estimators in the First Order Autoregressive Moving Average Model under a General Error Distribution

The quasi maximum likelihood estimator (QMLE) of parameters in the first order autoregressive moving average model (ARMA(1, 1)) can be biased in finite samples. This paper discusses bias properties of the QMLE of the ARMA(1,1) model up to order $O(T^{-1})$ by applying the stochastic expansion and the formula and sheds light on the bias correction for the parameter estimation in applied works. The analytical bias expression of the QMLE suggests that the bias is robust to nonnormality and the simulation results show that the bias corrected QML estimators is better even when sample size increased to a moderate size.

## 2.1 Introduction

Extensive literature has been focused on the finite-sample bias property of stable time-series models. Tanaka (1984) gives the finite sample bias correction formula using Edgeworth type asymptotic expansion for AR(1), AR(2), MA(1), MA(2) as well as ARMA(1, 1) model under normally distributed error terms. Bao and Ullah (2007) and Bao (2007) studied the finite sample bias for the QML estimator of MA(1) model and the OLS estimator of AR(1) model, the bias correction formula is given in both papers under a general distributed error term.

## 2.2 Main Results

Consider the pure first-order autoregressive moving average ARMA(1,1) model:

$$y_t = \rho y_{t-1} - \phi \varepsilon_{t-1} + \varepsilon_t \tag{2.1}$$

and the model with intercept:

$$y_t = \alpha + \rho y_{t-1} - \phi \varepsilon_{t-1} + \varepsilon_t \tag{2.2}$$

where $\rho, \phi \in (-1, 1)$, $\phi \neq \rho$, $\alpha$ is a constant of order $O(1)$, $\varepsilon_t \sim$ i.i.d.$(0, \sigma^2)$ and no further distributional assumptions is imposed on $\varepsilon_t$. The population parameters to be estimated is denoted by $\beta$, where $\beta = (\phi, \rho, \sigma^2)$ for the pure model and $\beta = (\alpha, \phi, \rho, \sigma^2)$ for the intercept model. In this paper, we consider the bias properties for the quasi maximum likelihood (QML) estimator of $\beta$ by assuming normality.

Conditional on $\varepsilon_0 = 0$ and $y_0$ given, either fixed or stochastic, we can rewrite the above pure model as $y_{-1} = Fy_0 + A\varepsilon$, where $y_{-1} = (y_0, y_2, ...y_{T-1})'$, $\varepsilon = (\varepsilon_1, \varepsilon_2, ...\varepsilon_T)'$, $F = (1, \rho, ..., \rho^{T-1})'$, and $A$ is defined by:

$$A = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \rho - \phi & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{T-3}(\rho - \phi) & \cdots & \rho - \phi & 1 & 0 \end{pmatrix}.$$

And the intercept model can be rewritten as $y_{-1} = \alpha M_1 \iota + y_0 F + A\varepsilon$, where $\iota$ is a vector of ones with dimension $T$, $M_1 = C^{-1}A$, where $C = I - B$, $I$ is a $T \times T$ identity matrix, and $B$ is a $T \times T$ matrix with $(i, j)$th element being $\phi$ if $i - j = 1$ and otherwise zero.

The average quasi log likelihood function assuming normality and given information up to time $T$ is

$$\mathcal{L}(\phi, \rho, \sigma^2) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{T}\frac{\varepsilon'\varepsilon}{2\sigma^2}$$

where $\varepsilon$ can be rewritten recursively as $\varepsilon = \iota_T' y_T + Dy_{-1} + B\varepsilon$ for the pure model, where $\iota_T'$ is $T$-dimensional vector with $T$th element equal to 1 and others zero, $D$ is a upper triangular $T \times T$ matrix with diagonal entries equal to $-\rho$ and $(i, j)$th upper off-diagonal element being 1 if $j - i = 1$, otherwise zero. For the intercept model, $\varepsilon$ can be rewritten recursively as $\varepsilon = \iota_T' y_T - \alpha\iota + Dy_{-1} + B\varepsilon$.

Following Bao and Ullah(2007, 2009), the stochastic expansion of the QML estimator $\hat{\beta}$ can be written as $\hat{\beta} - \beta = a_{-1/2} + a_{-1} + a_{-3/2} + o_P(T^{-3/2})$, where $a_{-i/2}$ is of order $O_p(T^{-i/2})$ for $i = 1, 2, 3$, and

$$\begin{aligned} a_{-1/2} &= -Q\psi_T, \quad a_{-1} = -QV_1 a_{-1/2} - \frac{1}{2}Q\bar{H}_2(a_{-1/2} \otimes a_{-1/2}), \\ a_{-3/2} &= -QV_1 a_{-1} - \frac{1}{2}QV_2(a_{-1/2} \otimes a_{-1/2}) - \frac{1}{2}Q\bar{H}_2(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2}) \\ &\quad + \frac{1}{6}Q\bar{H}_3(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}) \end{aligned} \tag{2.3}$$

where $\psi_T$ is the score function, $\bar{X} = \mathbb{E}(X)$, $Q = \bar{H}_1^{-1}$, $V_i = H_i - \bar{H}_i$ for $i = 1, 2, 3$, and $\otimes$ represents the Kronecker product. Applying the above expansion, the bias of the QML estimator $\hat{\beta}$ up to $O(T^{-1})$ is given by $B(\hat{\beta}) = \mathbb{E}(\hat{\beta} - \beta)$.

To get the bias of the estimator, we only need to check up to $a_{-1}$. To get the expectation of $a_{-1/2}$ and $a_{-1}$, we follow Ullah(2004, p.187) and Bao and Ullah(2010) for the expectation of quadratic forms, that for any matrices $N_1$, $N_2$ and $\varepsilon \sim$ i.i.d.$(0, \sigma^2)$, $\mathbb{E}(\varepsilon' N_1 \varepsilon) = \sigma^2 \mathbf{tr}(N_1)$, $\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_2 \varepsilon) = \sigma^4[\gamma_2 \mathbf{tr}(N_2 \odot N_1) + \mathbf{tr}(N_1)\mathbf{tr}(N_2) + \mathbf{tr}(N_1 N_2) + \mathbf{tr}(N_1' N_2)]$, and $\mathbb{E}(\varepsilon\varepsilon' N_1 \varepsilon) = \sigma^3 \gamma_1 (I \odot N_1)\iota$, where $\odot$ denotes the Hadamard product operator and $\gamma_1$ and $\gamma_2$ are the Pearson's measures of skewness and excess kurtosis of $\varepsilon$. Substitute the score function and Hessian matrix in appendix I for the pure model and appendix II for the intercept model to $a_{-1/2}$ in (2.3), it can be verified that $\mathbb{E}(a_{-1/2}) = \mathbb{E}(-Q\psi_T) = 0$, so $B(\hat{\beta}) = \mathbb{E}(a_{-1})$ for both models.

Notice that $F_1'F_1$, $F_1'F_2$, $F_1'F_3$, $F_1'N_1F_1$, $a'F_1$, $b'F_1$, $a'F_3$ and $b'N_1^*F_1$ are all of order $O(1)$, and also we can verify the following terms of order $O(T)$:

$$\mathbf{tr}(N_2) = \frac{T}{1-\phi^2}, \quad \mathbf{tr}(N_3) = \frac{6T\phi}{(1-\phi^2)^2}, \quad \mathbf{tr}(M_2) = \frac{T}{1-\rho^2}, \quad \mathbf{tr}(M_3) = \frac{T}{1-\phi\rho},$$

$$\mathbf{tr}(M_1M_3)=\frac{T\phi}{(1-\rho\phi)^2},\quad \mathbf{tr}(M_4)=\frac{T(\phi+\rho)}{(1-\phi^2)(1-\rho\phi)^2},\quad \mathbf{tr}(M_5)=\frac{T\rho}{(1-\rho\phi)(1-\rho^2)},$$
$$\mathbf{tr}(M_1N_2)=\frac{T\phi}{(1-\phi^2)(1-\rho\phi)},\quad \mathbf{tr}(M_1'N_1N_1)=\frac{T\rho}{(1-\rho\phi)^2},\quad \mathbf{tr}(M_1M_2)=\frac{T\rho}{(1-\rho^2)^2},$$
$$\mathbf{tr}(N_1N_2)=\frac{T\phi}{(1-\phi^2)^2},\quad \mathbf{tr}(N_1M_2)=\frac{T\rho}{(1-\rho^2)(1-\phi\rho)},$$
$$a'a=\frac{T}{(1-\phi)^2},\quad a'b=\frac{\alpha T}{(1-\phi)^2(1-\rho)},\quad b'b=\frac{\alpha^2T}{(1-\phi)^2(1-\rho)^2},\quad a'N_1a=\frac{T}{(1-\phi)^3},$$
$$a'M_1a=\frac{T}{(1-\phi)^2(1-\rho)},\quad b'N_1b=\frac{\alpha^2T}{(1-\phi)^3(1-\rho)^2},\quad b'M_1b=\frac{\alpha^2T}{(1-\phi)^2(1-\rho)^3},$$
$$a'N_1^*b=\frac{2\alpha T}{(1-\phi)^3(1-\rho)},\quad a'M_1b=b'M_1a=\frac{\alpha T}{(1-\rho)^2(1-\phi)^2}.$$

where $N_2=N_1'N_1+2N_1^2$, $N_1=C^{-1}B_1$, $B_1=\frac{\partial B}{\partial\phi}$, $N_3=6N_1'N_1^2+6N_1^3$, $M_2=M_1'M_1$, $M_3=N_1'M_1+N_1M_1$, $M_4=N_1'N_1M_1+M_1'N_1^2+N_1^2M_1$, $M_5=M_1'N_1M_1$, $F_1=C^{-1}F$, $F_2'=F_1'M_1$, $F_3'=F_1'N_1+F_1'N_1'$, $a=-C^{-1}\iota$, $b=\alpha C^{-1}M_1\iota$, $N_1^*=N_1+N_1'$. Applying the expectation of quadratic forms and substitute the above results into $a_{-1}$ in (2.3) with other results in the appendices, we have the following two propositions for the pure model and the intercept model respectively:

**Proposition 1**:*For the pure first order autoregressive moving average model with an initial condition $y_0$ and an initial error term $\varepsilon_0=0$, the approximate bias of the QML estimators, up to order $O(T^{-1})$, is given by:*

$$\begin{aligned}
B(\hat\phi) &= \frac{1}{T(\rho-\phi)^4}\Big(\phi^5(-6\rho^4+5\rho^2+1)+\phi^4(8\rho^3-9\rho)+\phi^3(5\rho^4-\rho^2)+\phi^2(-11\rho^3+5\rho)\\
&\quad +\phi(2\rho^4+2\rho^2)-\rho^3\Big)\\
B(\hat\rho) &= \frac{1}{T(\rho-\phi)^4}\Big(\rho^5(-6\phi^4+3\phi^2-1)+\rho^4(6\phi^3+6\phi)+\rho^3(9\phi^4-20\phi^2-1)\\
&\quad +\rho^2(2\phi^3+2\phi)+\rho(-5\phi^4+5\phi^2)\Big)\\
B(\hat\sigma^2) &= -\frac{2\sigma^2}{T} \qquad (2.4)
\end{aligned}$$

*provided $\phi, \rho \in (-1,1)$ and $\phi \neq \rho$. And the above results is robust to nonnormality.*

**Proposition 2**:*For the first order autoregressive moving average model with intercept and with an initial condition $y_0$ and an initial error term $\varepsilon_0 = 0$, the approximate bias of the QML estimators, up to order $O(T^{-1})$, is given by:*

$$
\begin{aligned}
B(\hat{\alpha}) &= \frac{\alpha}{T(1-\rho)(\rho-\phi)^4}\Big(\rho^5(6\phi^4 - 3\phi^2 - \phi + 1) + \rho^4(-6\phi^3 + 3\phi^2 - 7\phi + 1) + \rho^3(-9\phi^4 \\
&\quad -3\phi^3 + 23\phi^2 - 3\phi + 2)\rho^2(\phi^4 - 5\phi^3 + 3\phi^2 - 5\phi) + \rho(6\phi^4 - \phi^3 - 2\phi^2) - \phi^3\Big) \\
B(\hat{\phi}) &= \frac{1}{T(\rho-\phi)^4}\Big(\phi^5(-6\rho^4 + 5\rho^2 - \rho + 1) + \phi^4(8\rho^3 + 3\rho^2 - 10\rho + 1) + \phi^3(5\rho^4 - 3\rho^3 \\
&\quad +2\rho^2 - 3\rho + 1) + \phi^2(\rho^4 - 14\rho^3 + 3\rho^2 + 2\rho) + \phi(3\rho^4 - \rho^3 + 5\rho^2) - 2\rho^3\Big) \\
B(\hat{\rho}) &= \frac{1}{T(\rho-\phi)^4}\Big(\rho^5(-6\phi^4 + 3\phi^2 + \phi - 1) + \rho^4(6\phi^3 - 3\phi^2 + 7\phi - 1) + \rho^3(9\phi^4 + 3\phi^3 \\
&\quad -23\phi^2 + 3\phi - 2) - \rho^2(\phi^4 - 5\phi^3 + 3\phi^2 - 5\phi) + \rho(-6\phi^4 + \phi^3 + 2\phi^2) + \phi^3\Big) \\
B(\hat{\sigma}^2) &= -\frac{3\sigma^2}{T} \qquad (2.5)
\end{aligned}
$$

*provided $\phi, \rho \in (-1,1)$ and $\phi \neq \rho$. And the above results is robust to nonnormality.*

Remark 1. Notice that if $\rho = 0$ in the pure ARMA(1,1) model, it reduces to the MA(1) model, and from Bao and Ullah (2007), the bias of the QML estimators are $B(\hat{\phi}) = \frac{\phi}{T}$, $B(\hat{\sigma}^2) = -\frac{\sigma^2}{T}$, which could also be verified by deleting the corresponding elements with $\rho$ in the score, Hessian and $H_2$ matrix for the pure ARMA(1,1) model in the appendix. However, by substituting $\rho = 0$ into (3.3), we have $B(\hat{\phi}) = \frac{\phi}{T}$, $B(\hat{\rho}) = 0$ ,$B(\hat{\sigma}^2) = -\frac{2\sigma^2}{T}$, which indicates that if the true model is MA(1) but we misspecified as ARMA(1,1), the bias of the QML estimators for $\phi$ and $\rho$ will be the same, yet the bias of the variance term is twice as large in absolute value for the misspecified model.

Remark 2. On the other hand, if we assume $\phi = 0$ in the pure ARMA(1,1) model, it degenerates to the AR(1) model, and from Bao and Ullah (2007) and Bao (2007), the bias of the least squares estimators of the AR(1) model is given by $B(\hat{\rho}) = -\frac{2\rho}{T}$, and we can verify this is also the case for the QML estimators by deleting the corresponding rows and columns in the score function, Hessian and the $H_2$ matrix for the pure ARMA(1,1) model. Moreover, we can verify in this case, $B(\hat{\sigma}^2) = -\frac{\sigma^2}{T}$. However, if we substitute $\phi = 0$ into (3.3), we have: $B(\hat{\phi}) = -\frac{1}{T\rho}$, $B(\hat{\rho}) = -\frac{1}{T}(\rho + \frac{1}{\rho})$, and $B(\hat{\sigma}^2) = -\frac{2\sigma^2}{T}$. This implies that if the true model is AR(1) yet misspecified as ARMA(1,1), the bias for both of the QML estimators of $\phi$ and $\rho$ is larger in absolute value, may even be huge when $\rho$ is close to zero. Also, the bias of the variance term is twice as large in absolute value for the misspecified model.

Remark 3. For the pure model, when $\phi = 0$, the difference between the two bias in the misspecified ARMR(1,1) model is $B(\hat{\rho}) - B(\hat{\phi}) = -\frac{\rho}{T}$, implies that when the bias of $\hat{\rho}$ is bigger than the bias of $\hat{\phi}$ when $\rho < 0$, and opposite when $\rho > 0$ when the AR(1) model is misspecified as the ARMA(1,1) model. Moreover, since $\frac{dB(\hat{\phi})}{d\rho} = \frac{1}{T\rho^2} > 0$, the bias of $\hat{\phi}$ is always increasing when $\rho$ increases. And since $\frac{d^2B(\hat{\phi})}{d\rho^2} = -\frac{2}{T\rho^3}$, so the bias of $\hat{\phi}$ is concave in $\rho$ when $\rho > 0$ and convex in $\rho$ when $\rho < 0$. In addition, $\frac{dB(\hat{\rho})}{d\rho} = \frac{1-\rho^2}{T\rho^2} > 0$, means the bias of $\hat{\rho}$ increases when $\rho$ increases, and $\frac{d^2B(\hat{\phi})}{d\rho^2} = -\frac{2}{T\rho^3}$, so similarly, the bias of $\hat{\rho}$ is concave when $\rho > 0$ and convex when $\rho < 0$.

Remark 4. Another interesting thing for the intercept ARMA(1,1) model is that the bias for $\hat{\alpha}$ is equal to $-\frac{\alpha}{1-\rho}B(\hat{\rho})$. And for the MA(1) model, from Bao, Ullah and Zhang (2012), the bias of $\hat{\alpha}$ is equal to zero, which could also be verified by deleting the corresponding rows and columns of the score, Hessian and $H_2$ matrix related to $\rho$ for the

ARMA(1,1) model. For the AR(1) case, we can verify that the bias of $\alpha$ is equal to $\frac{\alpha(1+3\rho)}{T(1-\rho)}$, which is also equal to $-\frac{\alpha}{1-\rho}B(\hat{\rho})$, where $B(\hat{\rho}) = -\frac{1+3\rho}{T}$ from Bao (2007).

Remark 5. For both models, the initial condition $y_0$ does not affect the estimator bias up to order $O(T^{-1})$, however, $y_0$ does affect the bias of order $O(T^{-2})$. Moreover, the skewness and excess kurtosis of the error terms also enters the bias term of order $O(T^{-2})$. This is similar to the results of the AR(1) model, as in Bao (2007).

Remark 6. In general, for any ARMA(1,1) process, since the bias of $\hat{\sigma}^2$ comes both from AR(1) and MA(1) process, it is not surprising to see that it is twice as large as the bias in both AR(1) and MA(1). And similar to those two models, the bias of $\hat{\sigma}^2$ is increasing in $\sigma^2$ in absolute value. Moreover, the bias for $\hat{\phi}$ and $\hat{\rho}$ is bigger than both models, implying that it is affected by the correlation between the two parameters, since the Hessian matrix is no longer diagonal, as in the AR(1) and MA(1) models.

Remark 7. Tanaka (1984) derived the bias formula for both pure and intercept ARMA(1,1) model using Edgeworth type asymptotic expansion under normal distributed error terms. The bias formula given in this paper under general distributed error terms is slightly different from that paper. The simulation results below show that when sample size is small (for example, $T = 30$), the two results gives very similar bias correction, however, as sample size increases (for example, $T = 100$), the bias correction formula given in this paper tend to achieve the true parameter value while the formula given in Tanaka (1984) shows slower rate of convergence to the true value.

## 2.3 Simulation Results

This part uses simulation to check the effect of bias correction of the QML estimators applying (3.3) and (2.5) in the pure and intercept ARMA(1,1) model respectively. To check robustness to nonnormality, we allow the error term to have different distributions, including normal, uniform, exponential, mixture of of two normals N(-3,1) and N(3,1) with half probability for each, and student-$t$ with 5 degrees of freedom. The error terms are standardized to zero mean and with variance equal to 1 or 2 in different cases. The sample size is equal to 30 and 100. Since the model assumes that $\phi \neq \rho$, for the pure model, $(\phi, \rho) \in \{-0.9, -0.5, -0.2, 0.2, 0.5, 0.9\}$ and with opposite sign for each pair chosen; for the intercept model, $\alpha \in (0.2, 0.5, 1)$, and $(\phi, \rho) \in \{-0.8, -0.5, -0.2, 0.2, 0.5, 0.8\}$ to ensure convergence. The QML estimators are reported as $\hat{\alpha}$, $\hat{\phi}$, $\hat{\rho}$, $\hat{\sigma}^2$, the bias corrected estimators are reported as $\ddot{\alpha}$, $\ddot{\phi}$, $\ddot{\rho}$ and $\ddot{\sigma}^2$ by deducting the bias in (3.3) and (2.5) from the corresponding QML estimators, and the feasible bias corrected estimators are reported as $\tilde{\alpha}$, $\tilde{\phi}$, $\tilde{\rho}$ and $\tilde{\sigma}^2$ by substituting the QML estimators to the bias representation in the bias corrected estimators.

Table 1 to 4 report the results for the pure ARMA(1, 1) model, only the case when $\rho$ is positive and $\phi$ is negative is reported since the results for $\rho$ negative and $\phi$ positive is quite similar. Table 1 and 2 is for sample size is 30 and $\sigma^2$ is equal to 1 and 2 respectively. Table 3 and 4 are results for sample size equal to 100. From Table 1 to 4, we can see that for small sample size, when $\rho$ is equal to 0.5 or 0.9 and $\phi$ is equal to -0.2 or -0.5, both the bias corrected estimators for $\phi$ and $\rho$ give smaller bias than the QMLE estimators. And when $\rho$ is equal to 0.5 or 0.9 and $\phi$ is equal to -0.9, the bias corrected estimators for $\rho$ is better yet the QML estimator for $\phi$ is better. However, when $\rho$ is equal to 0.2, all the QML estimators for $\phi$ and $\rho$ are better no matter what value $\phi$ is. This implies that when the true value of $\rho$ increases from moderate to large, the bias corrected estimators work better than the QML estimators, which means in applied works the bias corrected estimator will outperform the QML estimator since $\rho$ always has larger value in real time series data. The bias corrected estimators for $\sigma^2$ is always better than QML estimators. And when sample size increase to 100, almost all the bias corrected estimators are better than the QML estimators, which implies the bias correction works even better when sample size is larger.

Table 5 to 10 show results for the intercept model, with 5 to 7 for sample size equal to 30 and 8 to 10 for sample size equal to 100. Here only cases for $\sigma^2 = 1$ and $\rho > 0$ are reported since other cases have similar results. The results show that for almost all cases, the bias corrected estimator outperform the QML estimator, and when sample size increase to 100, the results for bias corrected estimators are even better.

## 2.4 Concluding Remarks

The quasi maximum likelihood estimator (QMLE) of parameters in the first order autoregressive moving average model can be biased in finite samples. In this paper, we develop the bias of the QMLE of the pure ARMA(1,1) model up to order $O(T^{-1})$. And the formula represented in the proposition can be used as bias correction for parameter estimation in applied works. Our analytical bias expression of the QMLE suggests that the bias is robust to nonnormality since no special assumptions about the distribution of the error terms is imposed. Our simulation results show that the bias corrected QML estimators is better than the without correction, and it works even when sample size increased from 30 to 100. The comparison between the ARMA(1,1) model with AR(1) and MA(1) model shows that the bias for all the estimators in the ARMA(1,1) model is larger, due to the interaction between $\phi$ and $\rho$, as well as the variance coming from both AR(1) and MA(1) processes. When there is model misspecification, the bias will be larger, especially when the true model is AR(1).

Table 2.1: Bias Correction for pure ARMA(1,1) model, $\sigma^2 = 1, T = 30$

| ($\phi$, $\rho$) | $\hat{\phi}$ | $\ddot{\phi}$ | $\tilde{\phi}$ | $\hat{\rho}$ | $\ddot{\rho}$ | $\tilde{\rho}$ | $\hat{\sigma}$ | $\ddot{\sigma}$ | $\tilde{\sigma}$ |
|---|---|---|---|---|---|---|---|---|---|
| (-0.9, 0.9) | -0.9131 | -0.8809 | -0.8810 | 0.8439 | 0.9013 | 0.8962 | 0.9311 | 0.9978 | 0.9932 |
| | -0.9115 | -0.8792 | -0.8794 | 0.8495 | 0.9069 | 0.9023 | 0.9403 | 1.0070 | 1.0030 |
| | -0.9126 | -0.8803 | -0.8804 | 0.8545 | 0.9119 | 0.9078 | 0.9437 | 1.0104 | 1.0066 |
| | -0.9135 | -0.8812 | -0.8813 | 0.8440 | 0.9014 | 0.8963 | 0.9362 | 1.0029 | 0.9986 |
| | -0.9131 | -0.8808 | -0.8809 | 0.8475 | 0.9049 | 0.9001 | 0.9373 | 1.0040 | 0.9998 |
| (-0.5, 0.9) | -0.5348 | -0.5039 | -0.5050 | 0.8425 | 0.9029 | 0.8979 | 0.9332 | 0.9999 | 0.9954 |
| | -0.5320 | -0.5012 | -0.5023 | 0.8406 | 0.9010 | 0.8958 | 0.9300 | 0.9967 | 0.9920 |
| | -0.5349 | -0.5040 | -0.5050 | 0.8479 | 0.9083 | 0.9037 | 0.9378 | 1.0044 | 1.0003 |
| | -0.5345 | -0.5036 | -0.5046 | 0.8423 | 0.9027 | 0.8977 | 0.9333 | 1.0000 | 0.9956 |
| | -0.5369 | -0.5060 | -0.5070 | 0.8440 | 0.9044 | 0.8995 | 0.9369 | 1.0036 | 0.9994 |
| (-0.2, 0.9) | -0.2408 | -0.2071 | -0.2089 | 0.8370 | 0.9009 | 0.8962 | 0.9323 | 0.9989 | 0.9944 |
| | -0.2436 | -0.2099 | -0.2119 | 0.8306 | 0.8946 | 0.8894 | 0.9308 | 0.9975 | 0.9929 |
| | -0.2375 | -0.2038 | -0.2055 | 0.8399 | 0.9039 | 0.8995 | 0.9513 | 1.0179 | 1.0147 |
| | -0.2408 | -0.2071 | -0.2093 | 0.8292 | 0.8932 | 0.8882 | 0.9319 | 0.9986 | 0.9940 |
| | -0.2377 | -0.2040 | -0.2062 | 0.8374 | 0.9014 | 0.8965 | 0.9430 | 1.0097 | 1.0059 |
| (-0.9, 0.5) | -0.9104 | -0.8825 | -0.8812 | 0.4735 | 0.4955 | 0.4951 | 0.9387 | 1.0053 | 1.0012 |
| | -0.9091 | -0.8811 | -0.8799 | 0.4767 | 0.4987 | 0.4985 | 0.9422 | 1.0089 | 1.0051 |
| | -0.9076 | -0.8796 | -0.8783 | 0.4883 | 0.5102 | 0.5108 | 0.9405 | 1.0071 | 1.0032 |
| | -0.9100 | -0.8820 | -0.8807 | 0.4761 | 0.4980 | 0.4978 | 0.9377 | 1.0044 | 1.0003 |
| | -0.9117 | -0.8837 | -0.8824 | 0.4779 | 0.4999 | 0.4997 | 0.9374 | 1.0041 | 0.9999 |
| (-0.5, 0.5) | -0.5380 | -0.5233 | -0.5172 | 0.4556 | 0.4815 | 0.4818 | 0.9326 | 0.9993 | 0.9948 |
| | -0.5331 | -0.5184 | -0.5121 | 0.4615 | 0.4874 | 0.4885 | 0.9292 | 0.9959 | 0.9912 |
| | -0.5339 | -0.5192 | -0.5139 | 0.4655 | 0.4914 | 0.4914 | 0.9535 | 1.0201 | 1.0170 |
| | -0.5347 | -0.5200 | -0.5128 | 0.4587 | 0.4846 | 0.4869 | 0.9298 | 0.9964 | 0.9917 |
| | -0.5300 | -0.5153 | -0.5093 | 0.4627 | 0.4886 | 0.4895 | 0.9335 | 1.0001 | 0.9957 |
| (-0.2, 0.5) | -0.2487 | -0.2193 | -0.2142 | 0.4426 | 0.4894 | 0.5022 | 0.9318 | 0.9985 | 0.9939 |
| | -0.2506 | -0.2212 | -0.1835 | 0.4395 | 0.4863 | 0.5389 | 0.9326 | 0.9993 | 0.9948 |
| | -0.2459 | -0.2165 | -0.2235 | 0.4469 | 0.4937 | 0.4893 | 0.9377 | 1.0044 | 1.0003 |
| | -0.2593 | -0.2299 | -0.2401 | 0.4359 | 0.4827 | 0.4795 | 0.9277 | 0.9944 | 0.9896 |
| | -0.2571 | -0.2278 | -0.2548 | 0.4389 | 0.4857 | 0.4688 | 0.9341 | 1.0007 | 0.9963 |
| (-0.9, 0.2) | -0.9037 | -0.8772 | -0.8749 | 0.2041 | 0.2060 | 0.2118 | 0.9421 | 1.0087 | 1.0049 |
| | -0.9082 | -0.8817 | -0.8793 | 0.1997 | 0.2016 | 0.2072 | 0.9385 | 1.0052 | 1.0011 |
| | -0.8993 | -0.8729 | -0.8711 | 0.2178 | 0.2197 | 0.2251 | 0.9383 | 1.0049 | 1.0008 |
| | -0.9073 | -0.8808 | -0.8783 | 0.2031 | 0.2050 | 0.2108 | 0.9381 | 1.0047 | 1.0006 |
| | -0.9062 | -0.8798 | -0.8775 | 0.2068 | 0.2087 | 0.2140 | 0.9433 | 1.0100 | 1.0062 |
| (-0.5, 0.2) | -0.5251 | -0.5302 | -0.4342 | 0.1835 | 0.1713 | 0.2896 | 0.9320 | 0.9986 | 0.9941 |
| | -0.5284 | -0.5335 | -0.4640 | 0.1819 | 0.1697 | 0.2571 | 0.9277 | 0.9944 | 0.9896 |
| | -0.5231 | -0.5281 | -0.4797 | 0.1851 | 0.1730 | 0.2323 | 0.9387 | 1.0054 | 1.0013 |
| | -0.5360 | -0.5411 | -0.4365 | 0.1790 | 0.1668 | 0.2926 | 0.9267 | 0.9933 | 0.9884 |
| | -0.5238 | -0.5289 | -0.4623 | 0.1836 | 0.1715 | 0.2536 | 0.9282 | 0.9948 | 0.9900 |
| (-0.2, 0.2) | -0.2496 | -0.2613 | -0.0067 | 0.1735 | 0.1667 | 0.5810 | 0.9440 | 1.0106 | 1.0069 |
| | -0.2638 | -0.2755 | -0.0299 | 0.1609 | 0.1542 | 0.5773 | 0.9353 | 1.0020 | 0.9977 |
| | -0.2531 | -0.2647 | -0.0222 | 0.1663 | 0.1595 | 0.5685 | 0.9470 | 1.0137 | 1.0101 |
| | -0.2705 | -0.2821 | 0.0386 | 0.1557 | 0.1489 | 0.6356 | 0.9328 | 0.9995 | 0.9950 |
| | -0.2591 | -0.2708 | -0.0892 | 0.1679 | 0.1612 | 0.4947 | 0.9427 | 1.0093 | 1.0055 |

Note: All the estimators are the average value over 5,000 simulations. For each parameter value pair, the five rows in each column correspond to normal, uniform, exponential, mixture of two normals and student-$t$ distributions. All the above distributions are standardized to zero mean and unit variance.

Table 2.2: Bias Correction for pure ARMA(1,1) model, $\sigma^2 = 2, T = 30$

| ($\phi$, $\rho$) | $\hat{\phi}$ | $\ddot{\phi}$ | $\tilde{\phi}$ | $\hat{\rho}$ | $\ddot{\rho}$ | $\tilde{\rho}$ | $\hat{\sigma}$ | $\ddot{\sigma}$ | $\tilde{\sigma}$ |
|---|---|---|---|---|---|---|---|---|---|
| (-0.9, 0.9) | -0.9110 | -0.8787 | -0.8789 | 0.8484 | 0.9058 | 0.9011 | 1.8804 | 2.0138 | 2.0058 |
| | -0.9121 | -0.8798 | -0.8799 | 0.8473 | 0.9047 | 0.8999 | 1.8769 | 2.0103 | 2.0021 |
| | -0.9123 | -0.8800 | -0.8801 | 0.8535 | 0.9109 | 0.9067 | 1.8830 | 2.0163 | 2.0085 |
| | -0.9115 | -0.8792 | -0.8794 | 0.8451 | 0.9025 | 0.8975 | 1.8743 | 2.0076 | 1.9992 |
| | -0.9125 | -0.8802 | -0.8803 | 0.8489 | 0.9063 | 0.9016 | 1.9036 | 2.0369 | 2.0305 |
| (-0.5, 0.9) | -0.5381 | -0.5072 | -0.5082 | 0.8424 | 0.9028 | 0.8977 | 1.8757 | 2.0090 | 2.0007 |
| | -0.5423 | -0.5114 | -0.5125 | 0.8401 | 0.9005 | 0.8953 | 1.8687 | 2.0021 | 1.9933 |
| | -0.5382 | -0.5073 | -0.5084 | 0.8465 | 0.9069 | 0.9021 | 1.8847 | 2.0180 | 2.0104 |
| | -0.5361 | -0.5052 | -0.5062 | 0.8408 | 0.9012 | 0.8960 | 1.8576 | 1.9910 | 1.9815 |
| | -0.5345 | -0.5036 | -0.5046 | 0.8442 | 0.9046 | 0.8997 | 1.8770 | 2.0103 | 2.0021 |
| (-0.2, 0.9) | -0.2377 | -0.2040 | -0.2059 | 0.8353 | 0.8992 | 0.8943 | 1.8523 | 1.9857 | 1.9758 |
| | -0.2398 | -0.2061 | -0.2079 | 0.8347 | 0.8987 | 0.8939 | 1.8550 | 1.9883 | 1.9786 |
| | -0.2366 | -0.2029 | -0.2045 | 0.8405 | 0.9045 | 0.9002 | 1.8626 | 1.9959 | 1.9868 |
| | -0.2422 | -0.2086 | -0.2106 | 0.8363 | 0.9003 | 0.8954 | 1.8654 | 1.9987 | 1.9897 |
| | -0.2386 | -0.2049 | -0.2067 | 0.8408 | 0.9047 | 0.9002 | 1.8944 | 2.0278 | 2.0207 |
| (-0.9, 0.5) | -0.9074 | -0.8795 | -0.8782 | 0.4801 | 0.5021 | 0.5022 | 1.8759 | 2.0092 | 2.0010 |
| | -0.9085 | -0.8805 | -0.8793 | 0.4752 | 0.4972 | 0.4969 | 1.8797 | 2.0130 | 2.0050 |
| | -0.9096 | -0.8816 | -0.8803 | 0.4901 | 0.5121 | 0.5127 | 1.8947 | 2.0280 | 2.0210 |
| | -0.9115 | -0.8835 | -0.8822 | 0.4726 | 0.4946 | 0.4942 | 1.8756 | 2.0089 | 2.0006 |
| | -0.9105 | -0.8825 | -0.8812 | 0.4793 | 0.5013 | 0.5012 | 1.9051 | 2.0385 | 2.0322 |
| (-0.5, 0.5) | -0.5334 | -0.5187 | -0.5120 | 0.4609 | 0.4868 | 0.4884 | 1.8711 | 2.0045 | 1.9959 |
| | -0.5359 | -0.5212 | -0.5143 | 0.4628 | 0.4887 | 0.4903 | 1.8607 | 1.9941 | 1.9848 |
| | -0.5348 | -0.5201 | -0.5149 | 0.4655 | 0.4914 | 0.4910 | 1.8717 | 2.0050 | 1.9964 |
| | -0.5347 | -0.5200 | -0.5125 | 0.4590 | 0.4849 | 0.4872 | 1.8641 | 1.9975 | 1.9884 |
| | -0.5361 | -0.5214 | -0.5058 | 0.4597 | 0.4856 | 0.4962 | 1.8601 | 1.9935 | 1.9841 |
| (-0.2, 0.5) | -0.2567 | -0.2273 | -0.2201 | 0.4340 | 0.4808 | 0.4994 | 1.8629 | 1.9962 | 1.9871 |
| | -0.2521 | -0.2227 | -0.2188 | 0.4429 | 0.4897 | 0.5024 | 1.8617 | 1.9951 | 1.9859 |
| | -0.2405 | -0.2111 | -0.2127 | 0.4560 | 0.5027 | 0.5061 | 1.8512 | 1.9845 | 1.9746 |
| | -0.2463 | -0.2169 | -0.2345 | 0.4414 | 0.4882 | 0.4797 | 1.8613 | 1.9947 | 1.9854 |
| | -0.2468 | -0.2174 | -0.2366 | 0.4457 | 0.4925 | 0.4814 | 1.8735 | 2.0068 | 1.9984 |
| (-0.9, 0.2) | -0.9068 | -0.8804 | -0.8781 | 0.2018 | 0.2037 | 0.2091 | 1.8756 | 2.0089 | 2.0006 |
| | -0.9041 | -0.8776 | -0.8753 | 0.2059 | 0.2078 | 0.2135 | 1.8751 | 2.0084 | 2.0001 |
| | -0.9007 | -0.8743 | -0.8722 | 0.2145 | 0.2163 | 0.2219 | 1.9129 | 2.0462 | 2.0404 |
| | -0.9072 | -0.8808 | -0.8784 | 0.2062 | 0.2081 | 0.2137 | 1.8761 | 2.0094 | 2.0012 |
| | -0.9065 | -0.8801 | -0.8778 | 0.2039 | 0.2058 | 0.2112 | 1.9086 | 2.0419 | 2.0359 |
| (-0.5, 0.2) | -0.5273 | -0.5324 | -0.4655 | 0.1791 | 0.1669 | 0.2497 | 1.8667 | 2.0001 | 1.9912 |
| | -0.5293 | -0.5344 | -0.4625 | 0.1811 | 0.1689 | 0.2600 | 1.8632 | 1.9965 | 1.9874 |
| | -0.5225 | -0.5275 | -0.4832 | 0.1941 | 0.1819 | 0.2358 | 1.8755 | 2.0088 | 2.0005 |
| | -0.5264 | -0.5315 | -0.4538 | 0.1821 | 0.1699 | 0.2670 | 1.8555 | 1.9888 | 1.9792 |
| | -0.5304 | -0.5355 | -0.4613 | 0.1826 | 0.1704 | 0.2606 | 1.8835 | 2.0168 | 2.0091 |
| (-0.2, 0.2) | -0.2619 | -0.2735 | 0.0368 | 0.1586 | 0.1518 | 0.6550 | 1.8555 | 1.9889 | 1.9792 |
| | -0.2731 | -0.2847 | 0.0807 | 0.1541 | 0.1473 | 0.7003 | 1.8593 | 1.9926 | 1.9832 |
| | -0.2494 | -0.2611 | 0.0650 | 0.1728 | 0.1660 | 0.6571 | 1.8862 | 2.0195 | 2.0120 |
| | -0.2732 | -0.2849 | 0.0367 | 0.1539 | 0.1472 | 0.6341 | 1.8638 | 1.9971 | 1.9880 |
| | -0.2687 | -0.2803 | 0.0489 | 0.1606 | 0.1538 | 0.6623 | 1.8923 | 2.0257 | 2.0185 |

Note: All the estimators are the average value over 5,000 simulations. For each parameter value pair, the five rows in each column correspond to normal, uniform, exponential, mixture of two normals and student-$t$ distributions. All the above distributions are standardized to zero mean and unit variance.

