Purdue University Economics Department Working Paper No **1318**

Generalized Forecast Averaging in Autoregressions with a Near Unit Root

Mohitosh Kejriwal*

Purdue University

Xuewen Yu†

Purdue University

November 12, 2019

Abstract

This paper develops a new approach to forecasting a highly persistent time series that employs feasible generalized least squares (FGLS) estimation of the deterministic components in conjunction with Mallows model averaging. Within a local-to-unity asymptotic framework, we derive analytical expressions for the asymptotic mean squared error and one-step ahead mean squared forecast risk of the proposed estimator and show that the optimal FGLS weights are different from their ordinary least squares (OLS) counterparts. We also provide theoretical justification for a generalized Mallows averaging estimator that incorporates lag order uncertainty in the construction of the forecast. Monte Carlo simulations demonstrate that the proposed procedure yields considerably lower finite sample forecast risk relative to OLS averaging, with the improvements being particularly pronounced when the model includes a deterministic trend. An application to US macroeconomic time series illustrates the efficacy of the advocated method in practice and finds that both persistence and lag order uncertainty have important implications for the accuracy of forecasts.

Keywords: Model averaging, local to unity, generalized least squares, forecast combination

JEL Classification: C22

*Krannert School of Management, Purdue University, 403 West State Street, West Lafayette IN 47907 (mkejriwa@purdue.edu).

†Krannert School of Management, Purdue University, 403 West State Street, West Lafayette IN 47907 (yu656@purdue.edu).

1 Introduction

Over the past few decades, a variety of methods has been developed in both the statistics and econometrics literatures for estimation and inference with highly persistent time series. Following Chan and Wei (1987) and Phillips (1987), a highly persistent time series is typically modeled as one with an autoregressive root local to unity ($\alpha = 1 + c/T$), thereby permitting analysis of the stationary ($|\alpha| < 1$) and the nonstationary ($\alpha = 1$) cases within a unified asymptotic framework. Local-to-unity limit theory has been fruitfully employed to develop efficient unit root tests (e.g., Elliott et al., 1996), uniformly valid confidence intervals in autoregressive models (e.g., Hansen, 1999; Mikusheva, 2007) and robust inferential methods in predictive regressions (e.g., Phillips, 2014; 2015). The primary technical difficulty in this modeling framework arises from the fact that the noncentrality parameter c cannot be consistently estimated.

While a substantial body of work has addressed issues related to estimation and inference, the problem of forecasting a highly persistent time series has received relatively less attention. The essence of the forecasting problem lies in the bias-variance tradeoff whereby imposing a unit root reduces estimation uncertainty at the expense of potential model misspecification while unrestricted estimation can lead to high forecast risk due to variance inflation. Franses and Kleibergen (1996) apply the restricted and unrestricted models to the Nelson-Plosser dataset and argue that the restricted model is preferred in a variety of sample sizes and forecast horizons. Diebold and Kilian (2000) suggest that unit root pretesting improves forecast accuracy relative to restricted or unrestricted estimation. Kim (2001, 2003) and Clements and Kim (2007) investigate the impact of various bias correction methods on point forecasts and prediction intervals for univariate autoregressive models and find that bias correction delivers considerable gains in forecast accuracy for unit root or near-unit root autoregressive models.

Forecast combination, pioneered by the work of Bates and Granger (1969) and Granger and Ramanathan (1984), provides a useful, practical approach to constructing forecasts that can effectively capture the bias-variance tradeoff inherent in the individual forecasts. In the present context of forecasting a highly persistent time series, Hansen (2010) suggests combining forecasts from the restricted and unrestricted models with weights determined by optimizing a Mallows criterion, designed to provide an approximately unbiased estimator of the in-sample asymptotic mean squared error. Hansen's (2010) results strongly caution against using the pretesting method, which exhibits high risk over a range of persistence levels (values of c), while simulations show his Mallows model averaging forecast performs well relative to a number of commonly employed methods and dominates the unrestricted forecast uniformly in terms of finite sample forecast risk.

In the standard stationary framework, the classic result of Grenander and Rosenblatt (1957) shows that generalized least squares (GLS) and ordinary least squares (OLS) estimation of the deterministic components are asymptotically equivalent so that no efficiency gains are available from employing the former, at least in large samples. In a local-to-unity setup, however, the situation is different. Phillips and Lee (1996) and Canjels and Watson (1997) document the reduction in asymptotic variance afforded by GLS estimation while its implications for forecasting are explored in Stock (1996) and Ng and Vogelsang (2002).

Motivated by these findings, this paper develops a new approach to forecasting a highly persistent time series that employs feasible generalized least squares (FGLS) estimation of the deterministic components in conjunction with Mallows model averaging.[1] Within a local-to-unity asymptotic framework, we derive analytical expressions for the in-sample asymptotic mean squared error (AMSE) and one-step ahead mean squared forecast risk (MSFE) of the proposed estimator and show that the optimal FGLS weights are different from their OLS counterparts. We also provide theoretical justification for a generalized Mallows averaging estimator that incorporates lag order uncertainty in the construction of the forecast. Specifically, the generalized Mallows criterion follows from an asymptotic framework where the coefficients of the lagged differences are modeled as local to zero simultaneously with the largest autoregressive root being modeled as local to unity. Monte Carlo simulations illustrate that the proposed procedure yields considerably lower finite sample forecast risk relative to OLS averaging, with the improvements being particularly pronounced when the model includes a deterministic trend. Finally, a comparative out-of-sample forecasting exercise applied to US macroeconomic time series demonstrates the potential of the advocated method and finds that both persistence and lag order uncertainty have important implications for the accuracy of forecasts.

The remainder of the paper is organized as follows. Section 2 presents the model setup and FGLS estimation procedures. Section 3 introduces our FGLS Mallows model averaging estimator. Section 4 discusses general Mallow averaging strategies with both OLS and FGLS estimation. Monte Carlo simulation results and comparisons are provided in section 5. Section 6 presents the empirical application and section 7 concludes. Proofs of the theorems

[1] In related work, Liu et al. (2016) propose model averaging based on feasible GLS to account for the presence of heteroskedastic errors in a standard stationary regression framework.

are provided in Appendix A.

2 Model and Estimation

We consider an observed time series composed of deterministic and stochastic components as in Hansen (2010):

$$\begin{aligned} y_t &= m_t + u_t \\ m_t &= \beta_0 + \beta_1 t + ... + \beta_p t^p \\ u_t &= \alpha u_{t-1} + \alpha_1 \Delta u_{t-1} + \cdots + \alpha_k \Delta u_{t-k} + e_t \\ \alpha &= 1 + \frac{ac}{T}, \quad a = 1 - \alpha_1 - \cdots - \alpha_k, \quad c \leq 0 \end{aligned} \tag{1}$$

where $p \in \{0, 1\}$ is the order of the trend component and the stochastic component u_t follows an autoregressive process of order $(k+1)$ process driven by the innovations e_t. The persistence parameter α is modeled as local to unity with $c = 0$ corresponding to the unit root case and $c < 0$ to the stationary case. The true lag order k is assumed known in this section. Lag order uncertainty will be addressed in section 4. The initial observations are set at $u_0, u_{-1}, \cdots, u_{-k} = O_p(1)$[2]. Our analysis is based on the following assumptions:

- **Assumption A1** The sequence $\{e_t\}$ is a martingale difference sequence with $E(e_t|e_{t-1}, e_{t-2,...}) = 0$ and $E(e_t^2|e_{t-1}, e_{t-2,...}) = \sigma^2$.
- **Assumption A2** All roots of $A(L) = 1 - \sum_{i=1}^{k} \alpha_i L^i$ lie outside the unit circle.

Assumptions A1 and A2 are standard and made in Hansen (2010) thereby allowing comparison with his analysis. We denote the optimal (infeasible) mean squared error minimizing one-step ahead forecast as $y_{t+1|t}$. It is the conditional mean μ_{t+1} given the true parameter values, namely,

$$\mu_{t+1} = m_{t+1} + \alpha(y_t - m_t) + \alpha_1(\Delta y_t - \Delta m_t) + \cdots + \alpha_k(\Delta y_{t-k+1} - \Delta m_{t-k+1}) \tag{2}$$

While μ_{t+1} is unique, its feasible counterpart is not. Estimation of the conditional mean is associated with two important sources of uncertainty. The first emanates from uncertainty regarding the nature of persistence given that the parameter c is unknown and cannot be consistently estimated. Unrestricted estimation (i.e., simple OLS) avoids omitted variable

[2]The conclusion for the subsequent analysis will not be affected as long as the initial observations are $o_p(T^{1/2})$.

bias while restricted ($c = 0$) regression offers the possibility to achieve variance reduction. The local-to-unity parameterization ensures that squared model biases and estimator variances have the same order of magnitude. In order to optimize the bias-variance tradeoff, Hansen (2010) proposes averaging the unrestricted and restricted estimators with weights determined according to the Mallows criterion, which is designed to provide an approximately unbiased estimate of the in-sample AMSE. He derives analytical expressions for the AMSE and MSFE of unrestricted, restricted, pretest and the Mallows model averaging (MMA) estimators and finds they are functions only of c, which facilitates graphical comparisons and provides the evolving patterns of the forecast risk of alternative methods with respect to c. His theoretical and numerical results support the use of the MMA estimator relative to its competitors.