Table 2.3: Bias Correction for pure ARMA(1,1) model, $\sigma^2 = 1, T = 100$

| ($\phi$,$\rho$) | $\hat{\phi}$ | $\ddot{\phi}$ | $\tilde{\phi}$ | $\hat{\rho}$ | $\ddot{\rho}$ | $\tilde{\rho}$ | $\hat{\sigma}$ | $\ddot{\sigma}$ | $\tilde{\sigma}$ |
|---|---|---|---|---|---|---|---|---|---|
| (-0.9, 0.9) | -0.9102 | -0.9006 | -0.9006 | 0.8823 | 0.8995 | 0.8990 | 0.9814 | 1.0014 | 1.0010 |
| | -0.9093 | -0.8996 | -0.8996 | 0.8827 | 0.8999 | 0.8994 | 0.9793 | 0.9993 | 0.9989 |
| | -0.9103 | -0.9006 | -0.9006 | 0.8829 | 0.9001 | 0.8996 | 0.9794 | 0.9994 | 0.9990 |
| | -0.9105 | -0.9008 | -0.9008 | 0.8822 | 0.8994 | 0.8989 | 0.9790 | 0.9990 | 0.9986 |
| | -0.9100 | -0.9003 | -0.9003 | 0.8835 | 0.9007 | 0.9002 | 0.9767 | 0.9967 | 0.9963 |
| (-0.5, 0.9) | -0.5127 | -0.5034 | -0.5036 | 0.8825 | 0.9006 | 0.9001 | 0.9811 | 1.0011 | 1.0008 |
| | -0.5094 | -0.5001 | -0.5003 | 0.8812 | 0.8994 | 0.8989 | 0.9793 | 0.9993 | 0.9988 |
| | -0.5093 | -0.5000 | -0.5001 | 0.8827 | 0.9008 | 0.9004 | 0.9730 | 0.9930 | 0.9925 |
| | -0.5110 | -0.5017 | -0.5019 | 0.8815 | 0.8997 | 0.8992 | 0.9804 | 1.0004 | 1.0000 |
| | -0.5097 | -0.5004 | -0.5005 | 0.8827 | 0.9009 | 0.9004 | 0.9847 | 1.0047 | 1.0044 |
| (-0.2, 0.9) | -0.2107 | -0.2006 | -0.2007 | 0.8807 | 0.8999 | 0.8995 | 0.9813 | 1.0013 | 1.0009 |
| | -0.2091 | -0.1989 | -0.1990 | 0.8801 | 0.8992 | 0.8989 | 0.9807 | 1.0007 | 1.0003 |
| | -0.2097 | -0.1995 | -0.1996 | 0.8828 | 0.9020 | 0.9017 | 0.9737 | 0.9937 | 0.9932 |
| | -0.2124 | -0.2023 | -0.2024 | 0.8810 | 0.9002 | 0.8999 | 0.9821 | 1.0021 | 1.0018 |
| | -0.2112 | -0.2011 | -0.2012 | 0.8812 | 0.9004 | 0.9001 | 0.9830 | 1.0030 | 1.0027 |
| (-0.9, 0.5) | -0.9091 | -0.9007 | -0.9006 | 0.4909 | 0.4975 | 0.4974 | 0.9805 | 1.0005 | 1.0001 |
| | -0.9088 | -0.9004 | -0.9002 | 0.4907 | 0.4972 | 0.4971 | 0.9791 | 0.9991 | 0.9987 |
| | -0.9075 | -0.8991 | -0.8989 | 0.4928 | 0.4994 | 0.4993 | 0.9826 | 1.0026 | 1.0022 |
| | -0.9085 | -0.9001 | -0.9000 | 0.4911 | 0.4977 | 0.4976 | 0.9789 | 0.9989 | 0.9985 |
| | -0.9091 | -0.9007 | -0.9005 | 0.4899 | 0.4965 | 0.4964 | 0.9821 | 1.0021 | 1.0017 |
| (-0.5, 0.5) | -0.5092 | -0.5048 | -0.5044 | 0.4874 | 0.4952 | 0.4949 | 0.9808 | 1.0008 | 1.0004 |
| | -0.5075 | -0.5031 | -0.5027 | 0.4881 | 0.4959 | 0.4956 | 0.9798 | 0.9998 | 0.9994 |
| | -0.5101 | -0.5057 | -0.5053 | 0.4882 | 0.4960 | 0.4957 | 0.9787 | 0.9987 | 0.9983 |
| | -0.5091 | -0.5047 | -0.5043 | 0.4884 | 0.4962 | 0.4959 | 0.9808 | 1.0008 | 1.0004 |
| | -0.5088 | -0.5043 | -0.5040 | 0.4913 | 0.4990 | 0.4989 | 0.9840 | 1.0040 | 1.0037 |
| (-0.2, 0.5) | -0.2121 | -0.2033 | -0.2042 | 0.4814 | 0.4954 | 0.4937 | 0.9811 | 1.0011 | 1.0007 |
| | -0.2131 | -0.2042 | -0.2053 | 0.4832 | 0.4973 | 0.4954 | 0.9792 | 0.9992 | 0.9987 |
| | -0.2158 | -0.2070 | -0.2081 | 0.4813 | 0.4953 | 0.4935 | 0.9828 | 1.0028 | 1.0024 |
| | -0.2146 | -0.2058 | -0.2068 | 0.4848 | 0.4988 | 0.4971 | 0.9791 | 0.9991 | 0.9986 |
| | -0.2096 | -0.2008 | -0.2017 | 0.4868 | 0.5008 | 0.4993 | 0.9760 | 0.9960 | 0.9955 |
| (-0.9, 0.2) | -0.9085 | -0.9006 | -0.9003 | 0.1976 | 0.1981 | 0.1986 | 0.9791 | 0.9991 | 0.9987 |
| | -0.9074 | -0.8995 | -0.8992 | 0.1969 | 0.1974 | 0.1979 | 0.9787 | 0.9987 | 0.9983 |
| | -0.9069 | -0.8990 | -0.8987 | 0.1987 | 0.1993 | 0.1997 | 0.9799 | 0.9999 | 0.9995 |
| | -0.9073 | -0.8994 | -0.8991 | 0.1975 | 0.1980 | 0.1985 | 0.9806 | 1.0006 | 1.0002 |
| | -0.9094 | -0.9014 | -0.9011 | 0.1982 | 0.1988 | 0.1993 | 0.9841 | 1.0041 | 1.0038 |
| (-0.5, 0.2) | -0.5107 | -0.5123 | -0.5077 | 0.1904 | 0.1868 | 0.1917 | 0.9763 | 0.9963 | 0.9958 |
| | -0.5090 | -0.5105 | -0.5061 | 0.1935 | 0.1898 | 0.1945 | 0.9789 | 0.9989 | 0.9985 |
| | -0.5110 | -0.5125 | -0.5083 | 0.1905 | 0.1869 | 0.1913 | 0.9849 | 1.0049 | 1.0046 |
| | -0.5132 | -0.5147 | -0.5098 | 0.1864 | 0.1827 | 0.1878 | 0.9803 | 1.0003 | 0.9999 |
| | -0.5086 | -0.5101 | -0.5057 | 0.1954 | 0.1917 | 0.1966 | 0.9843 | 1.0043 | 1.0040 |
| (-0.2, 0.2) | -0.2090 | -0.2125 | -0.1713 | 0.1866 | 0.1846 | 0.2464 | 0.9803 | 1.0003 | 0.9999 |
| | -0.2143 | -0.2178 | -0.1935 | 0.1844 | 0.1824 | 0.2240 | 0.9794 | 0.9994 | 0.9990 |
| | -0.2099 | -0.2134 | -0.1771 | 0.1900 | 0.1879 | 0.2379 | 0.9748 | 0.9948 | 0.9943 |
| | -0.2168 | -0.2203 | -0.1784 | 0.1830 | 0.1809 | 0.2401 | 0.9799 | 0.9999 | 0.9995 |
| | -0.2153 | -0.2188 | -0.1952 | 0.1822 | 0.1802 | 0.2201 | 0.9793 | 0.9993 | 0.9989 |

Note: All the estimators are the average value over 5,000 simulations. For each parameter value pair, the five rows in each column correspond to normal, uniform, exponential, mixture of two normals and student-$t$ distributions. All the above distributions are standardized to zero mean and unit variance.

Table 2.4: Bias Correction for pure ARMA(1,1) model, $\sigma^2 = 2, T = 100$

| ($\phi$, $\rho$) | $\hat{\phi}$ | $\ddot{\phi}$ | $\tilde{\phi}$ | $\hat{\rho}$ | $\ddot{\rho}$ | $\tilde{\rho}$ | $\hat{\sigma}$ | $\ddot{\sigma}$ | $\tilde{\sigma}$ |
|---|---|---|---|---|---|---|---|---|---|
| (-0.9, 0.9) | -0.9092 | -0.8995 | -0.8996 | 0.8817 | 0.8990 | 0.8984 | 1.9649 | 2.0049 | 2.0042 |
| | -0.9100 | -0.9003 | -0.9003 | 0.8830 | 0.9002 | 0.8998 | 1.9577 | 1.9977 | 1.9969 |
| | -0.9090 | -0.8993 | -0.8993 | 0.8836 | 0.9008 | 0.9003 | 1.9704 | 2.0104 | 2.0098 |
| | -0.9103 | -0.9006 | -0.9006 | 0.8827 | 0.9000 | 0.8995 | 1.9590 | 1.9990 | 1.9982 |
| | -0.9112 | -0.9015 | -0.9015 | 0.8819 | 0.8991 | 0.8986 | 1.9534 | 1.9934 | 1.9924 |
| (-0.5, 0.9) | -0.5096 | -0.5004 | -0.5005 | 0.8810 | 0.8992 | 0.8987 | 1.9562 | 1.9962 | 1.9953 |
| | -0.5090 | -0.4997 | -0.4999 | 0.8800 | 0.8981 | 0.8977 | 1.9592 | 1.9992 | 1.9984 |
| | -0.5105 | -0.5012 | -0.5014 | 0.8810 | 0.8991 | 0.8987 | 1.9663 | 2.0063 | 2.0056 |
| | -0.5116 | -0.5023 | -0.5025 | 0.8808 | 0.8989 | 0.8984 | 1.9603 | 2.0003 | 1.9995 |
| | -0.5107 | -0.5014 | -0.5015 | 0.8830 | 0.9012 | 0.9007 | 1.9608 | 2.0008 | 2.0000 |
| (-0.2, 0.9) | -0.2097 | -0.1996 | -0.1997 | 0.8803 | 0.8995 | 0.8992 | 1.9616 | 2.0016 | 2.0009 |
| | -0.2113 | -0.2012 | -0.2014 | 0.8803 | 0.8995 | 0.8992 | 1.9573 | 1.9973 | 1.9965 |
| | -0.2128 | -0.2027 | -0.2028 | 0.8825 | 0.9017 | 0.9014 | 1.9642 | 2.0042 | 2.0034 |
| | -0.2122 | -0.2021 | -0.2022 | 0.8794 | 0.8986 | 0.8983 | 1.9620 | 2.0020 | 2.0013 |
| | -0.2104 | -0.2003 | -0.2004 | 0.8809 | 0.9001 | 0.8998 | 1.9603 | 2.0003 | 1.9995 |
| (-0.9, 0.5) | -0.9094 | -0.9010 | -0.9009 | 0.4888 | 0.4954 | 0.4952 | 1.9595 | 1.9995 | 1.9987 |
| | -0.9086 | -0.9002 | -0.9000 | 0.4901 | 0.4967 | 0.4966 | 1.9553 | 1.9953 | 1.9944 |
| | -0.9084 | -0.9000 | -0.8998 | 0.4926 | 0.4992 | 0.4991 | 1.9679 | 2.0079 | 2.0072 |
| | -0.9088 | -0.9004 | -0.9002 | 0.4894 | 0.4959 | 0.4958 | 1.9585 | 1.9985 | 1.9976 |
| | -0.9092 | -0.9008 | -0.9006 | 0.4907 | 0.4973 | 0.4972 | 1.9802 | 2.0202 | 2.0198 |
| (-0.5, 0.5) | -0.5088 | -0.5044 | -0.5041 | 0.4872 | 0.4950 | 0.4948 | 1.9591 | 1.9991 | 1.9983 |
| | -0.5067 | -0.5023 | -0.5019 | 0.4879 | 0.4957 | 0.4954 | 1.9619 | 2.0019 | 2.0012 |
| | -0.5101 | -0.5056 | -0.5053 | 0.4865 | 0.4943 | 0.4939 | 1.9643 | 2.0043 | 2.0036 |
| | -0.5079 | -0.5035 | -0.5031 | 0.4868 | 0.4946 | 0.4943 | 1.9617 | 2.0017 | 2.0009 |
| | -0.5091 | -0.5047 | -0.5044 | 0.4892 | 0.4970 | 0.4968 | 1.9747 | 2.0147 | 2.0142 |
| (-0.2, 0.5) | -0.2139 | -0.2050 | -0.2061 | 0.4844 | 0.4984 | 0.4967 | 1.9629 | 2.0029 | 2.0021 |
| | -0.2149 | -0.2061 | -0.2073 | 0.4800 | 0.4940 | 0.4921 | 1.9595 | 1.9995 | 1.9987 |
| | -0.2124 | -0.2036 | -0.2046 | 0.4846 | 0.4986 | 0.4969 | 1.9686 | 2.0086 | 2.0079 |
| | -0.2193 | -0.2104 | -0.2115 | 0.4782 | 0.4923 | 0.4904 | 1.9553 | 1.9953 | 1.9944 |
| | -0.2129 | -0.2040 | -0.2048 | 0.4820 | 0.4960 | 0.4946 | 1.9500 | 1.9900 | 1.9890 |
| (-0.9, 0.2) | -0.9086 | -0.9007 | -0.9004 | 0.1962 | 0.1968 | 0.1973 | 1.9567 | 1.9967 | 1.9959 |
| | -0.9084 | -0.9005 | -0.9002 | 0.1990 | 0.1996 | 0.2001 | 1.9545 | 1.9945 | 1.9935 |
| | -0.9089 | -0.9009 | -0.9006 | 0.1988 | 0.1994 | 0.1998 | 1.9672 | 2.0072 | 2.0065 |
| | -0.9077 | -0.8998 | -0.8995 | 0.1997 | 0.2003 | 0.2007 | 1.9606 | 2.0006 | 1.9999 |
| | -0.9098 | -0.9019 | -0.9016 | 0.1987 | 0.1993 | 0.1997 | 1.9679 | 2.0079 | 2.0073 |
| (-0.5, 0.2) | -0.5120 | -0.5135 | -0.5090 | 0.1907 | 0.1871 | 0.1918 | 1.9579 | 1.9979 | 1.9971 |
| | -0.5073 | -0.5088 | -0.5041 | 0.1946 | 0.1909 | 0.1961 | 1.9581 | 1.9981 | 1.9973 |
| | -0.5091 | -0.5106 | -0.5065 | 0.1935 | 0.1899 | 0.1941 | 1.9578 | 1.9978 | 1.9970 |
| | -0.5058 | -0.5073 | -0.5027 | 0.1957 | 0.1920 | 0.1971 | 1.9588 | 1.9988 | 1.9980 |
| | -0.5115 | -0.5130 | -0.5086 | 0.1907 | 0.1871 | 0.1917 | 1.9647 | 2.0047 | 2.0040 |
| (-0.2, 0.2) | -0.2138 | -0.2173 | -0.1822 | 0.1854 | 0.1834 | 0.2358 | 1.9545 | 1.9945 | 1.9936 |
| | -0.2152 | -0.2187 | -0.1903 | 0.1817 | 0.1797 | 0.2257 | 1.9592 | 1.9992 | 1.9984 |
| | -0.2226 | -0.2261 | -0.1936 | 0.1779 | 0.1758 | 0.2204 | 1.9665 | 2.0065 | 2.0058 |
| | -0.2220 | -0.2255 | -0.1875 | 0.1770 | 0.1750 | 0.2296 | 1.9557 | 1.9957 | 1.9948 |
| | -0.2181 | -0.2216 | -0.1854 | 0.1807 | 0.1787 | 0.2328 | 1.9738 | 2.0138 | 2.0133 |

Note: All the estimators are the average value over 5,000 simulations. For each parameter value pair, the five rows in each column correspond to normal, uniform, exponential, mixture of two normals and student-$t$ distributions. All the above distributions are standardized to zero mean and unit variance.

Table 2.5: Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.2$, $\sigma^2 = 1, T = 30$

| ($\alpha$, $\phi$) | $\hat{\alpha}$ | $\ddot{\alpha}$ | $\tilde{\alpha}$ | $\hat{\phi}$ | $\ddot{\phi}$ | $\tilde{\phi}$ | $\hat{\rho}$ | $\ddot{\rho}$ | $\tilde{\rho}$ | $\hat{\sigma}^2$ | $\ddot{\sigma}^2$ | $\tilde{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (0.2, -0.2) | 0.2307 | 0.2064 | 0.0327 | -0.3054 | -0.2477 | 0.2251 | 0.0811 | 0.1783 | 0.7921 | 0.8990 | 0.9990 | 0.9889 |
| | 0.2186 | 0.1943 | 0.0596 | -0.2957 | -0.2380 | 0.1121 | 0.0920 | 0.1893 | 0.6722 | 0.8972 | 0.9972 | 0.9869 |
| | 0.2275 | 0.2032 | 0.1268 | -0.2990 | -0.2413 | -0.0761 | 0.0856 | 0.1829 | 0.4855 | 0.9049 | 1.0049 | 0.9954 |
| | 0.2242 | 0.1999 | 0.0782 | -0.2986 | -0.2409 | 0.1892 | 0.0851 | 0.1823 | 0.7377 | 0.8943 | 0.9943 | 0.9838 |
| | 0.2185 | 0.1942 | 0.1447 | -0.2895 | -0.2318 | 0.0021 | 0.0974 | 0.1947 | 0.5656 | 0.9018 | 1.0018 | 0.9920 |
| (0.5, -0.2) | 0.5602 | 0.4994 | 0.3561 | -0.2946 | -0.2369 | -0.1110 | 0.0949 | 0.1922 | 0.4326 | 0.8988 | 0.9988 | 0.9887 |
| | 0.5555 | 0.4947 | 0.2110 | -0.2719 | -0.2142 | 0.1790 | 0.1071 | 0.2043 | 0.7078 | 0.8949 | 0.9949 | 0.9844 |
| | 0.5742 | 0.5134 | 0.2532 | -0.3030 | -0.2453 | -0.0059 | 0.0782 | 0.1755 | 0.5446 | 0.9021 | 1.0021 | 0.9923 |
| | 0.5631 | 0.5023 | 0.3397 | -0.2904 | -0.2327 | -0.0861 | 0.1017 | 0.1989 | 0.4414 | 0.8925 | 0.9925 | 0.9818 |
| | 0.5604 | 0.4997 | 0.3516 | -0.2925 | -0.2348 | -0.0544 | 0.0934 | 0.1907 | 0.4865 | 0.8913 | 0.9913 | 0.9804 |
| (1.0, -0.2) | 1.0907 | 0.9691 | 0.6157 | -0.2528 | -0.1951 | 0.0763 | 0.1209 | 0.2182 | 0.5776 | 0.8936 | 0.9936 | 0.9830 |
| | 1.0774 | 0.9558 | 0.9068 | -0.2454 | -0.1877 | -0.1282 | 0.1293 | 0.2266 | 0.3602 | 0.8875 | 0.9875 | 0.9762 |
| | 1.0751 | 0.9536 | 0.9543 | -0.2464 | -0.1887 | -0.2526 | 0.1309 | 0.2281 | 0.2585 | 0.8897 | 0.9897 | 0.9787 |
| | 1.1017 | 0.9802 | 0.7632 | -0.2606 | -0.2029 | -0.0522 | 0.1140 | 0.2112 | 0.4428 | 0.8861 | 0.9861 | 0.9747 |
| | 1.0837 | 0.9622 | 1.0558 | -0.2532 | -0.1955 | -0.2716 | 0.1263 | 0.2236 | 0.2232 | 0.9021 | 1.0021 | 0.9923 |
| (0.2, -0.5) | 0.2191 | 0.2064 | 0.1841 | -0.5531 | -0.5320 | -0.4458 | 0.1225 | 0.1732 | 0.2729 | 0.8998 | 0.9998 | 0.9897 |
| | 0.2258 | 0.2132 | 0.1950 | -0.5675 | -0.5464 | -0.4691 | 0.1157 | 0.1664 | 0.2598 | 0.8960 | 0.9960 | 0.9856 |
| | 0.2260 | 0.2133 | 0.1942 | -0.5643 | -0.5432 | -0.4834 | 0.1149 | 0.1656 | 0.2325 | 0.9059 | 1.0059 | 0.9965 |
| | 0.2196 | 0.2070 | 0.1671 | -0.5707 | -0.5496 | -0.4148 | 0.1093 | 0.1599 | 0.3171 | 0.8949 | 0.9949 | 0.9844 |
| | 0.2188 | 0.2062 | 0.1844 | -0.5610 | -0.5399 | -0.4543 | 0.1194 | 0.1701 | 0.2699 | 0.9035 | 1.0035 | 0.9939 |
| (0.5, -0.5) | 0.5457 | 0.5140 | 0.4806 | -0.5645 | -0.5434 | -0.4778 | 0.1164 | 0.1671 | 0.2458 | 0.8976 | 0.9976 | 0.9874 |
| | 0.5474 | 0.5158 | 0.4921 | -0.5617 | -0.5406 | -0.5042 | 0.1190 | 0.1697 | 0.2190 | 0.8953 | 0.9953 | 0.9848 |
| | 0.5430 | 0.5114 | 0.4740 | -0.5551 | -0.5340 | -0.4861 | 0.1264 | 0.1771 | 0.2311 | 0.9028 | 1.0028 | 0.9931 |
| | 0.5468 | 0.5151 | 0.4681 | -0.5713 | -0.5502 | -0.4745 | 0.1134 | 0.1641 | 0.2539 | 0.8916 | 0.9916 | 0.9808 |
| | 0.5450 | 0.5133 | 0.4741 | -0.5659 | -0.5448 | -0.4771 | 0.1182 | 0.1689 | 0.2474 | 0.9048 | 1.0048 | 0.9953 |
| (1.0, -0.5) | 1.0711 | 1.0078 | 0.9405 | -0.5552 | -0.5341 | -0.4740 | 0.1282 | 0.1788 | 0.2472 | 0.8958 | 0.9958 | 0.9854 |
| | 1.0766 | 1.0132 | 0.9358 | -0.5535 | -0.5324 | -0.4720 | 0.1297 | 0.1804 | 0.2493 | 0.8945 | 0.9945 | 0.9840 |
| | 1.0716 | 1.0082 | 0.9536 | -0.5549 | -0.5338 | -0.4926 | 0.1324 | 0.1831 | 0.2275 | 0.9010 | 1.0010 | 0.9911 |
| | 1.0863 | 1.0230 | 0.9411 | -0.5540 | -0.5329 | -0.4756 | 0.1262 | 0.1769 | 0.2465 | 0.8905 | 0.9905 | 0.9796 |
| | 1.0772 | 1.0139 | 0.9945 | -0.5489 | -0.5278 | -0.5120 | 0.1311 | 0.1817 | 0.2047 | 0.8951 | 0.9951 | 0.9846 |
| (0.2, -0.8) | 0.2177 | 0.2064 | 0.2044 | -0.8332 | -0.8063 | -0.7994 | 0.1521 | 0.1975 | 0.2046 | 0.8950 | 0.9950 | 0.9845 |
| | 0.2149 | 0.2036 | 0.2010 | -0.8319 | -0.8049 | -0.7978 | 0.1565 | 0.2019 | 0.2093 | 0.8963 | 0.9963 | 0.9859 |
| | 0.2128 | 0.2015 | 0.1990 | -0.8309 | -0.8040 | -0.7983 | 0.1560 | 0.2013 | 0.2062 | 0.9034 | 1.0034 | 0.9938 |
| | 0.2135 | 0.2022 | 0.2004 | -0.8332 | -0.8063 | -0.7992 | 0.1520 | 0.1973 | 0.2046 | 0.8958 | 0.9958 | 0.9853 |
| | 0.2130 | 0.2016 | 0.1997 | -0.8332 | -0.8063 | -0.7992 | 0.1529 | 0.1983 | 0.2052 | 0.9026 | 1.0026 | 0.9928 |
| (0.5, -0.8) | 0.5251 | 0.4968 | 0.4923 | -0.8343 | -0.8074 | -0.8006 | 0.1545 | 0.1998 | 0.2068 | 0.8986 | 0.9986 | 0.9885 |
| | 0.5199 | 0.4915 | 0.4880 | -0.8382 | -0.8113 | -0.8049 | 0.1561 | 0.2015 | 0.2074 | 0.8999 | 0.9999 | 0.9899 |
| | 0.5234 | 0.4951 | 0.4917 | -0.8340 | -0.8071 | -0.8013 | 0.1557 | 0.2011 | 0.2062 | 0.9033 | 1.0033 | 0.9937 |
| | 0.5214 | 0.4931 | 0.4879 | -0.8346 | -0.8077 | -0.8005 | 0.1589 | 0.2043 | 0.2121 | 0.8976 | 0.9976 | 0.9874 |
| | 0.5197 | 0.4913 | 0.4882 | -0.8387 | -0.8117 | -0.8050 | 0.1477 | 0.1931 | 0.1994 | 0.8994 | 0.9994 | 0.9893 |
| (1.0, -0.8) | 1.0219 | 0.9652 | 0.9583 | -0.8343 | -0.8074 | -0.8013 | 0.1659 | 0.2112 | 0.2174 | 0.8990 | 0.9990 | 0.9889 |
| | 1.0318 | 0.9751 | 0.9704 | -0.8370 | -0.8101 | -0.8055 | 0.1625 | 0.2078 | 0.2124 | 0.8960 | 0.9960 | 0.9856 |
| | 1.0420 | 0.9853 | 0.9792 | -0.8360 | -0.8090 | -0.8036 | 0.1515 | 0.1968 | 0.2009 | 0.8975 | 0.9975 | 0.9872 |
| | 1.0312 | 0.9745 | 0.9693 | -0.8359 | -0.8090 | -0.8032 | 0.1580 | 0.2034 | 0.2086 | 0.8923 | 0.9923 | 0.9815 |
| | 1.0300 | 0.9733 | 0.9682 | -0.8390 | -0.8121 | -0.8068 | 0.1600 | 0.2053 | 0.2104 | 0.9101 | 1.0101 | 1.0011 |

Note: All the estimators are the average value over 5,000 simulations. For each parameter value pair, the five rows in each column correspond to normal, uniform, exponential, mixture of two normals and student-$t$ distributions. All the above distributions are standardized to zero mean and unit variance.

Table 2.6: Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.5$, $\sigma^2 = 1, T = 30$

| ($\alpha$, $\phi$) | $\hat{\alpha}$ | $\ddot{\alpha}$ | $\tilde{\alpha}$ | $\hat{\phi}$ | $\ddot{\phi}$ | $\tilde{\phi}$ | $\hat{\rho}$ | $\ddot{\rho}$ | $\tilde{\rho}$ | $\hat{\sigma}^2$ | $\ddot{\sigma}^2$ | $\tilde{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (0.2, -0.2) | 0.2572 | 0.2070 | 0.1984 | -0.3005 | -0.2292 | -0.2456 | 0.3504 | 0.4758 | 0.4653 | 0.9049 | 1.0049 | 0.9954 |
| | 0.2471 | 0.1969 | 0.2052 | -0.3079 | -0.2366 | -0.2467 | 0.3511 | 0.4765 | 0.4704 | 0.8953 | 0.9953 | 0.9849 |
| | 0.2484 | 0.1982 | 0.1761 | -0.2975 | -0.2262 | -0.1959 | 0.3524 | 0.4778 | 0.5097 | 0.8964 | 0.9964 | 0.9860 |
| | 0.2527 | 0.2026 | 0.2150 | -0.3103 | -0.2390 | -0.2622 | 0.3466 | 0.4719 | 0.4520 | 0.8944 | 0.9944 | 0.9838 |
| | 0.2489 | 0.1988 | 0.2086 | -0.2928 | -0.2215 | -0.2086 | 0.3628 | 0.4882 | 0.5022 | 0.8895 | 0.9895 | 0.9784 |
| (0.5, -0.2) | 0.6242 | 0.4988 | 0.5538 | -0.2973 | -0.2260 | -0.2722 | 0.3618 | 0.4872 | 0.4468 | 0.8958 | 0.9958 | 0.9854 |
| | 0.6128 | 0.4874 | 0.5190 | -0.2933 | -0.2220 | -0.2374 | 0.3674 | 0.4927 | 0.4742 | 0.8945 | 0.9945 | 0.9840 |
| | 0.6273 | 0.5019 | 0.5348 | -0.2990 | -0.2277 | -0.2557 | 0.3608 | 0.4862 | 0.4571 | 0.9056 | 1.0056 | 0.9962 |
| | 0.6210 | 0.4956 | 0.5633 | -0.2969 | -0.2256 | -0.2844 | 0.3609 | 0.4863 | 0.4293 | 0.8942 | 0.9942 | 0.9836 |
| | 0.6215 | 0.4962 | 0.5730 | -0.2921 | -0.2208 | -0.2775 | 0.3673 | 0.4927 | 0.4387 | 0.9080 | 1.0080 | 0.9988 |
| (1.0, -0.2) | 1.1883 | 0.9376 | 1.3010 | -0.2630 | -0.1917 | -0.3597 | 0.3918 | 0.5172 | 0.3442 | 0.8957 | 0.9957 | 0.9853 |
| | 1.2009 | 0.9502 | 1.4370 | -0.2731 | -0.2018 | -0.4121 | 0.3834 | 0.5087 | 0.3042 | 0.8925 | 0.9925 | 0.9817 |
| | 1.1813 | 0.9306 | 1.1427 | -0.2584 | -0.1871 | -0.2860 | 0.3973 | 0.5226 | 0.4212 | 0.8944 | 0.9944 | 0.9838 |
| | 1.2070 | 0.9563 | 1.2537 | -0.2730 | -0.2017 | -0.3400 | 0.3869 | 0.5123 | 0.3778 | 0.8921 | 0.9921 | 0.9813 |
| | 1.1859 | 0.9352 | 1.1791 | -0.2554 | -0.1841 | -0.2921 | 0.3962 | 0.5215 | 0.4138 | 0.8925 | 0.9925 | 0.9817 |
| (0.2, -0.5) | 0.2390 | 0.2037 | 0.2133 | -0.5609 | -0.5254 | -0.5271 | 0.3954 | 0.4839 | 0.4719 | 0.9008 | 1.0008 | 0.9908 |
| | 0.2436 | 0.2082 | 0.2103 | -0.5729 | -0.5374 | -0.5325 | 0.3876 | 0.4760 | 0.4699 | 0.8996 | 0.9996 | 0.9895 |
| | 0.2417 | 0.2064 | 0.2071 | -0.5663 | -0.5307 | -0.5282 | 0.3907 | 0.4791 | 0.4699 | 0.9003 | 1.0003 | 0.9903 |
| | 0.2382 | 0.2029 | 0.2054 | -0.5669 | -0.5314 | -0.5278 | 0.3909 | 0.4793 | 0.4715 | 0.8937 | 0.9937 | 0.9831 |
| | 0.2343 | 0.1990 | 0.2031 | -0.5635 | -0.5279 | -0.5233 | 0.3908 | 0.4792 | 0.4721 | 0.9083 | 1.0083 | 0.9991 |
| (0.5, -0.5) | 0.5922 | 0.5038 | 0.5312 | -0.5671 | -0.5315 | -0.5448 | 0.3902 | 0.4786 | 0.4552 | 0.8923 | 0.9923 | 0.9815 |
| | 0.5903 | 0.5019 | 0.5079 | -0.5627 | -0.5272 | -0.5241 | 0.4013 | 0.4897 | 0.4821 | 0.8955 | 0.9955 | 0.9850 |
| | 0.5861 | 0.4977 | 0.5034 | -0.5570 | -0.5215 | -0.5217 | 0.3972 | 0.4856 | 0.4747 | 0.9017 | 1.0017 | 0.9919 |
| | 0.5853 | 0.4969 | 0.4996 | -0.5618 | -0.5263 | -0.5194 | 0.3968 | 0.4852 | 0.4813 | 0.8942 | 0.9942 | 0.9836 |
| | 0.5864 | 0.4979 | 0.5184 | -0.5610 | -0.5255 | -0.5326 | 0.3986 | 0.4870 | 0.4705 | 0.9014 | 1.0014 | 0.9916 |
| (1.0, -0.5) | 1.1522 | 0.9754 | 0.9933 | -0.5537 | -0.5181 | -0.5180 | 0.4110 | 0.4994 | 0.4897 | 0.8937 | 0.9937 | 0.9830 |
| | 1.1567 | 0.9799 | 1.0116 | -0.5587 | -0.5231 | -0.5277 | 0.4095 | 0.4979 | 0.4833 | 0.8915 | 0.9915 | 0.9806 |
| | 1.1568 | 0.9800 | 0.9906 | -0.5613 | -0.5258 | -0.5246 | 0.4098 | 0.4982 | 0.4889 | 0.8976 | 0.9976 | 0.9874 |
| | 1.1599 | 0.9831 | 1.0130 | -0.5594 | -0.5239 | -0.5263 | 0.4076 | 0.4960 | 0.4830 | 0.8896 | 0.9896 | 0.9785 |
| | 1.1470 | 0.9701 | 0.9898 | -0.5567 | -0.5212 | -0.5200 | 0.4139 | 0.5023 | 0.4928 | 0.8987 | 0.9987 | 0.9886 |
| (0.2, -0.8) | 0.2192 | 0.1889 | 0.1925 | -0.8367 | -0.8063 | -0.8048 | 0.4207 | 0.4965 | 0.4895 | 0.8968 | 0.9968 | 0.9865 |
| | 0.2231 | 0.1928 | 0.1951 | -0.8343 | -0.8040 | -0.8024 | 0.4208 | 0.4967 | 0.4898 | 0.8946 | 0.9946 | 0.9840 |
| | 0.2248 | 0.1945 | 0.1959 | -0.8352 | -0.8048 | -0.8035 | 0.4237 | 0.4996 | 0.4925 | 0.9021 | 1.0021 | 0.9923 |
| | 0.2131 | 0.1828 | 0.1867 | -0.8334 | -0.8030 | -0.8014 | 0.4190 | 0.4948 | 0.4880 | 0.8938 | 0.9938 | 0.9831 |
| | 0.2191 | 0.1888 | 0.1912 | -0.8337 | -0.8034 | -0.8019 | 0.4225 | 0.4984 | 0.4916 | 0.8998 | 0.9998 | 0.9897 |
| (0.5, -0.8) | 0.5543 | 0.4784 | 0.4851 | -0.8352 | -0.8048 | -0.8033 | 0.4292 | 0.5050 | 0.4989 | 0.8982 | 0.9982 | 0.9881 |
| | 0.5537 | 0.4779 | 0.4861 | -0.8413 | -0.8110 | -0.8099 | 0.4214 | 0.4972 | 0.4900 | 0.8928 | 0.9928 | 0.9821 |
| | 0.5597 | 0.4839 | 0.4899 | -0.8364 | -0.8060 | -0.8047 | 0.4251 | 0.5009 | 0.4940 | 0.9065 | 1.0065 | 0.9972 |
| | 0.5586 | 0.4828 | 0.4889 | -0.8378 | -0.8074 | -0.8058 | 0.4269 | 0.5028 | 0.4965 | 0.8942 | 0.9942 | 0.9836 |
| | 0.5528 | 0.4770 | 0.4933 | -0.8375 | -0.8072 | -0.8107 | 0.4255 | 0.5013 | 0.4904 | 0.8977 | 0.9977 | 0.9875 |
| (1.0, -0.8) | 1.0954 | 0.9437 | 0.9568 | -0.8393 | -0.8089 | -0.8083 | 0.4349 | 0.5107 | 0.5042 | 0.8923 | 0.9923 | 0.9816 |
| | 1.0856 | 0.9340 | 0.9456 | -0.8416 | -0.8112 | -0.8096 | 0.4387 | 0.5145 | 0.5090 | 0.8926 | 0.9926 | 0.9819 |
| | 1.0934 | 0.9417 | 0.9527 | -0.8390 | -0.8087 | -0.8073 | 0.4361 | 0.5119 | 0.5059 | 0.9002 | 1.0002 | 0.9903 |
| | 1.0816 | 0.9300 | 0.9444 | -0.8400 | -0.8097 | -0.8083 | 0.4360 | 0.5119 | 0.5057 | 0.8905 | 0.9905 | 0.9796 |
| | 1.0772 | 0.9255 | 0.9399 | -0.8359 | -0.8055 | -0.8051 | 0.4391 | 0.5150 | 0.5087 | 0.8997 | 0.9997 | 0.9897 |

Note: All the estimators are the average value over 5,000 simulations. For each parameter value pair, the five rows in each column correspond to normal, uniform, exponential, mixture of two normals and student-$t$ distributions. All the above distributions are standardized to zero mean and unit variance.

Table 2.7: Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.8$, $\sigma^2 = 1, T = 30$

| ($\alpha$, $\phi$) | $\hat{\alpha}$ | $\ddot{\alpha}$ | $\tilde{\alpha}$ | $\hat{\phi}$ | $\ddot{\phi}$ | $\tilde{\phi}$ | $\hat{\rho}$ | $\ddot{\rho}$ | $\tilde{\rho}$ | $\hat{\sigma}^2$ | $\ddot{\sigma}^2$ | $\tilde{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (0.2, -0.2) | 0.3226 | 0.1920 | 0.1969 | -0.2834 | -0.2187 | -0.2278 | 0.6431 | 0.7737 | 0.7566 | 0.8913 | 0.9913 | 0.9804 |
| | 0.3247 | 0.1941 | 0.2071 | -0.2890 | -0.2242 | -0.2302 | 0.6355 | 0.7661 | 0.7527 | 0.8885 | 0.9885 | 0.9774 |
| | 0.3143 | 0.1837 | 0.1738 | -0.2780 | -0.2133 | -0.2197 | 0.6531 | 0.7836 | 0.7693 | 0.8943 | 0.9943 | 0.9837 |
| | 0.3191 | 0.1885 | 0.2080 | -0.2910 | -0.2262 | -0.2395 | 0.6384 | 0.7690 | 0.7482 | 0.8839 | 0.9839 | 0.9723 |
| | 0.3169 | 0.1864 | 0.2019 | -0.2754 | -0.2107 | -0.2148 | 0.6562 | 0.7868 | 0.7785 | 0.8869 | 0.9869 | 0.9756 |
| (0.5, -0.2) | 0.7578 | 0.4313 | 0.4880 | -0.2601 | -0.1954 | -0.2128 | 0.6699 | 0.8005 | 0.7765 | 0.8928 | 0.9928 | 0.9820 |
| | 0.7534 | 0.4269 | 0.4601 | -0.2566 | -0.1919 | -0.2031 | 0.6692 | 0.7998 | 0.7810 | 0.8827 | 0.9827 | 0.9710 |
| | 0.7389 | 0.4124 | 0.4314 | -0.2583 | -0.1936 | -0.2038 | 0.6777 | 0.8083 | 0.7918 | 0.8836 | 0.9836 | 0.9719 |
| | 0.7588 | 0.4323 | 0.4845 | -0.2613 | -0.1966 | -0.2211 | 0.6673 | 0.7979 | 0.7691 | 0.8870 | 0.9870 | 0.9757 |
| | 0.7434 | 0.4170 | 0.5094 | -0.2582 | -0.1934 | -0.2237 | 0.6737 | 0.8042 | 0.7703 | 0.8884 | 0.9884 | 0.9772 |
| (1.0, -0.2) | 1.3386 | 0.6857 | 0.8425 | -0.2179 | -0.1532 | -0.1809 | 0.7163 | 0.8469 | 0.8152 | 0.8837 | 0.9837 | 0.9721 |
| | 1.3425 | 0.6896 | 0.8673 | -0.2201 | -0.1554 | -0.1923 | 0.7159 | 0.8464 | 0.8071 | 0.8894 | 0.9894 | 0.9784 |
| | 1.3213 | 0.6684 | 0.7739 | -0.2132 | -0.1484 | -0.1691 | 0.7211 | 0.8517 | 0.8273 | 0.8824 | 0.9824 | 0.9707 |
| | 1.3332 | 0.6803 | 0.8128 | -0.2149 | -0.1502 | -0.1737 | 0.7171 | 0.8477 | 0.8197 | 0.8881 | 0.9881 | 0.9769 |
| | 1.3096 | 0.6567 | 0.9502 | -0.2091 | -0.1444 | -0.2066 | 0.7228 | 0.8534 | 0.7836 | 0.8920 | 0.9920 | 0.9812 |
| (0.2, -0.5) | 0.3032 | 0.1852 | 0.1923 | -0.5625 | -0.5166 | -0.5205 | 0.6667 | 0.7847 | 0.7700 | 0.8857 | 0.9857 | 0.9743 |
| | 0.2929 | 0.1749 | 0.1900 | -0.5626 | -0.5168 | -0.5208 | 0.6629 | 0.7809 | 0.7658 | 0.8882 | 0.9882 | 0.9770 |
| | 0.3111 | 0.1931 | 0.1857 | -0.5619 | -0.5160 | -0.5203 | 0.6673 | 0.7853 | 0.7704 | 0.9059 | 1.0059 | 0.9965 |
| | 0.3019 | 0.1838 | 0.1991 | -0.5601 | -0.5143 | -0.5182 | 0.6642 | 0.7822 | 0.7672 | 0.8856 | 0.9856 | 0.9742 |
| | 0.2922 | 0.1742 | 0.1896 | -0.5602 | -0.5143 | -0.5184 | 0.6640 | 0.7821 | 0.7670 | 0.9020 | 1.0020 | 0.9921 |
| (0.5, -0.5) | 0.7307 | 0.4356 | 0.4588 | -0.5518 | -0.5059 | -0.5092 | 0.6793 | 0.7973 | 0.7842 | 0.8911 | 0.9911 | 0.9802 |
| | 0.7330 | 0.4380 | 0.4590 | -0.5563 | -0.5104 | -0.5139 | 0.6781 | 0.7961 | 0.7822 | 0.8884 | 0.9884 | 0.9772 |
| | 0.7309 | 0.4358 | 0.4499 | -0.5582 | -0.5123 | -0.5160 | 0.6806 | 0.7986 | 0.7853 | 0.8942 | 0.9942 | 0.9836 |
| | 0.7273 | 0.4323 | 0.4581 | -0.5523 | -0.5064 | -0.5104 | 0.6783 | 0.7963 | 0.7826 | 0.8837 | 0.9837 | 0.9720 |
| | 0.7210 | 0.4260 | 0.4555 | -0.5550 | -0.5091 | -0.5133 | 0.6833 | 0.8013 | 0.7873 | 0.8968 | 0.9968 | 0.9865 |
| (1.0, -0.5) | 1.3193 | 0.7292 | 0.7855 | -0.5385 | -0.4926 | -0.4992 | 0.7147 | 0.8327 | 0.8190 | 0.8881 | 0.9881 | 0.9770 |
| | 1.3360 | 0.7460 | 0.7846 | -0.5398 | -0.4939 | -0.4960 | 0.7126 | 0.8306 | 0.8209 | 0.8897 | 0.9897 | 0.9787 |
| | 1.3348 | 0.7447 | 0.7753 | -0.5341 | -0.4882 | -0.4906 | 0.7155 | 0.8336 | 0.8241 | 0.8939 | 0.9939 | 0.9833 |
| | 1.3160 | 0.7259 | 0.7621 | -0.5344 | -0.4886 | -0.4905 | 0.7161 | 0.8341 | 0.8246 | 0.8892 | 0.9892 | 0.9781 |
| | 1.3129 | 0.7229 | 0.7727 | -0.5370 | -0.4911 | -0.4959 | 0.7183 | 0.8363 | 0.8246 | 0.8807 | 0.9807 | 0.9688 |
| (0.2, -0.8) | 0.2735 | 0.1628 | 0.1778 | -0.8342 | -0.7979 | -0.7993 | 0.6888 | 0.7994 | 0.7862 | 0.8892 | 0.9892 | 0.9781 |
| | 0.2712 | 0.1605 | 0.1717 | -0.8403 | -0.8040 | -0.8058 | 0.6872 | 0.7979 | 0.7843 | 0.8867 | 0.9867 | 0.9754 |
| | 0.2874 | 0.1767 | 0.1712 | -0.8403 | -0.8040 | -0.8057 | 0.6905 | 0.8012 | 0.7879 | 0.9059 | 1.0059 | 0.9965 |
| | 0.2730 | 0.1623 | 0.1754 | -0.8357 | -0.7993 | -0.8009 | 0.6870 | 0.7977 | 0.7840 | 0.8906 | 0.9906 | 0.9797 |
| | 0.2651 | 0.1544 | 0.1733 | -0.8372 | -0.8009 | -0.8025 | 0.6873 | 0.7979 | 0.7844 | 0.8972 | 0.9972 | 0.9869 |
| (0.5, -0.8) | 0.6649 | 0.3882 | 0.4224 | -0.8383 | -0.8020 | -0.8035 | 0.6962 | 0.8069 | 0.7945 | 0.8904 | 0.9904 | 0.9794 |
| | 0.6601 | 0.3835 | 0.4077 | -0.8362 | -0.7999 | -0.8012 | 0.7022 | 0.8129 | 0.8012 | 0.8915 | 0.9915 | 0.9807 |
| | 0.6688 | 0.3922 | 0.4084 | -0.8415 | -0.8052 | -0.8066 | 0.6999 | 0.8106 | 0.7985 | 0.8935 | 0.9935 | 0.9829 |
| | 0.6642 | 0.3875 | 0.4220 | -0.8402 | -0.8039 | -0.8054 | 0.6977 | 0.8084 | 0.7961 | 0.8858 | 0.9858 | 0.9744 |
| | 0.6673 | 0.3906 | 0.4183 | -0.8432 | -0.8069 | -0.8084 | 0.7012 | 0.8119 | 0.7999 | 0.9054 | 1.0054 | 0.9960 |
| (1.0, -0.8) | 1.2196 | 0.6664 | 0.7140 | -0.8399 | -0.8036 | -0.8045 | 0.7297 | 0.8404 | 0.8317 | 0.8841 | 0.9841 | 0.9725 |
| | 1.2252 | 0.6720 | 0.7091 | -0.8347 | -0.7984 | -0.7993 | 0.7299 | 0.8406 | 0.8322 | 0.8855 | 0.9855 | 0.9740 |
| | 1.2504 | 0.6971 | 0.7207 | -0.8358 | -0.7995 | -0.8005 | 0.7295 | 0.8402 | 0.8316 | 0.9009 | 1.0009 | 0.9910 |
| | 1.2378 | 0.6845 | 0.7308 | -0.8413 | -0.8049 | -0.8060 | 0.7281 | 0.8387 | 0.8299 | 0.8860 | 0.9860 | 0.9746 |
| | 1.2259 | 0.6726 | 0.7161 | -0.8352 | -0.7989 | -0.7998 | 0.7309 | 0.8416 | 0.8332 | 0.8891 | 0.9891 | 0.9780 |

Note: All the estimators are the average value over 5,000 simulations. For each parameter value pair, the five rows in each column correspond to normal, uniform, exponential, mixture of two normals and student-$t$ distributions. All the above distributions are standardized to zero mean and unit variance.

Table 2.8: Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.2$, $\sigma^2 = 1, T = 100$

| ($\alpha$, $\phi$) | $\hat{\alpha}$ | $\ddot{\alpha}$ | $\tilde{\alpha}$ | $\hat{\phi}$ | $\ddot{\phi}$ | $\tilde{\phi}$ | $\hat{\rho}$ | $\ddot{\rho}$ | $\tilde{\rho}$ | $\hat{\sigma}^2$ | $\ddot{\sigma}^2$ | $\tilde{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (0.2, -0.2) | 0.2123 | 0.2050 | 0.1862 | -0.2473 | -0.2299 | -0.1746 | 0.1425 | 0.1717 | 0.2471 | 0.9693 | 0.9993 | 0.9984 |
| | 0.2116 | 0.2043 | 0.1878 | -0.2406 | -0.2233 | -0.1831 | 0.1510 | 0.1802 | 0.2389 | 0.9704 | 1.0004 | 0.9995 |
| | 0.2151 | 0.2078 | 0.1926 | -0.2438 | -0.2265 | -0.1748 | 0.1440 | 0.1732 | 0.2396 | 0.9729 | 1.0029 | 1.0021 |
| | 0.2128 | 0.2055 | 0.1872 | -0.2442 | -0.2269 | -0.1604 | 0.1460 | 0.1752 | 0.2621 | 0.9696 | 0.9996 | 0.9987 |
| | 0.2132 | 0.2059 | 0.1833 | -0.2374 | -0.2201 | -0.1548 | 0.1563 | 0.1855 | 0.2667 | 0.9689 | 0.9989 | 0.9980 |
| (0.5, -0.2) | 0.5357 | 0.5175 | 0.4693 | -0.2467 | -0.2294 | -0.1714 | 0.1447 | 0.1739 | 0.2535 | 0.9691 | 0.9991 | 0.9981 |
| | 0.5277 | 0.5095 | 0.4717 | -0.2378 | -0.2205 | -0.1771 | 0.1524 | 0.1816 | 0.2429 | 0.9685 | 0.9985 | 0.9976 |
| | 0.5298 | 0.5116 | 0.4792 | -0.2371 | -0.2198 | -0.1823 | 0.1508 | 0.1800 | 0.2315 | 0.9709 | 1.0009 | 1.0000 |
| | 0.5276 | 0.5093 | 0.4624 | -0.2355 | -0.2182 | -0.1595 | 0.1541 | 0.1833 | 0.2593 | 0.9707 | 1.0007 | 0.9998 |
| | 0.5284 | 0.5101 | 0.4763 | -0.2318 | -0.2145 | -0.1833 | 0.1557 | 0.1848 | 0.2359 | 0.9729 | 1.0029 | 1.0021 |
| (1.0, -0.2) | 1.0560 | 1.0195 | 0.9718 | -0.2365 | -0.2192 | -0.1960 | 0.1552 | 0.1843 | 0.2233 | 0.9711 | 1.0011 | 1.0002 |
| | 1.0606 | 1.0242 | 0.9421 | -0.2354 | -0.2181 | -0.1692 | 0.1521 | 0.1813 | 0.2496 | 0.9690 | 0.9990 | 0.9980 |
| | 1.0465 | 1.0100 | 0.9640 | -0.2295 | -0.2122 | -0.1868 | 0.1599 | 0.1891 | 0.2259 | 0.9680 | 0.9980 | 0.9970 |
| | 1.0470 | 1.0105 | 0.9372 | -0.2289 | -0.2116 | -0.1715 | 0.1618 | 0.1910 | 0.2480 | 0.9693 | 0.9993 | 0.9984 |
| | 1.0464 | 1.0099 | 0.9566 | -0.2316 | -0.2143 | -0.1830 | 0.1590 | 0.1881 | 0.2350 | 0.9733 | 1.0033 | 1.0025 |
| (0.2, -0.5) | 0.2077 | 0.2039 | 0.2028 | -0.5189 | -0.5126 | -0.5078 | 0.1685 | 0.1837 | 0.1881 | 0.9672 | 0.9972 | 0.9963 |
| | 0.2085 | 0.2047 | 0.2034 | -0.5140 | -0.5077 | -0.5026 | 0.1759 | 0.1911 | 0.1961 | 0.9688 | 0.9988 | 0.9979 |
| | 0.2057 | 0.2019 | 0.2009 | -0.5121 | -0.5058 | -0.5014 | 0.1777 | 0.1929 | 0.1970 | 0.9701 | 1.0001 | 0.9992 |
| | 0.2053 | 0.2015 | 0.2002 | -0.5178 | -0.5115 | -0.5064 | 0.1749 | 0.1901 | 0.1950 | 0.9684 | 0.9984 | 0.9975 |
| | 0.2038 | 0.2000 | 0.1989 | -0.5175 | -0.5112 | -0.5065 | 0.1741 | 0.1893 | 0.1938 | 0.9680 | 0.9980 | 0.9970 |
| (0.5, -0.5) | 0.5115 | 0.5020 | 0.4994 | -0.5166 | -0.5102 | -0.5056 | 0.1739 | 0.1891 | 0.1935 | 0.9716 | 1.0016 | 1.0007 |
| | 0.5188 | 0.5093 | 0.5063 | -0.5203 | -0.5139 | -0.5088 | 0.1685 | 0.1837 | 0.1885 | 0.9677 | 0.9977 | 0.9968 |
| | 0.5169 | 0.5074 | 0.5049 | -0.5193 | -0.5129 | -0.5085 | 0.1731 | 0.1883 | 0.1923 | 0.9707 | 1.0007 | 0.9998 |
| | 0.5146 | 0.5051 | 0.5015 | -0.5154 | -0.5091 | -0.5034 | 0.1746 | 0.1898 | 0.1955 | 0.9689 | 0.9989 | 0.9980 |
| | 0.5204 | 0.5109 | 0.5082 | -0.5213 | -0.5150 | -0.5103 | 0.1692 | 0.1844 | 0.1887 | 0.9774 | 1.0074 | 1.0067 |
| (1.0, -0.5) | 1.0242 | 1.0052 | 0.9998 | -0.5159 | -0.5096 | -0.5050 | 0.1768 | 0.1920 | 0.1963 | 0.9695 | 0.9995 | 0.9986 |
| | 1.0300 | 1.0110 | 1.0055 | -0.5173 | -0.5110 | -0.5063 | 0.1753 | 0.1905 | 0.1949 | 0.9698 | 0.9998 | 0.9989 |
| | 1.0297 | 1.0107 | 1.0060 | -0.5180 | -0.5117 | -0.5075 | 0.1719 | 0.1871 | 0.1909 | 0.9621 | 0.9921 | 0.9909 |
| | 1.0258 | 1.0068 | 1.0013 | -0.5136 | -0.5073 | -0.5026 | 0.1787 | 0.1939 | 0.1983 | 0.9681 | 0.9981 | 0.9972 |
| | 1.0249 | 1.0059 | 1.0006 | -0.5131 | -0.5067 | -0.5023 | 0.1783 | 0.1935 | 0.1978 | 0.9731 | 1.0031 | 1.0023 |
| (0.2, -0.8) | 0.2032 | 0.1998 | 0.1997 | -0.8143 | -0.8062 | -0.8056 | 0.1793 | 0.1929 | 0.1933 | 0.9712 | 1.0012 | 1.0003 |
| | 0.2028 | 0.1994 | 0.1993 | -0.8143 | -0.8062 | -0.8056 | 0.1811 | 0.1948 | 0.1951 | 0.9702 | 1.0002 | 0.9993 |
| | 0.2093 | 0.2059 | 0.2057 | -0.8118 | -0.8037 | -0.8031 | 0.1848 | 0.1984 | 0.1988 | 0.9776 | 1.0076 | 1.0069 |
| | 0.1996 | 0.1962 | 0.1961 | -0.8138 | -0.8057 | -0.8051 | 0.1834 | 0.1970 | 0.1974 | 0.9686 | 0.9986 | 0.9977 |
| | 0.2036 | 0.2002 | 0.2001 | -0.8131 | -0.8051 | -0.8045 | 0.1813 | 0.1949 | 0.1953 | 0.9701 | 1.0001 | 0.9992 |
| (0.5, -0.8) | 0.5089 | 0.5004 | 0.5002 | -0.8139 | -0.8058 | -0.8052 | 0.1818 | 0.1954 | 0.1958 | 0.9696 | 0.9996 | 0.9987 |
| | 0.5128 | 0.5043 | 0.5040 | -0.8123 | -0.8042 | -0.8037 | 0.1816 | 0.1952 | 0.1956 | 0.9699 | 0.9999 | 0.9989 |
| | 0.5076 | 0.4991 | 0.4990 | -0.8113 | -0.8032 | -0.8027 | 0.1824 | 0.1960 | 0.1964 | 0.9652 | 0.9952 | 0.9942 |
| | 0.5162 | 0.5077 | 0.5074 | -0.8132 | -0.8051 | -0.8045 | 0.1811 | 0.1947 | 0.1950 | 0.9691 | 0.9991 | 0.9982 |
| | 0.5096 | 0.5011 | 0.5009 | -0.8129 | -0.8048 | -0.8042 | 0.1823 | 0.1959 | 0.1963 | 0.9760 | 1.0060 | 1.0053 |
| (1.0, -0.8) | 1.0264 | 1.0094 | 1.0090 | -0.8156 | -0.8075 | -0.8069 | 0.1798 | 0.1934 | 0.1937 | 0.9681 | 0.9981 | 0.9971 |
| | 1.0219 | 1.0049 | 1.0045 | -0.8127 | -0.8047 | -0.8041 | 0.1809 | 0.1945 | 0.1949 | 0.9679 | 0.9979 | 0.9970 |
| | 1.0241 | 1.0071 | 1.0067 | -0.8132 | -0.8051 | -0.8045 | 0.1816 | 0.1952 | 0.1955 | 0.9705 | 1.0005 | 0.9997 |
| | 1.0171 | 1.0001 | 0.9997 | -0.8142 | -0.8062 | -0.8056 | 0.1833 | 0.1969 | 0.1973 | 0.9678 | 0.9978 | 0.9969 |
| | 1.0174 | 1.0004 | 0.9999 | -0.8110 | -0.8029 | -0.8023 | 0.1831 | 0.1967 | 0.1971 | 0.9709 | 1.0009 | 1.0000 |

Note: All the estimators are the average value over 5,000 simulations. For each parameter value pair, the five rows in each column correspond to normal, uniform, exponential, mixture of two normals and student-$t$ distributions. All the above distributions are standardized to zero mean and unit variance.

Table 2.9: Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.5$, $\sigma^2 = 1, T = 100$

| $(\alpha, \phi)$ | $\hat{\alpha}$ | $\ddot{\alpha}$ | $\tilde{\alpha}$ | $\hat{\phi}$ | $\ddot{\phi}$ | $\tilde{\phi}$ | $\hat{\rho}$ | $\ddot{\rho}$ | $\tilde{\rho}$ | $\hat{\sigma}^2$ | $\ddot{\sigma}^2$ | $\tilde{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (0.2, -0.2) | 0.2174 | 0.2023 | 0.2034 | -0.2317 | -0.2103 | -0.2114 | 0.4539 | 0.4915 | 0.4891 | 0.9712 | 1.0012 | 1.0003 |
| | 0.2176 | 0.2025 | 0.2035 | -0.2263 | -0.2049 | -0.2062 | 0.4576 | 0.4952 | 0.4925 | 0.9700 | 1.0000 | 0.9991 |
| | 0.2167 | 0.2017 | 0.2029 | -0.2256 | -0.2042 | -0.2059 | 0.4577 | 0.4953 | 0.4924 | 0.9644 | 0.9944 | 0.9933 |
| | 0.2157 | 0.2007 | 0.2019 | -0.2300 | -0.2086 | -0.2102 | 0.4543 | 0.4919 | 0.4890 | 0.9686 | 0.9986 | 0.9976 |
| | 0.2153 | 0.2002 | 0.2015 | -0.2284 | -0.2070 | -0.2087 | 0.4566 | 0.4942 | 0.4913 | 0.9736 | 1.0036 | 1.0028 |
| (0.5, -0.2) | 0.5413 | 0.5037 | 0.5065 | -0.2267 | -0.2053 | -0.2067 | 0.4563 | 0.4939 | 0.4912 | 0.9707 | 1.0007 | 0.9998 |
| | 0.5447 | 0.5071 | 0.5102 | -0.2306 | -0.2092 | -0.2109 | 0.4524 | 0.4900 | 0.4870 | 0.9689 | 0.9989 | 0.9979 |
| | 0.5388 | 0.5012 | 0.5039 | -0.2255 | -0.2041 | -0.2057 | 0.4600 | 0.4976 | 0.4949 | 0.9702 | 1.0002 | 0.9993 |
| | 0.5390 | 0.5014 | 0.5042 | -0.2271 | -0.2057 | -0.2072 | 0.4596 | 0.4972 | 0.4944 | 0.9682 | 0.9982 | 0.9972 |
| | 0.5398 | 0.5022 | 0.5051 | -0.2255 | -0.2041 | -0.2057 | 0.4575 | 0.4951 | 0.4923 | 0.9668 | 0.9968 | 0.9958 |
| (1.0, -0.2) | 1.0727 | 0.9975 | 1.0024 | -0.2194 | -0.1980 | -0.1994 | 0.4627 | 0.5003 | 0.4978 | 0.9661 | 0.9961 | 0.9950 |
| | 1.0797 | 1.0045 | 1.0097 | -0.2269 | -0.2055 | -0.2070 | 0.4595 | 0.4971 | 0.4945 | 0.9680 | 0.9980 | 0.9971 |
| | 1.0742 | 0.9990 | 1.0042 | -0.2246 | -0.2032 | -0.2047 | 0.4619 | 0.4995 | 0.4969 | 0.9713 | 1.0013 | 1.0004 |
| | 1.0696 | 0.9943 | 0.9986 | -0.2168 | -0.1954 | -0.1966 | 0.4660 | 0.5036 | 0.5014 | 0.9679 | 0.9979 | 0.9969 |
| | 1.0760 | 1.0008 | 1.0057 | -0.2238 | -0.2024 | -0.2037 | 0.4607 | 0.4983 | 0.4958 | 0.9747 | 1.0047 | 1.0039 |
| (0.2, -0.5) | 0.2122 | 0.2016 | 0.2023 | -0.5206 | -0.5099 | -0.5100 | 0.4654 | 0.4920 | 0.4906 | 0.9686 | 0.9986 | 0.9977 |
| | 0.2156 | 0.2050 | 0.2054 | -0.5197 | -0.5090 | -0.5091 | 0.4659 | 0.4924 | 0.4911 | 0.9685 | 0.9985 | 0.9976 |
| | 0.2099 | 0.1992 | 0.1998 | -0.5179 | -0.5072 | -0.5074 | 0.4674 | 0.4939 | 0.4926 | 0.9676 | 0.9976 | 0.9966 |
| | 0.2093 | 0.1986 | 0.1992 | -0.5172 | -0.5065 | -0.5065 | 0.4691 | 0.4957 | 0.4945 | 0.9694 | 0.9994 | 0.9984 |
| | 0.2147 | 0.2040 | 0.2046 | -0.5211 | -0.5105 | -0.5106 | 0.4646 | 0.4911 | 0.4897 | 0.9706 | 1.0006 | 0.9997 |
| (0.5, -0.5) | 0.5328 | 0.5063 | 0.5072 | -0.5146 | -0.5039 | -0.5040 | 0.4700 | 0.4966 | 0.4954 | 0.9711 | 1.0011 | 1.0003 |
| | 0.5338 | 0.5073 | 0.5085 | -0.5187 | -0.5080 | -0.5081 | 0.4664 | 0.4930 | 0.4917 | 0.9680 | 0.9980 | 0.9970 |
| | 0.5299 | 0.5034 | 0.5047 | -0.5186 | -0.5079 | -0.5080 | 0.4686 | 0.4951 | 0.4939 | 0.9696 | 0.9996 | 0.9987 |
| | 0.5321 | 0.5056 | 0.5069 | -0.5172 | -0.5065 | -0.5066 | 0.4656 | 0.4922 | 0.4909 | 0.9700 | 1.0000 | 0.9991 |
| | 0.5330 | 0.5064 | 0.5078 | -0.5204 | -0.5098 | -0.5099 | 0.4669 | 0.4934 | 0.4921 | 0.9701 | 1.0001 | 0.9992 |
| (1.0, -0.5) | 1.0603 | 1.0073 | 1.0096 | -0.5169 | -0.5062 | -0.5063 | 0.4687 | 0.4953 | 0.4941 | 0.9679 | 0.9979 | 0.9970 |
| | 1.0582 | 1.0052 | 1.0073 | -0.5165 | -0.5058 | -0.5058 | 0.4705 | 0.4970 | 0.4959 | 0.9699 | 0.9999 | 0.9990 |
| | 1.0561 | 1.0031 | 1.0058 | -0.5186 | -0.5079 | -0.5080 | 0.4682 | 0.4948 | 0.4935 | 0.9623 | 0.9923 | 0.9911 |
| | 1.0563 | 1.0032 | 1.0057 | -0.5172 | -0.5065 | -0.5066 | 0.4694 | 0.4959 | 0.4947 | 0.9691 | 0.9991 | 0.9982 |
| | 1.0571 | 1.0040 | 1.0060 | -0.5152 | -0.5045 | -0.5046 | 0.4717 | 0.4983 | 0.4972 | 0.9739 | 1.0039 | 1.0031 |
| (0.2, -0.8) | 0.2091 | 0.2000 | 0.2003 | -0.8135 | -0.8044 | -0.8043 | 0.4732 | 0.4959 | 0.4951 | 0.9697 | 0.9997 | 0.9988 |
| | 0.2026 | 0.1935 | 0.1942 | -0.8140 | -0.8049 | -0.8048 | 0.4729 | 0.4957 | 0.4948 | 0.9678 | 0.9978 | 0.9969 |
| | 0.2077 | 0.1986 | 0.1989 | -0.8132 | -0.8041 | -0.8040 | 0.4738 | 0.4966 | 0.4957 | 0.9686 | 0.9986 | 0.9977 |
| | 0.2103 | 0.2012 | 0.2017 | -0.8136 | -0.8045 | -0.8044 | 0.4691 | 0.4918 | 0.4908 | 0.9701 | 1.0001 | 0.9992 |
| | 0.2145 | 0.2054 | 0.2056 | -0.8140 | -0.8049 | -0.8048 | 0.4715 | 0.4942 | 0.4933 | 0.9685 | 0.9985 | 0.9975 |
| (0.5, -0.8) | 0.5270 | 0.5042 | 0.5049 | -0.8137 | -0.8045 | -0.8045 | 0.4750 | 0.4977 | 0.4969 | 0.9684 | 0.9984 | 0.9974 |
| | 0.5256 | 0.5028 | 0.5036 | -0.8132 | -0.8041 | -0.8040 | 0.4737 | 0.4965 | 0.4957 | 0.9689 | 0.9989 | 0.9980 |
| | 0.5252 | 0.5024 | 0.5032 | -0.8132 | -0.8041 | -0.8040 | 0.4749 | 0.4977 | 0.4969 | 0.9733 | 1.0033 | 1.0025 |
| | 0.5278 | 0.5050 | 0.5058 | -0.8136 | -0.8045 | -0.8044 | 0.4726 | 0.4953 | 0.4945 | 0.9689 | 0.9989 | 0.9980 |
| | 0.5183 | 0.4955 | 0.4964 | -0.8119 | -0.8028 | -0.8027 | 0.4758 | 0.4986 | 0.4978 | 0.9752 | 1.0052 | 1.0045 |
| (1.0, -0.8) | 1.0431 | 0.9976 | 0.9992 | -0.8125 | -0.8034 | -0.8033 | 0.4761 | 0.4989 | 0.4981 | 0.9689 | 0.9989 | 0.9980 |
| | 1.0472 | 1.0017 | 1.0031 | -0.8131 | -0.8040 | -0.8039 | 0.4765 | 0.4992 | 0.4985 | 0.9707 | 1.0007 | 0.9998 |
| | 1.0470 | 1.0015 | 1.0033 | -0.8145 | -0.8054 | -0.8053 | 0.4743 | 0.4970 | 0.4962 | 0.9709 | 1.0009 | 1.0000 |
| | 1.0388 | 0.9933 | 0.9951 | -0.8150 | -0.8059 | -0.8058 | 0.4760 | 0.4988 | 0.4980 | 0.9676 | 0.9976 | 0.9967 |
| | 1.0439 | 0.9984 | 1.0001 | -0.8143 | -0.8052 | -0.8051 | 0.4747 | 0.4974 | 0.4966 | 0.9687 | 0.9987 | 0.9977 |

Note: All the estimators are the average value over 5,000 simulations. For each parameter value pair, the five rows in each column correspond to normal, uniform, exponential, mixture of two normals and student-$t$ distributions. All the above distributions are standardized to zero mean and unit variance.

Table 2.10: Bias Correction for ARMA(1,1) model with intercept, $\rho = 0.8$, $\sigma^2 = 1, T = 100$

| $(\alpha, \phi)$ | $\hat{\alpha}$ | $\ddot{\alpha}$ | $\tilde{\alpha}$ | $\hat{\phi}$ | $\ddot{\phi}$ | $\tilde{\phi}$ | $\hat{\rho}$ | $\ddot{\rho}$ | $\tilde{\rho}$ | $\hat{\sigma}^2$ | $\ddot{\sigma}^2$ | $\tilde{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (0.2, -0.2) | 0.2427 | 0.2035 | 0.2045 | -0.2227 | -0.2033 | -0.2037 | 0.7557 | 0.7949 | 0.7939 | 0.9707 | 1.0007 | 0.9999 |
| | 0.2400 | 0.2008 | 0.2015 | -0.2212 | -0.2018 | -0.2022 | 0.7588 | 0.7979 | 0.7971 | 0.9700 | 1.0000 | 0.9991 |
| | 0.2415 | 0.2023 | 0.2027 | -0.2192 | -0.1998 | -0.2001 | 0.7605 | 0.7996 | 0.7988 | 0.9766 | 1.0066 | 1.0059 |
| | 0.2386 | 0.1995 | 0.2010 | -0.2218 | -0.2023 | -0.2027 | 0.7570 | 0.7962 | 0.7953 | 0.9680 | 0.9980 | 0.9970 |
| | 0.2422 | 0.2030 | 0.2036 | -0.2227 | -0.2033 | -0.2037 | 0.7571 | 0.7963 | 0.7953 | 0.9734 | 1.0034 | 1.0026 |
| (0.5, -0.2) | 0.5943 | 0.4964 | 0.4980 | -0.2182 | -0.1987 | -0.1991 | 0.7611 | 0.8003 | 0.7995 | 0.9667 | 0.9967 | 0.9957 |
| | 0.5968 | 0.4989 | 0.5013 | -0.2207 | -0.2013 | -0.2016 | 0.7588 | 0.7980 | 0.7971 | 0.9695 | 0.9995 | 0.9986 |
| | 0.5908 | 0.4928 | 0.4954 | -0.2193 | -0.1999 | -0.2002 | 0.7604 | 0.7996 | 0.7988 | 0.9625 | 0.9925 | 0.9914 |
| | 0.5975 | 0.4996 | 0.5014 | -0.2206 | -0.2012 | -0.2016 | 0.7601 | 0.7992 | 0.7984 | 0.9674 | 0.9974 | 0.9965 |
| | 0.5916 | 0.4936 | 0.4959 | -0.2200 | -0.2006 | -0.2009 | 0.7615 | 0.8007 | 0.7998 | 0.9646 | 0.9946 | 0.9935 |
| (1.0, -0.2) | 1.1557 | 0.9598 | 0.9627 | -0.2137 | -0.1943 | -0.1945 | 0.7675 | 0.8067 | 0.8061 | 0.9685 | 0.9985 | 0.9976 |
| | 1.1554 | 0.9596 | 0.9633 | -0.2164 | -0.1970 | -0.1972 | 0.7670 | 0.8062 | 0.8056 | 0.9702 | 1.0002 | 0.9993 |
| | 1.1628 | 0.9669 | 0.9704 | -0.2153 | -0.1958 | -0.1961 | 0.7656 | 0.8047 | 0.8041 | 0.9780 | 1.0080 | 1.0073 |
| | 1.1601 | 0.9642 | 0.9670 | -0.2125 | -0.1931 | -0.1932 | 0.7665 | 0.8057 | 0.8051 | 0.9683 | 0.9983 | 0.9974 |
| | 1.1607 | 0.9648 | 0.9686 | -0.2170 | -0.1976 | -0.1979 | 0.7658 | 0.8050 | 0.8043 | 0.9654 | 0.9954 | 0.9944 |
| (0.2, -0.5) | 0.2318 | 0.1964 | 0.1986 | -0.5179 | -0.5041 | -0.5046 | 0.7598 | 0.7952 | 0.7939 | 0.9690 | 0.9990 | 0.9981 |
| | 0.2357 | 0.2003 | 0.2015 | -0.5158 | -0.5020 | -0.5024 | 0.7622 | 0.7976 | 0.7964 | 0.9687 | 0.9987 | 0.9978 |
| | 0.2356 | 0.2002 | 0.2011 | -0.5168 | -0.5030 | -0.5035 | 0.7636 | 0.7990 | 0.7978 | 0.9726 | 1.0026 | 1.0018 |
| | 0.2318 | 0.1964 | 0.1985 | -0.5187 | -0.5050 | -0.5054 | 0.7616 | 0.7970 | 0.7957 | 0.9681 | 0.9981 | 0.9971 |
| | 0.2354 | 0.2000 | 0.2013 | -0.5165 | -0.5027 | -0.5031 | 0.7621 | 0.7975 | 0.7963 | 0.9706 | 1.0006 | 0.9997 |
| (0.5, -0.5) | 0.5934 | 0.5049 | 0.5076 | -0.5160 | -0.5023 | -0.5027 | 0.7612 | 0.7966 | 0.7953 | 0.9682 | 0.9982 | 0.9973 |
| | 0.5926 | 0.5041 | 0.5069 | -0.5162 | -0.5024 | -0.5028 | 0.7619 | 0.7973 | 0.7961 | 0.9668 | 0.9968 | 0.9958 |
| | 0.5830 | 0.4945 | 0.4977 | -0.5168 | -0.5030 | -0.5035 | 0.7647 | 0.8001 | 0.7989 | 0.9647 | 0.9947 | 0.9936 |
| | 0.5846 | 0.4961 | 0.4994 | -0.5145 | -0.5008 | -0.5012 | 0.7629 | 0.7983 | 0.7971 | 0.9688 | 0.9988 | 0.9978 |
| | 0.5831 | 0.4945 | 0.4978 | -0.5164 | -0.5027 | -0.5031 | 0.7637 | 0.7991 | 0.7979 | 0.9716 | 1.0016 | 1.0008 |
| (1.0, -0.5) | 1.1526 | 0.9756 | 0.9810 | -0.5147 | -0.5010 | -0.5013 | 0.7676 | 0.8030 | 0.8019 | 0.9659 | 0.9959 | 0.9949 |
| | 1.1561 | 0.9791 | 0.9851 | -0.5136 | -0.4998 | -0.5002 | 0.7664 | 0.8018 | 0.8007 | 0.9690 | 0.9990 | 0.9981 |
| | 1.1567 | 0.9797 | 0.9861 | -0.5168 | -0.5030 | -0.5034 | 0.7663 | 0.8017 | 0.8006 | 0.9694 | 0.9994 | 0.9985 |
| | 1.1494 | 0.9724 | 0.9781 | -0.5120 | -0.4982 | -0.4985 | 0.7676 | 0.8030 | 0.8020 | 0.9674 | 0.9974 | 0.9964 |
| | 1.1587 | 0.9817 | 0.9881 | -0.5155 | -0.5018 | -0.5021 | 0.7655 | 0.8009 | 0.7998 | 0.9701 | 1.0001 | 0.9992 |
| (0.2, -0.8) | 0.2325 | 0.1993 | 0.2009 | -0.8135 | -0.8026 | -0.8028 | 0.7636 | 0.7968 | 0.7954 | 0.9686 | 0.9986 | 0.9977 |
| | 0.2388 | 0.2056 | 0.2063 | -0.8131 | -0.8022 | -0.8024 | 0.7631 | 0.7963 | 0.7950 | 0.9672 | 0.9972 | 0.9962 |
| | 0.2308 | 0.1976 | 0.1986 | -0.8127 | -0.8018 | -0.8020 | 0.7681 | 0.8013 | 0.8001 | 0.9714 | 1.0014 | 1.0005 |
| | 0.2344 | 0.2012 | 0.2019 | -0.8145 | -0.8036 | -0.8038 | 0.7653 | 0.7985 | 0.7972 | 0.9684 | 0.9984 | 0.9974 |
| | 0.2290 | 0.1958 | 0.1981 | -0.8134 | -0.8025 | -0.8027 | 0.7634 | 0.7966 | 0.7952 | 0.9688 | 0.9988 | 0.9979 |
| (0.5, -0.8) | 0.5776 | 0.4946 | 0.4982 | -0.8134 | -0.8025 | -0.8027 | 0.7654 | 0.7986 | 0.7973 | 0.9648 | 0.9948 | 0.9937 |
| | 0.5720 | 0.4890 | 0.4928 | -0.8129 | -0.8020 | -0.8022 | 0.7673 | 0.8005 | 0.7993 | 0.9695 | 0.9995 | 0.9986 |
| | 0.5761 | 0.4931 | 0.4964 | -0.8127 | -0.8018 | -0.8020 | 0.7668 | 0.8000 | 0.7987 | 0.9704 | 1.0004 | 0.9995 |
| | 0.5846 | 0.5017 | 0.5058 | -0.8152 | -0.8043 | -0.8045 | 0.7619 | 0.7951 | 0.7937 | 0.9694 | 0.9994 | 0.9985 |
| | 0.5791 | 0.4961 | 0.4987 | -0.8138 | -0.8029 | -0.8031 | 0.7673 | 0.8005 | 0.7992 | 0.9702 | 1.0002 | 0.9993 |
| (1.0, -0.8) | 1.1245 | 0.9585 | 0.9655 | -0.8135 | -0.8026 | -0.8028 | 0.7710 | 0.8042 | 0.8031 | 0.9681 | 0.9981 | 0.9971 |
| | 1.1309 | 0.9649 | 0.9717 | -0.8145 | -0.8036 | -0.8038 | 0.7704 | 0.8036 | 0.8025 | 0.9691 | 0.9991 | 0.9982 |
| | 1.1328 | 0.9668 | 0.9728 | -0.8142 | -0.8033 | -0.8034 | 0.7709 | 0.8041 | 0.8030 | 0.9725 | 1.0025 | 1.0017 |
| | 1.1379 | 0.9719 | 0.9779 | -0.8127 | -0.8018 | -0.8019 | 0.7699 | 0.8031 | 0.8020 | 0.9682 | 0.9982 | 0.9973 |
| | 1.1293 | 0.9633 | 0.9696 | -0.8127 | -0.8018 | -0.8019 | 0.7711 | 0.8043 | 0.8033 | 0.9757 | 1.0057 | 1.0050 |

Note: All the estimators are the average value over 5,000 simulations. For each parameter value pair, the five rows in each column correspond to normal, uniform, exponential, mixture of two normals and student-$t$ distributions. All the above distributions are standardized to zero mean and unit variance.

# Bibliography

[1] Bao, Y. "The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution. *Ecomometric Theory*, 23:1013-1021, 2007a.

[2] Bao, Y. "Finite-sample properties of forecasts from the stationary first-order autoregressive model under a general error distribution. *Ecomometric Theory*, 23:767-773, 2007b.

[3] Bao, Y. and A. Ullah. "Expectation of quadratic forms in normal and nonnormal variables with applications. *Journal of Statistical Planning and Inference*, 140:1193-1205, 2010.

[4] Bao, Y. and A. Ullah. "The second-order bias and mean squared error of estimators in time-series models. *Journal of Econometrics*, 140:650-669, 2007.

[5] Bao, Y. and A. Ullah. "Higher-order bias and MSE of nonlinear estimators. *Pakistan Journal of Statistics*, 25:587-594, 2009.

[6] Ullah, A. *Finite Sample Ecomometrics*. Oxford Universit Press, New York, 2004.

# Chapter 3

# Estimation Bias and Feasible Conditional Forecasts from the First-Order Moving Average Model

The quasi maximum likelihood estimator (QMLE) of parameters in the first-order moving average model can be biased in finite samples. We develop the second-order analytical bias of the QMLE and investigate whether this estimation bias can lead to biased feasible optimal forecasts conditional on the available sample observations. We find that the feasible multiple-step-ahead forecasts are unbiased under any nonnormal distribution and the one-step-ahead forecast is unbiased under symmetric distributions.

## 3.1 Introduction

Forecasts are made to guide decisions in many fields and the reduced-from time-series models have been used commonly. It is well known that under a quadratic loss function, the optimal conditional point forecast is the conditional expectation. Typically, optimal forecasts are made feasible by replacing unknown model parameters with estimated parameters based on the sample data. One can argue that parameter estimation uncertainty should vanish as the sample size grows. Yet, it is not uncommon that for many economic time series, the sample size can be quite limited. Thus, one would naturally wonder how the model estimation uncertainty will affect the constructed feasible forecasts in finite samples. For example, one might ask whether the bias in parameter estimation will produce a biased feasible forecast.

There has been some literature on the finite-sample issues associated with forecasts based on the autoregressive (AR) models, see Phillips (1979) and Bao (2007) and references therein. Schmidt (1977) discussed the small-sample properties of dynamic forecasts from AR models with exogenous variables. To our best knowledge, however, the literature has been silent about forecasts based on the moving average (MA) models. Compared with the AR models, the MA models can be used to model and forecast economic variables of less persistence and shorter memory. A prominent example is from Stock and Watson (2007), who found that the simple MA model of order 1 (MA(1)) works really well in describing the inflation rate change for the US economy.

Tanaka (1984) and Cordeiro and Klein (1994) derived the approximate bias of the maximum likelihood estimator of the MA parameter under the assumption of normally distributed data. Bao and Ullah (2007) considered the case when the data might be non-normally distributed, but restricted to a zero-mean MA model. These authors found that the estimation bias is inversely proportional to the sample size. Then immediately one may ask whether this bias problem will carry over to the forecasts.

The purpose of this paper is twofold. First, we derive the approximate bias of the conditional quasi maximum likelihood estimator (QMLE) of the parameters in an invertible MA(1) model with a possible nonzero mean and nonnormally distributed data. For most economic data, nonnormality is more a norm than an exception. So an interesting issue is how the distribution assumption will affect parameter estimation in finite samples. Second, we investigate whether the parameter estimation bias will lead to biased conditional forecasts in finite samples. It is found that the feasible multiple-step-ahead forecasts are unbiased under any nonnormal distribution and the one-step-ahead forecast is unbiased under symmetric distributions. Our theoretical results regarding the estimation and forecast biases are confirmed by a simulation study.

Throughout, $\boldsymbol{\iota}$ is a vector of ones, $\boldsymbol{e}_i$ is a null vector except its $i$th element is one, $\boldsymbol{I}$ is the identity matrix, and $\mathbf{0}$ is a null vector. The dimensions of vectors/matrices are to be read from the context, and thus we suppress the dimension subscripts in our notation. For a square matrix $\boldsymbol{A}$, we use $\boldsymbol{A}^*$ to denote $\boldsymbol{A} + \boldsymbol{A}'$.

## 3.2 Main Results

Consider the first-order moving average MA(1) model:

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \tag{3.1}$$

where $\varepsilon_t \sim$i.i.d.$(0, \sigma^2)$ and $|\theta| < 1$. Note that here we do not assume that $\varepsilon_t$ is normal, so the observable data $\boldsymbol{y} = (y_1, \cdots, y_T)'$ is in general nonnormally distributed. Let $\boldsymbol{\beta} = (\mu, \theta, \sigma^2)'$ denote the population parameters to be estimated.[1] Conditional on the observable sample $\boldsymbol{y}$, we are interested in forecasting the future values of $y$. Under a quadratic loss function, the conditional optimal one-step-ahead forecast is $y_{T+1|T} = \mu + \theta\varepsilon_T$ and the optimal forecasts beyond that are nothing but the unconditional mean of the process, $y_{T+h|T} = \mu$, $\forall h > 1$, see Hamilton (1994). In practice, we need to replace the unknown population parameters in the forecast formulae with their sample estimates to make the forecasts feasible. Moreover, for the one-step-ahead forecast, we also need to replace $\varepsilon_T$ with its estimate that can be inferred from the sample data. We use $\hat{y}_{T+h|T}$ to denote the feasible conditional forecasts. In particular, $\hat{y}_{T+1|T} = \hat{\mu} + \hat{\theta}\hat{\varepsilon}_T$ and $\hat{y}_{T+h|T} = \hat{\mu}$ for any $h > 1$, where $\hat{\mu}$ and $\hat{\theta}$ are the estimated parameters and $\hat{\varepsilon}_T$ is defined recursively as $\hat{\varepsilon}_T = y_T - \hat{\mu} - \hat{\theta}\hat{\varepsilon}_{T-1}$. Typically, we estimate $\boldsymbol{\beta}$ by the method of quasi maximum likelihood (QML) by maximizing a Gaussian likelihood function even though the true process is nonnormal.