A second source of uncertainty results from estimating the deterministic component with highly persistent errors. Grenander and Rosenblatt (1957) show that OLS and GLS estimates of the trend component are asymptotically equivalent in the standard stationary framework ($|\alpha| < 1, \alpha$ fixed). In the local-to-unity framework, however, Phillips and Lee (1996) and Canjels and Watson (1997) establish that GLS can be asymptotically more efficient than OLS with respect to estimation of the trend parameters while Ng and Vogelsang (2002) provide analytical and simulation evidence comparing OLS with two different FGLS estimators, namely those based on the Cochrane-Orcutt (CO) and Prais-Winsten (PW) transformations, and find that FGLS based on the latter transformation generally dominates the others in terms of forecast accuracy.

Our paper aims to integrate FGLS estimation with Mallows model averaging to investigate if further improvements in forecasting performance can be achieved in the presence of the two aforementioned sources of uncertainty. Specifically, we propose an averaging strategy combining unrestricted and restricted FGLS estimators, whose weights are determined by a Mallows criterion. In what follows, the unrestricted and restricted FGLS estimates of μ_t are denoted by $\hat{\mu}_t$ and $\tilde{\mu}_t$, respectively. We first state how unrestricted FGLS estimation works for model (1). For brevity, we enumerate the steps only for $p = 1$, with obvious modifications in place for $p = 0$.

1. Estimate by OLS the regression

$$y_t = z_t'\beta^* + \alpha y_{t-1} + \alpha_1 \Delta y_{t-1} + \cdots + \alpha_k \Delta y_{t-k} + \epsilon_t, \qquad t = k+2, k+3, ..., T \qquad (3)$$

where $z_t = (1, t)'$, $\beta^* = (\beta_0^*, \beta_1^*)'$. Denote the estimate of α by $\dot{\alpha}$.

2. Consider the Prais-Winsten (PW) transformation to quasi-difference y_t and z_t: for $t = 2, 3, ..., T$, $y_t^+ = y_t - \dot{\alpha} y_{t-1}$, $z_t^+ = z_t - \dot{\alpha} z_{t-1}$, with $y_1^+ = y_1$ and $z_1^+ = z_1$.

3. Regress quasi-differenced data y^+ on z^+ to get trend estimates: $\ddot{\beta} = (z^{+\prime} z^+)^{-1}(z^{+\prime} y^+)$.

4. Construct detrended data $\hat{u}_t = y_t - z_t'\ddot{\beta}$, regress detrended data on its lags:

$$\hat{u}_t = \alpha \hat{u}_{t-1} + \alpha_1 \Delta \hat{u}_{t-1} + \cdots + \alpha_k \Delta \hat{u}_{t-k} + \xi_t, \qquad t = k+2, ..., T \tag{4}$$

   Obtain autoregressive parameter estimates $\ddot{\alpha}, \ddot{\alpha}_1, \cdots, \ddot{\alpha}_k$.

5. Construct the feasible one-step ahead forecast $\hat{y}_{T+1|T} = \hat{\mu}_{T+1} = z_{T+1}'\ddot{\beta} + \ddot{\alpha}(y_T - z_T'\ddot{\beta}) + \ddot{\alpha}_1(\Delta y_T - \Delta z_T'\ddot{\beta}) + \cdots + \ddot{\alpha}_k(\Delta y_{T-k+1} - \Delta z_{T-k+1}'\ddot{\beta})$.

To obtain the restricted FGLS estimate $\tilde{\mu}_t$, the procedure is same as that outlined above except that we impose $\alpha = 1$ in each step.[3] For $p = 0$, the restricted FGLS estimate is identical to the restricted OLS estimate in Hansen (2010) while the two are asymptotically equivalent for $p = 1$. The difference in finite samples for the latter case arises from the difference between one-step estimation (Hansen, 2010) and two-step estimation (detrend first and then estimate the lag parameters separately). The large sample analysis for the restricted FGLS estimate thus directly follows from Hansen (2010) and is not repeated here to save space.

To evaluate the quality of the unrestricted FGLS estimator, we derive expressions for the in-sample AMSE and one-step ahead MSFE as in Hansen (2010). To this end, define $\ddot{c} = \lim_{T\to\infty} T(\ddot{\alpha} - 1)/a$ with $\ddot{c}^0$ and $\ddot{c}^1$ denoting the limits in the $p = 0$ and $p = 1$ cases, respectively. Next, define the stochastic process

$$U_p(c, a, r) = \begin{cases} (\ddot{c}^0 - c)J_c(r) & for \quad p = 0 \\ \gamma_1(1 - cr) + (\ddot{c}^1 - c)P(r) & for \quad p = 1 \end{cases}$$

where $\gamma_1[P(.)]$ is a random variable [stochastic process] depending on a and c. Explicit expressions for these quantities are provided in Appendix A. We then have the following result:

Theorem 1 *Under Assumptions A1-A2,*
[a] (AMSE) $m_1(c, a, p, k) = \lim_{T\to\infty} \frac{1}{\sigma^2} \sum_{t=1}^{T} E(\hat{\mu}_t - \mu_t)^2 = E\left[\int_0^1 U_p(c, a, r)^2 dr\right] + k \equiv$

[3] The procedures were also implemented using the Roy and Fuller (2001) bias correction. The results were found to be qualitatively similar and hence not reported. They are available upon request.

$m_1(c,a,p)+k$
[b] $\lim_{c\to-\infty} m_1(c,a,p,k) = 1+p+k$
[c] (MSFE) $f_1(c,a,p,k) = \lim_{T\to\infty} \frac{T}{\sigma^2} E(\hat{\mu}_{T+1} - \mu_{T+1})^2 = E\left[U_p(c,a,1)^2\right] + k$

Remark 1 *Hansen (2010, Theorem 1) shows that the in-sample AMSE of the unconstrained OLS estimate is* $2+p+k$*, while our result shows that in the FGLS case, the in-sample AMSE is only* $1+p+k$*. This reduction reflects the fact that FGLS effectively eliminates the uncertainty about the unknown mean. This result thus directly quantifies the improvement from FGLS forecasting in local-to-unity models. Moreover, Theorem 1 extends Ng and Vogelsang's (2002) asymptotic analysis to models with more than one autoregressive lag.*

Remark 2 *The random process* $U_p(c,a,.)$ *not only depends on* c *but also on* a *(for* $p=1$*). This is different from the OLS case, where the in-sample AMSE of the deterministic component and the AR(1) component are independent of the short-run dynamics. Thus* $m_1(c,a,p)$ *depends on* a *for fixed* c *but becomes independent of* a *as* $c\to-\infty$*.*

With the restricted and unrestricted FGLS estimators in place, the GLS averaging estimator for a given weight vector $[w, 1-w],\ w\in[0,1]$ is defined as

$$\hat{\mu}_t(w) = w\hat{\mu}_t + (1-w)\tilde{\mu}_t$$

Define the stochastic process $V_p(c,.)$ as

$$V_p(c,r) = \begin{cases} -cJ_c(r) & for \quad p=0 \\ -c\bar{J}_c(r) + W(1) & for \quad p=1 \end{cases}$$

with the associated quantities $m_0(c,p) = E\left[\int_0^1 V_p(c,r)^2 dr\right]$, $m_{01}(c,a,p) = E\left[\int_0^1 U_p(c,a,r)V_p(c,r)dr\right]$, $f_1(c,a,p) = E[U_p(c,a,1)^2]$, $f_0(c,p) = E[V_p(c,1)^2]$, $f_{01}(c,a,p) = E[U_p(c,a,1)V_p(c,1)]$.