The QML estimator (QMLE) $\hat{\boldsymbol{\beta}}$ , though consistent under typical conditions, is biased in finite samples. Tanaka (1984) and Cordeiro and Klein (1994) derived the second-order bias of $\hat{\boldsymbol{\beta}}$ , up to $O(T^{-1})$, when the true distribution of $\varepsilon_t$ is Gaussian (and thus the QMLE is the maximum likelihood estimator (MLE)). Bao and Ullah (2007) relaxed the assumption of normality, but focused on a zero-mean MA(1) process.

[1]Throughout, we suppress the subscript 0 for the true parameter value for notational convenience.

### 3.2.1 Estimation Bias

Conditional on $\varepsilon_0 = 0$, the average Gaussian log likelihood function of the observable data $\boldsymbol{y}$ is

$$L(\boldsymbol{\beta}|\varepsilon_0 = 0) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2T\sigma^2}, \tag{3.2}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_T)'$ and $\varepsilon_t$ is defined recursively from $\varepsilon_t = y_t - \mu - \theta\varepsilon_{t-1}$ starting with $\varepsilon_0 = 0$. To derive the finite-sample bias of the QMLE, we can follow Bao and Ullah (2007) to implement a stochastic expansion $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{\beta}_{-1/2} + \boldsymbol{\beta}_{-1} + o_P(T^{-1})$ (see the appendix), where $\boldsymbol{\beta}_{-i/2} = O_P(T^{-i/2})$,

$$\boldsymbol{\beta}_{-1/2} = \begin{pmatrix} \frac{\boldsymbol{a}'\boldsymbol{\varepsilon}}{\boldsymbol{a}'\boldsymbol{a}} \\ \frac{\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)} \\ \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{T} - \sigma^2 \end{pmatrix},$$

$$\boldsymbol{\beta}_{-1} = \begin{pmatrix} -\frac{\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)\boldsymbol{a}'\boldsymbol{a}} + \frac{2\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)(\boldsymbol{a}'\boldsymbol{a})^2} \\ \frac{3\mathrm{tr}(\boldsymbol{A}_1^3+\boldsymbol{A}_1'\boldsymbol{A}_1^2)(\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon})^2}{\sigma^4\mathrm{tr}^3(\boldsymbol{A}_2)} - \frac{\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon}}{\sigma^4\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)\boldsymbol{a}'\boldsymbol{a}} + \frac{\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)(\boldsymbol{a}'\boldsymbol{a})^2} + \frac{\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)} \\ -\frac{\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}}{T\boldsymbol{a}'\boldsymbol{a}} - \frac{(\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon})^2}{T\sigma^2\mathrm{tr}(\boldsymbol{A}_2)} \end{pmatrix},$$

with $\boldsymbol{a} = \boldsymbol{C}^{-1}\boldsymbol{\iota}$, $\boldsymbol{A}_1 = \boldsymbol{C}^{-1}\boldsymbol{B}$, $\boldsymbol{A}_2 = 2\boldsymbol{A}_1^2 + \boldsymbol{A}_1'\boldsymbol{A}_1$, $\boldsymbol{B} = \partial\boldsymbol{C}/\partial\theta$, and $\boldsymbol{C}$ being a $T \times T$ tridiagonal matrix with main diagonal elements 1, super-diagonal elements 0, and sub-diagonal elements $\theta$. We can check that $\boldsymbol{A}_1$ is strictly lower triangular. Then immediately $\mathbb{E}(\boldsymbol{\beta}_{-1/2}) = \boldsymbol{0}$ and

$$\begin{aligned} \mathbb{E}(\hat{\mu} - \mu) &= -\frac{\boldsymbol{a}'\boldsymbol{A}_1^*\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon})}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)\boldsymbol{a}'\boldsymbol{a}} + \frac{2\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{a}'\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon})}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)(\boldsymbol{a}'\boldsymbol{a})^2} + o(T^{-1}) \\ &= 0 + o(T^{-1}), \end{aligned}$$

since $\boldsymbol{f}'\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}) = \mathbb{E}(\varepsilon_i^3)\boldsymbol{f}'\text{diag}(\boldsymbol{A}_1) = 0$ for any vector $\boldsymbol{f}$. For evaluating the biases of $\hat{\theta}$ and $\hat{\sigma}^2$, we need expectations of second-order quadratic forms in $\boldsymbol{\varepsilon}$. From Ullah (2004, p. 187), for any matrices $\boldsymbol{N}_1$ and $\boldsymbol{N}_2$, $\mathbb{E}(\boldsymbol{\varepsilon}'\boldsymbol{N}_1\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{N}_2\boldsymbol{\varepsilon}) = \sigma^4[\gamma_2\text{tr}(\boldsymbol{N}_1 \odot \boldsymbol{N}_2)+\text{tr}(\boldsymbol{N}_1)\text{tr}(\boldsymbol{N}_2)$ $+\text{tr}(\boldsymbol{N}_1\boldsymbol{N}_2)+\text{tr}(\boldsymbol{N}_1'\boldsymbol{N}_2)]$, where $\odot$ is the Hadamard (element by element) product operator and $\gamma_2$ is the excess kurtosis coefficient of the distribution of $\varepsilon_t$. Again, since $\boldsymbol{A}_1$ is strictly lower triangular, $\text{tr}(\boldsymbol{A}_1) = \text{tr}(\boldsymbol{A}_1\boldsymbol{A}_1) = \text{tr}(\boldsymbol{A}_1 \odot \boldsymbol{A}_1) = \text{tr}(\boldsymbol{A}_1 \odot \boldsymbol{A}_2) = 0$. This leads to

$$\begin{aligned}
\mathbb{E}(\hat{\theta} - \theta) &= \frac{3\text{tr}(\boldsymbol{A}_1^3 + \boldsymbol{A}_1'\boldsymbol{A}_1^2)\text{tr}(\boldsymbol{A}_1'\boldsymbol{A}_1)}{\text{tr}^3(\boldsymbol{A}_2)} - \frac{\text{tr}(\boldsymbol{A}_1^*\boldsymbol{A}_2^*)}{2\text{tr}^2(\boldsymbol{A}_2)} - \frac{\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}}{\text{tr}(\boldsymbol{A}_2)\boldsymbol{a}'\boldsymbol{a}} + o(T^{-1}), \\
\mathbb{E}(\hat{\sigma}^2 - \sigma^2) &= -\frac{\sigma^2}{T} - \frac{\sigma^2\text{tr}(\boldsymbol{A}_1'\boldsymbol{A}_1)}{T\text{tr}(\boldsymbol{A}_2)} + o(T^{-1}),
\end{aligned}$$

which suggests that up to order $O(T^{-1})$, $\mathbb{E}(\hat{\theta} - \theta)$ and $\mathbb{E}(\hat{\sigma}^2 - \sigma^2)$ are both robust to the distribution of the data. Utilizing the special structure of the matrix $\boldsymbol{A}_1$, we can verify

$$\begin{aligned}
\boldsymbol{a}'\boldsymbol{a} &= \frac{T}{(1+\theta)^2} + O(1), \\
\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a} &= \frac{T}{(1+\theta)^3} + O(1), \\
\text{tr}(\boldsymbol{A}_2) &= \frac{T}{1-\theta^2} + O(1), \\
\text{tr}(\boldsymbol{A}_1'\boldsymbol{A}_1) &= \frac{T}{1-\theta^2} + O(1), \\
\text{tr}(\boldsymbol{A}_1^*\boldsymbol{A}_2^*) &= -\frac{8T\theta}{(1-\theta^2)^2} + O(1), \\
\text{tr}(\boldsymbol{A}_1^3 + \boldsymbol{A}_1'\boldsymbol{A}_1^2) &= -\frac{T\theta}{(1-\theta^2)^2} + O(1).
\end{aligned}$$

Upon substitution,

$$\mathbb{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{T}\begin{pmatrix} 0 \\ -1 + 2\theta \\ -2\sigma^2 \end{pmatrix} + o(T^{-1}), \tag{3.3}$$

Tanaka (1984) and Cordeiro and Klein (1994) derived the second-order bias of the MLE in the MA(1) model with normally distributed $\varepsilon_t$. Not surprisingly, as indicated earlier, the bias of the QMLE is robust to the distribution, and (3.3) coincides with Tanaka (1984) and Cordeiro and Klein (1994).[2] Note that (3.3) indicates that up to the second order, $\hat{\mu}$ is unbiased, $\hat{\theta}$ is upward biased when $\theta > 0.5$ but downward biased when $\theta < 0.5$, and unbiased when $\theta = 0.5$, and $\hat{\sigma}^2$ is always downward biased.

### 3.2.2 Forecast Bias

We have already seen in the previous subsection that both $\hat{\mu}$ and $\hat{\theta}$ are biased in finite samples. Now we investigate whether their biases can be translated into forecast bias. (Note that the feasible forecasts depend on $\hat{\mu}$ and $\hat{\theta}$, but not $\hat{\sigma}^2$.) First we note that since $\hat{y}_{T+h|T} = \hat{\mu}$ for $h > 1$, then up to $O(T^{-1})$, the multiple-step-ahead forecasts $\hat{y}_{T+h|T}$ are unbiased under any distribution.

To derive the one-step-ahead feasible forecast bias, we first write

$$\begin{aligned}
\hat{y}_{T+1|T} - y_{T+1} &= \hat{\mu} - \mu + \hat{\theta}\hat{\varepsilon}_T - \varepsilon_{T+1} - \theta\varepsilon_T \\
&= \hat{\mu} - \mu + \hat{\theta}\sum_{i=0}^{T-1}(-\hat{\theta})^i(y_{T-i} - \hat{\mu}) - \varepsilon_{T+1} - \theta\varepsilon_T \\
&= \hat{\mu} - \mu + \hat{\theta}\sum_{i=0}^{T-1}(-\hat{\theta})^i(y_{T-i} - \mu) + \hat{\theta}(\mu - \hat{\mu})\sum_{i=0}^{T-1}[(-\hat{\theta})^i] \\
&\quad -\varepsilon_{T+1} - \theta\varepsilon_T. \qquad (3.4)
\end{aligned}$$

[2] In Tanaka (1984) and Cordeiro and Klein (1994), the MA(1) parameter is $-\theta$. One can see that $\mathbb{E}(\hat{\theta}-\theta) = -\mathbb{E}((-\hat{\theta}) - (-\theta))$. They gave $\mathbb{E}((-\hat{\theta}) - (-\theta)) = (1 + 2(-\theta))/T$.

Thus the feasible forecast bias is

$$\mathbb{E}(\hat{y}_{T+1|T}-y_{T+1})=\sum_{i=0}^{T-1}(-1)^i\mathbb{E}[\hat{\theta}^{i+1}(\varepsilon_{T-i}+\theta\varepsilon_{T-i-1})]-(\hat{\mu}-\mu)\sum_{i=0}^{T-1}(-1)^i\mathbb{E}(\hat{\theta}^{i+1})+o(T^{-1}).$$

Since $T$ terms are added up in the above bias expression, we need to expand $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}$ to order $O_P(T^{-2})$ for deriving the order $O(T^{-1})$ one-step-ahead feasible forecast bias. We also need expansions of $\hat{\theta}^{i+1}, i=0,\cdots,T-1$, to order $O_P(T^{-2})$. Suppose $\hat{\theta}^{i+1}=\theta^{i+1}+\theta_{-1/2}^{(i+1)}+\theta_{-1}^{(i+1)}+\theta_{-3/2}^{(i+1)}+\theta_{-2}^{(i+1)}+o_P(T^{-2})$, where $\theta_{-j/2}^{(i+1)}=O_P(T^{-j/2})$, then up to the order of approximation, we put

$$\begin{aligned}
\mathbb{E}(\hat{y}_{T+1|T}-y_{T+1}) &= \sum_{i=0}^{T-1}(-1)^i\sum_{j=1}^{4}\mathbb{E}\left[\theta_{-j/2}^{(i+1)}(\varepsilon_{T-i}+\theta\varepsilon_{T-i-1})\right] \\
&\quad -\sum_{i=0}^{T-1}(-1)^i\left[\theta^{i+1}\sum_{j=2}^{4}\mathbb{E}(\mu_{-j/2})+\sum_{j=1}^{3}\mathbb{E}(\theta_{-1/2}^{(i+1)}\mu_{-j/2})\right] \\
&\quad -\sum_{i=0}^{T-1}(-1)^i\left[\sum_{j=1}^{2}\mathbb{E}(\theta_{-1}^{(i+1)}\mu_{-j/2})+\mathbb{E}(\theta_{-3/2}^{(i+1)}\mu_{-1/2})\right]+o(T^{-1}),
\end{aligned}$$

where $\mu_{-j/2}=O_P(T^{-j/2})$, $j=1,\cdots,4$, are the stochastic terms in the order $O_P(T^{-2})$ expansion of $\hat{\mu}-\mu$.

Bao (2007) showed that for $q=i+1$, $\theta_{-j/2}^{(q)}$, $j=1,\cdots,4$, are as follows:

$$\begin{aligned}
\theta_{-1/2}^{(q)} &= q\theta^{q-1}\theta_{-1/2}, \\
\theta_{-1}^{(q)} &= q\theta^{q-1}\theta_{-1}+\frac{q(q-1)}{2}\theta^{q-2}\theta_{-1/2}^2, \\
\theta_{-3/2}^{(q)} &= q\theta^{q-1}\theta_{-3/2}+q(q-1)\theta^{q-2}\theta_{-1/2}\theta_{-1}+\frac{q(q-1)(q-2)}{6}\theta^{q-3}\theta_{-1/2}^3, \\
\theta_{-2}^{(q)} &= q\theta^{q-1}\theta_{-2}+q(q-1)\theta^{q-2}\theta_{-1/2}\theta_{-3/2}+\frac{q(q-1)}{2}\theta^{q-2}\theta_{-1}^2 \\
&\quad +\frac{q(q-1)(q-2)}{2}\theta^{q-3}\theta_{-1/2}^2\theta_{-1}+\frac{q(q-1)(q-2)(q-3)}{24}\theta^{q-4}\theta_{-1/2}^4,
\end{aligned}$$

where $\theta_{-j/2}=O_P(T^{-j/2})$, $j=1,\cdots,4$, are the stochastic terms in the order $O_P(T^{-2})$

expansion of $\hat{\theta}-\theta$. By writing $\varepsilon_{T-i}+\theta\varepsilon_{T-i-1} = (\boldsymbol{e}_{T-i}+\theta\boldsymbol{e}_{T-i-1})'\boldsymbol{\varepsilon}$ (when $i = T-1$, define $\boldsymbol{e}_0 = \mathbf{0}$ since $\varepsilon_0 = 0$) and defining

$$\begin{aligned}
\boldsymbol{f}_{1i} &= (-1)^i(i+1)\theta^i(\boldsymbol{e}_{T-i}+\theta\boldsymbol{e}_{T-i-1}),\\
\boldsymbol{f}_{2i} &= (-1)^i i(i+1)\theta^{i-1}(\boldsymbol{e}_{T-i}+\theta\boldsymbol{e}_{T-i-1}),\\
\boldsymbol{f}_{3i} &= (-1)^i i(i^2-1)\theta^{i-2}(\boldsymbol{e}_{T-i}+\theta\boldsymbol{e}_{T-i-1}),\\
\boldsymbol{f}_{4i} &= (-1)^i i(i^2-1)(i-2)\theta^{i-3}(\boldsymbol{e}_{T-i}+\theta\boldsymbol{e}_{T-i-1}),
\end{aligned}$$

then upon substituting $\theta^{(q)}_{-1/2}$, we have

$$\begin{aligned}
\mathbb{E}(\hat{y}_{T+1|T}-y_{T+1}) &= \sum_{i=0}^{T-1}\boldsymbol{f}'_{1i}\mathbb{E}[\boldsymbol{\varepsilon}(\theta_{-1/2}+\theta_{-1}+\theta_{-3/2}+\theta_{-2})]\\
&\quad+\frac{1}{2}\sum_{i=0}^{T-1}\boldsymbol{f}'_{2i}\mathbb{E}[\boldsymbol{\varepsilon}(\theta^2_{-1/2}+\theta^2_{-1}+2\theta_{-1/2}\theta_{-1}+2\theta_{-1/2}\theta_{-3/2})]\\
&\quad+\frac{1}{6}\sum_{i=0}^{T-1}\boldsymbol{f}'_{3i}\mathbb{E}[\boldsymbol{\varepsilon}(\theta^3_{-1/2}+3\theta^2_{-1/2}\theta_{-1})]+\frac{1}{24}\sum_{i=0}^{T-1}\boldsymbol{f}'_{4i}\mathbb{E}(\boldsymbol{\varepsilon}\theta^4_{-1/2})\\
&\quad-\sum_{j=3}^{4}\mathbb{E}(\mu_{-j/2})\sum_{i=0}^{T-1}(-1)^i\theta^{i+1}\\
&\quad-\mathbb{E}\left[\theta_{-1/2}\sum_{j=1}^{3}\mu_{-j/2}+\theta_{-1}\sum_{j=1}^{2}\mu_{-j/2}+\theta_{-3/2}\mu_{-1/2}\right]\sum_{i=0}^{T-1}(-1)^i(i+1)\theta^i\\
&\quad-\frac{1}{2}\mathbb{E}\left[\theta^2_{-1/2}\sum_{j=1}^{2}\mu_{-j/2}+2\theta_{-1/2}\theta_{-1}\mu_{-1/2}\right]\sum_{i=0}^{T-1}(-1)^i i(i+1)\theta^{i-1}\\
&\quad-\frac{1}{6}\mathbb{E}(\theta^3_{-1/2}\mu_{-1/2})\sum_{i=0}^{T-1}(-1)^i i(i+1)(i-1)\theta^{i-2}+o(T^{-1}).
\end{aligned}$$

Note that $\sum_{i=0}^{T-1}(-1)^i\theta^{i+1}$, $\sum_{i=0}^{T-1}(-1)^i(i+1)\theta^i$, $\sum_{i=0}^{T-1}(-1)^i(i+1)i\theta^{i-1}$, and $\sum_{i=0}^{T-1}(-1)^i i(i+1)(i-1)\theta^{i-2}$ are all of order $O(1)$, and $\mathbb{E}(\theta_{-1/2}\mu_{-1/2}) = [\sigma^2\mathrm{tr}(\boldsymbol{A}_2)\boldsymbol{a}'\boldsymbol{a}]^{-1}\boldsymbol{a}'\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}) = 0$.

Therefore,

$$\begin{aligned}\mathbb{E}(\hat{y}_{T+1|T} - y_{T+1}) &= \sum_{i=0}^{T-1} \boldsymbol{f}'_{1i}\mathbb{E}[\boldsymbol{\varepsilon}(\theta_{-1/2} + \theta_{-1} + \theta_{-3/2} + \theta_{-2})] \\ &+\frac{1}{2}\sum_{i=0}^{T-1} \boldsymbol{f}'_{2i}\mathbb{E}[\boldsymbol{\varepsilon}(\theta^2_{-1/2} + \theta^2_{-1} + 2\theta_{-1/2}\theta_{-1} + 2\theta_{-1/2}\theta_{-3/2})] \\ &+\frac{1}{6}\sum_{i=0}^{T-1} \boldsymbol{f}'_{3i}\mathbb{E}[\boldsymbol{\varepsilon}(\theta^3_{-1/2} + 3\theta^2_{-1/2}\theta_{-1})] \\ &+\frac{1}{24}\sum_{i=0}^{T-1} \boldsymbol{f}'_{4i}\mathbb{E}(\boldsymbol{\varepsilon}\theta^4_{-1/2}) + o(T^{-1}). \end{aligned} \quad (3.5)$$

When substituting $\theta_{-i/2}$, $i = 1, \cdots, 4$, from the appendix into (3.5), we notice that all the expectations are of the form: $\mathbb{E}[\boldsymbol{\varepsilon} \prod_{i=1}^{m}(\boldsymbol{\varepsilon}'\boldsymbol{A}_i\boldsymbol{\varepsilon})]$ for $m$ up to 4. Under a general non-normal distribution, the results for $m$ up to 3 can be found in Bao and Ullah (2010), but $\mathbb{E}[\boldsymbol{\varepsilon} \prod_{i=1}^{4}(\boldsymbol{\varepsilon}'\boldsymbol{A}_i\boldsymbol{\varepsilon})]$ has not been developed in the literature and deriving its analytical expression is beyond the scope of this paper. Nevertheless, for symmetric distributions, not necessarily normal, $\mathbb{E}[\boldsymbol{\varepsilon} \prod_{i=1}^{m}(\boldsymbol{\varepsilon}'\boldsymbol{A}_i\boldsymbol{\varepsilon})] = 0$. Then immediately, $\mathbb{E}(\hat{y}_{T+1|T} - y_{T+1}) = 0 + o(T^{-1})$, namely, the 1-step-ahead forecast is also unbiased. We summarize our results in the following proposition.[3]

**Proposition**: *For the first-order moving average model with an initial error term* $\varepsilon_0 = 0$, *the approximate bias of the QMLE* $\hat{\boldsymbol{\beta}} = (\hat{\mu}, \hat{\theta}, \hat{\sigma}^2)'$, *given by* $(0, -1 + 2\theta, -2\sigma^2)'/T$, *under any distribution of* $\varepsilon_t$. *Under a quadratic loss function, the feasible conditional optimal multiple-step-ahead forecasts based on the QMLE of model parameters are unbiased up to order* $O(T^{-1})$ *under any distribution, and the conditional one-step-ahead forecast is unbiased up*

[3]When $\mu = 0$ and we estimate the model $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, the bias of the QMLE is given by $(-\theta, -\sigma^2)'/T$ as derived in Bao and Ullah (2007). It is the same as the bias of the MLE derived in Tanaka (1984) and Cordeiro and Klein (1994). The unbiasedness of the feasible forecasts also holds.

*to order $O(T^{-1})$ under any symmetric distribution.*

### 3.2.3 A Simulation Experiment

This subsection reports a simulation experiment to illustrate the finite-sample properties of the QMLE. The results are based on 100,000 simulations. We set $\mu = 0.2, 0.5, 1$, $\theta = 0.2, 0.5, 0.9$, $\sigma^2 = 1$, $T = 30, 50, 100, 200$. We experimented with normal, exponential, uniform, scale mixtures of normals, and $t$ distribution with 10 degrees of freedom. To save space, we report only the results from the $t$ distribution, whereas the results under the other distributions, with similar findings, are available upon request from the corresponding author. In Tables 1 and 2, $\tilde{\theta} = \hat{\theta} - (-1 + 2\hat{\theta})/T$ and $\tilde{\sigma}^2 = \hat{\sigma}^2 - (-2\hat{\sigma}^2)/T$ denote the feasible bias-corrected $\hat{\theta}$ and $\hat{\sigma}^2$, respectively, FE1, FE2, and FE5 are the 1-, 2- and 5-step-ahead feasible forecast errors, with $\sqrt{\text{M1}}$, $\sqrt{\text{M2}}$, and $\sqrt{\text{M5}}$ being the corresponding square root of the mean squared forecast errors (MSFEs).[4]

We observe first that the bias behaviors of $\hat{\mu}, \hat{\theta}$, and $\hat{\sigma}^2$ match what our theory predicts: up to the second order, $\hat{\mu}$ is unbiased, $\hat{\theta}$ maybe upward or downward biased (depending on whether $\theta$ is greater than or less than 0.5), and $\hat{\sigma}^2$ is always downward biased. The approximate bias results given by (3.3) generally capture really well the true biases. In particular, $\hat{\theta}$ tends to be severely biased in small samples when $\theta$ is small (0.2), and $\tilde{\theta}$ corrects for the bias substantially. Second, in all cases, the forecast errors are very close to zero and this is consistent with the proposition that under any symmetric distribution, the

[4]In this paper we do not attempt to derive the analytical MSFE. For the one-step feasible forecast error (3.4), one can check that we need results on the expectation of nonnormal quadratic forms of order 5 like $\mathbb{E}[\prod_{i=1}^{5}(\boldsymbol{\varepsilon}'\boldsymbol{A}_i\boldsymbol{\varepsilon})]$. Unfortunately, an analytical expression for $\mathbb{E}[\prod_{i=1}^{5}(\boldsymbol{\varepsilon}'\boldsymbol{A}_i\boldsymbol{\varepsilon})]$ is not available and deriving such an expression is beyond the scope of this paper.

feasible forecasts are unbiased. Third, for a given sample size, the MSFEs do not vary much over the values of $\mu$ and $\theta$ when th forecast horizon is 1, but increase with $\theta$ for the 2- and 5-step forecasts. Moreover, the MSFEs do not necessarily increase with the forecast horizon, perhaps largely due to the presence of parameter estimation uncertainty. Analytical results regarding the MSFEs, which unfortunately are not available, might provide us with a better understanding of the finite-sample behavior of the MSFEs.

Table 3.1: Finite-Sample Properties of QMLE in MA(1) ($T = 30, 50$)

| $T$ | $\mu$ | $\theta$ | $\hat{\mu}$ | $\hat{\theta}$ | $\hat{\sigma}^2$ | $\tilde{\theta}$ | $\tilde{\sigma}^2$ | FE1 | FE2 | FE5 | $\sqrt{\text{M1}}$ | $\sqrt{\text{M2}}$ | $\sqrt{\text{M5}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 30 | 0.2 | 0.2 | 0.1978 | 0.1693 | 0.9552 | 0.1913 | 1.0189 | -0.0053 | -0.0044 | 0.0054 | 1.0541 | 1.0649 | 1.0546 |
| | | 0.5 | 0.2002 | 0.4978 | 0.9558 | 0.4979 | 1.0195 | 0.0054 | -0.0085 | 0.0066 | 1.0672 | 1.1704 | 1.1665 |
| | | 0.9 | 0.1954 | 0.9133 | 0.9580 | 0.8857 | 1.0218 | 0.0029 | -0.0065 | -0.0100 | 1.0581 | 1.4103 | 1.4028 |
| | 0.5 | 0.2 | 0.4984 | 0.1677 | 0.9581 | 0.1898 | 1.0220 | -0.0052 | -0.0109 | 0.0007 | 1.0554 | 1.0573 | 1.0540 |
| | | 0.5 | 0.4966 | 0.4973 | 0.9562 | 0.4975 | 1.0199 | -0.0004 | -0.0071 | 0.0063 | 1.0525 | 1.1633 | 1.1657 |
| | | 0.9 | 0.5019 | 0.9127 | 0.9578 | 0.8852 | 1.0217 | -0.0020 | -0.0049 | -0.0069 | 1.0413 | 1.3951 | 1.4020 |
| | 1 | 0.2 | 0.9985 | 0.1719 | 0.9553 | 0.1938 | 1.0190 | 0.0002 | 0.0028 | 0.0027 | 1.0599 | 1.0506 | 1.0568 |
| | | 0.5 | 0.9971 | 0.5001 | 0.9555 | 0.5001 | 1.0192 | 0.0070 | -0.0041 | -0.0001 | 1.0523 | 1.1707 | 1.1578 |
| | | 0.9 | 0.9964 | 0.9136 | 0.9591 | 0.8861 | 1.0231 | 0.0022 | -0.0037 | -0.0022 | 1.0404 | 1.4064 | 1.4075 |
| 50 | 0.2 | 0.2 | 0.1987 | 0.1833 | 0.9847 | 0.1960 | 1.0240 | 0.0028 | -0.0053 | 0.0157 | 1.0412 | 1.0510 | 1.0385 |
| | | 0.5 | 0.2010 | 0.5004 | 0.9853 | 0.5004 | 1.0247 | -0.0069 | -0.0159 | 0.0073 | 1.0381 | 1.1513 | 1.1594 |
| | | 0.9 | 0.2032 | 0.9137 | 0.9824 | 0.8971 | 1.0217 | -0.0165 | -0.0192 | -0.0210 | 1.0410 | 1.3885 | 1.3902 |
| | 0.5 | 0.2 | 0.5012 | 0.1855 | 0.9866 | 0.1981 | 1.0260 | -0.0030 | -0.0032 | 0.0060 | 1.0436 | 1.0487 | 1.0461 |
| | | 0.5 | 0.4994 | 0.4995 | 0.9896 | 0.4995 | 1.0291 | -0.0078 | -0.0079 | -0.0045 | 1.0491 | 1.1529 | 1.1475 |
| | | 0.9 | 0.5032 | 0.9143 | 0.9825 | 0.8977 | 1.0218 | -0.0024 | -0.0043 | -0.0273 | 1.0395 | 1.3948 | 1.3891 |
| | 1 | 0.2 | 1.0008 | 0.1847 | 0.9843 | 0.1973 | 1.0237 | -0.0008 | -0.0084 | -0.0024 | 1.0355 | 1.0400 | 1.0393 |
| | | 0.5 | 1.0031 | 0.4987 | 0.9844 | 0.4987 | 1.0238 | -0.0130 | -0.0137 | -0.0002 | 1.0406 | 1.1548 | 1.1547 |
| | | 0.9 | 1.0033 | 0.9138 | 0.9827 | 0.8972 | 1.0220 | -0.0121 | -0.0178 | -0.0079 | 1.0371 | 1.3887 | 1.3888 |

Table 3.2: Finite-Sample Properties of QMLE in MA(1) ($T = 100, 200$)

| $T$ | $\mu$ | $\theta$ | $\hat{\mu}$ | $\hat{\theta}$ | $\hat{\sigma}^2$ | $\tilde{\theta}$ | $\tilde{\sigma}^2$ | FE1 | FE2 | FE5 | $\sqrt{\text{M1}}$ | $\sqrt{\text{M2}}$ | $\sqrt{\text{M5}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 0.2 | 0.2 | 0.2005 | 0.1930 | 1.0064 | 0.1992 | 1.0265 | 0.2005 | 0.0011 | -0.0042 | 1.0209 | 1.0376 | 1.0467 |
| | | 0.5 | 0.2007 | 0.5002 | 1.0061 | 0.5002 | 1.0262 | 0.2007 | 0.0009 | -0.0102 | 1.0243 | 1.1401 | 1.1435 |
| | | 0.9 | 0.1996 | 0.9093 | 1.0042 | 0.9011 | 1.0243 | 0.1996 | -0.0013 | -0.0067 | 1.0200 | 1.3694 | 1.3796 |
| | 0.5 | 0.2 | 0.5004 | 0.1929 | 1.0058 | 0.1991 | 1.0260 | 0.5004 | 0.0091 | -0.0034 | 1.0245 | 1.0386 | 1.0474 |
| | | 0.5 | 0.5014 | 0.5007 | 1.0057 | 0.5007 | 1.0258 | 0.5014 | -0.0063 | 0.0002 | 1.0289 | 1.1457 | 1.1369 |
| | | 0.9 | 0.4998 | 0.9091 | 1.0046 | 0.9009 | 1.0247 | 0.4998 | -0.0100 | 0.0101 | 1.0204 | 1.3757 | 1.3825 |
| | 1 | 0.2 | 0.9992 | 0.1938 | 1.0063 | 0.1999 | 1.0265 | 0.9992 | -0.0005 | -0.0026 | 1.0228 | 1.0385 | 1.0437 |
| | | 0.5 | 1.0009 | 0.4997 | 1.0064 | 0.4997 | 1.0265 | 1.0009 | -0.0071 | -0.0017 | 1.0270 | 1.1394 | 1.1428 |
| | | 0.9 | 0.9981 | 0.9090 | 1.0038 | 0.9008 | 1.0239 | 0.9981 | -0.0062 | 0.0087 | 1.0098 | 1.3688 | 1.3789 |
| 200 | 0.2 | 0.2 | 0.1998 | 0.1965 | 1.0171 | 0.1995 | 1.0273 | 0.0018 | 0.0007 | -0.0038 | 1.0120 | 1.0534 | 1.0440 |
| | | 0.5 | 0.2006 | 0.5001 | 1.0177 | 0.5001 | 1.0279 | 0.0007 | 0.0079 | -0.0070 | 1.0118 | 1.1443 | 1.1516 |
| | | 0.9 | 0.1984 | 0.9053 | 1.0176 | 0.9012 | 1.0278 | -0.0061 | -0.0092 | 0.0134 | 1.0180 | 1.3748 | 1.3816 |
| | 0.5 | 0.2 | 0.4997 | 0.1967 | 1.0203 | 0.1997 | 1.0305 | 0.0014 | -0.0080 | -0.0126 | 1.0155 | 1.0473 | 1.0428 |
| | | 0.5 | 0.5001 | 0.5007 | 1.0169 | 0.5007 | 1.0271 | 0.0134 | 0.0026 | 0.0027 | 1.0169 | 1.1457 | 1.1507 |
| | | 0.9 | 0.5003 | 0.9051 | 1.0163 | 0.9011 | 1.0265 | 0.0166 | 0.0209 | -0.0034 | 1.0073 | 1.3698 | 1.3788 |
| | 1 | 0.2 | 0.9999 | 0.1968 | 1.0176 | 0.1998 | 1.0278 | 0.0088 | 0.0012 | -0.0030 | 1.0153 | 1.0427 | 1.0399 |
| | | 0.5 | 0.9989 | 0.4999 | 1.0180 | 0.4999 | 1.0281 | 0.0043 | -0.0070 | -0.0066 | 1.0119 | 1.1415 | 1.1395 |
| | | 0.9 | 0.9993 | 0.9047 | 1.0155 | 0.9007 | 1.0256 | -0.0016 | -0.0142 | 0.0116 | 1.0142 | 1.3629 | 1.3877 |

## 3.3 Concluding Remarks

In this paper, we have derived approximate bias of the QMLE of parameters in an invertible MA(1) model with possibly nonnormally distributed data. We then investigate whether the feasible conditional forecasts can be biased due to the bias in the QMLE in finite samples. It turns out that multiple-step-ahead forecasts are in fact unbiased, up to $O(T^{-1})$, under any distribution, and the one-step-ahead forecast is unbiased when the distribution is symmetric. This finding, together with the finding of similar properties from Bao (2007) regarding forecasts from AR processes, perhaps can partially alleviate for researchers the worry of parameter estimation uncertainty. Of course, one has yet to investigate the more general ARMA type models and other nonlinear time-series models before more affirmative conclusions can be made.

We have restricted ourselves to the invertible MA(1) model, but recently noncausal AR and noninvertible MA models have been used by empirical researchers. For example, the noncausal AR model was used by Lanne and Luoto (2012) to forecast the US inflation rate, and the noninvertible MA model was used by Huang and Pawitan (2000) and Breidt et al. (2001) to study the US unemployment rate and New Zealand/US exchange rate, respectively. Estimation and forecasting strategies in these cases are quite nonstandard, and studying the finite-sample properties of forecasts in this direction is a future subject of investigation.

# Bibliography


[1] Bao, Y. "Finite-sample properties of forecasts from the stationary first-order autoregressive model under a general error distribution". *Econometric Theory*, 23:767-773, 2007.

[2] Bao, Y. and A. Ullah. "The second-order bias and mean squared error of estimators in time series models". *Journal of Econometrics*, 140:650-669, 2007.

[3] Bao, Y. and A. Ullah. "Higher-order bias and MSE of nonlinear estimators". *Pakistan Journal of Statistics*, 25:587-594, 2009.

[4] Bao, Y. and A. Ullah. "Expectation of quadratic forms in normal and nonnormal variables with applications". *Journal of Statistical Planning and Inference*, 140:1193-1205, 2010.

[5] Breidt, J., R.A. Davis and A.A. Trindade. "Least absolute deviation estimation for all-pass time series models". *The Annals of Statistics*, 29:919-946, 2001.

[6] Cordeiro, G.M. and R. Klein. "Bias correction in ARMA models". *Statistics and Probability Letters*, 19:169-176, 1994.

[7] Hamilton, J. *Time Series Analysis*. Princeton University Press, New Jersey, 1994.

[8] Huang, J. and Y. Pawitan. "Quasi-likelihood estimation of noninvertible moving average processes". *Scandinavian Journal of Statistics*, 27:689-710, 2000.

[9] Lanne, M. and J. Luoto. "Has US inflations really become harder to forecast?". *Economics Letters*, 115:383-386, 2012.

[10] Phillips, P.C.B. "The sampling distribution of forecasts from a first-order autoregression". *Journal of Econometrics*, 9:241-261, 1979.

[11] Rilstone, P., V.K. Srivastava and A. Ullah. "The second-order bias and mean squared error of nonlinear estimators". *Journal of Econometrics*, 75:369-395, 1996.

[12] Schmidt, P. "Some small-sample evidence on the distribution of dynamic simulation forecasts". *Econometrica*, 45:997-1005, 1977.

[13] Stock, J.H. and M.W. Watson. "Why has U.S. inflation become harder to forecast?". *Journal of Money, Credit and Banking* 39:3-33, 2007.

[14] Tanaka, K. "An asymptotic expansion associated with the maximum likelihood estimators in ARMA models". *Journal of the Royal Statistical Society: Series B*, 46:58-67, 1984.

[15] Ullah, A. *Finite Sample Ecomometrics*. Oxford Universit Press, New York, 2004.

# Chapter 4

# Using Extreme Learning Machines for Out-of-Sample Prediction

In this paper, we apply the artificial neural network (ANN) model to out-of-sample prediction of financial return using a set of covariates. The main challenge in ANN model estimation is the multicolinearity between the large numbers of randomly generated hidden layers. We explore several methods to deal with the large dimension regressors, such as general inverse, ridge, pretest and principal components, which are also named extreme learning machines (ELM). We find that although the ELM methods sometimes fit perfectly for in-sample data, it has very poor out-of-sample forecast ability. We then introduce some modifications to the ELM method, which is a two step algorithm, where the first step uses ELM methods with some modifications to get a set of forecasts, and the second step combines the forecasts using principal components weighting scheme. Empirical results show that our method gives best forecast for annually aggregated equity premium among all the alternatives.

## 4.1 Introduction

The artificial neural network (ANN) model, as an flexible universal approximator, works efficiently in testing nonlinearity in the sense that it can approximate and thus detect any kind of nonlinear relationship, see Lee, White and Granger (1993, denoted by LWG thereafter), Lee, Xi and Zhang (2013, denoted by LXZ thereafter, 2014). In this paper, we apply the ANN model to forecast stock returns using a set of covariates. LWG uses a set randomly generated activation functions as the hidden layer in the ANN model and LXZ further extend the method to a lager set of randomly generated activation functions, which makes the test more robust to empirical applications. Both of LWG and LXZ apply the principal component (PC) method to solve the problem of multicolinearity between the randomly generated hidden layers, where a relatively small number of principal components are selected rather than including all the activation functions generated. In this paper, we first apply this method to forecast excess stock returns, and compare the results with some linear models as benchmarks. Results show that the this LWG method using a set of principal components does not give very good results for out-of-sample forecasting. Then we further explore several other methods rather than principal components to shrink the dimension of the activation function, these methods are named extreme learning machines (ELM) according to Huang, Wang and Lan (2011). Their ELM method does not reduce the dimension of the activation function, but instead use a general inverse or ridge estimator to estimate the ANN model with a large number of activation functions. Their paper shows that the ELM method can perfectly fit any nonlinear functions with the set of activation functions of dimension less than the total number of observations. We apply their methods to fit and forecast excess stock returns, and find that although their ELM methods using general inverse regression fit perfectly for in sample data, it has very poor forecast ability for out-of-sample data.

Based on these above results, we introduce some modifications to the ELM method following Huang ana Lee (2009), which we refer to as ELM-CFPC method. The ELM-CFPC method is a two step algorithm, the first step uses ELM methods with some modifications to get set of forecasts, and the second step combines the forecasts using principal components weighting scheme. We also compare this method with alternative models, including linear and nonlinear models, as well as forecast combination of ELM using other weighting schemes, including equal weight, Mellows criteria (Hansen, 2007, 2008). Empirical results show that the ELM-CFPC method gives best forecast for annual aggregated excess stock returns among all the linear and nonlinear alternative methods.

The rest of this paper is arranged as follows: part two reviews different extreme learning machines applying different shrinkage methods; part three introduces the two step algorithm which we call ELM-CFPC method; part four lists the alternative methods considered in this paper for our application of forecasting equity premium, and part five shows the results of application and part six concludes.

## 4.2 Extreme Learning Machines

The linear-augmented single hidden-layer feedforward ANN model has the following architecture:

$$y_t = f(\mathbf{x}_t, \theta) + \varepsilon_t := \mathbf{x}_t'\alpha + \sum_{j=1}^{q} \theta_j \psi\left(\mathbf{x}_t'\gamma_j\right) + \varepsilon_t, \tag{4.1}$$

where $t = 1, \ldots, n$, $\mathbf{x}_t = (x_{1,t}, \ldots x_{k,t})'$, $\theta = \left(\alpha', \beta', \gamma_1', \ldots, \gamma_q'\right)'$, $\alpha = (\alpha_1, \ldots, \alpha_k)'$, $\beta = (\beta_1, \ldots, \beta_q)'$, and $\gamma_j = (\gamma_{j,1}, \ldots, \gamma_{j,k})'$ for $j = 1, \ldots, q$, and $\psi(\cdot)$ is an activation function.[1] An example of the activation function is the logistic function $\psi(z) = (1 + \exp(z))^{-1}$. $\alpha$ is a column vector of connection strength from the input layer to the output layer; $\gamma_j$ is a conformable column vector of connection strength from the input layer to the hidden units, $j = 1, \ldots, q$; $\beta_j$ is a (scalar) connection strength from the hidden unit $j$ to the output unit, $j = 1, \ldots, q$; and $\psi$ is a squashing function (e.g., the logistic squasher) or a radial basis function. Input units $\mathbf{x}$ send signals to intermediate hidden units, then each of hidden unit produces an activation $\psi$ that then sends signals toward the output unit. The integer $q$ denotes the number of hidden units added to the affine (linear) network.