The in-sample AMSE and MSFE of the averaging estimator are given in the following corollary:

Corollary 1 *[a]* $m_w(c,a,p,k) = \lim_{T\to\infty} \frac{1}{\sigma^2}\sum_{t=1}^T E[\hat{\mu}_t(w) - \mu_t]^2 = w^2 m_1(c,a,p) + (1-w)^2 m_0(c,p) + 2w(1-w)m_{01}(c,a,p) + k$
[b] $f_w(c,a,p,k) = \lim_{T\to\infty} \frac{T}{\sigma^2} E(\hat{\mu}_{T+1}(w) - \mu_{T+1})^2 = w^2 f_1(c,a,p) + (1-w)^2 f_0(c,p) + 2w(1-w)f_{01}(c,a,p) + k$
[c] $\lim_{c\to 0} m_{01}(c,a,p) = p$

As an alternative strategy, one can perform a pretest to choose between the restricted and unrestricted forecasts. Stock (1996) and Diebold and Kilian (2000) show that pretesting is useful for selection of forecasting models while Hansen's (2010) analysis cautions against pretesting due to high finite sample forecast risk for an intermediate range of the parameter (c) space. In the GLS framework, we adopt the Dickey-Fuller GLS (DF^{GLS}) t-test proposed by Elliott et al. (1996) with the lag length selected using the modified Akaike Information Criterion ($MAIC$) proposed by Ng and Perron (2001). We denote the pretest estimator $\hat{\mu}_t^{pt} = \hat{\mu}_t 1(DF^{GLS} \leq cv_p) + \tilde{\mu}_t 1(DF^{GLS} > cv_p)$. The critical values cv_p for $p = 0, 1$ are -1.98 and -2.91, respectively. Elliott et al. (1996) show that

$$DF^{GLS} \to \begin{cases} DF_0^{GLS} = \frac{1}{2}(J_c(1)^2 - 1)/(\int_0^1 J_c(r)^2 dr)^{1/2} & if \quad p = 0 \\ DF_1^{GLS} = \frac{1}{2}(V_c(1,\bar{c})^2 - 1)/(\int_0^1 V_c(r,\bar{c})^2 dr)^{1/2} & if \quad p = 1 \end{cases}$$

where

$$V_c(r,\bar{c}) = J_c(r) - r[\lambda J_c(1) + 3(1-\lambda)\int_0^1 sJ_c(s)ds]$$
$$\lambda = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3), \ \ \bar{c} = -7[1(p=0)] - 13.5[1(p=1)]$$

The in-sample AMSE and one-step MSFE of the DF^{GLS} pretest estimator is summarized in the following corollary:

Corollary 2 *[a]* $m_{pt}(c,a,p,k) = \lim_{T\to\infty} \frac{1}{\sigma^2}\sum_{t=1}^{T} E(\hat{\mu}_t^{pt} - \mu_t)^2 = E\left[\int_0^1 U_p(c,a,r)^2 dr I(DF_p^{GLS} \leq cv_p)\right]$
$+E\left[\int_0^1 V_p(c,r)^2 dr I(DF_p^{GLS} > cv_p)\right] + k$
[b] $f_{pt}(c,a,p,k) = \lim_{T\to\infty} \frac{T}{\sigma^2} E(\hat{\mu}_{T+1}^{pt} - \mu_{T+1})^2 = E\left[U_p(c,a,1)^2 I(DF_p^{GLS} \leq cv_p)\right] +$
$E\left[V_p(c,1)^2 I(DF_p^{GLS} > cv_p)\right] + k$

Figures 1 and 2 present the in-sample AMSE and MSFE of various OLS/GLS estimators for $p = 0$ and $p = 1$, respectively.[4] These include the FGLS pretest estimator (Pretest-GLS), unrestricted FGLS estimator (Unres-GLS), FGLS Mallows averaging estimator (GLS-Ave) and GLS optimal (infeasible) averaging estimator (GLS-Ave-Opt), as well as their OLS counterparts with corresponding labeling. It is clear that for each type of estimator (unrestricted, pretest, averaging), FGLS performs better than its OLS counterpart for $p = 0, 1$ in terms of both in-sample AMSE and MSFE, except for the pretest estimator at values of

[4] These figures are computed on a grid of 101 evenly-spaced points from -20 to 0 for an AR(1) model ($a = 1$). We approximate the limiting distributions by simulating the random variables/processes using $T = 1000$. The number of replications is 500,000.

c close to 0, where OLS and FGLS are comparable to each other. The relative performance among different FGLS estimators is similar to that of the OLS estimators as analyzed in Hansen (2010). Further, while pretesting continues to incur high risk even when employing the more efficient unit root test, FGLS averaging leads to uniformly lower risk compared to OLS averaging. Finally, the ranking of the estimators is invariant to whether evaluation is according to AMSE or MSFE.

3 FGLS Mallows Averaging

The Mallows (1973) criterion was originally designed as an information criterion for the purpose of model selection which provides an unbiased estimate of the in-sample AMSE. The seminal work of Hansen (2007, 2008) has spawned a vast literature that employs Mallows model averaging for estimation and forecasting. The Mallows criteria for the unrestricted and restricted models based on the FGLS estimates are as follows:

$$M_0(c, a, p, k) = T\tilde{\sigma}^2 + 2\hat{\sigma}^2(m_{01}(c, a, p) + k) \tag{5}$$

$$M_1(c, a, p, k) = T\hat{\sigma}^2 + 2\hat{\sigma}^2(m_1(c, a, p) + k) \tag{6}$$

where $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ are, respectively, the estimates of σ^2 from the unrestricted and restricted models, i.e., $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(y_t - \hat{\mu}_t)^2$, $\tilde{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(y_t - \tilde{\mu}_t)^2$.

As in Hansen (2010), the dependence of M_0 and M_1 on the unknown parameter c make them infeasible in practice. Moreover, unlike OLS, the expressions in the FGLS case are now complicated by dependence on the short-run dynamics through the parameter a. We suggest obtaining feasible rules by taking limits of these expressions. In particular, we have

$$M_0 = T\tilde{\sigma}^2 + 2\hat{\sigma}^2(\lim_{c\to 0} m_{01}(c, a, p) + k)$$
$$M_1 = T\hat{\sigma}^2 + 2\hat{\sigma}^2(\lim_{c\to -\infty} m_1(c, a, p) + k).$$

Fortunately, from Theorem 1 and Corollary 1, we can obtain specific values for these criteria that are independent of a. Specifically, using $\lim_{c\to 0} m_{01}(c, a, p) = p$ and $\lim_{c\to -\infty} m_1(c, a, p) = 1 + p$, we have

$$M_0 = T\tilde{\sigma}^2 + 2\hat{\sigma}^2(p + k) \tag{7}$$

$$M_1 = T\hat{\sigma}^2 + 2\hat{\sigma}^2(1 + p + k) \tag{8}$$

A Mallows selection estimator is then easily obtained as the rule of picking the unrestricted model when $F_T = T(\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2}) \geq 2$. The following result shows that the criteria $M_0(c, a, p, k)$ and $M_1(c, a, p, k)$ are asymptotically unbiased estimates of the AMSE after normalization and evaluating the quantities at the limits of c.

Theorem 2 *Under Assumptions A1-A2,*

$$\lim_{c \to 0} \lim_{T \to \infty} \frac{EM_0(c, a, p, k)}{\sigma^2} - T = \lim_{c \to 0} m_0(c, p) + k$$
$$\lim_{c \to -\infty} \lim_{T \to \infty} \frac{EM_1(c, a, p, k)}{\sigma^2} - T = \lim_{c \to -\infty} m_1(c, a, p) + k$$

For a given weight vector $[w, 1-w]$, we construct the Mallows criterion for the averaging estimator as

$$M_w(c) = T\hat{\sigma}^2(w) + 2\hat{\sigma}^2[w\{m_1(c, a, p) + k\} + (1-w)\{m_{01}(c, a, p) + k\}]$$

, where $\hat{\sigma}^2(w) = T^{-1} \sum_{t=1}^{T} [y_t - \hat{\mu}_t(w)]^2$. The feasible version of this criterion, using the previous results, is

$$M_w = T\hat{\sigma}^2(w) + 2\hat{\sigma}^2(w + p + k) \tag{9}$$

The Mallows selected weight $\hat{w}$ is derived from minimizing (9) over $w \in [0, 1]$. The solution is

$$\hat{w} = \begin{cases} 1 - 1/F_T & if \quad F_T > 1 \\ 0 & otherwise \end{cases}$$

The Mallows averaging estimator is then defined as

$$\hat{\mu}_t^a = \hat{w}\hat{\mu}_t + (1-\hat{w})\tilde{\mu}_t = \begin{cases} \tilde{\mu}_t & if \quad F_T \leq 1 \\ (1 - \frac{1}{F_T})\hat{\mu}_t + \frac{1}{F_T}\tilde{\mu}_t & otherwise \end{cases} \tag{10}$$

4 General Mallows Averaging [GMA]

The foregoing analysis assumes that the true lag order k is known. In practice, lag order uncertainty needs to be addressed since omitting relevant lags will contribute to misspecification bias while including too many lags would lead to variance inflation. The traditional approach has been to employ model selection rules such as standard information criteria to choose the number of lags. Hansen (2010) proposes an alternative approach that averages over different lag orders in addition to averaging over the unit root restriction. In section 4.1, we first provide theoretical justification for Hansen's general Mallows averaging (GMA) criterion that incorporates both lag order uncertainty and persistence uncertainty. The analysis is subsequently extended to the FGLS setting in section 4.2.

4.1 GMA for OLS

To obtain Hansen's (2010) GMA criterion, we adopt a local asymptotic framework which models the coefficients of the short-run dynamics in a $O(T^{-1/2})$-neighborhood around zero in addition to the $O(T^{-1})$ local-to-unity parameterization for the persistence parameter α, i.e., $\alpha_i = \frac{\delta_i}{\sqrt{T}}$ for $i = 1, ..., k$ where $\delta = (\delta_1, ..., \delta_k)'$ is fixed and independent of T. This particular rate ensures that the squared bias from omitting relevant lags is of the same order as the variance from estimating additional lags. In contrast, a fixed specification for the lagged coefficients would imply that the misspecification bias diverges to infinity with the sample size. The use of local asymptotic analysis in the frequentist model averaging literature was pioneered by Hjort and Claeskens (2003).

We consider restricted regression (setting $c = 0$) and unrestricted regression, each with l lags. We include sub-models with $l \in \{0, 1, ..., K\}$, $K \geq k$, with the corresponding unrestricted and restricted estimates denoted by $\breve{\mu}_t(l)$ and $\tilde{\mu}_t(l)$, respectively.[5] This gives a total of $2(K+1)$ sub-models. We first analyze the unrestricted regression with l lags:

$$y_t = z_t'\beta^* + \alpha y_{t-1} + \alpha_1 \Delta y_{t-1} + \cdots + \alpha_l \Delta y_{t-l} + \epsilon_t, \qquad t = l+2, ..., T \tag{11}$$

The feasible forecast is $\breve{\mu}_t(l) = z_t'\breve{\beta}^* + \breve{\alpha} y_{t-1} + \breve{\alpha}_1 \Delta y_{t-1} + \cdots + \breve{\alpha}_l \Delta y_{t-l}$. Define the quantities

$$m_{0K}^{ols}(c, \delta, p, l) = \lim_{T\to\infty} \frac{1}{\sigma^2} \sum_{t=1}^{T} E(\tilde{\mu}_t(l) - \mu_t)(\breve{\mu}_t(K) - \mu_t)$$
$$m_{1K}^{ols}(c, \delta, p, l) = \lim_{T\to\infty} \frac{1}{\sigma^2} \sum_{t=1}^{T} E(\breve{\mu}_t(l) - \mu_t)(\breve{\mu}_t(K) - \mu_t)$$

The Mallows criteria for restricted and unrestricted OLS estimators are constructed as

$$M_0^{ols}(c, \delta, p, l) = T\tilde{\sigma}_l^2 + 2\breve{\sigma}_K^2 m_{0K}^{ols}(c, \delta, p, l)$$
$$M_1^{ols}(c, \delta, p, l) = T\breve{\sigma}_l^2 + 2\breve{\sigma}_K^2 m_{1K}^{ols}(c, \delta, p, l)$$

[5] We use the notation $\tilde{\mu}_t$ to denote both restricted OLS and restricted FGLS estimation, although it must be borne in mind that they are equivalent in finite samples only for $p = 0$, while the equivalence holds asymptotically for $p = 1$.

where $\breve{\sigma}_j^2 = T^{-1}\sum_{t=1}^{T}(y_t - \breve{\mu}_t(j))^2$, $j = l, K$ and $\tilde{\sigma}_l^2 = T^{-1}\sum_{t=1}^{T}(y_t - \tilde{\mu}_t(l))^2$. The following theorem establishes that the criteria M_0^{ols} and M_1^{ols} are asymptotically unbiased estimates of the AMSE after normalization:

Theorem 3 *Let* $m_0^{ols}(c,\delta,p,l) = \lim_{T\to\infty}\frac{1}{\sigma^2}\sum_{t=1}^{T}E(\tilde{\mu}_t(l)-\mu_t)^2$, $m_1^{ols}(c,\delta,p,l) = \lim_{T\to\infty}\frac{1}{\sigma^2}\sum_{t=1}^{T}E(\breve{\mu}_t(l)-\mu_t)^2$. *Then we have, under Assumptions A1-A2,*

$$\frac{EM_0^{ols}(c,\delta,p,l)}{\sigma^2} - T \to m_0^{ols}(c,\delta,p,l)$$
$$\frac{EM_1^{ols}(c,\delta,p,l)}{\sigma^2} - T \to m_1^{ols}(c,\delta,p,l)$$

As shown in the proof of Theorem 3, the quantities m_{0K}^{ols} and m_{1K}^{ols} are infeasible as they depend on c. To obtain their feasible versions, we evaluate them at the limits of c which gives us the following result:

Theorem 4 *Under Assumptions A1-A2,*

$$\lim_{c\to 0} m_{0K}^{ols}(c,\delta,p,l) = p + l$$
$$\lim_{c\to -\infty} m_{1K}^{ols}(c,\delta,p,l) = 2 + p + l$$

The feasible Mallows criteria are then obtained as

$$M_0^{ols}(p,l) = T\tilde{\sigma}_l^2 + 2\breve{\sigma}_K^2(p+l)$$
$$M_1^{ols}(p,l) = T\breve{\sigma}_l^2 + 2\breve{\sigma}_K^2(2+p+l)$$

Now, the averaging estimator over all $2(K+1)$ sub-models can be constructed as

$$\breve{\mu}_t^a(w) = \sum_{l=0}^{K}(w_{0l}\breve{\mu}_t(l) + w_{1l}\tilde{\mu}_t(l)) \tag{12}$$

where the weights are non-negative and sum to one: $w_{1l} \geq 0, w_{0l} \geq 0, \sum_{l=0}^{K}(w_{0l} + w_{1l}) = 1$. Hence the feasible Mallows averaging criterion is obtained as

$$M_w^{ols}(p,K) = T\breve{\sigma}^2(w) + 2\breve{\sigma}_K^2(\sum_{l=0}^{K}[w_{0l}l + w_{1l}(2+l)] + p)$$

where $\breve{\sigma}^2(w) = T^{-1}\sum_{t=1}^{T}(y_t - \breve{\mu}_t^a(w))^2$.

4.2 GMA for FGLS

We now develop the asymptotics of GMA for FGLS. Let the unrestricted FGLS estimate from the sub-model with l lags be denoted $\hat{\mu}_t(l),\ l \in \{0, 1, ..., K\}$. Our goal is to combine the estimates $\tilde{\mu}_t(l)$ with $\hat{\mu}_t(l)$ for each l and average over all the sub-models. The procedure for unrestricted FGLS estimation with l lags is exactly the same as that outlined in section 2. Analogous to the OLS case, define the quantities

$$m_{0K}^{gls}(c, a, \delta, p, l) = \lim_{T\to\infty} \frac{1}{\sigma^2} \sum_{t=1}^{T} E(\tilde{\mu}_t(l) - \mu_t)(\hat{\mu}_t(K) - \mu_t)$$
$$m_{1K}^{gls}(c, a, \delta, p, l) = \lim_{T\to\infty} \frac{1}{\sigma^2} \sum_{t=1}^{T} E(\hat{\mu}_t(l) - \mu_t)(\hat{\mu}_t(K) - \mu_t)$$

The Mallows criteria based on FGLS estimation are constructed as

$$M_0^{gls}(c, a, \delta, p, l) = T\tilde{\sigma}_l^2 + 2\hat{\sigma}_K^2 m_{0K}^{gls}(c, a, \delta, p, l)$$
$$M_1^{gls}(c, a, \delta, p, l) = T\hat{\sigma}_l^2 + 2\hat{\sigma}_K^2 m_{1K}^{gls}(c, a, \delta, p, l)$$

where $\hat{\sigma}_j^2 = T^{-1} \sum_{t=1}^{T} (y_t - \hat{\mu}_t(j))^2,\ j = l, K$. The asymptotic unbiasedness of M_0^{gls} and M_1^{gls} for the AMSE are established in the following result:

Theorem 5 *Let* $m_0^{gls}(c, a, \delta, p, l) = \lim_{T\to\infty} \frac{1}{\sigma^2} \sum_{t=1}^{T} E(\tilde{\mu}_t(l) - \mu_t)^2 = m_0^{ols}(c, \delta, p, l),\ m_1^{gls}(c, a, \delta, p, l) = \lim_{T\to\infty} \frac{1}{\sigma^2} \sum_{t=1}^{T} E(\hat{\mu}_t(l) - \mu_t)^2$. *Then we have, under Assumptions A1-A2,*

$$\lim_{c\to 0} \lim_{T\to\infty} \frac{EM_0^{gls}(c, a, \delta, p, l)}{\sigma^2} - T = \lim_{c\to 0} m_0^{gls}(c, a, \delta, p, l)$$
$$\lim_{c\to -\infty} \lim_{T\to\infty} \frac{EM_1^{gls}(c, a, \delta, p, l)}{\sigma^2} - T = \lim_{c\to -\infty} m_1^{gls}(c, a, \delta, p, l)$$

The feasible versions of m_{0K}^{gls} and m_{1K}^{gls} are obtained from their respective limits:

Theorem 6 *Under Assumptions A1-A2,*

$$\lim_{c\to 0} m_{0K}^{gls}(c, a, \delta, p, l) = p + l$$
$$\lim_{c\to -\infty} m_{1K}^{gls}(c, a, \delta, p, l) = 1 + p + l$$

The feasible Mallows criteria are then obtained as

$$M_0^{gls}(p,l) = T\tilde{\sigma}_l^2 + 2\hat{\sigma}_K^2(p+l)$$
$$M_1^{gls}(p,l) = T\hat{\sigma}_l^2 + 2\hat{\sigma}_K^2(1+p+l)$$

The averaging estimator $\hat{\mu}_t^a(w)$ over all $2(K+1)$ sub-models is constructed in the same way as in (12) except that $\hat{\mu}(l)$ replaces $\breve{\mu}(l)$. Hence the feasible Mallows averaging criterion is obtained as

$$M_w^{gls}(p,K) = T\hat{\sigma}^2(w) + 2\hat{\sigma}_K^2(\sum_{l=0}^{K}[w_{0l}l + w_{1l}(1+l)] + p)$$

where $\hat{\sigma}^2(w) = T^{-1}\sum_{t=1}^{T}(y_t - \hat{\mu}_t^a(w))^2$.