When then hidden layers of the ANN model does not need turning, the algorithm is called extreme learning machine (ELM) according to Huang, Wang and Lan (2011). And as a special case of ANN model, the set of $\gamma_j$'s are randomly generated in ELM, which is also proved in Bierens (1982). When the set of $\gamma_j$ is randomly generated, the hidden layers $\psi(x_t'\gamma_j)$ can be determined given a realization of $\gamma_j$, then the ANN model can be regarded as a linear function of the the covariants $x_t$ and the activation functions $\psi$. To estimate those linear coefficients, we can just apply the least squares method. However, as

[1] '$a := b$' means that $a$ is defined by $b$, while '$a =: b$' means that $b$ is defined by $a$.

the set of $\gamma_j$ is randomly generated with large dimension, there may exist multicolinearity problem. To solve this, several estimation methods could be used. Huang, Wang and Lan (2011) introduced using the general inverse in the OLS regression as well as ridge regression, referred to as basic ELM and ridge based ELM in this paper, respectively. In this paper, we further extend their methods.

### 4.2.1 Basic ELM

The above $N$ equations for the ANN model can be rewritten compactly as:

$$y = \Psi(x)\beta + \varepsilon \tag{4.2}$$

where $y = (y_1, ..., y_n)$, $\Psi(x) = [\Psi'(x_1'), ..., \Psi'(x_n')]'$, in which $\Psi(x_t') = [x_t', \psi(x_t'\gamma_1), ..., \psi(x_t'\gamma_q)]$, $\Psi(x)$ contains both the linear part and the nonlinear activation functions of $x_t$, $\beta = [\alpha', \theta_1, ..., \theta_q]'$.

Given randomly generated $\gamma$, the above system is linear in $\beta$. The basic ELM solves the linear system using the least squares method. However, in case of multicolinearity, the matrix of $\Psi(x)'\Psi(x)$ may not be full rank, as an alternative, the general inverse could be applied instead. The basic ELM method estimates the model by:

$$\hat{\beta}^b_{ELM} = \Psi(x)^\dagger y \tag{4.3}$$

where $\Psi(x)^\dagger$ is the Moore-Penrose generalized inverse of $\Psi(x)$.

The fact that least square estimators can give good approximation and can fit well for any functional form is shown in Huang and Chen (2007, 2008) and summarized theorem 2.4 in Huang, Whang and Lan (2011), that for any nonconstant piecewise continuous activation function, if the hidden layers are dense in $L^2$, then for any randomly generated $\gamma$ with dimension $q$, the norm of the approximation error goes to zero with probability one as $q$ goes to infinity if $\beta$ is determined by ordinary least square method. Nevertheless, their proof is for in-sample estimation only, we will apply the method to both in-sample regression and out-of-sample forecasting for excess stock returns in the next section.

### 4.2.2 Ridge based ELM

In the existence of multicolinearity between the $\psi(x_t'\gamma_j)$'s, various shrinkage methods could be used. In particular, the $L^2$ shrinkage method, also known as ridge regression, adds a penalty of the sum of squares of the coefficients to the least squares residual sum of squares, in this case, $\hat{\beta}$ solves the following minimization problem:

$$\hat{\beta}^r_{ELM} = \arg\min_{\beta} \left\{ (y - \Psi(x)\beta)'(y - \Psi(x)\beta) + \lambda\beta'\beta \right\} \tag{4.4}$$

the solution to the above equation can be written explicitly as $\hat{\beta}_{ELMr} = (\Psi(x)'\Psi(x) + \lambda I)^{-1}\Psi(x)'y$, where $\lambda$ is a small number, chosen to be $10^{-6}$ in this paper. We refer to this method ridge based ELM according Huang, Wang and Lan (2011). They suggested that the ridge based ELM works better than the basic ELM for in-sample fitting. We will check whether this method also works for out-of-sample forecasting in the next section as well.

### 4.2.3 PC based ELM

Instead of doing model shrinkage through optimization such as the ridge method, an alternative way to solve the multicolinearity problem between the $\Psi$ functions is to use the principal components in the regression. According to LWG and LXZ, the shrinkage by principal component method works well for nonlinearity testing with ANN models, so we also consider this method for forecasting in this paper. After getting the $\Psi$ functions of dimension $q$ from given randomly generated $\gamma_j$'s, a relatively smaller number of principal components are selected from $\Psi(x)$, with dimension of $q^*$, where $q^* < q$, and those principal components are used in the least squares regression. We refer this method as principal component based ELM, and will apply it to both in-sample fitting and out-of-sample forecast as well.

### 4.2.4 Pretest based ELM

In addition to selection by coefficients shrinkage or by combination of the information in the $\Psi$ functions through principal components, we can also select $\Psi$ functions directly through setting a threshold criterion. In particular, we consider the pretest method, which selects the significant $\Psi$ functions using the $t$-statistic with a chosen significance level, such as 0.01. Then the selected set of $\Psi$ functions is used in least squares regression both in-sample fitting and out-of-sample forecast.

## 4.3 Forecast Combination using ELM

Huang and Lee (2010) shows that with many explanatory variables, the combination of forecasts (CF) is better than combination of information (CI), where CI refers to model shrinkage such as by principal components, and the combination weight scheme can be based on equal weight, principal component, or other optimization methods. From empirical results, we find that (reported in next section) both the above ELM and LWG methods does not work well for out-of-sample forecasting for stock returns, so we consider apply forecast combination methods to the ELM models. That is, we modify the ELM methods into a two step algorithm, where in the first step, a set of forecasts are made using variations of ELM models, and then in the second step, we combine those forecasts using certain weighting schemes.

In particular, for the weighting scheme in the second step, we will consider a principal component based weighting scheme, we will compare this weighing method with equal weight weighting scheme as well as the Mallows criteria based weighting scheme. That is, to combine the forecast of $y_t$ under the principal component based weighting scheme, the weights in forecast combination is calculated according to the principal component coefficients. And in order to get the forecasts of $y_t$ in the first step, we consider three different methods, based on what methods to use in order to deal with the multicollinearity problem of activation functions and get each individual forecasts: one-at-a-time regression, ridge regression, and principal components regression.

### 4.3.1 Two Step Algorithm

In stead of using all the activation functions with the existence of multicollinearity in the first step regression, we can regress $y_t$ on the linear part $x_t$ and each activation function $\psi(x_t'\gamma_j)$ at a time for $j = 1, ...q$. And make forecasts of $y_t$ based on each individual models. Then in the second step, we combine these $q$ individual forecasts using certain weighting scheme. We call this method one-at-a-time then combine.

Alternatively, for the first step of individual forecast model, in stead of doing the regression one activation function at a time, we can also use ridge estimation to get rid of the multicollinearity problem. We refer this method as ridge-then-combine method. In particular, in the first step, we regress $y_t$ on the linear part $x_t$ and a set of activation functions using ridge estimator, and then repeat for $m$ times, and then get a series of forecasts $\hat{y}_{t+h} = (\hat{y}_{t+h}^{(1)}, \hat{y}_{t+h}^{(2)}, ..., \hat{y}_{t+h}^{(m)})'$. Then in the second step, we apply some weighting scheme to these $m$ forecasts.

On the other hand, since ridge estimator does not reduce the dimension of the activation functions, we further consider reducing dimension through principal components, that is, we use only a few principal components of the activation functions for regression in the first step. In particular, we regress $y_t$ on the linear part $x_t$ and $q^*$ principal components of the activation functions for each model in the first step, and then repeat for $m$ times. The second step is the same as above again where a certain weighting scheme can be applied. We call this method principal component then combine.

### 4.3.2 Weighting Scheme

To use the principal component weighting scheme in the one-at-a-time then combine method, we first divide the in-sample data into two parts, the in-sample regression part and in-sample prediction part. Then use in-sample regression data, we can produce $h$-step-ahead forecasts of $y_t$, each using one activation function and denote these $q$ forecasts as $\hat{y}_{t+h} = (\hat{y}_{t+h}^{(1)}, \hat{y}_{t+h}^{(2)}, ..., \hat{y}_{t+h}^{(q)})'$, where $\hat{y}_{t+h}^{(i)}$ is the forecast from $i$th model. We then combine them through the principal components weighting scheme, that we first get the principal components of $\hat{y}_{t+h}$, denoted as $f_{t+h}$, where $f_{t+h} = \Lambda \hat{y}_{t+h}$, and $\Lambda$ contains the first $q^*$

eigenvectors of $\hat{y}_{t+h}$ corresponding to the first $q^*$ largest eigenvalues. And then estimate the weight from the OLS regression of $y_{t+h}$ on the principal components $f_{t+h}$ using the in-sample prediction data:

$$y_{t+h} = \alpha f_{t+h} + v_{t+h} \tag{4.5}$$

Then the principal components weighing forecast combination is given by the weighted principal components of $\hat{y}_{T+h}$:

$$\hat{y}_{T+h}^{CFPC} = \hat{\alpha} f_{T+h} = \hat{\alpha}\Lambda\hat{y}_{T+h} \tag{4.6}$$

where $f_{T+h} = \Lambda\hat{y}_{T+h}$. Notice that this forecast combination can be recovered in terms of $\hat{y}_{T+h}$, with weight equal to $\hat{\alpha}\Lambda$. We will refer the principal component weighing scheme as ELM-CFPC in the following.

To use the principal component weighting scheme in the ridge then combine and principal component then combine method, the first step is similar that the data is divided into in-sample regression and prediction parts and we use ridge estimation or selected principal components on the in-sample regression part and repeat for $m$ times. And then in the second step, we apply the above principal components weighting scheme to these $m$ forecasts.

As a comparison to the principal components weighting scheme, we also use the equal weight weighing scheme. That is, in the second step, in stead of estimating weights by principal components of $\hat{y}_{t+h}$, we assign each individual forecast $(\hat{y}_{t+h}^{(1)}, \hat{y}_{t+h}^{(2)}, ..., \hat{y}_{t+h}^{(m)})'$ with a weight equal to $1/m$ ($m = q$ for one-at-a-time then combine method).

As another weighting scheme comparison, we also apply the Mallows weighing criteria for forecast combination, according to Hansen (2007, 2008). That is, we first use a set of nested models to get individual forecasts, each of which contains the first $m$ activation functions for $m = 1, ..., q$, and then combine these $q$ models where the weight minimizes the penalized mean squared error in terms of the Mallow's criteria:

$$C(w) = \sum_{t=1}^{n}(y_t - \hat{y}_t' w)^2 + 2\sum_{m=1}^{q} w(m)k(m)s^2 = w'\hat{e}'\hat{e}w + 2w'Ks^2 \tag{4.7}$$

where $\hat{e} = (\hat{e}(1), ..., \hat{e}(q))$, is the vector of mean squared errors for individual models, $K = (1, ...q)'$ is the vector of dimension for each model and $s^2 = \hat{e}'(q)\hat{e}(q)/(n-q)$ is the sample variance for the largest model with all the activation fuctions.

## 4.4 Alternative Models

As comparison, we also include some simple linear and nonlinear models as benchmarks. The linear models include:

1. Martingale Difference (MD)

   The model assumes that excess stock return is a martingale difference process, that the return at time $t$ is equal to a random error, given by:

   $$y_t = \varepsilon_t \tag{4.8}$$

   so the one-step-ahead forecast at any time $T$ is equal to zero, denoted by $\hat{y}_{T+1}^{MD} = 0$.

2. Historical Mean (HM)

   This model assumes the excess stock return at any time $T$ is equal to the average of all historical returns up to this time, so the one-step ahead forecast uses all the past information of $y_t$:

$$\hat{y}_{T+1}^{HM} = \frac{1}{T}\sum_{t=1}^{T} y_t \tag{4.9}$$

In this paper, we make the forecast using rolling window method with the window width fixed.

3. Autoregressive of Order One (AR(1))

   In this model, we assume the excess stock return is an AR(1) process:

$$y_t = \beta y_{t-1} + \varepsilon_t \tag{4.10}$$

   where $|\beta| < 1$. Then the one-step-ahead forecast is $\hat{y}_{T+1}^{AR} = \hat{\beta} y_T$.

4. Random Walk (RW)

   When the parameter $\beta$ in the AR(1) is equal to one, we have a nonstationary random walk process.

$$y_t = y_{t-1} + \varepsilon_t \tag{4.11}$$

   The one-step-ahead forecast at any time $T$ is equal to the previous time return $\hat{y}_{T+1}^{RW} = y_t$.

5. Linear Regression (LR)

   The above models are all univariate. The stock return can also be explained by a set of covariates using the linear model, given by:

$$y_t = x_t\beta + \varepsilon_t \tag{4.12}$$

where the regressor $x_t$ is $k$-dimensional. And the one-step-ahead forecast is given by $\hat{y}^{LR}_{T+1} = x_t\hat{\beta}$, where $\hat{\beta}$ is estimated using least squares method.

6. Linear Principal Component (PC)

   Assume we fit the stock return using the linear model, yet the regressors maybe correlated. As a result, using a relatively smaller number of principal components of the regressors would solve the problem. The model is written as:

$$y_t = x_t^*\beta + \varepsilon_t \tag{4.13}$$

   where $x_t^*$ is a $k^*$-dimension principal component of $x_t$ where $k^* < k$. In this paper, we choose $k^* = 5$.

The two alternative nonlinear models as a comparison to the nonlinear ELM models in this paper are from Bai and Ng (2008):

1. Principal Component of $X^2$ (QPC)

   As a comparison to the linear PC model, following Bai and Ng (2008), we use a quadratic principal component model, where the principal components is composed from $x_t$ and $x_t^2$.

2. Nonlinear Principal Component (PCSQ)

   An alternative way to construct the nonlinear PC model, is to assume excess stock returns is a nonlinear function of the principal components of $x_t$, in particular, a second order polynomial function.

$$y_t = x_t^{**}\beta + \varepsilon_t \tag{4.14}$$

where $x_t^{**} = [x_t^*, x_t^{*2}]$. This model is called principal component square model (PCSQ), following Bai and Ng (2008).

## 4.5 Results

### 4.5.1 Data and Variables

We use the data set provided by Goyal and Welch (2008). The stock return in this paper is the $k$ period excess return, which is calculated from the S&P 500 Index ($P_t$), with $k = 1, 2, 3$ representing the annualized monthly data, quarterly aggregated monthly data and the monthly data respectively. To get excess return, we use the compound return minus the risk free rate:

$$Q_t(1) = R_t(1) - R_f$$

where $Q_t(1)$ is the monthly excess return, $R_t(1)$ is the one-period simple return, calculated from $R_t(1) = \frac{P_{t+1}}{P_t} - 1$, and $R_f$ is the risk free rate, which is the Treasury-bill rate. Following Campbell and Thompson (2008), the $k$ period excess return is calculated by:

$$Q_t(k) = (Q_{t+k-1}(1) + 1) \times .... \times (Q_t(1) + 1) - 1 = \prod_{j=1}^{k}(Q_{t+k-j}(1) + 1) - 1$$

where $k = 1, 3, 12$. We call them monthly, quarterly aggregated and annualized excess returns respectively in this paper.

The covariates in this paper include 13 variables: dividend price ratio (d/p), dividend yield (d/y), earnings price ratio (e/p), dividend payment ratio (d/e), stock variance (svar), cross-sectional premium (csp), book to market ratio (b/m), net equity expansion (ntis), treasury bills (tbl), long term yield (lty), long term rate of returns (ltr), default yield spread (dfy), inflation (infl). For detailed description of each variables, refer to Goyal and Welch (2008). The time period we choose is form May 1937 to December 2002, during which we have the balanced data for all the 13 variables. In addition, we divide the whole period into 3 subsamples in order to check robustness of each method. We group the whole sample into the first half, 25 to 75 percentile, and last half of the data as three subsamples, called subsample 1, 2 and 3 respectively. In each sample and subsample, we use the rolling window method to do forecasting, and according to each model, the ratio for estimation and prediction period is equal to 1.

### 4.5.2 Main Results

One issue for the ELM models is to choose the number randomly generated set of $\gamma_j$'s, or alternatively, the number of activation functions, denoted by $q$. According to Huang, Zhu and Siew (2006) and also summarized in theorem 2.1 and 2.2 in Huang, Wang and Lan (2011), the number of activation functions needed to make good approximation is smaller or equal to the number of observations. That is, when $q \leq N$, the ELM method is good enough to approximate any functional forms for any randomly generated $\gamma$ in the sense that the norm of the approximation error goes to zero with probability one. In particular, when $q = N$, the norm of the approximation error could achieve zero with probability one for any randomly generated $\gamma$. However, their proof is valid theoretically since the probability to have multicolinear $\Psi$ functions is zero when $\gamma_j$ is randomly generated. However, in practice, for any given realization of $\gamma_j$'s, the $\Psi$ functions may be correlated, and thus making the estimation questionable. Moreover, their theorem is for in sample approximation rather than time series forecasting, so in this paper, we will consider both the case when $q = N$ and $q >> N$ as well as both in-sample and out-of-sample.

We consider two cases for the dimension of the activation functions, $q = N$ and $q >> N$, and for $q >> N$, we choose $q = 1000$. And for PC based ELM, $q^* = 3$, as well as in the PC weighting forecasting combination. For all the following tables, MD, HM, AR, RW, LIN denotes for martingale difference, historical mean, AR(1), random walk, linear model, respectively, and LINPC, PCSQ, QPC denote for linear principal component model, principal component of $X$ and $X^2$, as well as nonlinear principal component model, as in Bai and Ng (2008). ELMB, ELMR ELMPC, ELMPT denote for basic ELM, ridge based ELM, principal component based ELM and pretest based ELM. ELMCFPC denotes for one-at-a-time then combine using principal component weighting scheme, ELMCFPC2 denotes for principal component then combine using principal component weighting scheme, ELMCFR denotes for ridge then combine using principal component weighting scheme, ELMCFEW denotes for one-at-a-time then combine using equal weighting scheme, ELMCFEW2 denotes for principal component then combine using equal weighting scheme, ELMCFEW3 denotes for ridge then combine using equal weighting scheme, ELMMMA denotes for using MMA to combine nested models. For MMA weighting, since the models are nested and requires number of variables less than number of observations, so we only consider $q = N$. And for the principle components and then combine as well as ridge then combine method, since the first step needs regressors to have large dimension, and in order to compare the total sample with sub samples, we only consider the case for $q = 1000$.

Table 1 summarizes the in-sample fitting mean squared error using various ELM methods as well as the alternative models. From the table we can observe that the basic ELM method works best among all the models with the mean squared error very close to zero. Moreover, we can see that the results for $q = N$ and $q >> N$ are quite similar, verifies that $q = N$ is enough to make a good fit. In addition, for the annually aggregated data, all the ELM models fit better than the linear and quadratic pc models, suggesting ANN model can make a perfect fit for the nonlinear feature of the aggregated data. And for quarterly aggregated and monthly data, the results are similar between other ELM methods rather than the basic ELM and the linear methods.

Table 2 summarizes the out-of-sample mean square forecast error (MSFE) of all the above models. Among the linear models, martingale difference model works best for all annualized, quarterly aggregated and monthly data, so we will use it as the linear benchmark for comparison. From the results, we can see that for the annually aggregated data, basic ELM and PC based ELM give very poor forecasts, implies that the basic ELM method can give good approximation of nonlinear functions only for in-sample fitting rather than out-of-sample forecast. On the other hand, the principal component weighting scheme forecast combination gives best results compared with all the alternative linear and nonlinear models in terms of a lower MSFE. This implies that the forecast combination works well for out-of-sample forecasting of nonlinear data, and moreover, the principal component weighting scheme outperforms the equal weight weighting scheme, meaning that the "supervision" works in the sense that the choice of weight takes into account the forecast ability of individual models, rather than giving equal weight for every individual forecast.

For quarterly aggregated and especially monthly data, most models give similar results, yet ELMCFPC an ELMCFPC2 give similar or better forecasts. Further, the results from the three subsamples has similar properties as the full sample for all the annualized, quarterly aggregated and monthly data, indicating the above conclusion is robust.

### 4.5.3 Comparison Criterion

To compare between different models, one way is to compare their mean squared forecast error (MSFE), yet this may not be enough to see if there is significant improvement. In this paper, we choose three criteria for forecast comparision: the out of sample goodness of fit ($R^2_{OOS}$), the adjusted DM statistic, and the first or second order stochastic dominance (FOSD/SOSD).

The out of sample goodness of fit statistic follows from Campbell and Thompson (2008), which is computed as:

$$R^2_{OS} = 1 - \frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{T}(y_t - \bar{y}_t)^2}$$

where $y_t$ is the true excess return, $\bar{y}_t$ is the forecast from benchmark model and $\hat{y}_t$ is the forecast from alternative model. The out of sample $R^2$ could be positive, zero, or negative, depending on whether alternative model forecast outperforms, equals to, or worse than the benchmark model forecast.

Diebold and Mariano construct an test statistic for comparing predictive accuracy, which is usually referred as the DM statistic. It is computed from:

$$S_t = \frac{\bar{d}}{\hat{\text{avar}}(\bar{d})^{1/2}} = \frac{\bar{d}}{(\hat{\text{avar}}(d_t)/P)^{1/2}}$$

where $\bar{d} = \frac{1}{P}\sum_{t=1}^{P} d_t$, $R$ is the number of out of sample period, and $d_t$ is the difference of benchmark loss function and alternative loss function. $a\hat{var}d_t$ is the consistent estimator of the asymptotic variance of $d_t$, and in this paper, we use the Newey-West estimator from Newey and West (1987).

However, when the benchmark model and the alternative model is nested, DM statistic is not appropriate since the MSFE from the parsimonious model is expected to be smaller than the larger model. Following Clark and West (2007), we use the adjusted DM statistic, which is calculated from:

$$S_t^* = \frac{\bar{d}^*}{\text{a}\hat{\text{var}}(\bar{d}^*)^{1/2}} = \frac{\bar{d}^*}{(\text{a}\hat{\text{var}}(d_t^*)/P)^{1/2}}$$

where $\bar{d}^* = \bar{d} + \text{adj}$, and when we have squared loss function, the adjustment term is given by $\text{adj} = \frac{1}{P}\sum_{t=1}^{P}(e_{tb} - e_{ta})^2$, where $e_{tb}$ is the forecast error of the benchmark model and $e_{ta}$ is the forecast error from alternative model. Since the forecast comparision in this paper is based on nested model, we will use the adjusted DM statistic, and the asymptotic variance is estimated using the Newey-West estimator.

The rationality to use the adjusted DM statistic is that it is equivalent to the encompassing test in Stock and Watson (2002), which tests the coefficient in the following forecast combination regression:

$$y_t = \alpha \hat{y}_t^A + (1-\alpha)\hat{y}_t^B + u_t$$

where $y_t$ is the true excess return, $\hat{y}_t^B$ is the forecast of $y_t$ using benchmark model and $\hat{y}_t^A$ is the forecast of $y_t$ using alternative model. If the test $H_0 : \alpha = 0$ versus $H_1 : \alpha \neq 0$ is rejected, then it implies that the benchmark model does not dominate the alternative model. On the other hand, if the test $H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$ is rejected, it implies the alternative model does not dominate the benchmark model. It can be shown that the encompassing test is equivalent to the adjusted DM statistic that if the encompassing test cannot reject $H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$, which means the alternative model dominates, then the adjusted DM statistic is also significant.

We could also apply the above adjustment to the out of sample $R^2$, as did in Hillebrand, Lee and Medeiros (2012), where the adjusted $R^2_{OS}$, $R^{2*}_{OS}$ is defined by:

$$R^{2*}_{OS} = 1 - \frac{\sum_{t=1}^{T}[(y_t - \hat{y}_t)^2 - (\bar{y}_t - \hat{y}_t)^2]}{\sum_{t=1}^{T}(y_t - \bar{y}_t)^2}$$

The previous two criteria for forecast comparison are both based on the mean value of the MSFE, however, a lower average MSFE does not necessary lead to a better forecast since the whole distribution of the MSFE is not considered. The stochastic dominance comparison criteria takes into account this problem through comparing the two distributions of the MSFE of the benchmark and alternative models. By definition, given the error distribution functions for benchmark model ($F^B(e)$) and the alternative model ($F^A(e)$), then the alternative model first order stochastic dominate (FOSD) the benchmark model if $F^A(e) - F^B(e) > 0$ for all $e$. And the alternative model second order stochastic dominate the benchmark model (SOSD) up to $r$ if $\int_0^r [F^A(e) - F^B(e)]de > 0$, for $r > 0$. Since FOSD implies SOSD, we will say alternative model is better if alternative model FOSD benchmark model, or alternative model outperforms benchmark model up to point $r$ if alternative model SOSD benchmark model up to $r$.

Since martingale difference is the best among all the linear models and basic ELM model is the best among all nonlinear models for annualized data, we will set the benchmark as the martingale difference model and alternative as the basic ELM model. Table 3 summarizes the comparison results for the out of sample goodness of fit as well as the DM statistic, both adjusted and unadjusted, as well as the results for the stochastic dominance between ELMCFPC for $q = N$, $q >> N$, ELMCFPC2 for $q = N$ compared with benchmark. From this table, for annually aggregated data and quarterly aggregated data both out of sample $R^2$ and the DM statistic shows ELMCFPC and ELMCFPC2 significantly improves upon the martingale difference model. Yet for monthly data, martingale difference model is significantly better than the ELM models. Figure 1 shows the graph for FOSD, SOSD and the distribution of the squared errors for the benchmark and the ELMCFPC2 model. The conclusion is that in most cases, ELMCFPC2 second order stochastic dominate the martingale difference model for annualized and quarterly aggregated data, while martingale difference model second order stochastic dominate basic ELM model. All the above results show that the conclusion from table 2 is significant.

## 4.6 Conclusion

In this paper, we apply the artificial neural network (ANN) model to out-of-sample prediction of financial return using a set of covariates. The main challenge in ANN model estimation is the multicolinearity between the large numbers of randomly generated hidden layers. We explore several methods of extreme learning machines to deal with the large dimension regressors. We find that the dimension shrinkage methods such as general inverse, ridge, pretest and principal components sometimes fit perfectly for in-sample data, however, it has very poor out-of-sample forecast ability. We then introduce some modifications to the ELM method, which is a two step forecast combination algorithm, where the first step uses ELM methods with some modifications to get a set of forecasts, and the second step combines the forecasts using principal components weighting scheme. We compared our methods with some alternative linear and nonlinear models as well as applying other weighting schemes in forecast combination. Empirical results show that our method gives best forecast for annually aggregated equity premium among all the alternatives, which indicates that the two step ELM-CFPC algorism is the method to use for out-of-sample forecasting of the ANN model.

Table 4.1: In-sample Mean Squared Error

| | $Q_t(12)$ | | | | $Q_t(3)$ | | | | $Q_t(1)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | S1 | S2 | S3 | All | S1 | S2 | S3 | All | S1 | S2 | S3 |
| MD | 0.0272 | 0.0279 | 0.0255 | 0.0246 | 0.0062 | 0.0067 | 0.0051 | 0.0059 | 0.0021 | 0.0023 | 0.0016 | 0.0020 |
| HM | 0.0257 | 0.0261 | 0.0254 | 0.0241 | 0.0060 | 0.0063 | 0.0051 | 0.0059 | 0.0021 | 0.0022 | 0.0016 | 0.0020 |
| AR | 0.0040 | 0.0039 | 0.0029 | 0.0039 | 0.0035 | 0.0039 | 0.0026 | 0.0033 | 0.0021 | 0.0023 | 0.0016 | 0.0020 |
| RW | 0.0041 | 0.0041 | 0.0030 | 0.0041 | 0.0043 | 0.0047 | 0.0030 | 0.0040 | 0.0044 | 0.0049 | 0.0031 | 0.0040 |
| LIN | 0.0207 | 0.0138 | 0.0100 | 0.0200 | 0.0058 | 0.0054 | 0.0037 | 0.0052 | 0.0020 | 0.0021 | 0.0015 | 0.0019 |
| LINPC | 0.0235 | 0.0185 | 0.0132 | 0.0213 | 0.0061 | 0.0058 | 0.0040 | 0.0057 | 0.0021 | 0.0022 | 0.0015 | 0.0020 |
| PCSQ | 0.0235 | 0.0185 | 0.0129 | 0.0209 | 0.0060 | 0.0056 | 0.0040 | 0.0055 | 0.0021 | 0.0022 | 0.0015 | 0.0020 |
| QPC | 0.0237 | 0.0182 | 0.0140 | 0.0225 | 0.0061 | 0.0058 | 0.0042 | 0.0058 | 0.0021 | 0.0022 | 0.0015 | 0.0020 |
| (q=1000) | | | | | | | | | | | | |
| ELMB | 4.62E-17 | 7.05E-20 | 1.64E-20 | 1.95E-05 | 9.41E-17 | 1.61E-20 | 1.48E-20 | 2.71E-21 | 4.59E-17 | 3.17E-21 | 1.88E-20 | 5.71E-22 |
| ELMR | 0.0043 | 0.0032 | 0.0026 | 0.0026 | 0.0024 | 0.0020 | 0.0017 | 0.0015 | 0.0013 | 0.0011 | 0.0008 | 0.0010 |
| ELMPC | 0.0193 | 0.0130 | 0.0095 | 0.0183 | 0.0056 | 5.30E-03 | 0.0036 | 0.0049 | 0.0020 | 0.0021 | 0.0014 | 0.0019 |
| ELMPT | 0.0207 | 0.0138 | 0.0100 | 0.0139 | 0.0058 | 0.0054 | 0.0037 | 0.0052 | 0.0020 | 0.0021 | 0.0015 | 0.0019 |
| (q=N) | | | | | | | | | | | | |
| ELMB | 3.34E-05 | 1.38E-05 | 1.78E-05 | 2.59E-21 | 2.10E-05 | 2.43E-05 | 1.28E-05 | 1.13E-05 | 1.48E-05 | 1.16E-05 | 1.67E-05 | 8.67E-06 |
| ELMR | 0.0050 | 0.0041 | 0.0031 | 0.0020 | 0.0026 | 0.0024 | 0.0020 | 0.0021 | 0.0013 | 0.0013 | 0.0010 | 0.0012 |
| ELMPC | 0.0235 | 0.0129 | 0.0085 | 0.0182 | 0.0056 | 0.0053 | 0.0036 | 0.0050 | 0.0020 | 0.0021 | 0.0014 | 0.0019 |
| ELMPT | 0.0172 | 0.0138 | 0.0093 | 0.0200 | 0.0051 | 0.0053 | 0.0037 | 0.0044 | 0.0020 | 0.0021 | 0.0014 | 0.0017 |

Note: $Q_t(12)$, $Q_t(3)$, $Q_t(1)$ denote for the annualized, quarterly aggregated and monthly data respectively. All, S1, S2, S3 denotes for the full sample period, subsample 1, 2, 3 respectively. The first part of the table is the MSFE for the linear and nonlinear alternative models, and the middle part is the ELM methods with $q = 1000$, and last part is the ELM methods with $q = N$.

Table 4.2: Out-of-sample Mean Squared Forecast Error

| | $Q_t(12)$ | | | | $Q_t(3)$ | | | | $Q_t(1)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | S1 | S2 | S3 | All | S1 | S2 | S3 | All | S1 | S2 | S3 |
| MD | 0.0255 | 0.0278 | 0.0246 | 0.0252 | 0.0060 | 0.0043 | 0.0060 | 0.0063 | 0.0021 | 0.0013 | 0.0020 | 0.0022 |
| HM | 0.0269 | 0.0293 | 0.0271 | 0.0272 | 0.0061 | 0.0043 | 0.0061 | 0.0064 | 0.0021 | 0.0013 | 0.0020 | 0.0022 |
| AR | 0.0477 | 0.0499 | 0.0577 | 0.0363 | 0.0087 | 0.0051 | 0.0079 | 0.0100 | 0.0021 | 0.0013 | 0.0020 | 0.0023 |
| RW | 0.0509 | 0.0527 | 0.0611 | 0.0391 | 0.0118 | 0.0067 | 0.0103 | 0.0138 | 0.0041 | 0.0024 | 0.0037 | 0.0045 |
| LIN | 0.0534 | 0.0476 | 0.0556 | 0.0496 | 0.0075 | 0.0041 | 0.0070 | 0.0094 | 0.0022 | 0.0014 | 0.0021 | 0.0025 |
| LINPC | 0.0391 | 0.0313 | 0.0469 | 0.0264 | 0.0071 | 0.0041 | 0.0066 | 0.0073 | 0.0022 | 0.0014 | 0.0021 | 0.0024 |
| PCSQ | 0.0400 | 0.0320 | 0.0502 | 0.0264 | 0.0073 | 0.0041 | 0.0071 | 0.0071 | 0.0022 | 0.0014 | 0.0021 | 0.0024 |
| QPC | 0.0370 | 0.0311 | 0.0445 | 0.0263 | 0.0068 | 0.0040 | 0.0069 | 0.0077 | 0.0022 | 0.0013 | 0.0021 | 0.0025 |
| (q=1000) | | | | | | | | | | | | |
| ELMB | 43.5205 | 0.2195 | 52.2015 | 0.0245 | 8.2129 | 0.2074 | 13.3428 | 0.6081 | 7.7495 | 0.0841 | 46.5603 | 1.4765 |
| ELMR | 0.0751 | 0.0562 | 0.0802 | 0.0301 | 0.0214 | 0.0075 | 0.0603 | 0.0352 | 0.0079 | 0.0027 | 0.0169 | 0.0114 |
| ELMPC | 22.2340 | 15.1789 | 72.3221 | 9.8265 | 1.4840 | 0.9757 | 3.3956 | 1.7257 | 0.1786 | 0.0872 | 0.4583 | 0.2867 |
| ELMPT | 0.0534 | 0.0476 | 0.0556 | 0.0496 | 0.0075 | 0.0041 | 0.0070 | 0.0094 | 0.0022 | 0.0014 | 0.0021 | 0.0025 |
| ELMCFPC | 0.0232 | 0.0282 | 0.0236 | 0.0231 | 0.0059 | 0.0042 | 0.0061 | 0.0066 | 0.0025 | 0.0013 | 0.0023 | 0.0024 |
| ELMCFPC2 | 0.0228 | 0.0314 | 0.0256 | 0.0229 | 0.0061 | 0.0042 | 0.0060 | 0.0059 | 0.0021 | 0.0013 | 0.0020 | 0.0023 |
| ELMCFR | 0.0279 | 0.0267 | 0.0242 | 0.0267 | 0.0062 | 0.0041 | 0.0063 | 0.0063 | 0.0021 | 0.0013 | 0.0020 | 0.0023 |
| ELMCFEW | 0.0515 | 0.0493 | 0.0571 | 0.0417 | 0.0073 | 0.0041 | 0.0070 | 0.0095 | 0.0022 | 0.0014 | 0.0023 | 0.0025 |
| ELMCFEW2 | 13.3596 | 102.6214 | 20.9627 | 26.6212 | 1.5188 | 4.4847 | 7.1449 | 1.8509 | 0.2143 | 1.3407 | 2.0209 | 0.3659 |
| ELMCFEW3 | 0.0545 | 0.0544 | 0.0606 | 0.0572 | 0.0122 | 0.0062 | 0.0138 | 0.0178 | 0.0036 | 0.0020 | 0.0051 | 0.0091 |
| (q=N) | | | | | | | | | | | | |
| ELMB | 323.8111 | 11.0306 | 2029.7000 | 26.8175 | 1655.3000 | 5.0791 | 161.1933 | 32.6020 | 392.6649 | 2.8299 | 2722.9000 | 15.4184 |
| ELMR | 0.0616 | 0.0542 | 0.0633 | 0.0455 | 0.0147 | 0.0058 | 0.0166 | 0.0279 | 0.0056 | 0.0020 | 0.0064 | 0.0098 |
| ELMPC | 20.7082 | 26.5227 | 49.8685 | 7.2852 | 1.5345 | 1.8209 | 4.6028 | 2.0563 | 0.2249 | 0.2334 | 0.9325 | 0.2914 |
| ELMPT | 0.0538 | 0.0502 | 0.0622 | 0.0440 | 0.0075 | 0.0042 | 0.0081 | 0.0099 | 0.0039 | 0.0015 | 0.0033 | 0.0026 |
| ELMCFPC | 0.0230 | 0.0281 | 0.0232 | 0.0227 | 0.0059 | 0.0042 | 0.0062 | 0.0066 | 0.0025 | 0.0013 | 0.0025 | 0.0025 |
| ELMCFEW | 0.0513 | 0.0493 | 0.0574 | 0.0418 | 0.0072 | 0.0041 | 0.0070 | 0.0094 | 0.0022 | 0.0014 | 0.0023 | 0.0025 |
| ELMMMA | 0.0534 | 0.0481 | 0.0574 | 0.0421 | 0.0075 | 0.0040 | 0.0070 | 0.0093 | 0.0022 | 0.0014 | 0.0021 | 0.0024 |

Note: $Q_t(12)$, $Q_t(3)$, $Q_t(1)$ denote for the annualized, quarterly aggregated and monthly data respectively. All, S1, S2, S3 denotes for the full sample period, subsample 1, 2, 3 respectively. The first part of the table is the MSFE for the linear and nonlinear alternative models, and the middle part is the ELM methods with $q = 1000$, and last part is the ELM methods with $q = N$.

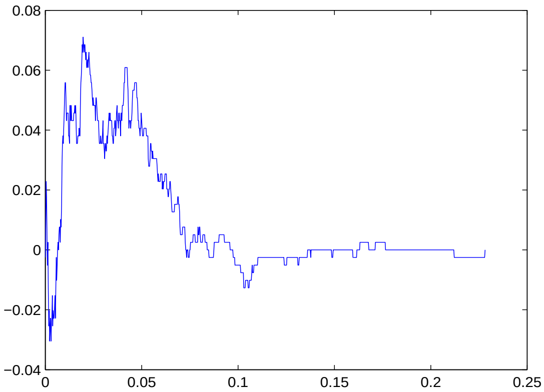
Table 4.3: Forecast Comparison

| | $Q_t(12)$ | | | | $Q_t(3)$ | | | | $Q_t(1)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | S1 | S2 | S3 | All | S1 | S2 | S3 | All | S1 | S2 | S3 |
| ELMCFPC(q=1000) | | | | | | | | | | | | |
| $R^2_{OS}$ | 0.0891 | -0.0116 | 0.0417 | 0.0850 | 0.0238 | 0.0278 | -0.0211 | -0.0521 | -0.2134 | -0.0165 | -0.2992 | -0.1282 |
| $R^{2*}_{OS}$ | 0.2171 | 0.1296 | 0.1527 | 0.5470 | 0.2143 | 0.1173 | 0.0644 | 0.0247 | 0.0336 | -0.0062 | -0.0851 | -0.0531 |
| $S_t$ | 2.3856 | -0.2453 | 0.7054 | 1.1184 | 0.3702 | 0.5412 | -0.3931 | -1.1160 | -0.9507 | -1.1466 | -1.3022 | -1.2074 |
| $S_t^*$ | 4.9083 | 2.8948 | 2.8484 | 3.4680 | 3.4829 | 2.0903 | 1.0038 | 0.6081 | 0.3810 | -0.4447 | -0.7982 | -0.9956 |
| FOSD | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| SOSD | A | B | A | A | A | A | NA | B | NA | NA | B | B |
| ELMCFPC2(q=1000) | | | | | | | | | | | | |
| $R^2_{OS}$ | 0.1037 | -0.1288 | -0.0371 | 0.0941 | 0.0370 | 0.0184 | -0.0107 | 0.0685 | -0.0227 | -0.0311 | -0.0294 | -0.0092 |
| $R^{2*}_{OS}$ | 0.5877 | 0.6496 | 0.1956 | 0.8574 | 0.1259 | 0.2048 | 0.0600 | 0.1783 | -0.0009 | 0.0796 | 0.0040 | 0.0196 |
| $S_t$ | 1.1134 | -0.9921 | -0.3930 | 0.5350 | 0.8756 | 0.2607 | -0.1761 | 1.1817 | -1.2639 | -0.5696 | -0.8426 | -0.3878 |
| $S_t^*$ | 4.9428 | 3.2041 | 1.8305 | 4.6749 | 2.6107 | 2.6071 | 0.8965 | 2.7397 | -0.0479 | 1.4574 | 0.1109 | 0.8631 |
| FOSD | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| SOSD | A | B | NA | A | NA | A | NA | A | B | NA | NA | NA |
| ELMCFPC(q=N) | | | | | | | | | | | | |
| $R^2_{OS}$ | 0.0965 | -0.0112 | 0.0568 | 0.0998 | 0.0263 | 0.0225 | -0.0347 | -0.0478 | -0.1975 | -0.0127 | -0.2794 | -0.1306 |
| $R^{2*}_{OS}$ | 0.2291 | 0.1174 | 0.1584 | 0.5182 | 0.2064 | 0.1177 | 0.0413 | 0.0323 | 0.0157 | -0.0027 | -0.0503 | -0.0547 |
| $S_t$ | 2.5821 | -0.2366 | 1.0168 | 1.3441 | 0.4222 | 0.4145 | -0.7227 | -0.8666 | -0.9174 | -0.8834 | -1.4276 | -1.2982 |
| $S_t^*$ | 5.0173 | 2.6338 | 2.9823 | 3.5341 | 3.6310 | 2.0411 | 0.7935 | 0.6858 | 0.1759 | -0.2006 | -0.5806 | -1.0362 |
| FOSD | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| SOSD | A | NA | A | A | A | A | NA | NA | NA | NA | B | B |

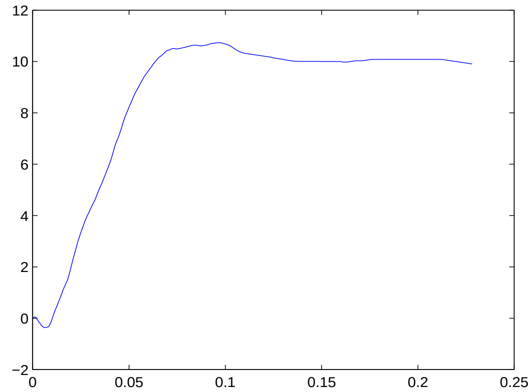
Note: $Q_t(12)$, $Q_t(3)$, $Q_t(1)$ denote for the annualized, quarterly aggregated and monthly data respectively. All, S1, S2, S3 denotes for the full sample period, subsample 1, 2, 3 respectively. Benchmark model is martingale difference, alternative model is basic ELM. The first part of the table is results for ELM methods with $q = 1000$, and the second part is results for the ELM methods with $q = N$. B, A, NA for FOSD, SOSD denote for benchmark model dominates, alternative model dominates, and non of them dominates the other, respectively.

Figure 4.1: Stochastic Dominance and Distribution of Squared Errors for $Q_t(12)$

(a) FOSD for $Q_t(12)$




(b) SOSD for $Q_t(12)$




(c) DIST for $Q_t(12)$

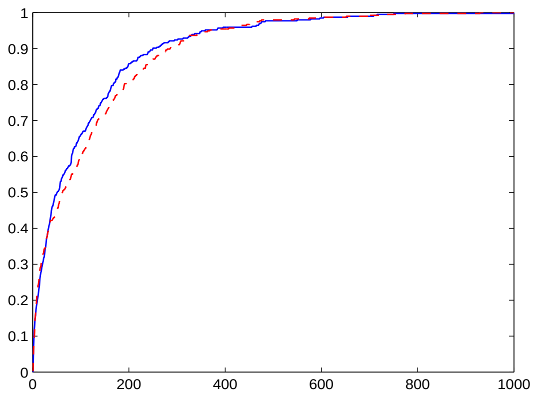
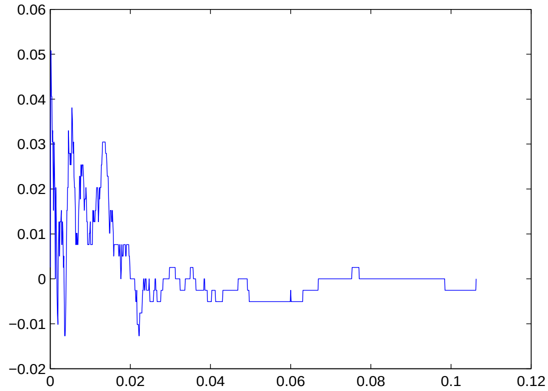
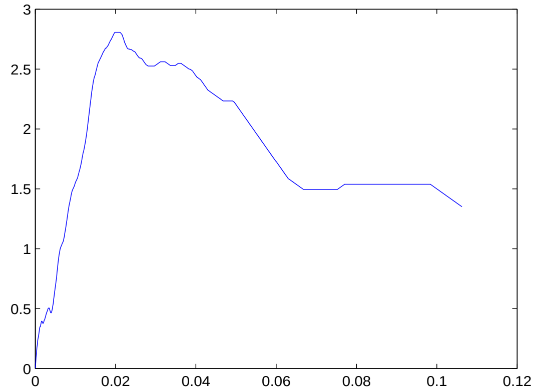


Figure 4.2: Stochastic Dominance and Distribution of Squared Errors for $Q_t(3)$

(a) FOSD for $Q_t(3)$




(b) SOSD for $Q_t(3)$




(c) DIST for $Q_t(3)$

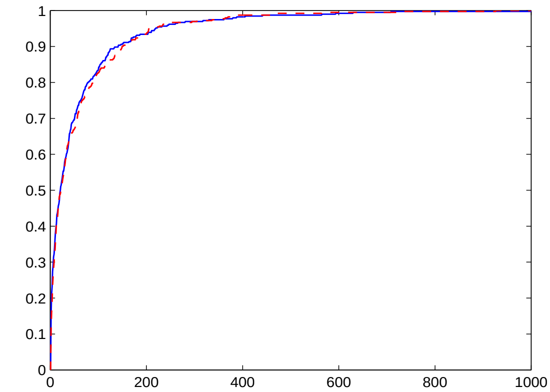
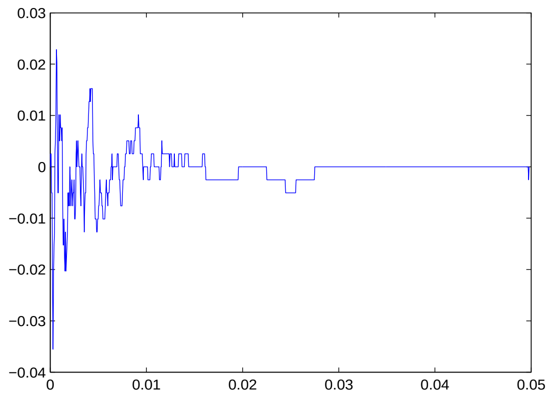
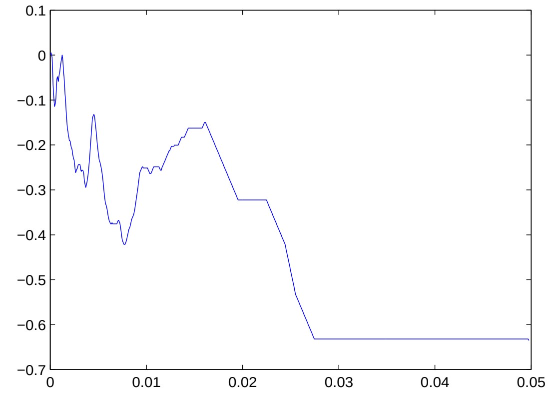


Figure 4.3: Stochastic Dominance and Distribution of Squared Errors for $Q_t(1)$

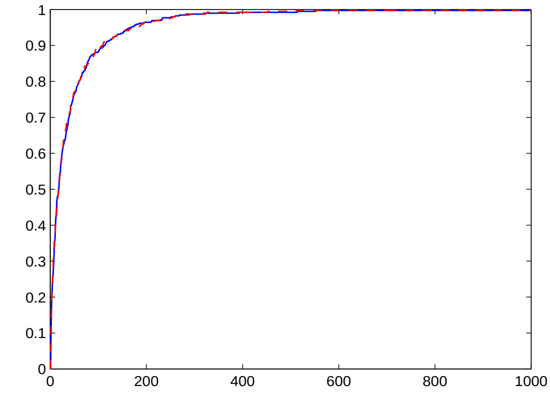
(a) FOSD for $Q_t(1)$




(b) SOSD for $Q_t(1)$




(c) DIST for $Q_t(1)$



# Bibliography


[1] Bai, J. and S. Ng. "Forecasting economic time series using targeted predictors". *Journal of Econometrics*, 146:304-317, 2008.

[2] Campbell, J.Y. and S. Thompson. "Predicting the equity premium out of sample: Can anything beat the historical average?". *Review of Financial Studies*, 21(4):1511-1531, 2008.

[3] Clark, T.E. and K.D. West. "Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis". *Journal of Econometrics*, 135:155-186, 2006.

[4] Goyal, A. and I. Welch. "A comprehensive look at the empirical performance of equity premium prediction".*Review of Financial Studies*, 21(4):1455-1508, 2008.

[5] Huang, G. and L. Chen. "Convex incremental extreme learning machine". *Neurocomputing*, 70:3056-3062, 2007.

[6] Huang, G. and L. Chen. "Enhanced random search based incremental extreme learning machine". *Neurocomputing*, 71:3460-3468, 2008.

[7] Huang, H. and T. Lee. "To Combine Forecasts or to Combine Information?". *Econometric Reviews*, 29(5-6):534-570, 2010.

[8] Huang, G., D. Wang and Y. Lan. "Extreme learning machine: a survey". *International Journal of Machine Leaning and Cybernetics*, 2(2):107-122, 2011.

[9] Huang, G. B., Q. Y. Zhu and C. K. Siew. "Extreme learning machine: theory and applications". *Neurocomputing* , 70:489-501, 2006.

[10] Kock, A. B. and T. Teräsvirta. "Forecasting the Finnish Consumer Price Inflation Using Artificial Neural Network Models and Three Automated Model Selection Techniques". *Finnish Economic Papers*, 26:13-24, 2013.

[11] Lee, T., H. White, and C. W. J. Granger. "Testing for neglected nonlinearity in time series models: a comparison of neural network methods and alternative tests". *Journal of Ecomometrics*, 56:269-290, 1993.

[12] Lee, T., Z. Xi and R. Zhang. "Testing for Neglected Nonlinearity Using Artificial Neural Networks with Many Randomized Hidden Unit Activations". *Journal of Time Series Econometrics*, 5(1):61-86, 2013.

[13] Lee, T., Z. Xi and R. Zhang. "Testing for Neglected Nonlinearity Using Regularized Artificial Neural Networks", *Recent Advances in Estimating Nonlinear Models*, Ma, J. and Wohar, M. E. (ed.), Springer Publishers, 2014.

[14] Stock, J. H. and M. W. Watson. "Macroeconomic forecasting using diffusion indexes". *Journal of Business and Economic Statistics*, 20(2):147-162, 2002.

# Chapter 5

# Mallows Model Averaging in the Presence of Multicollinearity

A challenge with large dimensional data in regression is the collinearity among covariates. A common solution to this problem is to apply principal component analysis (PCA). Yet one needs to select the number of principal components. Many studies have focused on finding the optimal number of principal components assuming the linear factor model is correctly specified. In this paper, we do not assume that the data generating process (DGP) is a linear factor model and thus there is no true number of factors. Under this circumstance, we can combine several principal component regressions with different numbers of principal components through the Mallows criteria. Under certain conditions, the model averaging estimator is minimax such that the estimation risk is smaller. We show that the Mallow model averaging estimator can improve the estimation efficiency.

## 5.1 Introduction

In empirical studies of macroeconomics and finance, the problem of how to deal with a large number of variables, especially in the presence of multicollinearity is an important issue. A common way to solve the above large dimension problem is to use the principal component analysis for a factor model, where the variations of the large set of variables can be modeled by a small number of reference variables. To efficiently use a factor model, the predetermined number of factors ($k$) has become a major concern, since the regression results is sensitive to the choice of $k$. Many studies including Bai and Ng (2002), Onatski (2009), Ahn and Horenstein (2013), among others, developed various criteria to choose (estimate) $k$ assuming the factor model is a true data generating model and there is a true value for $k$. Much of these papers are about consistently estimating $k$.

However, some of the above criteria are complicated to use in empirical work and also may not generate satisfying results due to estimation error, model instability, and structural breaks. In this research, we first consider a two-step Stein-Mallows Model Averaging (Stein-MMA) method to use factor model without choosing $k$, and we aim to show that such procedure improves upon the results for any choice of $k$ in estimation and prediction especially when there is no true value of $k$ in the underlying data generating process such as a nonlinear functional form. For a chosen $k$ factors, a Stein shrinkage estimator (Hill and Judge 1987) could be used to combine the full model that includes all ($K$) factors and the model including $k$ factors to improve the risk of estimators and predictor. The main point of the paper is that the above process can be replicated for different values of $k$, ranging from 1 to $K-1$. Then, for the second step, as the Mallows

Model Average (MMA) method (Hansen 2007) can be used to these $K-1$ models. This two-step procedure can be shown to improve upon any single factor model for a chosen $k$. Alternatively, we skip the first step of Stein shrinkage, and directly apply the MMA to the factor model with each $k$ to obtain the MMA forecast. In this case $k$ ranges from 1 to $K$.

These MMA procedures can be extended to out-of-sample forecast combination of factor models. This method is easy to use, much easier than choosing a $k$ using complicated model selection methods or cross validation. Applications of this paper can be in asset returns forecasts, in portfolio performance evaluation, in prediction of inflation and in monetary policy analysis, where the interested variable can be modeled as a function of a number of factors extracted from a large set of predictors.

## 5.2 OLS estimator

Consider we have the following model of $y$ as a function of some predictor matrix $X$:

$$y = m(X) + \varepsilon, \tag{5.1}$$

where $y$ is a $T \times 1$ time series, $X$ is a $T \times K$ matrix of $K$ predictors, and $\varepsilon$ is a $T \times 1$ error time series with the conditional mean zero $E(\varepsilon|X) = 0$ and conditional variance $E(\varepsilon\varepsilon'|X) = \sigma^2 I_K$. Note that we do not assume normality of $\varepsilon$. The relation between $Y$ and $X$, which is denoted by function $m(X)$, is possibly linear or nonlinear. Our interest is to forecast $y$ when the number $K$ of predictors in $X$ is large using a linear model. That is, we use the following linear model in estimation:

$$y = X\beta + \varepsilon, \tag{5.2}$$

where $\beta$ is a $K \times 1$ parameter vector. The location vector $\beta$ is unknown and the objective is to estimate it by $\beta(y, X)$. We consider three estimators for $\beta$ in this paper: (i) the ordinary least squares (OLS) estimator denoted $\hat{\beta}$, (ii) the principal component (PC) estimator denoted $\hat{\beta}^*$, and (iii) the combined estimator of $\hat{\beta}$ and $\hat{\beta}^*$, using the Mallows model averaging criteria.

The OLS estimator of the linear model is written as:

$$\hat{\beta} = \left(X'X\right)^{-1} X'y, \tag{5.3}$$

When the dimension of $X$ is large, the major concern of the OLS model is the existence of multicollinearity. To deal with this problem, in this paper, we consider using the orthogonalized variables from $X$. Let $V$ be the $K \times K$ matrix of eigenvectors of $X'X$, then $X'X = T\,V\Lambda V'$, where $\Lambda$ is the diagonal matrix of eigenvalues of $X$ in descending order, and $T$ is a scalar, $V$ is orthognormal such that $V'V = VV' = I_K$ and

$$y = X\beta + \varepsilon = XV\Lambda^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\sqrt{T}\Lambda^{\frac{1}{2}}V'\beta + \varepsilon = P\delta + \varepsilon, \tag{5.4}$$

where $P = XV\Lambda^{-\frac{1}{2}}\frac{1}{\sqrt{T}}$, $\delta = \sqrt{T}\Lambda^{\frac{1}{2}}V'\beta$. This is the principal components representation of the OLS regression, and $P$ contains all the $K$ principal components of $X$, where

$$P'P = \frac{1}{\sqrt{T}}\Lambda^{-\frac{1}{2}}V'X'XV\Lambda^{-\frac{1}{2}}\frac{1}{\sqrt{T}} = \frac{1}{\sqrt{T}}\Lambda^{-\frac{1}{2}}V'TV\Lambda V'V\Lambda^{-\frac{1}{2}}\frac{1}{\sqrt{T}} = I_K \tag{5.5}$$

$$\hat{\delta} = \left(P'P\right)^{-1} P'y = P'y = \frac{1}{\sqrt{T}}\Lambda^{-\frac{1}{2}}V'X'y \tag{5.6}$$

Note that

$$PP' = \frac{1}{\sqrt{T}} XV\Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}V'X'\frac{1}{\sqrt{T}} = \frac{1}{T}XV\Lambda^{-1}V'X' = \frac{1}{T}X(V\Lambda V')^{-1}X' = X(X'X)^{-1}X$$

The OLS estimator $\hat{\beta}$ can be written using the PC representation, since $\hat{y} = P\hat{\delta} + \hat{\varepsilon} = XV\Lambda^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\hat{\delta} + \hat{\varepsilon}$, so the OLS estimator can be written as:

$$\hat{\beta} = \frac{1}{\sqrt{T}} V\Lambda^{-\frac{1}{2}}\hat{\delta} \tag{5.7}$$

## 5.3 Principal Component Estimator

Next we consider shrinking the dimension of the model through selecting some of the principal components. The number of all the $K$ principal components of $X'X$ can be decomposed into two parts, $K = K_1 + K_2$, where $K_1$ is the number of eigenvalues that are relatively large and $K_2$ is the number of eigenvalues that are relatively close to zero. For the model selection, we consider discarding the $K_2$ principal components which correspond to the small eigenvalues while maintaining the $K_1$ principal components correspond to larger eigenvalues. Write $P = (P_1, P_2)$, $\delta = (\delta_1', \delta_2')'$, $V = (V_1, V_2)$, then the model becomes

$$\begin{aligned} y = P\delta + \varepsilon = (P_1\ P_2)\begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} + \varepsilon = P_1\delta_1 + P_2\delta_2 + \varepsilon, \\ = X(V_1\ V_2)\Lambda^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\sqrt{T}\Lambda^{\frac{1}{2}}(V_1\ V_2)'\beta + \varepsilon \\ = XV_1\Lambda_1^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\sqrt{T}\Lambda_1^{\frac{1}{2}}V_1'\beta + XV_2\Lambda_2^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\sqrt{T}\Lambda_2^{\frac{1}{2}}V_2'\beta + \varepsilon, \end{aligned}$$

where $\Lambda_1$ and $\Lambda_2$ are of dimension $K_1$ and $K_2$ diagonal matrices, that contain the corresponding the $K_1$ largest and $K_2$ smallest eigenvalues respectively, and $P_2\delta_2 = XV_2\Lambda_2^{-\frac{1}{2}}\Lambda_2^{\frac{1}{2}}V_2'\beta$

is deleted in the PC method. Therefore, principal components regression with deleted components is equivalent to OLS estimation with the restriction:

$$\delta_2 = R\beta = \sqrt{T}\Lambda_2^{\frac{1}{2}}V_2'\beta = 0, \tag{5.8}$$

where $R = \sqrt{T}\Lambda_2^{\frac{1}{2}}V_2'$ imposes $K_2$ linear restrictions on $\beta$. Note that $R = \sqrt{T}\Lambda_2^{\frac{1}{2}}V_2'$ is stochastic depending on $X$. And $P_1 = \frac{1}{\sqrt{T}}XV_1\Lambda_1^{-\frac{1}{2}}$, and we can verify that $P_1'P_1 = I_{K_1}$, where $I_{K_1}$ is identity matrix with dimension $K_1$.

The principal components estimator of $\delta$ with $K_2$ deleted components, corresponding to the restrictions $\delta_2 = 0$, is

$$\hat{\delta}_1 = (P_1'P_1)^{-1}P_1'y = P_1'y = \frac{1}{\sqrt{T}}\Lambda_1^{-\frac{1}{2}}V_1'X'y. \tag{5.9}$$

And note that $P_1P_1' = \frac{1}{T}XV_1\Lambda_1^{-1}V_1'X'$, $V_1'V_1 = I_{K_1}$, $V_2'V_2 = I_{K_2}$, $V_1'V_2 = 0_{K_2\times K_1}$, $V_2'V_1 = 0_{K_1\times K_2}$.

Using the estimator $\hat{\delta}_1$ and setting $\delta_2 = 0$, the fitted value of $y$ is given by:

$$\hat{y} = P_1\hat{\delta}_1 + \hat{\varepsilon} = \frac{1}{\sqrt{T}}X\,V_1\Lambda_1^{-\frac{1}{2}}\hat{\delta}_1 + \hat{\varepsilon}, \tag{5.10}$$

and the principal components estimator of $\beta$ is:

$$\hat{\beta}^* = \frac{1}{\sqrt{T}}V_1\Lambda_1^{-\frac{1}{2}}\hat{\delta}_1. \tag{5.11}$$

## 5.4 Mallows Model Average Estimator with a Fixed $k$

Given the OLS estimation using all the $K$ principal components and the shrinkage model using only $k = K_1$ principal components, there is always ways to improve both models, such as model averaging. Model averaging is an alternative to model selection. The advantage of model averaging is that rather than selecting a criterion to pick up a particular model, the model averaging method gives the best fit by choosing a weight to combine all the candidate models. Model averaging can reduce estimation variance and at the same time controlling the bias. Model averaging is more flexible compared to model selection since the latter is just a special case of model averaging if putting the weights to extreme values 0 and 1. It has been proved that the model averaging estimator would give a lower loss than any of the individual model, see Hansen (2007, 2008). In this sense, the procedure of model selection using different criterion could be regarded as a special case of model averaging by setting the weights to the extreme values with weight equal to 1 for the particular selected model and 0 for the models discarded. The model selection procedure is a discrete selection of individual models, while the model averaging method finds a proper weight that is a smooth function of all the individual models. In this section, we will consider the Mallows model averaging (MMA) estimator over the PC and OLS models.

First we can rewrite the OLS model and the PC model denoted by $\hat{g}_1$ and $\hat{g}_2$ respectively:

$$\hat{g}_1 = P\hat{\delta} = PP'y = B_1 y$$

$$\hat{g}_2 = P_1\hat{\delta}_1 = P_1P_1'y = B_2 y$$

where $B_1 = PP'$, $B_2 = P_1P_1'$. Then the model averaging estimation of the two models is:

$$\hat{g} = w\hat{g}_1 + (1-w)\hat{g}_2 = wB_1y + (1-w)B_2y = B(w)y$$

where $B(w) = wB_1 + (1-w)B_2 = wPP' + (1-w)P_1P_1'$, $w$ is the combination weight, and

$0 \leq w \leq 1$. The estimation error from the OLS model, PC model and the averaging model are denoted respectively by $\hat{e}_1$, $\hat{e}_2$ and $\hat{e}(w)$:

$$\hat{e}_1 = y - \hat{g}_1 = y - B_1 y = (I_T - B_1)y = (I_T - PP')y$$

$$\hat{e}_2 = y - \hat{g}_2 = y - B_2 y = (I_T - B_2)y = (I_T - P_1 P_1')y$$

$$\hat{e}(w) = w\hat{e}_1 + (1-w)\hat{e}_2 = y - B(w)y = (I_T - wPP' - (1-w)P_1 P_1')y$$

By the Mallows criteria for model averaging, we need to choose the weight $w$ such that the loss form the model averaging is minimized, where the loss from the model is a function of the squared error terms, penalized by the dimension of the model, that is, we need to minimize the following loss function $C(w)$ with respect to $w$:

$$\begin{aligned} C(w) &= \hat{e}(w)'\hat{e}(w) + 2\hat{\sigma}^2 \mathrm{tr}(B(w)) \\ &= y'(I_T - wPP' - (1-w)P_1P_1')'(I_T - wPP' - (1-w)P_1P_1')y + 2\hat{\sigma}^2 \mathrm{tr}(B(w)) \end{aligned}$$

where $tr(B(w)) = wtr(PP') + (1-w)tr(P_1P_1') = wK + (1-w)K_1 = K_1 + wK_2$. Penalty term is added such that both the estimation error and the parsimoniousness of the model is considered when choosing a proper weight vector $w$. Also we can verify that $P_2'P_1 = 0_{K_2 \times K_1}$, and

$$P'P_1 = \begin{pmatrix} P_1' \\ P_2' \end{pmatrix} P_1 = \begin{pmatrix} I_{K_1} \\ 0_{K_2} \end{pmatrix},$$

$$PP'P_1P_1' = P \begin{pmatrix} I_{K_1} \\ 0_{K_2} \end{pmatrix} P_1' = (P_1 P_2) \begin{pmatrix} P_1' \\ 0 \end{pmatrix} = P_1P_1',$$

$$P_1P_1'PP' = P_1(I_{K_1} 0) \begin{pmatrix} P_1' \\ P_2' \end{pmatrix} = (P_1 0) \begin{pmatrix} P_1' \\ P_2' \end{pmatrix} = P_1P_1'.$$

Substitute to $C(w)$ and simplify, we have,

$$C(w) = y'y - 2wy'PP'y - 2(1-w)y'P_1P_1'y + w^2y'PP'y + 2w(1-w)y'P_1P_1'y$$
$$+(1-w)^2y'P_1P_1'y + 2\hat{\sigma}^2(K_1 + wK_2)$$

Take the first order condition with respect to $w$ and set to zero, we can then solve the optimal weight which minimizes the loss function:

$$w = \frac{y'PP'y - y'P_1P_1'y - \hat{\sigma}^2K_2}{yPP'y - y'P_1P_1'y} = \frac{y'P_2P_2'y - \hat{\sigma}^2K_2}{y'P_2P_2'y}$$

since $PP' = (P_1P_2)(P_1P_2)' = P_1P_1' + P_2P_2'$, so the optimal weight from MMA can be written as:

$$w = 1 - \frac{\hat{\sigma}^2K_2}{y'P_2P_2'y} = 1 - \frac{K_2}{\frac{y'P_2P_2'y}{\hat{\sigma}^2}} = 1 - \frac{K_2}{\frac{\hat{\delta}_2'\hat{\delta}_2}{\hat{\sigma}^2}}$$

if $K_2 < \frac{\hat{\delta}_2'\hat{\delta}_2}{\hat{\sigma}^2}$, since $\hat{\delta}_2 = P_2'y$, and $w = 0$ if $K_2 > \frac{\hat{\delta}_2'\hat{\delta}_2}{\hat{\sigma}^2}$. We will show below that this optimal weight from MMA can be written equivalently as:

$$w = 1 - \frac{1}{F_{K_2,T-K}}$$

if $F_{K_2,T-K} > 1$, and $w = 0$ if $F_{K_2,T-K} < 1$, where $F_{K_2,T-K}$ is the $F$-statistic from the null hypothesis $H_0 : \delta_2 = 0$ in the OLS model $y = P_1\delta_1 + P_2\delta_2 + \varepsilon$. Since we can write the F-statistic in the above test with $K_2$ constraints as:

$$F = \frac{(RRSS - URSS)/K_2}{URSS/(T-K)}$$

where $RRSS$ is the restricted sum of squares under the null (PC model), and $URSS$ is the unrestricted sum of squares from the OLS model. So we have

$$\begin{aligned}
F &= \frac{\Big(\hat\delta - \begin{pmatrix}\hat\delta_1\\ 0\end{pmatrix}\Big)' P'P \Big(\hat\delta - \begin{pmatrix}\hat\delta_1\\ 0\end{pmatrix}\Big)/K_2}{\hat e_1'\hat e_1/(T-K)} \\
&= \frac{\Big(P'y - \begin{pmatrix}P_1'\\ 0\end{pmatrix} y\Big)' \Big(P'y - \begin{pmatrix}P_1'\\ 0\end{pmatrix} y\Big)/K_2}{\hat\sigma^2} \\
&= \frac{(0\ \hat\delta_2')P'P \begin{pmatrix}0\\ \hat\delta_2'\end{pmatrix}/K_2}{\hat\sigma^2} \\
&= \frac{\hat\delta_2'\hat\delta_2/K_2}{\hat\sigma^2} \sim F_{K_2,T-K}
\end{aligned}$$

Substitute the optimal weight back to the MMA, we have, when $F_{K2,T-K} > 1$,

$$\begin{aligned}
\hat g &= (wPP' + (1-w)P_1P_1')y = (1 - \frac{1}{F_{K_2,T-K}})PP'y + \frac{1}{F_{K_2,T-K}}P_1P_1'y \\
&= (1 - \frac{1}{F_{K_2,T-K}})P\hat\delta + \frac{1}{F_{K_2,T-K}}P_1\hat\delta_1
\end{aligned}$$

and since $PP'y = P\hat\delta = X\hat\beta$, and $P_1P_1'y = P_1\hat\delta_1 = X\hat\beta^*$, we have,

$$\hat g = wX\hat\beta + (1-w)X\hat\beta^* = X(w\hat\beta + (1-w)\hat\beta^*) = X\hat\beta^{**}$$

where $\hat\beta^{**}$ is denoted as the MMA estimator. From the above formula, we can see that the MMA estimator $\beta^{**}$ can be written as a linear combination of the OLS and PC estimator, with weight equal to $w$ and $1-w$ respectively. And when $F_{K_2,T-K} < 1$, the model averaging estimator just shrink to the PC estimator $\hat\beta^*$.

Hill and Judge (1987, 1990) propose a Stein-rule estimator $\tilde\beta$ that shrinks the standard OLS estimator $\hat{\beta}$ towards the principal components estimator $\hat{\beta}^*$:

$$\begin{aligned}\tilde{\beta} &= \hat{\beta}^* + \psi(\hat{\beta} - \hat{\beta}^*) \\ &= \psi\hat{\beta} + (1-\psi)\hat{\beta}^*\end{aligned}$$

where $\psi = 1 - \frac{a^*}{F}$, and $a^* = \frac{(T-k)a}{K_2}$ and $a$ is a constant. The Stein coefficient $\psi$ is the shrinkage from the OLS estimator $\hat{\beta}$ to the PC estimator $\hat{\beta}^*$, see Hillebrand and Lee (2012). We can see that when $a^* = 1$, the Stein rule estimator is equivalent to the MMA estimator, that $\tilde{\beta} = \hat{\beta}^{**}$, also see Hansen (2013). Since the Stein-rule estimator $\tilde{\beta}$ is minimax if

$$0 \leq a^* \leq \frac{2(T-K)(K_2-2)}{(T-K+2)K_2} \tag{5.12}$$

and $a^* = 1$ satisfies the above condition, we can conclude that the MMA estimator $\hat{\beta}^{**}$ is a special case of Stein-rule estimator that is minimax, in the sense that the MMA estimator will minimize the maximum of the model risk defined in Hill and Judge (1987, 1990).

## 5.5 Mallows Model Average Estimator with Many $k$

The above method of model shrinkage with a chosen $k = K$ principal components and model averaging through combining the OLS estimator with the K principal components and the PC model with only $K$ of them creates a question of how to determine the key parameter k in this method. Bai and Ng (2002), Onatski (2009), Ahn and Horenstein (2013), among others, developed various criteria to choose (estimate) $k$ assuming the factor model is a true data generating process and there is a true value for $k$. However, when the underlying data generating process is nonlinear, there may not exist a true value of $k$, in this case, the estimation of $k$ is biased and model selection and averaging based on $k$ principal components are not good approximation for the original model, since the restriction $\delta_2 = 0$ is not true. In this section, we consider a two-step model averaging method that avoids estimation the value of $k$.

Consider the PC model with the first $i$ principal components, where $i = 1, ..., K-1$, the estimator of $\beta$ of this model is given by $\hat{\beta}_i^*$ and the fitted value of $y$ is denoted by $\hat{f}_i = X\hat{\beta}_i^*$, where $\hat{\beta}_i^* = \frac{1}{\sqrt{T}} V_i \Lambda_i^{-\frac{1}{2}} \hat{\delta}_i$. Also consider the full model which uses all the $K$ principal components, denoted by $\hat{f}_K = X\hat{\beta}$, where $\hat{\beta}$ is the OLS estimator. Then the model averaging estimator $\hat{\beta}_i^{**}$ which combines the full model and the model using $i$ principal components is given by:

$$\hat{\beta}_i^{**} = w_i \hat{\beta} + (1 - w_i)\hat{\beta}_i^* \tag{5.13}$$

Using the MMA method to solve the optimal weight $w_i$, will give the weight that minimizes the penalized loss function, that $w_i = 1 - \frac{1}{F_{K_i, T-K}}$ if $F_{K_i,T-K} > 1$ and $w_i = 0$ if $F_{K_i,T-K} < 1$. And the combined estimation of the model using the first $i$ principal components of $X$ is given by

$$\hat{g}_i = X\hat{\beta}_i^{**} = X(w_i\hat{\beta} + (1 - w_i)\hat{\beta}_i^*) = w_i\hat{f}_K + (1 - w_i)\hat{f}_i \tag{5.14}$$

where $i = 1, ..., K - 1$ and the models are nested. And each model $\hat{g}_i$ is a combination of the model using the first $i$ principal components and the full model. For the case where there is true value of $k$ and $k = i$, the combined estimation will do no worse than $\hat{g}_i$.

For the case when there is no true value of $k$, we consider another step of model averaging using the Mallows criteria to combine all the models of $\hat{g}_i$ for $i = 1, ..., K - 1$, assuming the weight for model $\hat{g}_i$ is equal to $v_i$, where $0 \leq v_i \leq 1$, and $\sum_{i=1}^{K-1} v_i = 1$, the combined model can be written as:

$$
\begin{aligned}
\hat{h} &= \sum_{i=1}^{K-1} v_i \hat{g}_i = \sum_{i=1}^{K-1} v_i [w_i \hat{f}_K + (1-w_i)\hat{f}_i] \\
&= \sum_{i=1}^{K-1} v_i [(1 - \frac{1}{F_{K-i,T-K}})\hat{f}_K + \frac{1}{F_{K-i,T-K}}\hat{f}_i] \\
&= \sum_{i=1}^{K-1} v_i (1 - \frac{1}{F_{K-i,T-K}})\hat{f}_K + \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}}\hat{f}_i \\
&= (1 - \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}})\hat{f}_K + \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}}\hat{f}_i
\end{aligned}
$$

since $F_{K-i,T-k}$ can be calculated from the first step model averaging, to get the model averaging estimation $\hat{h}$, we only need to calculate the optimal $v_i$ for $i = 1, ..., K-1$ from minimizing the loss function $C(\mathbf{v})$ using the Mallows criteria, where $\mathbf{v} = (v_1, ..., v_{K-1})$. Since we can written the model using first $i$ principal components and the full model as $\hat{f}_i = B_i y$, $\hat{f}_K = By$, where B is the same as before, and $B_i = P_i P_i'$. Then we can rewrite $\hat{h}$ as:

$$
\begin{aligned}
\hat{h} &= (1 - \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}})By + \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}} B_i y \\
&= [(1 - \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}})B + \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}} B_i]y = B(\mathbf{v})y
\end{aligned}
$$

where $B(\mathbf{v}) = (1 - \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}})B + \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}} B_i$, and the estimation error from the combined model is given by:

$$
\hat{e}(\mathbf{v}) = (I - B(\mathbf{v}))y = (I - (1 - \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}})B - \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}} B_i)y
$$

and the optimal weight $\mathbf{v} = (v_1, ..., v_{K-1})$ by the Mallows' model averaging is calculated by minimizing the following loss function:

$$
C(\mathbf{v}) = \hat{e}(\mathbf{v})'\hat{e}(\mathbf{v}) + 2\hat{\sigma}^2 \mathrm{tr}(B(\mathbf{v}))
$$

where $\text{tr}(B(\mathbf{v})) = \sum_{i=1}^{K-1} v_i K_i + (1 - \sum_{i=1}^{K-1} v_i)K$. Then apply quadratic programming, one can find the optimal weights for $v_i$, $i = 1, ..., K-1$.

The above two-step Mallows' model averaging method gives the combined estimation using all the information of $X$ without selecting $k$, this nests the case where there is true value of $k$, in which the optimal weight for $v_i$ will be one for $k = i$, and zero otherwise. This two step procedure can deal with the problem of multicollinearity of $X$ as well as the possibility that there is no true value of $k$.

Alternatively to this two step procedure, one can combine the two step calculation into one step applying the MMA method, since the combined model $\hat{h}$ can also be written as:

$$\hat{h} = \sum_{i=1}^{K} p_i \hat{f}_i = \sum_{i=1}^{K} p_i B_i y = B(\mathbf{p})y$$

where $\mathbf{p} = (p_1, ..., p_K)$, $B(\mathbf{p}) = \sum_{i=1}^{K} p_i B_i$. And the estimation error from the combined model can be rewritten as:

$$\hat{e}(\mathbf{p}) = (I - B(\mathbf{p}))y = (I - \sum_{i=1}^{K} p_i B_i)y$$

and the optima weight $\mathbf{p} = (p_1, ..., p_K)$ using the one step MMA method is calculated from minimizing the following loss function:

$$C(\mathbf{p}) = \hat{e}(\mathbf{p})'\hat{e}(\mathbf{p}) + 2\hat{\sigma}^2 \text{tr}(B(\mathbf{p}))$$

where $\text{tr}(B(\mathbf{p})) = \sum_{i=1}^{K} p_i K_i$. This one step procedure is essentially equivalent to the two step procedure, since we have $p_i = \frac{v_i}{F_{K-i,T-K}}$ for $i = 1, ..., K-1$, and $p_i = 1 - \sum_{i=1}^{K-1} \frac{v_i}{F_{K-i,T-K}}$ for $i = K$. In both methods, we can observe that the optimal weight is proportional to the reciprocal of the $F$ statistic for the $i$th model where the first $i$ principal components are applied, with restriction that the rest of the principal components are discarded.

## 5.6 Conclusion

This paper considers model estimation in the existence of multicollinearity with relatively large dimension of predictor $X$, and when there is possible no true value of predetermined number of factors, such as a nonlinear process. We showed that the model averaging applying the Mallows criteria is equivalent to the Stein-rule estimator under some conditions, and a special case of the latter in general. And to avoid selecting the number of $k$, one can apply a two step or one step MMA procedure, which uses information from all the principal components of $X$. The optimal weight is proportional to the reciprocal of the $F$ statistic for the $i$th model where the first $i$ principal components are applied, with restriction that the rest of the principal components are discarded. Applications of this method can be in asset returns forecasts, in portfolio performance evaluation, in prediction of inflation and in monetary policy analysis, where the interested variable can be modeled as a function of a number of factors extracted from a large set of predictors.

# Bibliography


[1] Ahn, S. C. and A. R. Horenstein. "Eigenvalue ratio test for the number of factors". *Econometrica*, 81(3):1203-1227, 2013.

[2] Bai, J. and S. Ng. "Determining the number of factors in approximate factor models". *Econometrica*, 70(1):191-221, 2002.

[3] Hansen, B.E. "Least squares model averaging". *Econometrica*, 75:1175-1189, 2007.

[4] Hansen, B.E. "Least squares forecast averaging". *Journal of Econometrics*, 146:342-350, 2008.

[5] Hansen, B.E. "Model Averaging, Asymptotic Risk, and Regressor Groups". *Quantitative Economics*, forthcoming, 2013.

[6] Hill, R. C. and G. Judge. "Improved prediction in the presence of multicollinearity". *Journal of Econometrics*, 35:83-100, 1987.

[7] Hill, R.C. and G.G. Judge. "Improved Estimation under Collinearity and Squared Error Loss". *Journal of Multivariate Analysis*, 32:296-312, 1990.

[8] Hillebrand, E. and T.H. Lee. "Stein-Rule Estimation and Generalized Shrinkage Methods for Forecasting Using Many Predictors". *Advances in Econometrics*, 30:171-196, Millimet, D. and Terrell, D. (ed.), Emerald Publishers, 2012.

[9] Onatski, A. "Testing hypotheses about the number of factors in large factor models". *Econometrica*, 77(5):1447-1479, 2009.

# Appendix A

# Appendix for Chapter 1

## A.1 Calculation of $E(X'X)^{-1}$ for model without intercept

For the stationary AR(1) process, notice that from the Nagar-type expansion of $\frac{1}{D}$, after taking expectation, we have $Ea_{-1} = \frac{1}{ED}$, and $Ea_{-2} = \frac{E(D-ED)^2}{(ED)^3} = \frac{ED^2}{(ED)^3}$. Rewrite the AR(1) process without intercept in vector representation as $X = Cv$, where $C$ is a strict lower triangular $T \times T$ matrix with the $ij$th element $\rho^{i-j-1}$, for $i > j$, so that $ED = E(X'X) = E(v'C'Cv)$, and $ED^2 = E(X'XX'X) = E(v'C'Cvv'C'Cv)$. Then apply the expectation of quadratic forms by Bao and Ullah (2010), Ullah(2004, page 187), we have: $ED = E(X'X) = E(v'C'Cv) = \sigma_v^2\text{tr}(C'C)$, and $ED^2 = E(X'XX'X) = E(v'C'Cvv'C'Cv) = \sigma^4(\gamma_2\text{tr}(C'C \bigodot C'C) + \text{tr}(C'C)\text{tr}(C'C) + 2\text{tr}(C'CC'C))$, where $\gamma_2$ is the excess kurtosis of $v_t$ which is zero in our case. Then check the traces of the respective matrices, we can verify that $E(a_{-1}) = \frac{1-\rho^2}{T\sigma_v^2}$, $E(a_{-2}) = \frac{2(1-\rho^2)}{T^2\sigma_v^2}$. The similar method follows for the unit root case, where only the traces of the matrices are different.

## A.2 Nagar Expansion of $\hat{\rho}$ for model without intercept

For the stationary AR(1) model without intercept, the OLS estimator of $\rho$ can be written as: $\hat{\rho} = \rho + \frac{x'_{-1}v}{x'_{-1}x_{-1}} = \rho + \frac{N}{D}$, where $N = x'_{-1}v$, $D = x'_{-1}x_{-1}$. Then the Nagar expansion for $\hat{\rho}$ can be written as:

$$\hat{\rho} - \rho = a_{-1/2} + a_{-1} + a_{-3/2} + o_P(T^{-3/2})$$

where $a_{-1/2} = N(ED)^{-1}$, $a_{-1} = -N(D - ED)(ED)^{-2}$, $a_{-3/2} = N(D - ED)^2(ED)^{-3}$, and $a_{-i/2}$ is of order $O_P(T^{-i/2})$ for $i = 1, 2, 3$.