5 Monte Carlo Simulations

This section reports the results of a set of Monte Carlo experiments to assess the adequacy of the asymptotic approximations in finite samples and evaluate the effectiveness of the proposed approach relative to existing methods. To facilitate a direct comparison, we adopt the same design as Hansen (2010). In particular, the sample size $T \in \{50, 200\}$, the innovations $e_t \overset{i.i.d}{\sim} N(0,1)$, the trend parameters are set at $\beta_0 = \beta_1 = 0$ and the true lag order $k \in \{0, 4, 8\}$. Results are presented for $p \in \{0, 1\}$.

5.1 Forecast Risk with Known Lag Order

The first two experiments assume knowledge of the true order k thereby enabling us to delineate the effect of persistence uncertainty on the forecasts. With reference to equation (1), the first data generating process (DGP) sets $\alpha_1 = \cdots = \alpha_k = 0$, varies c from -20 to 0, which implies a range for α of [0.6, 1] for $T = 50$ and a range of [0.9, 1] for $T = 200$. For each parameter configuration, the finite sample forecast risk $TE[(\hat{\mu}_{T+1} - \mu_{T+1})^2]$ is calculated for six estimators: unrestricted FGLS estimator, DF^{GLS} pretest estimator and FGLS Mallows averaging estimator together with their three OLS counterparts. The risk is calculated using 500,000 Monte Carlo replications.

Figures 3a and 3b present the results for the first DGP for $p = 0$ and $p = 1$, respectively. It is clear that FGLS incurs lower risk than OLS for all three types of estimators: unrestricted, pretest and averaging. This suggests that the efficiency gain of using FGLS not only lies in the unrestricted case, but is more broadly applicable to the pretesting and averaging schemes. Moreover, as in the OLS case illustrated by Hansen (2010), the FGLS pretest estimator exhibits high risk and the FGLS Mallows averaging estimator uniformly dominates the unrestricted FGLS estimator for $p = 1$.[6] For $p = 0$, the superiority of the proposed estimator over unrestricted FGLS estimation is only discernible for $c > -5$. In terms of comparison with OLS model averaging, the risk of the proposed estimator is uniformly smaller for $p = 1$ and nearly uniformly smaller for $p = 0$. Overall, our FGLS Mallows averaging estimator performs well and displays lowest risk among all estimators for $c < -5$ when $p = 1$.

The second DGP sets $\alpha_j = -(-\theta)^j$ for $j = 1, ..., k$ and $\theta = 0.6$. The results are presented in Figures 4a and 4b, which exhibit the same overall pattern as observed in Figures 3a and 3b ,respectively, i.e., the FGLS estimators dominate their OLS counterparts, and for a large range of c values (around $c < -3$), the FGLS averaging estimator has the smallest forecast risk among all estimators when the model includes a deterministic trend.

5.2 Forecast Risk with Unknown Lag Order

We next consider the situation where the number of autoregressive lags k is unknown. Three types of estimators are compared: (1) the Mallows selection estimator (denoted S-OLS/FGLS), which selects unrestricted models from AR(1) through AR($K + 1$), i.e., $\hat{\mu}_t(0)$ through $\hat{\mu}_t(K)$; (2) the Mallows averaging estimator (denoted PA-OLS/FGLS, PA abbreviating partial averaging) that averages over this set of unrestricted models; (3) the general averaging estimator (denoted GA-OLS/FGLS) which combines all models from $\{\hat{\mu}_t(l)\}$ and $\{\tilde{\mu}_t(l)\}$ for $l \in \{0, 1, ..., K\}$. Again, we set $\alpha_j = -(-\theta)^j$ for $j = 1, ..., k$ and $\theta = 0.6$.

Figures 5a and 5b present the results for the six forecast methods. All three types of FGLS estimators uniformly dominate their OLS counterparts. The risk reduction is substantial. Overall, FGLS general averaging achieves uniformly lowest risk among all averaging/selection strategies when $p = 1$ and is competitive with the best estimator (which turns out to be PA-FGLS for an intermediate range of c values when $T = 200$) for each value of c when $p = 0$. The results are very similar across all K and T.

6 Empirical Application

This section undertakes a pseudo out-of-sample forecasting exercise using a set of US macroeconomic time series to (i) evaluate the performance of the proposed approach relative to

[6] However, this is only observed in simulations; to have a concrete judgment, one might follow Zhang, Ullah and Zhao (2016) to derive sufficient conditions which involves sample size, the number of parameters and possibly the persistence parameter.

OLS-based methods; (ii) assess the relative contribution of persistence uncertainty and lag order uncertainty in determining the accuracy of forecasts. We employ the FRED-MD dataset compiled by McCracken and Ng (2016) and maintained/updated at the Federal Reserve Bank of St. Louis. Our analysis is based on 123 monthly time series over the period 1960:02-2018:12.[7] McCracken and Ng (2016) provide a set of seven codes in order to transform the series to stationarity: (1) no transformation; (2) Δy_t; (3) $\Delta^2 y_t$; (4) $\log(y_t)$; (5) $\Delta\log(y_t)$; (6) $\Delta^2\log(y_t)$; (7) $\Delta(y_t/y_{t-1} - 1)$. In order to ensure that the series fit our framework that allows for highly persistent time series with/without deterministic trends, we adopt the following modified codes: (1') no transformation; (2') y_t; (3') Δy_t; (4') $\log(y_t)$; (5') $\log(y_t)$; (6') $\Delta\log(y_t)$; (7') $(y_t/y_{t-1} - 1)$. For codes (1') and (4'), we use the forecasts from the model with no deterministic trend ($p = 0$) while for the remainder, we use the forecasts that allow for a deterministic trend ($p = 1$). In addition to analyzing the full set of time series, we also report results for eight core series as in Stock and Watson (2002a).

The out-of sample results are based on a rolling window scheme with an initial estimation period 1960:02-1969:12 (119 observations) so that the forecast evaluation period is 1970:01-2018:12 (588 observations). We compare eight different methods in terms of MSFE: (1) S-GLS: unconstrained FGLS with lag selection using the Mallows criterion; (2) PA-GLS: partial FGLS Mallows averaging over the number of lags only; (3) GA-GLS: general FGLS Mallows averaging over the unit root restriction and the number of lags; (4) PT-GLS: the pretest GLS estimator based on the Dickey-Fuller GLS t-statistic with lag selection using the MAIC criterion of Ng and Perron (2001); (5)-(8): S-OLS, PA-OLS, GA-OLS, PT-OLS: the OLS counterparts of methods (1)-(4). The maximum number of allowable lags in each method is set at $K = 12$.

Table 1 reports the percentage wins and losses based on MSFE for the 123 series, both pairwise and overall. In particular, it shows the percentage of the 123 series for which a method listed in a row outperforms a method listed in a column, as well as all methods (last column).[8] The results clearly illustrate the overall superior performance of the GLS-based methods which dominate their OLS versions in about 74% of the series. The GA-GLS

[7] As of 2018:12, the dataset consisted of 128 raw series of which 5 series had at least 30 observations missing and were dropped from the analysis. These are: (1) VXOCLSx (CBOE S&P 100 Volatility Index); (2) ACOGNOx (Real Value of Manufacturers' New Orders Consumer Goods Industries deflated by Core PCE); (3) ANDENOx (Real Value of Manufacturers' New Orders for Capital Goods: Nondefense Capital Goods Industries deflated by Core PCE); (4) UMCSENTx (University of Michigan: Consumer Sentiment); (5) TWEXMMTH (Trade Weighted U.S. Dollar Index: Major Currencies).

[8] The results for a large number of series can be succinctly summarized in this way as in Boot and Nibbering (2019).

estimator delivers the most accurate forecasts for the majority (about 53%) of the series, consistent with our theoretical and simulation results. The pairwise comparisons reveal some interesting patterns. First, comparing GA-GLS with PA-GLS (or GA-OLS with PA-OLS) indicates that accounting for persistence uncertainty by averaging over the unit root restriction results in considerable forecasting gains compared to using the unconstrained FGLS estimator. Second, comparing PA-GLS to S-GLS (or PA-OLS with S-OLS) shows that accounting for lag order uncertainty by averaging over the number of lags in contrast to lag selection using an information criterion delivers more accurate forecasts in more than 95% of the series. Third, comparing GA-GLS with GA-OLS (or PA-GLS with PA-OLS) suggests that trend estimation by FGLS relative to OLS offers a substantial improvement in forecasting performance. Fourth, in more than 90% of the series, the best forecasting method involves some kind of averaging, whether over the unit root restriction or the number of lags or both.

To further understand the performance of the different methods for various types of series, Figures 6 and 7 plot the MSFE according to the eight groups defined in McCracken and Ng (2016): (1) output and income; (2) labor market; (3) housing; (4) consumption, orders, and inventories; (5) money and credits; (6) interest and exchange rates; (7) prices; (8) stock market. Figure 6 shows that most of the improvements offered by FGLS relative to OLS are concentrated in groups 2,4,6,7. The top panel of Figure 7 compares GA-GLS to PA-GLS to identify those groups most sensitive to the unit root restriction. The plot indicates that the advantage of the former over the latter is discernible primarily for the series in groups 1,2,7,8. The bottom panel of Figure 7 compares PA-GLS with S-GLS in an attempt to uncover the types of series most susceptible to lag order uncertainty. Averaging over the number of lags as opposed to lag selection is found to be the dominant approach mainly for all series in group 8, 85% of the series in group 7 and 57% of the series in group 4 with improvements in at least some series in each of the other groups. Our analysis therefore suggests that addressing both sources of uncertainty through model averaging can be helpful in generating reliable forecasts.

Finally, we consider the relative predictive ability of the methods with respect to the eight core series analyzed in Stock and Watson (2002a): four real variables (industrial production, real personal income less transfers, real manufacturing and trade sales, number of employees on nonagricultural payrolls) and four price indices (the consumer price index, the personal consumption expenditure implicit price deflator, the consumer price index less food and energy, the producer price index for finished goods). Table 2 reports the MSFE of the eight different methods relative to that of the OLS estimator using twelve autoregressive lags of the first differences of the variable. Hence, a number less than one indicates a lower MSFE relative to the OLS benchmark and vice-versa. The best method for a given series is highlighted in bold. The GA-GLS estimator turns out to be the best method in seven out of the eight series, the exception being nonagricultural employment for which S-GLS dominates the other methods. These results further confirm the effectiveness of the proposed approach when forecasting US macroeconomic time series.

7 Conclusion

This paper is concerned with developing a new forecast combination approach for highly persistent univariate autoregressions which entails a feasible generalized least squares Mallows averaging estimator that combines the unrestricted and restricted estimators. Our contributions are three-fold. First, we derive analytical results for the in-sample AMSE and MSFE of the proposed estimator and show that the optimal averaging weights are different from the OLS weights studied in Hansen (2010). Second, our analysis fills a gap in the literature in terms of providing a theoretical basis for the generalized mallows averaging estimator by modeling the coefficients of the short-run dynamics as local to zero. Third, our simulation and empirical results indicate that the proposed approach yields considerable improvement over existing univariate methods in terms of finite sample forecast risk which should be appealing to practitioners. The new procedure can also potentially serve as a useful univariate benchmark for evaluating forecasts based on exploiting information in large datasets (e.g., the diffusion index methodology of Stock and Watson, 2002a,b). In terms of future research, our analysis assumes homoskedastic innovations so an interesting extension would be to the heteroskedastic case which would involve adapting the jackknife method of Hansen and Racine (2012) to the present context. It would also be useful to extend the theory to multi-step ahead forecasts using a cross-validation based criterion to select the weights.

References

Bates, J. M., Granger, C. W. (1969), "The combination of forecasts," *Journal of the Operational Research Society* 20, 451-468.

Boot, T., Nibbering, D. (2019), "Forecasting using random subspace methods," *Journal of Econometrics* 209, 391-406.

Canjels, E., Watson, M. W. (1997), "Estimating deterministic trends in the presence of serially correlated errors," *Review of Economics and Statistics* 79, 184-200.

Chan, N. H., Wei, C. Z. (1987), "Asymptotic inference for nearly nonstationary AR(1) processes," *Annals of Statistics* 15, 1050-1063.

Clements, M. P., Kim, J. H. (2007), "Bootstrap prediction intervals for autoregressive time series," *Computational statistics & data analysis* 51, 3580-3594.

Diebold, F. X., Kilian, L. (2000), "Unit-root tests are useful for selecting forecasting models," *Journal of Business & Economic Statistics* 18, 265-273.

Elliott, G., Rothenberg, T. J., Stock, J. H. (1996), "Efficient tests for an autoregressive unit root," *Econometrica* 64, 813-836.

Franses, P. H., Kleibergen, F. (1996), "Unit roots in the Nelson-Plosser data: Do they matter for forecasting?," *International Journal of Forecasting* 12, 283-288.

Granger, C. W., Ramanathan, R. (1984), "Improved methods of combining forecasts," *Journal of Forecasting* 3, 197-204.

Grenander, U., Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series.* New York: Wiley.

Hansen, B.E. (1999), "The grid bootstrap and the autoregressive model," *Review of Economics and Statistics* 81, 594-607.

Hansen, B.E. (2007), "Least squares model averaging," *Econometrica* 75, 1175-1189.

Hansen, B.E. (2008), "Least-squares forecast averaging," *Journal of Econometrics* 146, 342-350.

Hansen, B.E. (2010), "Averaging estimators for autoregressions with a near unit root," *Journal of Econometrics* 158, 142-155.

Hansen, B. E., Racine, J. S. (2012), "Jackknife model averaging," *Journal of Econometrics* 167, 38-46.

Hjort, N. L., Claeskens, G. (2003), "Frequentist model average estimators," *Journal of the American Statistical Association* 98, 879-899.

Kim, J. H. (2001), "Bootstrap-after-bootstrap prediction intervals for autoregressive models," *Journal of Business & Economic Statistics* 19, 117-128.

Kim, J. H. (2003), "Forecasting autoregressive time series with bias-corrected parameter estimators," *International Journal of Forecasting* 19, 493-502.

Liu, Q., Okui, R., Yoshimura, A. (2016), "Generalized least squares model averaging," *Econometric Reviews* 35, 1692-1752.

McCracken, M. W., Ng, S. (2016), "FRED-MD: A monthly database for macroeconomic research," *Journal of Business & Economic Statistics*34, 574-589.

Mallows, C. L. (1973), "Some comments on Cp," *Technometrics* 15, 661-675.

Mikusheva, A. (2007), "Uniform inference in autoregressive models," *Econometrica* 75, 1411-1452.

Ng, S., Perron, P. (2001), "Lag length selection and the construction of unit root tests with good size and power," *Econometrica* 69, 1519-1554.

Ng, S., Vogelsang, T. (2002), "Forecasting autoregressive time series in the presence of deterministic components," *Econometrics Journal* 5, 196-224.

Phillips, P. C. B. (1987), "Towards a unified asymptotic theory for autoregression," *Biometrika* 74, 535-547.

Phillips, P. C. B. (2014), "On confidence intervals for autoregressive roots and predictive regression," *Econometrica* 82, 1177-1195.

Phillips, P. C. B. (2015), "Pitfalls and possibilities in predictive regression," *Journal of Financial Econometrics* 13, 521-555.

Phillips, P. C. B., Lee, C. C. (1996), "Efficiency gains from quasi-differencing under non-stationarity," *Athens Conference on Applied Probability and Time Series Analysis* 300-314. Springer, New York, NY.

Roy, A., Fuller, W. A. (2001), "Estimation for autoregressive time series with a root near one,". *Journal of Business & Economic Statistics* 19, 482–493.

Stock, J. H. (1996), "VAR, error correction and pretest forecasts at long horizons," *Oxford Bulletin of Economics and Statistics* 58, 685-701.

Stock, J. H., Watson, M. W. (2002a), "Macroeconomic forecasting using diffusion indexes," *Journal of Business & Economic Statistics* 20, 147-162.

Stock, J. H., Watson, M. W. (2002b), "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association* 97, 1167-1179.

Zhang, X., Ullah, A., Zhao, S. (2018), "On the dominance of Mallows model averaging estimator over ordinary least squares estimator," *Economics Letters* 142, 69-73.