## A.3 Calculation of $E(X'X)^{-1}$ for model with intercept

For the stationary AR(1) process with intercept, notice that from the Nagar-type expansion of $\frac{1}{D}$, after take expectation, we have $Ea_{-1} = (ED)^{-1}$, and $Ea_{-2} = (ED)^{-1}(ED)^{-1}EDD'(ED')^{-1} - (ED')^{-1}$. Rewrite the AR(1) process with intercept in vector representation as $X = \theta C\iota + C\varepsilon$, where $C$ is defined the same as the no intercept case, $\iota$ is a vector of ones. Plug in the vector form representation into $Ea_{-1}$ and $Ea_{-2}$ and then apply the expectation of quadratic forms by Bao and Ullah (2010) and Ullah(2004, page 187) similar as the no intercept case. After some calculations with matrices traces and getting the $(2, 2)$th elements for the two terms respectively, the results follows.

## A.4 h-step-ahead forecast bias of AR(1) model with intercept

Consider the following stationary AR(1) model,

$$x_t = \theta + \rho x_{t-1} + v_t$$

where $|\rho| < 1$, $v_t \sim i.i.d.(0, \sigma^2)$, not necessary normal. We are interested in the forecast bias and MSFE of the feasible conditional h period ahead forecast of $x_{T+h}$, given information set at time $T$. The least squares estimator based, $h$-step-ahead feasible conditional forecast is given by:

$$\hat{x}_{T+h|T} = \hat{\theta} + \hat{\rho}\hat{x}_{T+h-1|T} + v_t$$

where $\hat{x}_{T+h-1|T} = x_T$ when $h = 1$, $\hat{\rho}$ and $\hat{\theta}$ are the least squares estimators of $\rho$ and $\theta$. Notice that since $x_{T+h} = \rho^h x_T + \theta \sum_{t=1}^{h} \rho^{t-1} + \sum_{t=1}^{h} \rho^{h-t-1} v_{T+t+1}$, the feasible forecast can be rewritten as: $\hat{x}_{T+h|T} = \hat{\rho}^h x_T + \hat{\theta} \sum_{t=1}^{h} \hat{\rho}^{t-1}$. Then the forecast error is given by: $e_{T+h|T} = x_{T+h} - \hat{x}_{T+h|T} = \sum_{t=1}^{h} (\theta\rho^{t-1} - \hat{\theta}\hat{\rho}^{t-1}) + (\rho^h - \hat{\rho}^h)x_T + \sum_{t=1}^{h} \rho^{h-t-1} v_{T+t+1}$. Therefore the forecast bias can be written as:

$$E(e_{T+h|T}) = \sum_{t=1}^{h} E(\theta\rho^{t-1} - \hat{\theta}\hat{\rho}^{t-1}) + E[(\rho^h - \hat{\rho}^h)x_T] \tag{A.1}$$

where the second term $E[(\rho^h - \hat{\rho}^h)x_T]$ is equal to the forecast bias for AR(1) model without intercept. And MSFE is:

$$\begin{aligned} E(e_{T+h|T}^2) &= \rho^{2h} E(x_T^2) + E(\hat{\rho}^{2h} x_T^2) - 2\rho^h E(\hat{\rho}^h y_T^2) + \sigma^2 \sum_{t=1}^{h} \rho^{2(t-1)} + \sum_{t=1}^{h} E(\theta\rho^{t-1} \\ &\quad -\hat{\theta}\hat{\rho}^{t-1})^2 + 2\sum_{t=1}^{h} E[(\theta\rho^{t-1} - \hat{\theta}\hat{\rho}^{t-1})(\rho^h - \hat{\rho}^h)x_T] \end{aligned} \tag{A.2}$$

where the first four terms $\rho^{2h}E(x_T^2) + E(\hat{\rho}^{2h}x_T^2) - 2\rho^h E(\hat{\rho}^h y_T^2) + \sigma^2 \sum_{t=1}^h \rho^{2(t-1)}$ are equal to the MSFE for AR(1) model without intercept case.

Consider the OLS estimator of $(\theta, \rho)'$, where

$$\begin{aligned} \hat{\rho} &= \rho + \frac{x_{-1}'Av}{x_{-1}'Ax_{-1}} = \rho + \frac{N}{D} \\ \hat{\theta} &= \bar{x} - \hat{\rho}\bar{x}_{-1} \end{aligned} \tag{A.3}$$

where $A = I_T - \frac{1}{T}\iota\iota'$, $x_{-1} = (x_0, ..., x_{T-1})'$, $v = (v_1, ..., v_T)'$, and $N = x_{-1}'Av$, $D = x_{-1}'Ax_{-1}$. We can also rewrite the above terms in vector form by substituting $x_{-1} = \theta C\iota + Cv$, where $C$ is defined the same as before.

Apply the Nagar expansion for $\hat{\rho}$ and $\hat{\theta}$ similar as before, we have,

$$\begin{aligned} \hat{\rho} - \rho &= a_{-1/2} + a_{-1} + a_{-3/2} + o_P(T^{-3/2}) \\ \hat{\theta} - \theta &= -\frac{\theta}{(1-\rho)}(a_{-1/2} + a_{-1} + a_{-3/2} + o_P(T^{-3/2})) \end{aligned}$$

where $a_{-1/2} = N(ED)^{-1}$, $a_{-1} = -N(D - ED)(ED)^{-2}$, $a_{-3/2} = N(D - ED)^2(ED)^{-3}$, and $a_{-i/2}$ is of order $O_P(T^{-i/2})$ for $i = 1, 2, 3$. Plug into the forecast bias in equation A.1, and notice that $E(a_{-1/2}) = 0$, then up to order $O(T^{-1})$, the forecast bias can be rewritten as:

$$\begin{aligned} E(e_{T+h|T}) &= \sum_{t=1}^{h}(\frac{\theta}{1-\rho}\rho^{h-1} - \theta(h-1)\rho^{h-2})E(a_{-1}) - h\rho^{h-1}E(a_{-1/2}x_T) \\ &+ \sum_{t=1}^{h}(\frac{\theta}{1-\rho}(h-1)\rho^{h-2} - \frac{\theta(h-1)(h-2)}{2}\rho^{h-3})E(a_{-1/2}^2) - h\rho^{h-1}E(a_{-1}x_T) \\ &- \frac{h(h-1)}{2}\rho^{h-2}E(a_{-1/2}^2 x_T) + o(T^{-1}) \end{aligned} \tag{A.4}$$

since up to order $O(T^{-1})$, we can verify that $E(a_{-1}) = -\frac{1+3\rho}{T}$, $E(a_{-1/2}^2) = \frac{1-\rho^2}{T}$, and in addition, notice that $x_T = \iota' X_T = \iota'(\theta\iota + \theta\rho C\iota + \rho Cv + v)$. Substitute to the above equation, and apply the expectation of quadratic forms, we can see that the $h$-period-ahead feasible forecast of the AR(1) model is unbiased up to order $O(T^{-1})$, that is,

$$E(e_{T+h|T}) = 0 + o_P(T^{-1}) \tag{A.5}$$

Similarity, substitute A.3 into A.2, we can rewrite the MSFE as:

$$\begin{aligned} E(e^2_{T+h|T}) &= \sum_{t=1}^{h}(\frac{\theta}{1-\rho}\rho^{h-1} - \theta(h-1)\rho^{h-2})^2 E(a^2_{-1/2}) \\ &\quad - \sum_{t=1}^{h}(\frac{\theta}{1-\rho}\rho^{h-1} - \theta(h-1)\rho^{h-2})(h\rho^{h-1})E(a^2_{-1/2}x_T) \\ &\quad + \frac{\sigma^2(1-\rho^{2h})}{1-\rho^2} + h^2\rho^{2h-2}E(a^2_{-1/2}x^2_T) + o(1) \end{aligned}$$

Similarly apply the expectations of quadratic forms, after simplification, the MSFE up to order one is given by:

$$E(e^2_{T+h|T}) = \frac{\sigma^2(1-\rho^{2h})}{1-\rho^2} + o(1)$$

# Appendix B

# Appendix for Chapter 2

## B.1 Expansion of $\hat{\beta} - \beta$ for the pure model

Following Bao and Ullah (2007, 2009), a stochastic expansion of the QMLE $\hat{\beta}$ can be written as:

$$\hat{\beta} - \beta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + o_P(T^{-3/2})$$

To derive the bias of $\hat{\beta}$ up to $O(T^{-1})$, we need the expansion up to $a_{-1}$. Given the quasi likelihood function, the score function is given by:

$$\psi_T = \begin{pmatrix} -\frac{\varepsilon' N_1 \varepsilon}{T\sigma^2}, & \frac{\varepsilon' M_1 \varepsilon + y_0 F_1' \varepsilon}{T\sigma^2}, & -\frac{1}{2\sigma^2} + \frac{\varepsilon'\varepsilon}{2T\sigma^4} \end{pmatrix}'$$

where $N_1 = C^{-1}B_1$, $C = I - B$, $B_1 = \frac{\partial B}{\partial \phi}$, and $M_1 = C^{-1}A$, $F_1 = C^{-1}F$.

The Hessian matrix $H_1$, which is symmetric $3 \times 3$ matrix, has the following elements:

$$\frac{\partial^2 L}{\partial \phi^2} = -\frac{\varepsilon' N_2 \varepsilon}{T\sigma^2}, \quad \frac{\partial^2 L}{\partial \phi \partial \rho} = \frac{\varepsilon' M_3 \varepsilon + y_0 F_3' \varepsilon}{T\sigma^2}, \quad \frac{\partial^2 L}{\partial \phi \partial \sigma^2} = \frac{\varepsilon' N_1 \varepsilon}{T\sigma^4},$$

$$\frac{\partial^2 L}{\partial \rho^2} = -\frac{\varepsilon' M_2 \varepsilon + y_0^2 F_1' F_1 + 2y_0 F_2' \varepsilon}{T\sigma^2}, \quad \frac{\partial^2 L}{\partial \rho \partial \sigma^2} = -\frac{\varepsilon' M_1 \varepsilon + y_0 F_1' \varepsilon}{T\sigma^4}, \quad \frac{\partial^2 L}{\partial \sigma^4} = \frac{1}{2\sigma^4} - \frac{\varepsilon'\varepsilon}{T\sigma^6}.$$

where $F_2' = F_1' M_1$, $F_3' = F_1' N_1 + F_1' N_1'$, $N_2 = N_1' N_1 + 2N_1^2$, $M_2 = M_1' M_1$, $M_3 = N_1' M_1 + N_1 M_1$.

For $H_2$, similarly, we have the following elements:

$$\frac{\partial^3 L}{\partial \phi^3} = -\frac{\varepsilon' N_3 \varepsilon}{T\sigma^2}, \quad \frac{\partial^3 L}{\partial \phi^2 \partial \rho} = \frac{2(\varepsilon' M_4 \varepsilon + y_0 F_4' \varepsilon)}{T\sigma^2}, \quad \frac{\partial^3 L}{\partial \phi \partial \rho^2} = -\frac{2(\varepsilon' M_5 \varepsilon + y_0^2 F_1' N_1 F_1 + y_0 F_5' \varepsilon)}{T\sigma^2},$$
$$\frac{\partial^3 L}{\partial \phi \partial \rho \partial \sigma^2} = -\frac{\varepsilon' M_3 \varepsilon + y_0 F_3' \varepsilon}{T\sigma^4}, \quad \frac{\partial^3 L}{\partial \phi^2 \partial \sigma^2} = \frac{\varepsilon' N_2 \varepsilon}{T\sigma^4}, \quad \frac{\partial^3 L}{\partial \phi \partial \sigma^4} = -\frac{2\varepsilon' N_1 \varepsilon}{T\sigma^6}, \quad \frac{\partial^3 L}{\partial \rho^3} = 0,$$
$$\frac{\partial^3 L}{\partial \rho^2 \partial \sigma^2} = \frac{\varepsilon' M_2 \varepsilon + y_0^2 F_1' F_1 + 2y_0 F_2' \varepsilon}{T\sigma^4}, \quad \frac{\partial^3 L}{\partial \rho \partial \sigma^4} = \frac{2(\varepsilon' M_1 \varepsilon + y_0 F_1' \varepsilon)}{T\sigma^6}, \quad \frac{\partial^3 L}{\partial \sigma^6} = -\frac{1}{\sigma^6} + \frac{3\varepsilon'\varepsilon}{T\sigma^8}.$$

where $F_4' = F_1' N_1' N_1 + F_1' N_1^2 + F_1' N_1'^2$, $F_5' = F_1' N_1' M_1 + F_1' N_1 M_1$, $N_3 = 6N_1' N_1^2 + 6N_1^3$, $M_4 = N_1' N_1 M_1 + M_1' N_1^2 + N_1^2 M_1$, $M_5 = M_1' N_1 M_1$.

Notice that $\mathbf{tr}(M_1) = \mathbf{tr}(N_1) = 0$, $y_0^2 F_1' F_1$, $y_0^2 F_1' N_1 F_1$, $\mathbb{E}(F_1' \varepsilon\varepsilon' N_2 \varepsilon)$, $\mathbb{E}(F_1' \varepsilon\varepsilon' M_2 \varepsilon)$, $\mathbb{E}(F_1' \varepsilon\varepsilon' M_3 \varepsilon)$ are all of order $O(1)$, then the expectation of $H_1$ and $H_2$ and matrix $Q$, up to order $O(1)$, is given as:

$$\bar{H}_1 = \begin{pmatrix} -\frac{\mathbf{tr}(N_2)}{T} & \frac{\mathbf{tr}(M_3)}{T} & 0 \\ \frac{\mathbf{tr}(M_3)}{T} & -\frac{\mathbf{tr}(M_2)}{T} & 0 \\ 0 & 0 & -\frac{1}{2\sigma^4} \end{pmatrix}$$

$$Q = \bar{H}_1^{-1} = \begin{pmatrix} -\frac{\mathbf{tr}(M_2)T}{\mathbf{tr}(M_2)\mathbf{tr}(N_2) - \mathbf{tr}^2(M_3)} & -\frac{\mathbf{tr}(M_3)T}{\mathbf{tr}(M_2)\mathbf{tr}(N_2) - \mathbf{tr}^2(M_3)} & 0 \\ -\frac{\mathbf{tr}(M_3)T}{\mathbf{tr}(M_2)\mathbf{tr}(N_2) - \mathbf{tr}^2(M_3)} & -\frac{\mathbf{tr}(N_2)T}{\mathbf{tr}(M_2)\mathbf{tr}(N_2) - \mathbf{tr}^2(M_3)} & 0 \\ 0 & 0 & -2\sigma^4 \end{pmatrix}$$

$$\bar{H}_2 = \begin{pmatrix} -\frac{\mathbf{tr}(N_3)}{T} & \frac{2\mathbf{tr}(M_4)}{T} & \frac{\mathbf{tr}(N_2)}{T\sigma^2} & \frac{2\mathbf{tr}(M_4)}{T} & -\frac{2\mathbf{tr}(M_5)}{T} & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} & \frac{\mathbf{tr}(N_2)}{T\sigma^2} & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} & 0 \\ \frac{2\mathbf{tr}(M_4)}{T} & -\frac{2\mathbf{tr}(M_5)}{T} & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} & -\frac{2\mathbf{tr}(M_5)}{T} & 0 & \frac{\mathbf{tr}(M_2)}{T\sigma^2} & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} & \frac{\mathbf{tr}(M_2)}{T\sigma^2} & 0 \\ \frac{\mathbf{tr}(N_2)}{T\sigma^2} & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} & 0 & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} & \frac{\mathbf{tr}(M_2)}{T\sigma^2} & 0 & 0 & 0 & \frac{2}{\sigma^6} \end{pmatrix}$$

Substitute the above results in to (2.3) and denote $\phi_{-i/2} = a_{-i/2,1}$, $\rho_{-i/2} = a_{-i/2,2}$ and $\sigma^2_{-i/2} = a_{-i/2,3}$, we have:

$$\phi_{-1/2} = -\frac{\mathbf{tr}(M_2)\varepsilon' N_1\varepsilon - \mathbf{tr}(M_3)(\varepsilon' M_1\varepsilon + y_0 F_1'\varepsilon)}{\sigma^2(\mathbf{tr}(M_2)\mathbf{tr}(N_2) - \mathbf{tr}^2(M_3))},$$
$$\rho_{-1/2} = -\frac{\mathbf{tr}(M_3)\varepsilon' N_1\varepsilon - \mathbf{tr}(N_2)(\varepsilon' M_1\varepsilon + y_0 F_1'\varepsilon)}{\sigma^2(\mathbf{tr}(M_2)\mathbf{tr}(N_2) - \mathbf{tr}^2(M_3))},$$
$$\sigma^2_{-1/2} = -\sigma^2 + \frac{\varepsilon'\varepsilon}{T}.$$

$$\begin{aligned}
\phi_{-1} \quad = \quad & \Big[ -\frac{1}{2}\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)\mathbf{tr}(N_3)(\varepsilon' M_1\varepsilon)^2 + 2\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}(M_4)\mathbf{tr}(N_2)(\varepsilon' M_1\varepsilon)^2 \\
& -\mathbf{tr}(M_2)\mathbf{tr}(M_5)\mathbf{tr}^2(N_2)(\varepsilon' M_1\varepsilon)^2 + \mathbf{tr}(M_4)\mathbf{tr}^3(M_3)(\varepsilon' M_1\varepsilon)^2 \\
& -2\mathbf{tr}(M_5)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)(\varepsilon' M_1\varepsilon)^2 + \mathbf{tr}^3(M_3)\mathbf{tr}(N_2)\varepsilon' M_1\varepsilon\varepsilon' M_2\varepsilon \\
& +\mathbf{tr}^2(M_2)\mathbf{tr}(M_3)\mathbf{tr}(N_3)\varepsilon' M_1\varepsilon\varepsilon' N_1\varepsilon - 2\mathbf{tr}^2(M_2)\mathbf{tr}(M_4)\mathbf{tr}(N_2)\varepsilon' M_1\varepsilon\varepsilon' N_1\varepsilon \\
& -4\mathbf{tr}(M_2)\mathbf{tr}(M_4)\mathbf{tr}^2(M_3)\varepsilon' M_1\varepsilon\varepsilon' N_1\varepsilon + 4\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}(M_5)\mathbf{tr}(N_2)\varepsilon' M_1\varepsilon\varepsilon' N_1\varepsilon \\
& +2\mathbf{tr}(M_5)\mathbf{tr}^3(M_3)\varepsilon' M_1\varepsilon\varepsilon' N_1\varepsilon + \mathbf{tr}^2(M_2)\mathbf{tr}^2(N_2)\varepsilon' M_1\varepsilon\varepsilon' M_3\varepsilon - \mathbf{tr}^4(M_3)\varepsilon' M_1\varepsilon\varepsilon' M_3\varepsilon \\
& -\mathbf{tr}^2(M_2)\mathbf{tr}(M_3)\mathbf{tr}(N_2)\varepsilon' M_1\varepsilon\varepsilon' N_2\varepsilon + \mathbf{tr}(M_2)\mathbf{tr}^3(M_3)\varepsilon' M_1\varepsilon\varepsilon' N_2\varepsilon \\
& -\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)\varepsilon' M_1\varepsilon\varepsilon' M_2\varepsilon - \frac{1}{2}\mathbf{tr}^3(M_2)\mathbf{tr}(N_3)(\varepsilon' N_1\varepsilon)^2 \\
& +3\mathbf{tr}(M_4)\mathbf{tr}^2(M_2)\mathbf{tr}(M_3)(\varepsilon' N_1\varepsilon)^2 + \mathbf{tr}^3(M_2)\mathbf{tr}(N_2)\varepsilon' N_1\varepsilon\varepsilon' N_2\varepsilon \\
& -3\mathbf{tr}(M_5)\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)(\varepsilon' N_1\varepsilon)^2 - \mathbf{tr}^2(M_2)\mathbf{tr}^2(M_3)\varepsilon' N_1\varepsilon\varepsilon' N_2\varepsilon \\
& +2\mathbf{tr}(M_2)\mathbf{tr}^3(M_3)\varepsilon' N_1\varepsilon\varepsilon' M_3\varepsilon - 2\mathbf{tr}^2(M_2)\mathbf{tr}(M_3)\mathbf{tr}(N_2)\varepsilon' N_1\varepsilon\varepsilon' M_3\varepsilon - \mathbf{tr}^4(M_3)\varepsilon' N_1\varepsilon\varepsilon' M_2\varepsilon \\
& +\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)\varepsilon' N_1\varepsilon\varepsilon' M_2\varepsilon + \sigma^2\mathbf{tr}^2(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)\varepsilon' M_1\varepsilon \\
& -2\sigma^2\mathbf{tr}(M_2)\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)\varepsilon' M_1\varepsilon - \frac{1}{2}y_0^2\mathbf{tr}(N_3)\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)(F_1'\varepsilon)^2 \\
& +\sigma^2\mathbf{tr}^5(M_3)\varepsilon' M_1\varepsilon - y_0^2 F_1' F_1\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)\varepsilon' M_1\varepsilon + y_0^2 F_1' F_1\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)\varepsilon' M_1\varepsilon \\
& -\sigma^2\mathbf{tr}(M_2)\mathbf{tr}^4(M_3)\varepsilon' N_1\varepsilon + 2\sigma^2\mathbf{tr}^2(M_2)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)\varepsilon' N_1\varepsilon - \sigma^2\mathbf{tr}^3(M_2)(\mathbf{tr}(N_2))^2\varepsilon' N_1\varepsilon \\
& +y_0^2 F_1' F_1'\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)\varepsilon' N_1\varepsilon - y_0^2 F_1' F_1'\mathbf{tr}^4(M_3)\varepsilon' N_1\varepsilon
\end{aligned}$$

$$
\begin{aligned}
&+2y_0^2\mathbf{tr}(M_4)\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}(N_2)(F_1'\varepsilon)^2 - y_0^2\mathbf{tr}(M_5)\mathbf{tr}(M_2)\mathbf{tr}^2(N_2)(F_1'\varepsilon)^2 \\
&+y_0^2\mathbf{tr}(M_4)\mathbf{tr}^3(M_3)(F_1'\varepsilon)^2 + 2y_0\mathbf{tr}(M_4)\mathbf{tr}^3(M_3)F_1'\varepsilon\varepsilon' M_1\varepsilon \\
&-2y_0^2\mathbf{tr}(M_5)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)(F_1'\varepsilon)^2 - y_0\mathbf{tr}(N_3)\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)F_1'\varepsilon\varepsilon' M_1\varepsilon \\
&+4y_0\mathbf{tr}(M_4)\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon' M_1\varepsilon - 2y_0\mathbf{tr}(M_5)\mathbf{tr}(M_2)\mathbf{tr}^2(N_2)F_1'\varepsilon\varepsilon' M_1\varepsilon \\
&-4y_0\mathbf{tr}(M_5)\mathbf{tr}^2(M3)\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon' M_1\varepsilon + y_0\mathbf{tr}(N_3)\mathbf{tr}^2(M_2)\mathbf{tr}(M_3)F_1'\varepsilon\varepsilon' N_1\varepsilon \\
&-2y_0\mathbf{tr}(M_4)\mathbf{tr}^2(M_2)\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon' N_1\varepsilon - 2y_0\sigma^2\mathbf{tr}(M_2)\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)F_1'\varepsilon \\
&-4y_0\mathbf{tr}(M_4)\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)F_1'\varepsilon\varepsilon' N_1\varepsilon + 4y_0\mathbf{tr}(M_5)\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon' N_1\varepsilon \\
&+2y_0\mathbf{tr}(M_5)\mathbf{tr}^3(M_3)F_1'\varepsilon\varepsilon' N_1\varepsilon + y_0\sigma^2\mathbf{tr}^2(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)F_1'\varepsilon \\
&-y_0^3F_1'F_1\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)F_1'\varepsilon + y_0\sigma^2\mathbf{tr}^5(M_3)F_1'\varepsilon + y_0^3F_1'F1\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)F_1'\varepsilon \\
&-y_0\mathbf{tr}^2(M_2)\mathbf{tr}(M_3)\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon' N_2\varepsilon + y_0\mathbf{tr}(M_2)\mathbf{tr}^3(M_3)F_1'\varepsilon\varepsilon' N_2\varepsilon \\
&+y_0^2\mathbf{tr}^2(M_2)\mathbf{tr}^2(N_2)\varepsilon' F_1F_3'\varepsilon - 2y_0^2\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)\varepsilon' F_1F_2'\varepsilon \\
&-y_0^2\mathbf{tr}^4(M_3)\varepsilon' F_1F_3'\varepsilon + y_0\mathbf{tr}^2(M_2)\mathbf{tr}^2(N_2)F_1'\varepsilon\varepsilon' M_3\varepsilon - y_0\mathbf{tr}^4(M_3)F_1'\varepsilon\varepsilon' M_3\varepsilon \\
&+2y_0^2\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)\varepsilon' F_1F_2'\varepsilon - y_0\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)F_1'\varepsilon\varepsilon' M_2\varepsilon \\
&+y_0\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon' M_2\varepsilon + 2y_0\mathbf{tr}(M_2)\mathbf{tr}^3(M_3)F_3'\varepsilon\varepsilon' N_1\varepsilon \\
&+y_0\mathbf{tr}^2(M_2)\mathbf{tr}^2(N_2)F_3'\varepsilon\varepsilon' M_1\varepsilon - y_0\mathbf{tr}^4(M_3)F_3'\varepsilon\varepsilon' M_1\varepsilon - 2y_0\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)F_2'\varepsilon\varepsilon' M_1\varepsilon \\
&+2y_0\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)F_2'\varepsilon\varepsilon' M_1\varepsilon - 2y_0\mathbf{tr}^2(M_2)\mathbf{tr}(M_3)\mathbf{tr}(N_2)F_3'\varepsilon\varepsilon' N_1\varepsilon \\
&-2y_0\mathbf{tr}^4(M_3)F_2'\varepsilon\varepsilon' N_1\varepsilon + 2y_0\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)F_2'\varepsilon\varepsilon' N_1\varepsilon\Big] \\
&/\Big[\sigma^4(\mathbf{tr}(N_2)\mathbf{tr}(M_2) - (\mathbf{tr}(M_3))^2)^3\Big]
\end{aligned}
$$

$$
\begin{aligned}
\rho_{-1} \quad = \quad &\Big[ -\frac{1}{2}\mathbf{tr}^3(M_3)\mathbf{tr}(N_3)(\varepsilon' M_1\varepsilon)^2 + 3\mathbf{tr}(M_4)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)(\varepsilon' M_1\varepsilon)^2 \\
&-3\mathbf{tr}(M_3)\mathbf{tr}(M_5)\mathbf{tr}^2(N_2)(\varepsilon' M_1\varepsilon)^2 + 4\mathbf{tr}(M_5)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)\varepsilon' M_1\varepsilon\varepsilon' N_1\varepsilon \\
&+\mathbf{tr}(N_3)\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)\varepsilon' M_1\varepsilon\varepsilon' N_1\varepsilon - 4\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}(M_4)\mathbf{tr}(N_2)\varepsilon' M_1\varepsilon\varepsilon' N_1\varepsilon \\
&+2\mathbf{tr}(M_2)\mathbf{tr}(M_5)\mathbf{tr}^2(N_2)\varepsilon' M_1\varepsilon\varepsilon' N_1\varepsilon - 2\mathbf{tr}(M_4)\mathbf{tr}^3(M_3)\varepsilon' M_1\varepsilon\varepsilon' N_1\varepsilon
\end{aligned}
$$

$$-\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)\varepsilon' M_1\varepsilon\varepsilon' N_2\varepsilon + \mathbf{tr}^4(M_3)\varepsilon' M_1\varepsilon\varepsilon' N_2\varepsilon - \mathbf{tr}(M_2)\mathbf{tr}^3(N_2)\varepsilon' M_1\varepsilon\varepsilon' M_2\varepsilon$$

$$+\mathbf{tr}^2(M_3)\mathbf{tr}^2(N_2)\varepsilon' M_1\varepsilon\varepsilon' M_2\varepsilon - 2\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)\varepsilon' M_1\varepsilon\varepsilon' M_3\varepsilon$$

$$+2\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)\varepsilon' M_1\varepsilon\varepsilon' M_3\varepsilon + 2\mathbf{tr}(M_2)\mathbf{tr}(M_4)\mathbf{tr}^2(M_3)(\varepsilon' N_1\varepsilon)^2$$

$$-\frac{1}{2}\mathbf{tr}(N_3)\mathbf{tr}^2(M_2)\mathbf{tr}(M_3)(\varepsilon' N_1\varepsilon)^2 + \mathbf{tr}(M_4)\mathbf{tr}^2(M_2)\mathbf{tr}(N_2)(\varepsilon' N_1\varepsilon)^2$$

$$-2\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}(M_5)\mathbf{tr}(N_2)(\varepsilon' N_1\varepsilon)^2 - \mathbf{tr}(M_5)\mathbf{tr}^3(M_3)(\varepsilon' N_1\varepsilon)^2$$

$$-\sigma^2\mathbf{tr}^2(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)\varepsilon' N_1\varepsilon - \mathbf{tr}(M_2)\mathbf{tr}^3(M_3)\varepsilon' N_1\varepsilon\varepsilon' N_2\varepsilon$$

$$+2\sigma^2\mathbf{tr}(M_2)\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)\varepsilon' N_1\varepsilon - \sigma^2\mathbf{tr}^5(M_3)\varepsilon' N_1\varepsilon + y_0^2F_1'F_1\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)\varepsilon' N_1\varepsilon$$

$$-y_0^2F_1'F_1\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)\varepsilon' N_1\varepsilon + (\mathbf{tr}(M_2))^2\mathbf{tr}(M_3)\mathbf{tr}(N_2)\varepsilon' N_1\varepsilon\varepsilon' N_2\varepsilon$$

$$+\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)\varepsilon' N_1\varepsilon\varepsilon' M_2\varepsilon - \mathbf{tr}^3(M_3)\mathbf{tr}(N_2)\varepsilon' N_1\varepsilon\varepsilon' M_2\varepsilon + \mathbf{tr}^4(M_3)\varepsilon' N_1\varepsilon\varepsilon' M_3\varepsilon$$

$$-\mathbf{tr}^2(M_2)\mathbf{tr}^2(N_2)\varepsilon' N_1\varepsilon\varepsilon' M_3\varepsilon + \sigma^2\mathbf{tr}^4(M_3)\mathbf{tr}(N_2)\varepsilon' M_1\varepsilon + \sigma^2\mathbf{tr}^2(M_2)\mathbf{tr}^3(N_2)\varepsilon' M_1\varepsilon$$

$$-2\sigma^2\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)\mathbf{tr}^2(N_2)\varepsilon' M_1\varepsilon - y_0^2F_1'F_1\mathbf{tr}(M_2)\mathbf{tr}^3(N_2)\varepsilon' M_1\varepsilon$$

$$+y_0^2F_1'F_1\mathbf{tr}^2(M_3)\mathbf{tr}^2(N_2)\varepsilon' M_1\varepsilon - 3y_0^2\mathbf{tr}(M_5)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)(F_1'\varepsilon)^2$$

$$-\frac{1}{2}y_0^2\mathbf{tr}(N_3)\mathbf{tr}^3(M_3)(F_1'\varepsilon)^2 + 3y_0^2\mathbf{tr}(M_4)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)(F_1'\varepsilon)^2$$

$$-y_0\mathbf{tr}(N_3)\mathbf{tr}^3(M_3)F_1'\varepsilon\varepsilon' M_1\varepsilon + 6y_0\mathbf{tr}(M_4)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon' M_1\varepsilon$$

$$-6y_0\mathbf{tr}(M_5)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)F_1'\varepsilon\varepsilon' M_1\varepsilon + 4y_0\mathbf{tr}(M_5)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon' N_1\varepsilon$$

$$+y_0\mathbf{tr}(N_3)\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)F_1'\varepsilon\varepsilon' N_1\varepsilon - 4y_0\mathbf{tr}(M_4)\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon' N_1\varepsilon$$

$$+2y_0\mathbf{tr}(M_5)\mathbf{tr}(M_2)\mathbf{tr}^2(N_2)F_1'\varepsilon\varepsilon' N_1\varepsilon - 2y_0\mathbf{tr}(M_4)\mathbf{tr}^3(M_3)F_1'\varepsilon\varepsilon' N_1\varepsilon$$

$$+y_0\sigma^2\mathbf{tr}^2(M_2)\mathbf{tr}^3(N_2)F_1'\varepsilon - y_0^3F_1'F_1\mathbf{tr}(M_2)\mathbf{tr}^3(N_2)F_1'\varepsilon - 2y_0\sigma^2\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)\mathbf{tr}^2(N_2)F_1'\varepsilon$$

$$+y_0\sigma^2\mathbf{tr}^4(M_3)\mathbf{tr}(N_2)F_1'\varepsilon + y_0^3F_1'F_1\mathbf{tr}^2(M_3)\mathbf{tr}^2(N_2)F_1'\varepsilon + 2y_0^2\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)\varepsilon' F_1F_3'\varepsilon$$

$$-2y_0^2\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)\varepsilon' F_1F_3'\varepsilon + 2y_0\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)F_1'\varepsilon\varepsilon' M_3\varepsilon$$

$$-2y_0\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon' M_3\varepsilon - 2y_0\mathbf{tr}(M_2)\mathbf{tr}^3(N_2)F_2'\varepsilon\varepsilon' M_1\varepsilon$$

$$-2y_0^2\mathbf{tr}(M_2)\mathbf{tr}^3(N_2)\varepsilon' F_1F_2'\varepsilon + 2y_0^2\mathbf{tr}^2(M_3)\mathbf{tr}^2(N_2)\varepsilon' F_1F_2'\varepsilon - y_0\mathbf{tr}(M_2)\mathbf{tr}^3(N_2)F_1'\varepsilon\varepsilon' M_2\varepsilon$$

$$+y_0\mathbf{tr}^2(M_3)\mathbf{tr}^2(N_2)F_1'\varepsilon\varepsilon'M_2\varepsilon + y_0\mathbf{tr}^4(M_3)F_1'\varepsilon\varepsilon'N_2\varepsilon - y_0\mathbf{tr}(M_2)\mathbf{tr}^2(M_3)\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon'N_2\varepsilon$$

$$+2y_0\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)F_3'\varepsilon\varepsilon'M_1\varepsilon - 2y_0\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)F_3'\varepsilon\varepsilon'M_1\varepsilon$$

$$+2y_0\mathbf{tr}^2(M_3)\mathbf{tr}^2(N_2)F_2'\varepsilon\varepsilon'M_1\varepsilon - y_0\mathbf{tr}^2(M_2)\mathbf{tr}^2(N_2)F_3'\varepsilon\varepsilon'N_1\varepsilon + y_0\mathbf{tr}^4(M_3)F_3'\varepsilon\varepsilon'N_1\varepsilon$$

$$+2y_0\mathbf{tr}(M_2)\mathbf{tr}(M_3)\mathbf{tr}^2(N_2)F_2'\varepsilon\varepsilon'N_1\varepsilon - 4y_0\mathbf{tr}^3(M_3)\mathbf{tr}(N_2)F_2'\varepsilon\varepsilon'N_1\varepsilon\Big]$$

$$/\Big[\sigma^4(\mathbf{tr}(M_2)\mathbf{tr}(N_2) - (\mathbf{tr}(M_3))^2)^3\Big]$$

$$\sigma^2_{-1} = -\Big[\mathbf{tr}(N_2)(\varepsilon'M_1\varepsilon)^2 - 2\mathbf{tr}(M_3)\varepsilon'M_1\varepsilon\varepsilon'N_1\varepsilon + \mathbf{tr}(M_2)(\varepsilon'N_1\varepsilon)^2 - 2y_0\mathbf{tr}(M_3)F_1'\varepsilon\varepsilon'N_1\varepsilon$$
$$+(F_1'\varepsilon)^2\mathbf{tr}(N_2)y_0^2 + 2F_1'\varepsilon\varepsilon'M_1\varepsilon\mathbf{tr}(N_2)y_0\Big]/\Big[T\sigma^2(\mathbf{tr}(N_2)\mathbf{tr}(M_2) - \mathbf{tr}^2(M_3))\Big]$$

## B.2 Expansion of $\hat{\beta} - \beta$ for the intercept model

Consider the ARMA(1,1) with intercept: $y_t = \alpha + \rho y_{t-1} - \phi\varepsilon_{t-1} + \varepsilon_t$, where $\alpha$ is a constant, $\alpha = O(1)$. In vector form, we have $y_{-1} = \alpha M_1\iota + y_0F + A\varepsilon$, where $\iota$ is a vector of ones with dimension $T$. Alternatively, we can rewrite the model as $\varepsilon = \iota_T' y_T - \alpha\iota + Dy_{-1} + B\varepsilon$. Now the parameter to be estimated is $\beta = (\alpha, \phi, \rho, \sigma^2)'$. Then given the quasi likelihood function by imposing normality assumption and following similar steps, we have the following score function:

$$\psi_T = \left( \frac{a'\varepsilon}{T\sigma^2}, \quad -\frac{\varepsilon'N_1\varepsilon}{T\sigma^2}, \quad \frac{\varepsilon'M_1\varepsilon + y_0F_1'\varepsilon + b'\varepsilon}{T\sigma^2}, \quad -\frac{1}{2\sigma^2} + \frac{\varepsilon'\varepsilon}{2T\sigma^4} \right)'$$

where $a = -C^{-1}\iota$, $b = \alpha C^{-1}M_1\iota$. The Hessian matrix $H_1$, is now $4 \times 4$ symmetric, and it has the following unique elements:

$$\frac{\partial^2 L}{\partial\alpha^2} = -\frac{a'a}{T\sigma^2}, \quad \frac{\partial^2 L}{\partial\alpha\partial\phi} = \frac{a'N_1^*\varepsilon}{T\sigma^2}, \quad \frac{\partial^2 L}{\partial\alpha\partial\rho} = -\frac{a'b + y_0a'F_1 + a'M_1\varepsilon}{T\sigma^2},$$

$$\frac{\partial^2 L}{\partial\alpha\partial\sigma^2} = -\frac{a'\varepsilon}{T\sigma^4}, \quad \frac{\partial^2 L}{\partial\phi^2} = -\frac{\varepsilon' N_2\varepsilon}{T\sigma^2}, \quad \frac{\partial^2 L}{\partial\phi\partial\rho} = \frac{\varepsilon' M_3\varepsilon + y_0 F_3'\varepsilon + b' N_1^*\varepsilon}{T\sigma^2},$$
$$\frac{\partial^2 L}{\partial\phi\partial\sigma^2} = \frac{\varepsilon' N_1\varepsilon}{T\sigma^4}, \quad \frac{\partial^2 L}{\partial\rho^2} = -\frac{\varepsilon' M_2\varepsilon + 2y_0 F_2'\varepsilon + y_0^2 F_1' F_1 + 2b' M_1\varepsilon + 2y_0 b' F_1 + b'b}{T\sigma^2},$$
$$\frac{\partial^2 L}{\partial\rho\partial\sigma^2} = -\frac{\varepsilon' M_1\varepsilon + y_0 F_1'\varepsilon + b'\varepsilon}{T\sigma^4}, \quad \frac{\partial^2 L}{\partial\sigma^4} = -\frac{1}{2\sigma^4} - \frac{\varepsilon'\varepsilon}{T\sigma^6}.$$

where for any matrix $X$, $X^* = X + X'$.