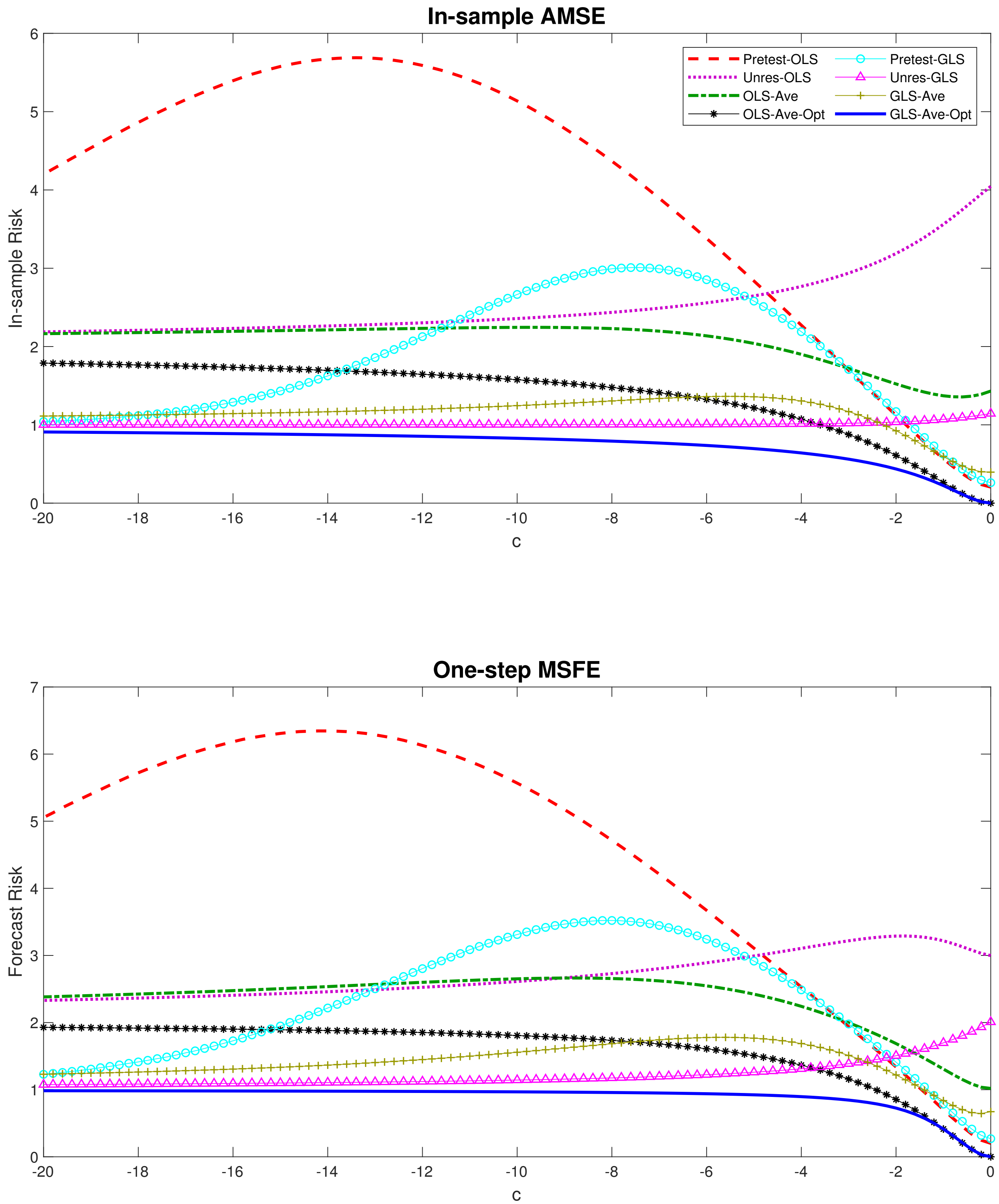

Figure 1: In-sample AMSE/MSFE of OLS and GLS estimators, $p = 0$

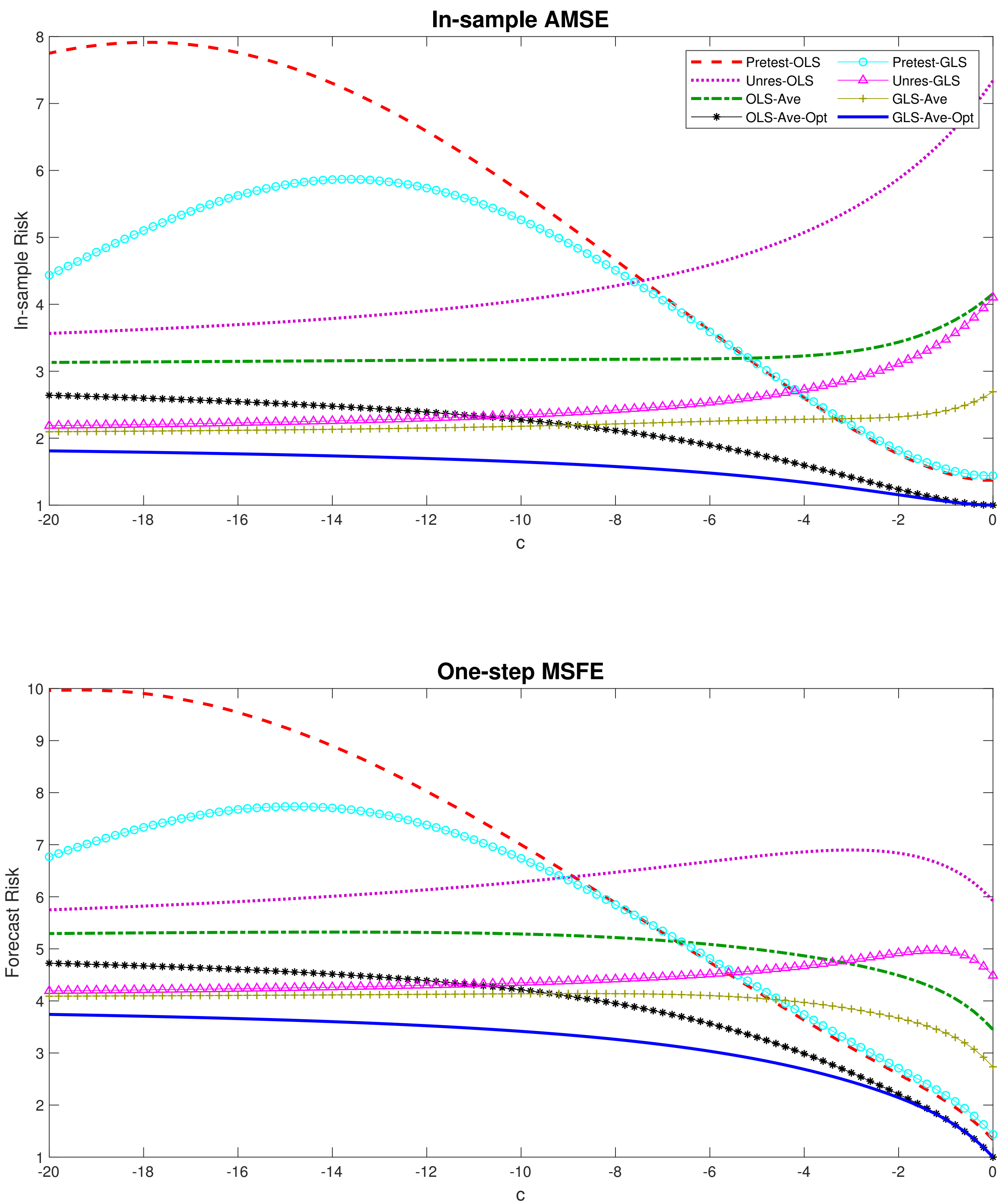

Figure 2: In-sample AMSE/MSFE of OLS and GLS estimators, $p = 1$

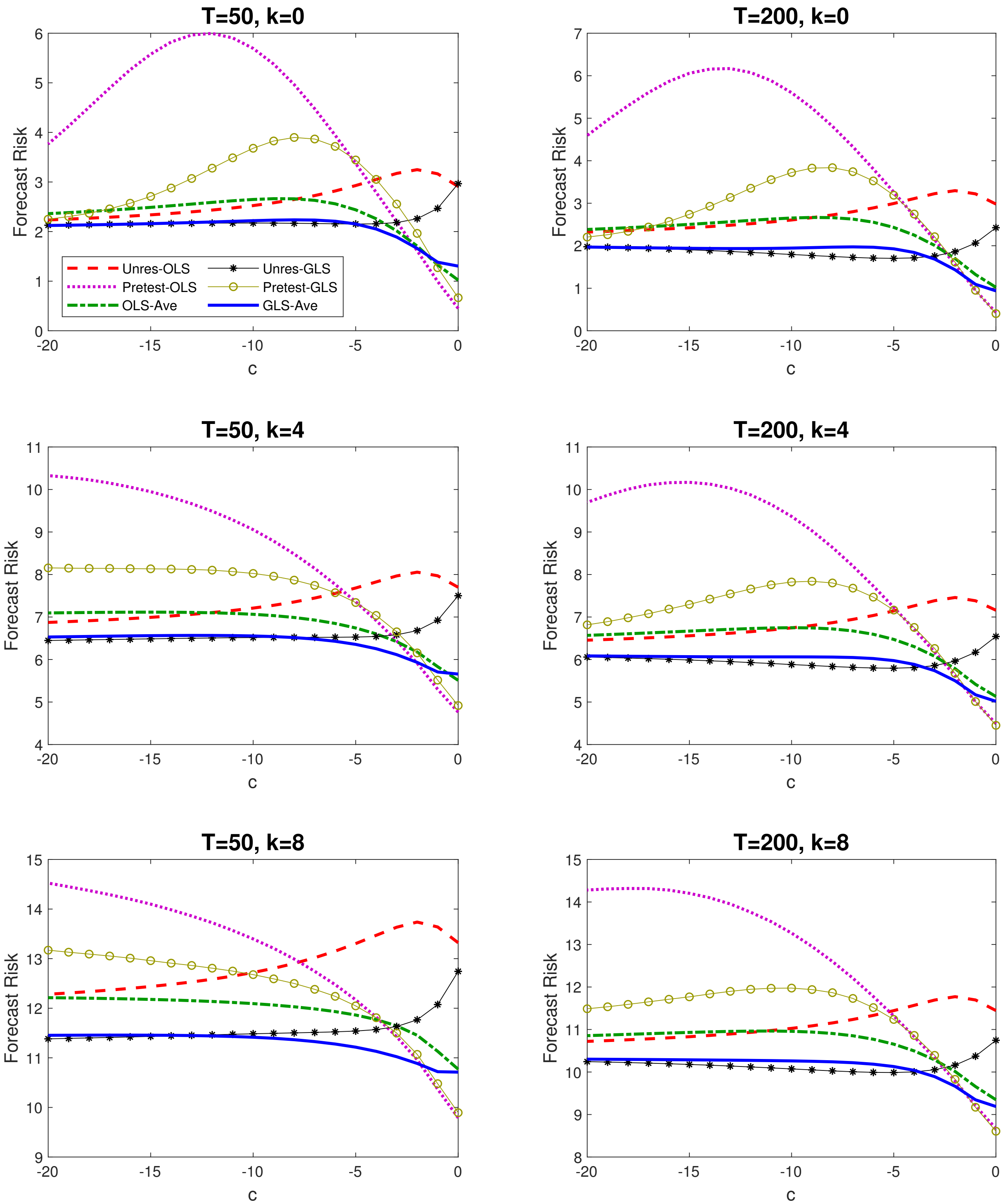