Then $H_2$, which is $16 \times 4$, has the following elements:

$$\frac{\partial^3 L}{\partial\alpha^3} = \frac{\partial^3 L}{\partial\alpha^2\partial\rho} = \frac{\partial^3 L}{\partial\alpha\partial\rho^2} = \frac{\partial^3 L}{\partial\rho^3} = 0, \quad \frac{\partial^3 L}{\partial\alpha^2\partial\phi} = -\frac{2a' N_1 a}{T\sigma^2}, \quad \frac{\partial^3 L}{\partial\alpha^2\partial\sigma^2} = \frac{a'a}{T\sigma^4},$$
$$\frac{\partial^3 L}{\partial\alpha\partial\phi^2} = \frac{a' N_2^*\varepsilon}{T\sigma^2}, \quad \frac{\partial^3 L}{\partial\alpha\partial\phi\partial\rho} = -\frac{a' N_1^* b + y_0 a' F_3 + a' M_3\varepsilon}{T\sigma^2}, \quad \frac{\partial^3 L}{\partial\alpha\partial\phi\partial\sigma^2} = -\frac{a' N_1^*\varepsilon}{T\sigma^4},$$
$$\frac{\partial^3 L}{\partial\alpha\partial\rho\partial\sigma^2} = \frac{a'b + y_0 a' F_1 + a' M_1\varepsilon}{T\sigma^4}, \quad \frac{\partial^3 L}{\partial\alpha\partial\sigma^4} = \frac{2a'\varepsilon}{T\sigma^6}, \quad \frac{\partial^3 L}{\partial\phi^3} = -\frac{\varepsilon' N_3\varepsilon}{T\sigma^2},$$
$$\frac{\partial^3 L}{\partial\phi^2\partial\rho} = \frac{2(y_0 F_4'\varepsilon + \varepsilon' M_4\varepsilon)}{T\sigma^2} + \frac{b' N_2^*\varepsilon}{T\sigma^2}, \quad \frac{\partial^3 L}{\partial\phi^2\partial\sigma^2} = \frac{\varepsilon' N_2\varepsilon}{T\sigma^4},$$
$$\frac{\partial^3 L}{\partial\phi\partial\rho^2} = -\frac{2(b' N_1 b + y_0 b' N_1^* F_1 + b' N_1^* M_1\varepsilon + y_0^2 F_1' N_1 F_1 + y_0 F_5'\varepsilon + \varepsilon' M_5\varepsilon)}{T\sigma^2},$$
$$\frac{\partial^3 L}{\partial\phi\partial\sigma^4} = -\frac{2\varepsilon' N_1\varepsilon}{T\sigma^6}, \quad \frac{\partial^3 L}{\partial\phi\partial\rho\partial\sigma^2} = -\frac{\varepsilon' M_3\varepsilon + y_0 F_3'\varepsilon + b' N_1^*\varepsilon}{T\sigma^4},$$
$$\frac{\partial^3 L}{\partial\rho^2\partial\sigma^2} = \frac{\varepsilon' M_2\varepsilon + 2y_0 F_2'\varepsilon + y_0^2 F_1' F_1 + 2b' M_1\varepsilon + 2y_0 b' F_1 + b'b}{T\sigma^4},$$
$$\frac{\partial^3 L}{\partial\rho\partial\sigma^4} = \frac{2(\varepsilon' M_1\varepsilon + y_0 F_1'\varepsilon + b'\varepsilon)}{T\sigma^6}, \quad \frac{\partial^3 L}{\partial\alpha\partial\sigma^4} = \frac{2a'\varepsilon}{T\sigma^6}, \quad \frac{\partial^3 L}{\partial\sigma^6} = -\frac{1}{\sigma^6} + \frac{3\varepsilon'\varepsilon}{T\sigma^8}.$$

Applying the expectations of quadratic forms to the Hessian and $H_2$ matrix, we have, up to order $O(1)$,

$$\bar{H}_1 = \begin{pmatrix} -\frac{a'a}{T\sigma^2} & 0 & -\frac{a'b}{T\sigma^2} & 0 \\ 0 & -\frac{\mathbf{tr}(N_2)}{T} & \frac{\mathbf{tr}(M_3)}{T} & 0 \\ -\frac{a'b}{T\sigma^2} & \frac{\mathbf{tr}(M_3)}{T} & -\frac{\mathbf{tr}(M_2)}{T} - \frac{b'b}{T\sigma^2} & 0 \\ 0 & 0 & 0 & -\frac{1}{2\sigma^4} \end{pmatrix}$$

$$
\bar{H}_2 = \begin{pmatrix} 0 & -\frac{2a'N_1a}{T\sigma^2} & 0 & \frac{a'a}{T\sigma^4} & -\frac{2a'N_1a}{T\sigma^2} & 0 & -\frac{a'N_1^*b}{T\sigma^2} & 0 \\ -\frac{2a'N_1a}{T\sigma^2} & 0 & -\frac{a'N_1^*b}{T\sigma^2} & 0 & 0 & -\frac{\mathbf{tr}(N_3)}{T} & \frac{2\mathbf{tr}(M_4)}{T} & \frac{\mathbf{tr}(N_2)}{T\sigma^2} \\ 0 & -\frac{a'N_1^*b}{T\sigma^2} & 0 & \frac{a'b}{T\sigma^4} & -\frac{a'N_1^*b}{T\sigma^2} & \frac{2\mathbf{tr}(M_4)}{T} & -\frac{2\mathbf{tr}(M_5)}{T} - \frac{2b'N_1b}{T\sigma^2} & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} \\ \frac{a'a}{T\sigma^4} & 0 & \frac{a'b}{T\sigma^4} & 0 & 0 & \frac{\mathbf{tr}(N_2)}{T\sigma^2} & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} & 0 \end{pmatrix}
$$

$$
\asymp \begin{pmatrix} 0 & -\frac{a'N_1^*b}{T\sigma^2} & 0 & \frac{a'b}{T\sigma^4} & \frac{a'a}{T\sigma^4} & 0 & \frac{a'b}{T\sigma^4} & 0 \\ -\frac{a'N_1^*b}{T\sigma^2} & \frac{2\mathbf{tr}(M_4)}{T} & -\frac{2\mathbf{tr}(M_5)}{T} - \frac{2b'N_1b}{T\sigma^2} & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} & 0 & \frac{\mathbf{tr}(N_2)}{T\sigma^2} & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} & 0 \\ 0 & -\frac{2\mathbf{tr}(M_5)}{T} - \frac{2b'N_1b}{T\sigma^2} & 0 & \frac{\mathbf{tr}(M_2)}{T\sigma^2} + \frac{b'b}{T\sigma^4} & \frac{a'b}{T\sigma^4} & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} & \frac{\mathbf{tr}(M_2)}{T\sigma^2} + \frac{b'b}{T\sigma^4} & 0 \\ \frac{a'b}{T\sigma^4} & -\frac{\mathbf{tr}(M_3)}{T\sigma^2} & \frac{\mathbf{tr}(M_2)}{T\sigma^2} + \frac{b'b}{T\sigma^4} & 0 & 0 & 0 & 0 & \frac{2}{\sigma^6} \end{pmatrix}
$$

where $\asymp$ denotes matrix horizontal concatenation.

Substitue the above results to (2.3), we have:

$$
\alpha_{-1/2} = \frac{a'b\mathbf{tr}(M_3)\varepsilon'N_1\varepsilon - \mathbf{tr}(N_2)(a'b\varepsilon'M_1\varepsilon + a'by_0F_1'\varepsilon + a'bb'\varepsilon - b'ba'\varepsilon) + \sigma^2a'\varepsilon(\mathbf{tr}(M_2)\mathbf{tr}(N_2) - \mathbf{tr}^2(M_3))}{\sigma^2a'a(\mathbf{tr}(M_2)\mathbf{tr}(N_2) - \mathbf{tr}^2(M_3)) + \mathbf{tr}(N_2)(a'ab'b - (a'b)^2)}
$$

$$
\phi_{-1/2} = -\frac{a'a\mathbf{tr}(M_2)\varepsilon'N_1\varepsilon - \mathbf{tr}(M_3)(a'a\varepsilon'M_1\varepsilon + a'ay_0F_1'\varepsilon + a'ab'\varepsilon - a'ba'\varepsilon) + \varepsilon'N_1\varepsilon(a'ab'b - (a'b)^2)/\sigma^2}{\sigma^2a'a(\mathbf{tr}(M_2)\mathbf{tr}(N_2) - \mathbf{tr}^2(M_3)) + \mathbf{tr}(N_2)(a'ab'b - (a'b)^2)},
$$

$$
\rho_{-1/2} = -\frac{a'a\mathbf{tr}(M_3)\varepsilon'N_1\varepsilon - \mathbf{tr}(N_2)(a'a\varepsilon'M_1\varepsilon + a'ay_0F_1'\varepsilon + a'ab'\varepsilon - a'ba'\varepsilon)}{\sigma^2a'a(\mathbf{tr}(M_2)\mathbf{tr}(N_2) - \mathbf{tr}^2(M_3)) + \mathbf{tr}(N_2)(a'ab'b - (a'b)^2)}, \quad \sigma^2_{-1/2} = -\sigma^2 + \frac{\varepsilon'\varepsilon}{T}.
$$

$$
\begin{aligned}
\sigma^2_{-1} = & -\Big[y_0^2\sigma^2a'a\mathbf{tr}(N_2)(F_1'\varepsilon)^2 - 2y_0\sigma^2a'b\mathbf{tr}(N_2)\varepsilon'aF_1'\varepsilon + 2y_0\sigma^2a'a\mathbf{tr}(N_2)\varepsilon'bF_1'\varepsilon \\
& +2y_0\sigma^2a'a\mathbf{tr}(N_2)F_1'\varepsilon\varepsilon'M_1\varepsilon - 2\sigma^2a'b\mathbf{tr}(N_2)\varepsilon'a\varepsilon'M_1\varepsilon \\
& -2y_0\sigma^2a'a\mathbf{tr}(M_3)F_1'\varepsilon\varepsilon'N_1\varepsilon - (a'b)^2(\varepsilon'N_1\varepsilon)^2 - 2\sigma^2a'b\mathbf{tr}(N_2)\varepsilon'ab'\varepsilon \\
& +2\sigma^2a'b\mathbf{tr}(M_3)a'\varepsilon\varepsilon'N_1\varepsilon - \sigma^4\mathbf{tr}^2(M_3)(a'\varepsilon)^2 + \sigma^4\mathbf{tr}(M_2)\mathbf{tr}(N_2)(a'\varepsilon)^2 \\
& +\sigma^2b'b\mathbf{tr}(N_2)(a'\varepsilon)^2 + \sigma^2a'a\mathbf{tr}(N_2)(\varepsilon'M_1\varepsilon)^2 \\
& +\sigma^2a'a\mathbf{tr}(N_2)(b'\varepsilon)^2 + 2\sigma^2a'a\mathbf{tr}(N_2)b'\varepsilon\varepsilon'M_1\varepsilon - 2\sigma^2a'a\mathbf{tr}(M_3)b'\varepsilon\varepsilon'N_1\varepsilon \\
& -2\sigma^2a'a\mathbf{tr}(M_3)\varepsilon'M_1\varepsilon\varepsilon'N_1\varepsilon + \sigma^2a'a\mathbf{tr}(M_2)(\varepsilon'N_1\varepsilon)^2 + a'ab'b(\varepsilon'N_1\varepsilon)^2\Big] \\
& /\big[T\sigma^2(\sigma^2a'a(\mathbf{tr}(M_2)\mathbf{tr}(N_2) - \mathbf{tr}^2(M_3)) + \mathbf{tr}(N_2)(a'ab'b - (a'b)^2))\big]
\end{aligned}
$$

The terms of $\alpha_{-1}$, $\phi_{-1}$ and $\rho_{-1}$ is much longer than the pure model and is not reported here but available upon request to the author.

# Appendix C

# Appendix for Chapter 3

## C.1 Expansion of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$

Let $\boldsymbol{\psi}$ denote the score function and $\boldsymbol{H}_i = \bigtriangledown^i \boldsymbol{\psi}$, all evaluated the true parameter vector.[1] Following Bao and Ullah (2007, 2009), a stochastic expansion of the QMLE $\hat{\boldsymbol{\beta}}$ can be written as

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{\beta}_{-1/2} + \boldsymbol{\beta}_{-1} + \boldsymbol{\beta}_{-3/2} + \boldsymbol{\beta}_{-2} + o_P(T^{-2}),$$

where $\boldsymbol{\beta}_{-i/2}$ are terms of orders $O_P(T^{-i/2})$, defined as

$$\boldsymbol{\beta}_{-1/2} = \boldsymbol{\Sigma}\boldsymbol{\psi},$$

$$\boldsymbol{\beta}_{-1} = \boldsymbol{\Sigma}\boldsymbol{V}_1\boldsymbol{\beta}_{-1/2} + \frac{1}{2}\boldsymbol{\Sigma}\mathbb{E}(\boldsymbol{H}_2)(\boldsymbol{\beta}_{-1/2} \otimes \boldsymbol{\beta}_{-1/2}),$$

$$\boldsymbol{\beta}_{-3/2} = \boldsymbol{\Sigma}\boldsymbol{V}_1\boldsymbol{a}_{-1} + \frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{V}_2(\boldsymbol{\beta}_{-1/2} \otimes \boldsymbol{\beta}_{-1/2}) + \frac{1}{2}\boldsymbol{\Sigma}\mathbb{E}(\boldsymbol{H}_2)(\boldsymbol{\beta}_{-1/2} \otimes \boldsymbol{\beta}_{-1} + \boldsymbol{\beta}_{-1} \otimes \boldsymbol{\beta}_{-1/2})$$
$$+ \frac{1}{6}\boldsymbol{\Sigma}\mathbb{E}(\boldsymbol{H}_3)(\boldsymbol{\beta}_{-1/2} \otimes \boldsymbol{\beta}_{-1/2} \otimes \boldsymbol{\beta}_{-1/2}),$$

$$\boldsymbol{\beta}_{-2} = \boldsymbol{\Sigma}\boldsymbol{V}_1\boldsymbol{\beta}_{-3/2} + \frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{V}_2(\boldsymbol{\beta}_{-1/2} \otimes \boldsymbol{\beta}_{-1} + \boldsymbol{\beta}_{-1} \otimes \boldsymbol{\beta}_{-1/2}) + \frac{1}{2}\boldsymbol{\Sigma}\mathbb{E}(\boldsymbol{H}_2)(\boldsymbol{\beta}_{-1/2} \otimes \boldsymbol{\beta}_{-3/2}$$

[1] The matrices $\boldsymbol{H}_i$ are defined recursively as in Rilstone et al. (1996)

$$+\boldsymbol{\beta}_{-3/2}\otimes\boldsymbol{\beta}_{-1/2}+\boldsymbol{\beta}_{-1}\otimes\boldsymbol{\beta}_{-1})+\frac{1}{6}\boldsymbol{\Sigma}\boldsymbol{V}_3(\boldsymbol{\beta}_{-1/2}\otimes\boldsymbol{\beta}_{-1/2}\otimes\boldsymbol{\beta}_{-1/2})$$
$$+\frac{1}{6}\boldsymbol{\Sigma}\mathbb{E}(\boldsymbol{H}_3)(\boldsymbol{\beta}_{-1/2}\otimes\boldsymbol{\beta}_{-1/2}\otimes\boldsymbol{\beta}_{-1}+\boldsymbol{\beta}_{-1/2}\otimes\boldsymbol{\beta}_{-1}\otimes\boldsymbol{\beta}_{-1/2}+\boldsymbol{\beta}_{-1}\otimes\boldsymbol{\beta}_{-1/2}\otimes\boldsymbol{\beta}_{-1/2})$$
$$+\frac{1}{24}\boldsymbol{\Sigma}\mathbb{E}(\boldsymbol{H}_4)\left(\boldsymbol{\beta}_{-1/2}\otimes\boldsymbol{\beta}_{-1/2}\otimes\boldsymbol{\beta}_{-1/2}\otimes\boldsymbol{\beta}_{-1/2}\right),$$

in which $\boldsymbol{\Sigma}=-[\mathbb{E}(\boldsymbol{H}_1)]^{-1}$ and $\boldsymbol{V}_i=\boldsymbol{H}_i-\mathbb{E}(\boldsymbol{H}_i)$. For deriving the second-order bias of $\hat{\boldsymbol{\beta}}$, we need only the expansion up to $\boldsymbol{\beta}_{-1}$. But for investigating the second-order bias of the one-step-ahead feasible forecast, as suggested in (3.5), we need the expansion up to $\boldsymbol{\beta}_{-2}$. For notational convenience, let $L_{i_1i_2\cdots i_r}$ denote the $r$th-order derivative of the quasi log likelihood function (3.2) with respect to the elements of $\boldsymbol{\beta}$ in the order of $i_1i_2\cdots i_r$.

Given the quasi log likelihood function (3.2), the score function can be written as

$$\boldsymbol{\psi}'=\left(\frac{\boldsymbol{a}'\boldsymbol{\varepsilon}}{T\sigma^2},\quad \frac{\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{T\sigma^2},\quad -\frac{1}{2\sigma^2}+\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2T\sigma^4}\right)'.$$

The Hessian matrix $\boldsymbol{H}_1$, which is $3\times 3$ and symmetric, has its unique elements

$$\begin{aligned} L_{11} &= -\frac{\boldsymbol{a}'\boldsymbol{a}}{T\sigma^2},\quad L_{12}=-\frac{\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}}{T\sigma^2},\quad L_{13}=-\frac{\boldsymbol{a}'\boldsymbol{\varepsilon}}{T\sigma^4},\\ L_{22} &= -\frac{\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon}}{T\sigma^2},\quad L_{23}=-\frac{\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{T\sigma^4},\quad L_{33}=-\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{T\sigma^6}+\frac{1}{2\sigma^4}, \end{aligned}$$

which gives

$$\boldsymbol{\Sigma}=\mathrm{diag}\left(\frac{T\sigma^2}{\boldsymbol{a}'\boldsymbol{a}},\quad \frac{T}{\mathrm{tr}(\boldsymbol{A}_2)},\quad 2\sigma^4\right).$$

Next, by taking derivatives and using the identities $\boldsymbol{\varepsilon}=\boldsymbol{C}^{-1}(\boldsymbol{y}-\mu\boldsymbol{\iota})$ and $\boldsymbol{y}=\mu\boldsymbol{\iota}+\boldsymbol{C}\boldsymbol{\varepsilon}$, we put the non-zero unique elements of higher-order derivative matrices $\boldsymbol{H}_i$, $i=2,3,4$, in the following, with the additional notation: $\boldsymbol{A}_3=\boldsymbol{A}_1'\boldsymbol{A}_1+\boldsymbol{A}_1'^2+\boldsymbol{A}_1^2$, $\boldsymbol{A}_4=\boldsymbol{A}_1^3+\boldsymbol{A}_1'\boldsymbol{A}_1^2$, $\boldsymbol{A}_5=\boldsymbol{A}_1'^3+\boldsymbol{A}_1^3+\boldsymbol{A}_1'\boldsymbol{A}_1^2+\boldsymbol{A}_1'^2\boldsymbol{A}_1$, $\boldsymbol{A}_6=2\boldsymbol{A}_1^4+2\boldsymbol{A}_1'\boldsymbol{A}_1^3+\boldsymbol{A}_1'^2\boldsymbol{A}_1^2$, $\boldsymbol{A}_7=\boldsymbol{A}_1^5+\boldsymbol{A}_1'\boldsymbol{A}_1^4+\boldsymbol{A}_1'^2\boldsymbol{A}_1^3$.

For $\boldsymbol{H}_2$,

$$\begin{aligned}
L_{112} &= \frac{2\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}}{T\sigma^2}, \quad L_{113} = \frac{\boldsymbol{a}'\boldsymbol{a}}{T\sigma^4}, \quad L_{122} = \frac{2\boldsymbol{a}'\boldsymbol{A}_3\boldsymbol{\varepsilon}}{T\sigma^2}, \\
L_{123} &= \frac{\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}}{T\sigma^4}, \quad L_{133} = \frac{2\boldsymbol{a}'\boldsymbol{\varepsilon}}{T\sigma^6}, \quad L_{222} = \frac{6\boldsymbol{\varepsilon}'\boldsymbol{A}_4\boldsymbol{\varepsilon}}{T\sigma^2}, \\
L_{223} &= \frac{\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon}}{T\sigma^4}, \quad L_{233} = \frac{2\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{T\sigma^6}, \quad L_{333} = \frac{3\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{T\sigma^8} - \frac{1}{\sigma^6};
\end{aligned}$$

for $\boldsymbol{H}_3$,

$$\begin{aligned}
L_{1122} &= -\frac{2\boldsymbol{a}'\boldsymbol{A}_2\boldsymbol{a}}{T\sigma^2}, \quad L_{1123} = -\frac{2\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}}{T\sigma^4}, \quad L_{1133} = -\frac{2\boldsymbol{a}'\boldsymbol{a}}{T\sigma^6}, \\
L_{1222} &= -\frac{6\boldsymbol{a}'\boldsymbol{A}_5\boldsymbol{\varepsilon}}{T\sigma^2}, \quad L_{1223} = -\frac{2\boldsymbol{a}'\boldsymbol{A}_3\boldsymbol{\varepsilon}}{T\sigma^4}, \quad L_{1233} = -\frac{2\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}}{T\sigma^6}, \\
L_{1333} &= -\frac{6\boldsymbol{a}'\boldsymbol{\varepsilon}}{T\sigma^8}, \quad L_{2222} = -\frac{12\boldsymbol{\varepsilon}'\boldsymbol{A}_6\boldsymbol{\varepsilon}}{T\sigma^2}, \quad L_{2223} = -\frac{6\boldsymbol{\varepsilon}'\boldsymbol{A}_4\boldsymbol{\varepsilon}}{T\sigma^4}, \\
L_{2233} &= -\frac{2\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon}}{T\sigma^6}, \quad L_{2333} = -\frac{6\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{T\sigma^8}, \quad L_{3333} = -\frac{12\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{T\sigma^{10}} + \frac{3}{\sigma^8};
\end{aligned}$$

and for $\boldsymbol{H}_4$

$$\begin{aligned}
L_{11222} &= \frac{12\boldsymbol{a}'\boldsymbol{A}_4\boldsymbol{a}}{T\sigma^2}, \quad L_{11223} = \frac{2\boldsymbol{a}'\boldsymbol{A}_2\boldsymbol{a}}{T\sigma^4}, \quad L_{11233} = \frac{4\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}}{T\sigma^6}, \\
L_{11333} &= \frac{6\boldsymbol{a}'\boldsymbol{a}}{T\sigma^8}, \quad L_{12222} = \frac{12\boldsymbol{a}'\boldsymbol{A}_6^*\boldsymbol{\varepsilon}}{T\sigma^2}, \quad L_{12223} = \frac{6\boldsymbol{a}'\boldsymbol{A}_5\boldsymbol{\varepsilon}}{T\sigma^4}, \\
L_{12233} &= \frac{4\boldsymbol{a}'\boldsymbol{A}_3\boldsymbol{\varepsilon}}{T\sigma^6}, \quad L_{12333} = \frac{6\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}}{T\sigma^8}, \quad L_{13333} = \frac{24\boldsymbol{a}'\boldsymbol{\varepsilon}}{T\sigma^{10}}, \\
L_{22222} &= \frac{120\boldsymbol{\varepsilon}'\boldsymbol{A}_7\boldsymbol{\varepsilon}}{T\sigma^2}, \quad L_{22223} = \frac{12\boldsymbol{\varepsilon}'\boldsymbol{A}_6\boldsymbol{\varepsilon}}{T\sigma^4}, \quad L_{22233} = \frac{12\boldsymbol{\varepsilon}'\boldsymbol{A}_4\boldsymbol{\varepsilon}}{T\sigma^6}, \\
L_{22333} &= \frac{6\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon}}{T\sigma^8}, \quad L_{23333} = \frac{24\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{T\sigma^{10}}, \quad L_{33333} = \frac{60\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{T\sigma^{12}} - \frac{12}{\sigma^{10}}.
\end{aligned}$$

Readers should keep in mind that the $\boldsymbol{H}_i$ are defined recursively as in Rilstone et al. (1996) and the dimensions of $\boldsymbol{H}_i$ are $3 \times 3^i$. The expectations of $\boldsymbol{H}_i$ can also be straightforwardly calculated as they involve only order-1 quadratic forms in $\boldsymbol{\varepsilon}$. By substituting $\boldsymbol{\psi}$, $\boldsymbol{\Sigma}$, $\boldsymbol{H}_i$, and $\mathbb{E}(\boldsymbol{H}_i)$ into the order $O_P(T^{-2})$ expansion of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$, we can derive the expansions for each parameter estimator. In the following, we present only those terms needed for developing our results:

$$\mu_{-1/2} = \frac{\boldsymbol{a'\varepsilon}}{\boldsymbol{a'a}},$$

$$\mu_{-1} = -\frac{\boldsymbol{a'A_1^*\varepsilon\varepsilon'A_1\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)\boldsymbol{a'a}} + \frac{2\boldsymbol{a'A_1aa'\varepsilon\varepsilon'A_1\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)(\boldsymbol{a'a})^2},$$

$$\sigma^2_{-1/2} = \frac{\boldsymbol{\varepsilon'\varepsilon}}{T} - \sigma^2,$$

$$\sigma^2_{-1} = -\frac{\boldsymbol{\varepsilon'aa'\varepsilon}}{T\boldsymbol{a'a}} - \frac{(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{T\sigma^2\mathrm{tr}(\boldsymbol{A}_2)},$$

$$\theta_{-1/2} = \frac{\boldsymbol{\varepsilon'A_1\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)},$$

$$\theta_{-1} = \frac{3\mathrm{tr}(\boldsymbol{A}_4)(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^4\mathrm{tr}^3(\boldsymbol{A}_2)} - \frac{\boldsymbol{\varepsilon'A_1\varepsilon\varepsilon'A_2\varepsilon}}{\sigma^4\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{\boldsymbol{\varepsilon'aa'A_1^*\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)\boldsymbol{a'a}} + \frac{\boldsymbol{a'A_1a\varepsilon'aa'\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)(\boldsymbol{a'a})^2} + \frac{\boldsymbol{\varepsilon'A_1\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)},$$

$$\begin{aligned}\theta_{-3/2} = {} & \frac{18\mathrm{tr}^2(\boldsymbol{A}_4)(\boldsymbol{\varepsilon'A_1\varepsilon})^3}{\sigma^6\mathrm{tr}^5(\boldsymbol{A}_2)} - \frac{2\mathrm{tr}(\boldsymbol{A}_6)(\boldsymbol{\varepsilon'A_1\varepsilon})^3}{\sigma^6\mathrm{tr}^4(\boldsymbol{A}_2)} - \frac{9\mathrm{tr}(\boldsymbol{A}_4)(\boldsymbol{\varepsilon'A_1\varepsilon})^2\boldsymbol{\varepsilon'A_2\varepsilon}}{\sigma^6\mathrm{tr}^4(\boldsymbol{A}_2)} + \frac{\boldsymbol{\varepsilon'A_1\varepsilon}(\boldsymbol{\varepsilon'A_2\varepsilon})^2}{\sigma^6\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{3(\boldsymbol{\varepsilon'A_1\varepsilon})^2\boldsymbol{\varepsilon'A_4\varepsilon}}{\sigma^6\mathrm{tr}^3(\boldsymbol{A}_2)} \\ & + \frac{\boldsymbol{\varepsilon'A_1^*aa'A_1^*\varepsilon\varepsilon'A_1\varepsilon}}{\sigma^4\mathrm{tr}^2(\boldsymbol{A}_2)\boldsymbol{a'a}} - \frac{4\boldsymbol{a'A_1a\varepsilon'aa'A_1^*\varepsilon\varepsilon'A_1\varepsilon}}{\sigma^4\mathrm{tr}^2(\boldsymbol{A}_2)(\boldsymbol{a'a})^2} + \frac{2\boldsymbol{\varepsilon'aa'A_3\varepsilon\varepsilon'A_1\varepsilon}}{\sigma^4\mathrm{tr}^2(\boldsymbol{A}_2)\boldsymbol{a'a}} + \frac{4(\boldsymbol{a'A_1a})^2\boldsymbol{\varepsilon'aa'\varepsilon\varepsilon'A_1\varepsilon}}{\sigma^4\mathrm{tr}^2(\boldsymbol{A}_2)(\boldsymbol{a'a})^3} \\ & - \frac{\boldsymbol{a'A_2a\varepsilon'aa'\varepsilon\varepsilon'A_1\varepsilon}}{\sigma^4\mathrm{tr}^2(\boldsymbol{A}_2)(\boldsymbol{a'a})^2} - \frac{6\mathrm{tr}(\boldsymbol{A}_4)\boldsymbol{\varepsilon'aa'A_1^*\varepsilon\varepsilon'A_1\varepsilon}}{\sigma^4\mathrm{tr}^3(\boldsymbol{A}_2)\boldsymbol{a'a}} + \frac{6\mathrm{tr}(\boldsymbol{A}_4)\boldsymbol{a'A_1a\varepsilon'aa'\varepsilon\varepsilon'A_1\varepsilon}}{\sigma^4\mathrm{tr}^3(\boldsymbol{A}_2)(\boldsymbol{a'a})^2} + \frac{6\mathrm{tr}(\boldsymbol{A}_4)(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^4\mathrm{tr}^3(\boldsymbol{A}_2)} \\ & + \frac{\boldsymbol{\varepsilon'aa'A_1^*\varepsilon\varepsilon'A_2\varepsilon}}{\sigma^4\mathrm{tr}^2(\boldsymbol{A}_2)\boldsymbol{a'a}} - \frac{\boldsymbol{a'A_1a\varepsilon'aa'\varepsilon\varepsilon'A_2\varepsilon}}{\sigma^4\mathrm{tr}^2(\boldsymbol{A}_2)(\boldsymbol{a'a})^2} - \frac{2\boldsymbol{\varepsilon'A_1\varepsilon\varepsilon'A_2\varepsilon}}{\sigma^4\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{\boldsymbol{\varepsilon'aa'A_1^*\varepsilon}}{\sigma^2\boldsymbol{a'a}\mathrm{tr}(\boldsymbol{A}_2)} + \frac{\boldsymbol{a'A_1a\varepsilon'aa'\varepsilon}}{\sigma^2(\boldsymbol{a'a})^2\mathrm{tr}(\boldsymbol{A}_2)} + \\ & \frac{\boldsymbol{\varepsilon'A_1\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)},\end{aligned}$$

$$\begin{aligned}\theta_{-2} = {} & \frac{135\mathrm{tr}^3(\boldsymbol{A}_4)(\boldsymbol{\varepsilon'A_1\varepsilon})^4}{\sigma^8\mathrm{tr}^7(\boldsymbol{A}_2)} - \frac{30\mathrm{tr}(\boldsymbol{A}_4)\mathrm{tr}(\boldsymbol{A}_6)(\boldsymbol{\varepsilon'A_1\varepsilon})^4}{\sigma^8\mathrm{tr}^6(\boldsymbol{A}_2)} + \frac{5\mathrm{tr}(\boldsymbol{A}_7)(\boldsymbol{\varepsilon'A_1\varepsilon})^4}{\sigma^8\mathrm{tr}^5(\boldsymbol{A}_2)} - \frac{90\mathrm{tr}^2(\boldsymbol{A}_4)(\boldsymbol{\varepsilon'A_1\varepsilon})^3\boldsymbol{\varepsilon'A_2\varepsilon}}{\sigma^8\mathrm{tr}^6(\boldsymbol{A}_2)} \\ & + \frac{8\mathrm{tr}(\boldsymbol{A}_6)(\boldsymbol{\varepsilon'A_1\varepsilon})^3\boldsymbol{\varepsilon'A_2\varepsilon}}{\sigma^8\mathrm{tr}^5(\boldsymbol{A}_2)} - \frac{\boldsymbol{\varepsilon'A_1\varepsilon}(\boldsymbol{\varepsilon'A_2\varepsilon})^3}{\sigma^8\mathrm{tr}^4(\boldsymbol{A}_2)} + \frac{18\mathrm{tr}(\boldsymbol{A}_4)(\boldsymbol{\varepsilon'A_1\varepsilon})^2(\boldsymbol{\varepsilon'A_2\varepsilon})^2}{\sigma^8\mathrm{tr}^5(\boldsymbol{A}_2)} + \frac{36\mathrm{tr}(\boldsymbol{A}_4)(\boldsymbol{\varepsilon'A_1\varepsilon})^3\boldsymbol{\varepsilon'A_4\varepsilon}}{\sigma^8\mathrm{tr}^5(\boldsymbol{A}_2)} \\ & - \frac{2(\boldsymbol{\varepsilon'A_1\varepsilon})^3\boldsymbol{\varepsilon'A_6\varepsilon}}{\sigma^8\mathrm{tr}^4(\boldsymbol{A}_2)} - \frac{9(\boldsymbol{\varepsilon'A_1\varepsilon})^2\boldsymbol{\varepsilon'A_2\varepsilon\varepsilon'A_4\varepsilon}}{\sigma^8\mathrm{tr}^4(\boldsymbol{A}_2)} - \frac{3\boldsymbol{\varepsilon'A_1^*aa'A_3\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6\boldsymbol{a'a}\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{3\boldsymbol{a'A_1a\varepsilon'A_1^*aa'A_1^*\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^2\mathrm{tr}^3(\boldsymbol{A}_2)} \\ & + \frac{6\boldsymbol{a'A_1a\varepsilon'aa'A_3\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^2\mathrm{tr}^3(\boldsymbol{A}_2)} - \frac{12(\boldsymbol{a'A_1a})^2\boldsymbol{\varepsilon'aa'A_1^*\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^3\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{3\boldsymbol{a'A_2a\varepsilon'aa'A_1^*\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^2\mathrm{tr}^3(\boldsymbol{A}_2)} \\ & - \frac{3\boldsymbol{\varepsilon'aa'A_5\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6\boldsymbol{a'a}\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{12(\boldsymbol{a'A_1a})^3\boldsymbol{\varepsilon'aa'\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^4\mathrm{tr}^3(\boldsymbol{A}_2)} - \frac{6\boldsymbol{a'A_1aa'A_2a\varepsilon'aa'\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^3\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{3\boldsymbol{a'A_4a\varepsilon'aa'\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^2\mathrm{tr}^3(\boldsymbol{A}_2)} \\ & + \frac{9\mathrm{tr}(\boldsymbol{A}_4)\boldsymbol{\varepsilon'A_1^*aa'A_1^*\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6\boldsymbol{a'a}\mathrm{tr}^4(\boldsymbol{A}_2)} + \frac{18\mathrm{tr}(\boldsymbol{A}_4)\boldsymbol{\varepsilon'aa'A_3\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6\boldsymbol{a'a}\mathrm{tr}^4(\boldsymbol{A}_2)} - \frac{36\mathrm{tr}(\boldsymbol{A}_4)\boldsymbol{a'A_1a\varepsilon'aa'A_1^*\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^2\mathrm{tr}^4(\boldsymbol{A}_2)} \\ & + \frac{36\mathrm{tr}(\boldsymbol{A}_4)(\boldsymbol{a'A_1a})^2\boldsymbol{\varepsilon'aa'\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^3\mathrm{tr}^4(\boldsymbol{A}_2)} - \frac{9\mathrm{tr}(\boldsymbol{A}_4)\boldsymbol{a'A_2a\varepsilon'aa'\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^2\mathrm{tr}^4(\boldsymbol{A}_2)} - \frac{54\mathrm{tr}^2(\boldsymbol{A}_4)\boldsymbol{\varepsilon'aa'A_1^*\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6\boldsymbol{a'a}\mathrm{tr}^5(\boldsymbol{A}_2)} \\ & + \frac{54\mathrm{tr}^2(\boldsymbol{A}_4)\boldsymbol{a'A_1a\varepsilon'aa'\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^2\mathrm{tr}^5(\boldsymbol{A}_2)} + \frac{6\mathrm{tr}(\boldsymbol{A}_6)\boldsymbol{\varepsilon'aa'A_1^*\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6\boldsymbol{a'a}\mathrm{tr}^4(\boldsymbol{A}_2)} - \frac{6\mathrm{tr}(\boldsymbol{A}_6)\boldsymbol{a'A_1a\varepsilon'aa'\varepsilon}(\boldsymbol{\varepsilon'A_1\varepsilon})^2}{\sigma^6(\boldsymbol{a'a})^2\mathrm{tr}^4(\boldsymbol{A}_2)} \\ & + \frac{54\mathrm{tr}^2(\boldsymbol{A}_4)(\boldsymbol{\varepsilon'A_1\varepsilon})^3}{\sigma^6\mathrm{tr}^5(\boldsymbol{A}_2)} - \frac{6\mathrm{tr}(\boldsymbol{A}_6)(\boldsymbol{\varepsilon'A_1\varepsilon})^3}{\sigma^6\mathrm{tr}^4(\boldsymbol{A}_2)} - \frac{2\boldsymbol{\varepsilon'A_1^*aa'A_1^*\varepsilon\varepsilon'A_1\varepsilon\varepsilon'A_2\varepsilon}}{\sigma^6\boldsymbol{a'a}\mathrm{tr}^3(\boldsymbol{A}_2)} - \frac{4\boldsymbol{\varepsilon'aa'A_3\varepsilon\varepsilon'A_1\varepsilon\varepsilon'A_2\varepsilon}}{\sigma^6\boldsymbol{a'a}\mathrm{tr}^3(\boldsymbol{A}_2)} \\ & + \frac{8\boldsymbol{a'A_1a\varepsilon'aa'A_1^*\varepsilon\varepsilon'A_1\varepsilon\varepsilon'A_2\varepsilon}}{\sigma^6(\boldsymbol{a'a})^2\mathrm{tr}^3(\boldsymbol{A}_2)} - \frac{8(\boldsymbol{a'A_1a})^2\boldsymbol{\varepsilon'aa'\varepsilon\varepsilon'A_1\varepsilon\varepsilon'A_2\varepsilon}}{\sigma^6(\boldsymbol{a'a})^3\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{2\boldsymbol{a'A_2a\varepsilon'aa'\varepsilon\varepsilon'A_1\varepsilon\varepsilon'A_2\varepsilon}}{\sigma^6(\boldsymbol{a'a})^2\mathrm{tr}^3(\boldsymbol{A}_2)}\end{aligned}$$

$$
\begin{aligned}
&+\frac{18\mathrm{tr}(\boldsymbol{A}_4)\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon}}{\sigma^6\boldsymbol{a}'\boldsymbol{a}\mathrm{tr}^4(\boldsymbol{A}_2)} - \frac{18\mathrm{tr}(\boldsymbol{A}_4)\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon}}{(\boldsymbol{a}'\boldsymbol{a})^2\sigma^6\mathrm{tr}^4(\boldsymbol{A}_2)} - \frac{27\mathrm{tr}(\boldsymbol{A}_4)(\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon})^2\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon}}{\sigma^6\mathrm{tr}^4(\boldsymbol{A}_2)} \\
&-\frac{\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}(\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon})^2}{\sigma^6\boldsymbol{a}'\boldsymbol{a}\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}(\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon})^2}{\sigma^6(\boldsymbol{a}'\boldsymbol{a})^2\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{3\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}(\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon})^2}{\sigma^6\mathrm{tr}^3(\boldsymbol{A}_2)} - \frac{6\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_4\boldsymbol{\varepsilon}}{\sigma^6\boldsymbol{a}'\boldsymbol{a}\mathrm{tr}^3(\boldsymbol{A}_2)} \\
&+\frac{6\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_4\boldsymbol{\varepsilon}}{\sigma^6(\boldsymbol{a}'\boldsymbol{a})^2\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{9(\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon})^2\boldsymbol{\varepsilon}'\boldsymbol{A}_4\boldsymbol{\varepsilon}}{\sigma^6\mathrm{tr}^3(\boldsymbol{A}_2)} - \frac{\boldsymbol{\varepsilon}'\boldsymbol{A}_1^*\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^2\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{2\boldsymbol{\varepsilon}'\boldsymbol{A}_1^*\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_3\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^2\mathrm{tr}^2(\boldsymbol{A}_2)} \\
&+\frac{5\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}(\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon})^2}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^3\mathrm{tr}^2(\boldsymbol{A}_2)} + \frac{2\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_3\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^3\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{8(\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a})^2\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^4\mathrm{tr}^2(\boldsymbol{A}_2)} + \frac{\boldsymbol{a}'\boldsymbol{A}_2\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^3\mathrm{tr}^2(\boldsymbol{A}_2)} \\
&+\frac{4(\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a})^3(\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon})^2}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^5\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_2\boldsymbol{a}(\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon})^2}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^4\mathrm{tr}^2(\boldsymbol{A}_2)} + \frac{3\mathrm{tr}(\boldsymbol{A}_4)(\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon})^2}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^2\mathrm{tr}^3(\boldsymbol{A}_2)} - \frac{6\mathrm{tr}(\boldsymbol{A}_4)\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^3\mathrm{tr}^3(\boldsymbol{A}_2)} \\
&+\frac{3\mathrm{tr}(\boldsymbol{A}_4)(\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a})^2(\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon})^2}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^4\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{2\boldsymbol{\varepsilon}'\boldsymbol{A}_1^*\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^4\boldsymbol{a}'\boldsymbol{a}\mathrm{tr}^2(\boldsymbol{A}_2)} + \frac{4\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_3\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^4\boldsymbol{a}'\boldsymbol{a}\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{8\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^2\mathrm{tr}^2(\boldsymbol{A}_2)} \\
&+\frac{8(\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a})^2\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^3\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{2\boldsymbol{a}'\boldsymbol{A}_2\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^2\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{12\mathrm{tr}(\boldsymbol{A}_4)\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^4\boldsymbol{a}'\boldsymbol{a}\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{12\mathrm{tr}(\boldsymbol{A}_4)\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^2\mathrm{tr}^3(\boldsymbol{A}_2)} \\
&+\frac{9\mathrm{tr}(\boldsymbol{A}_4)(\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon})^2}{\sigma^4\mathrm{tr}^3(\boldsymbol{A}_2)} + \frac{2\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon}}{\sigma^4\boldsymbol{a}'\boldsymbol{a}\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{2\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon}}{\sigma^4(\boldsymbol{a}'\boldsymbol{a})^2\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{3\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}_2\boldsymbol{\varepsilon}}{\sigma^4\mathrm{tr}^2(\boldsymbol{A}_2)} - \frac{\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{A}_1^*\boldsymbol{\varepsilon}}{\sigma^2\boldsymbol{a}'\boldsymbol{a}\mathrm{tr}(\boldsymbol{A}_2)} \\
&+\frac{\boldsymbol{a}'\boldsymbol{A}_1\boldsymbol{a}\boldsymbol{\varepsilon}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\varepsilon}}{\sigma^2(\boldsymbol{a}'\boldsymbol{a})^2\mathrm{tr}(\boldsymbol{A}_2)} + \frac{\boldsymbol{\varepsilon}'\boldsymbol{A}_1\boldsymbol{\varepsilon}}{\sigma^2\mathrm{tr}(\boldsymbol{A}_2)}.
\end{aligned}
$$