Figure 3a: Forecast risk of OLS averaging and GLS averaging, $p = 0$

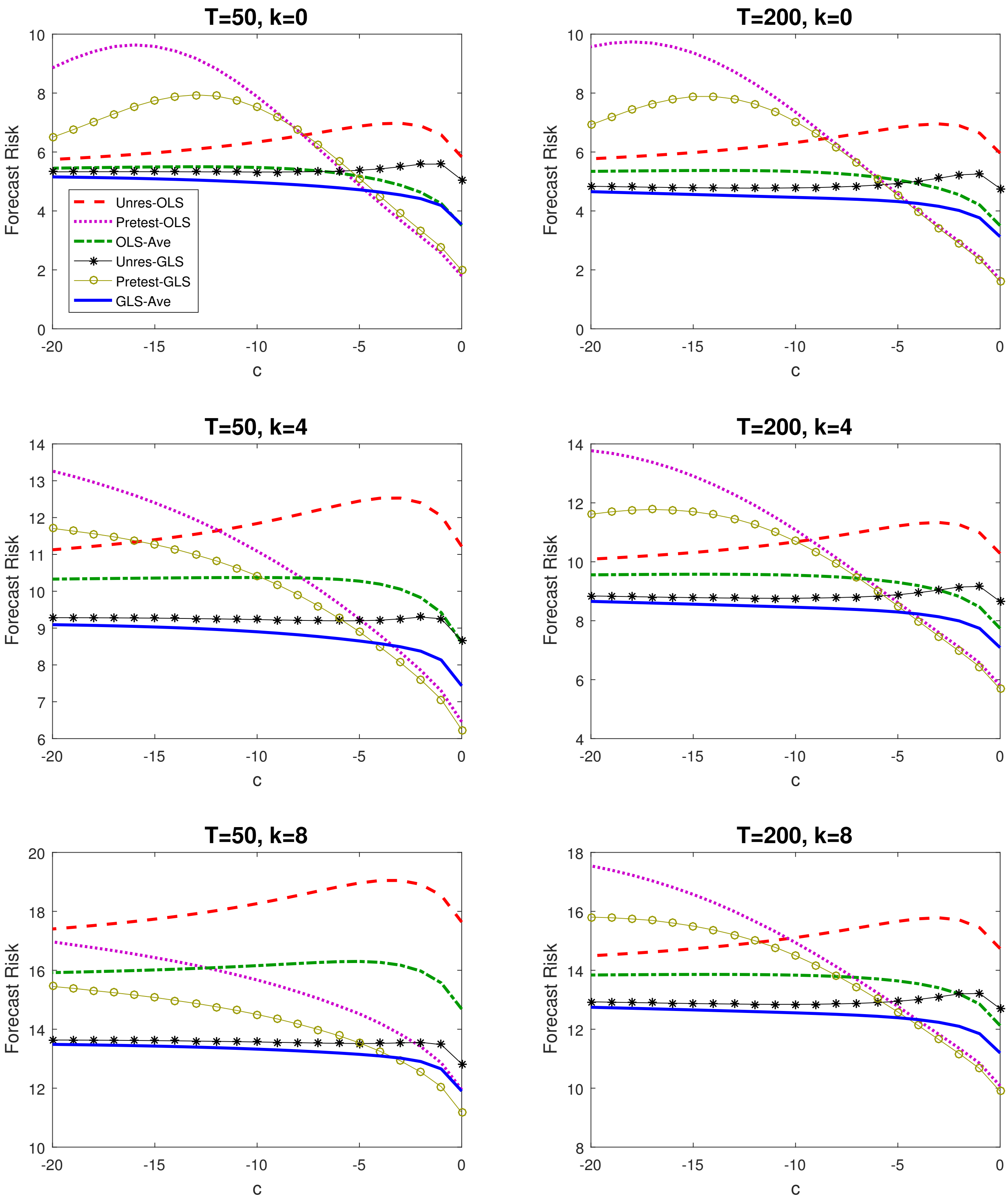

Figure 3b: Forecast risk of OLS averaging and GLS averaging, $p = 1$

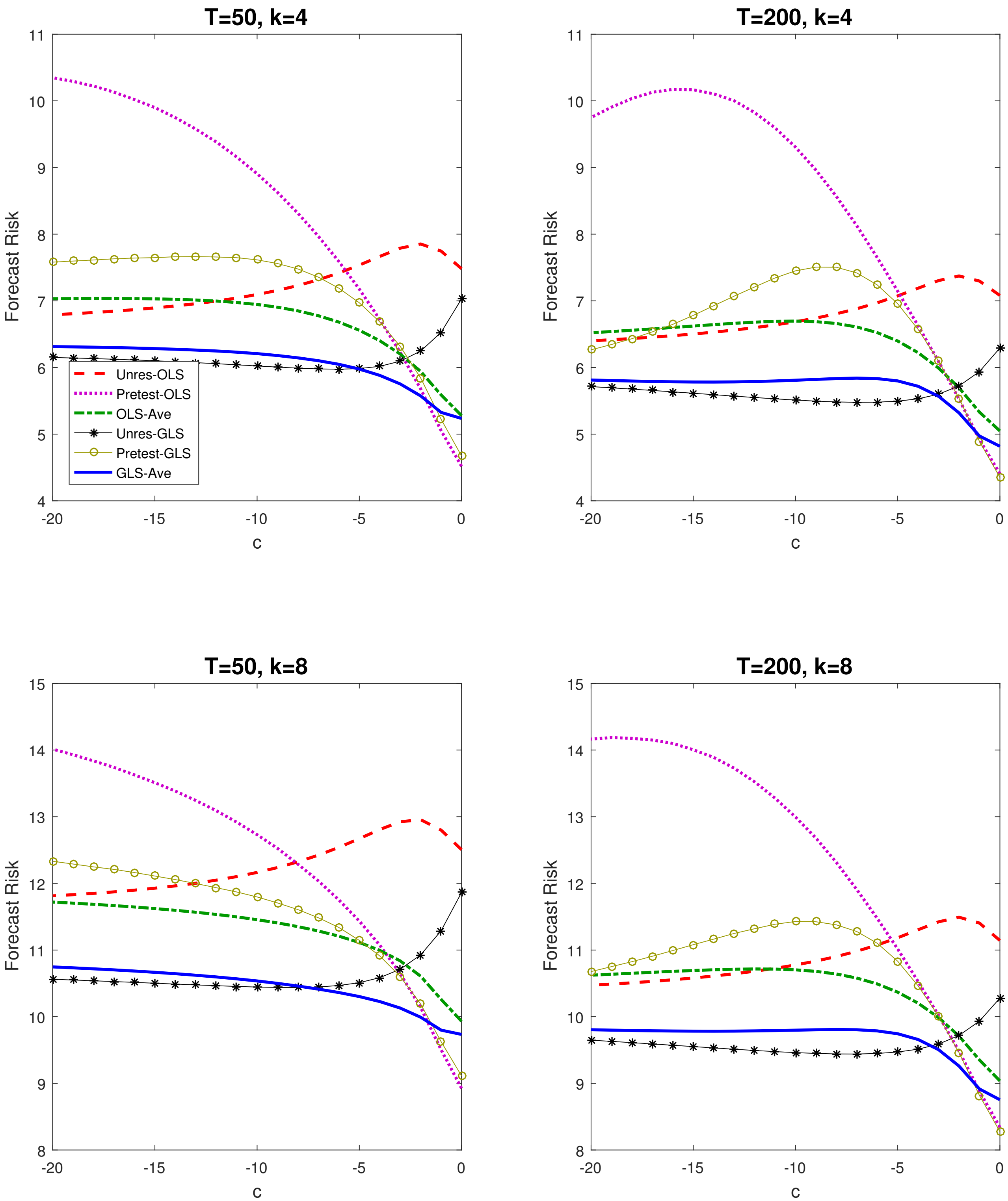

Figure 4a: Forecast risk of OLS averaging and GLS averaging, $p = 0$

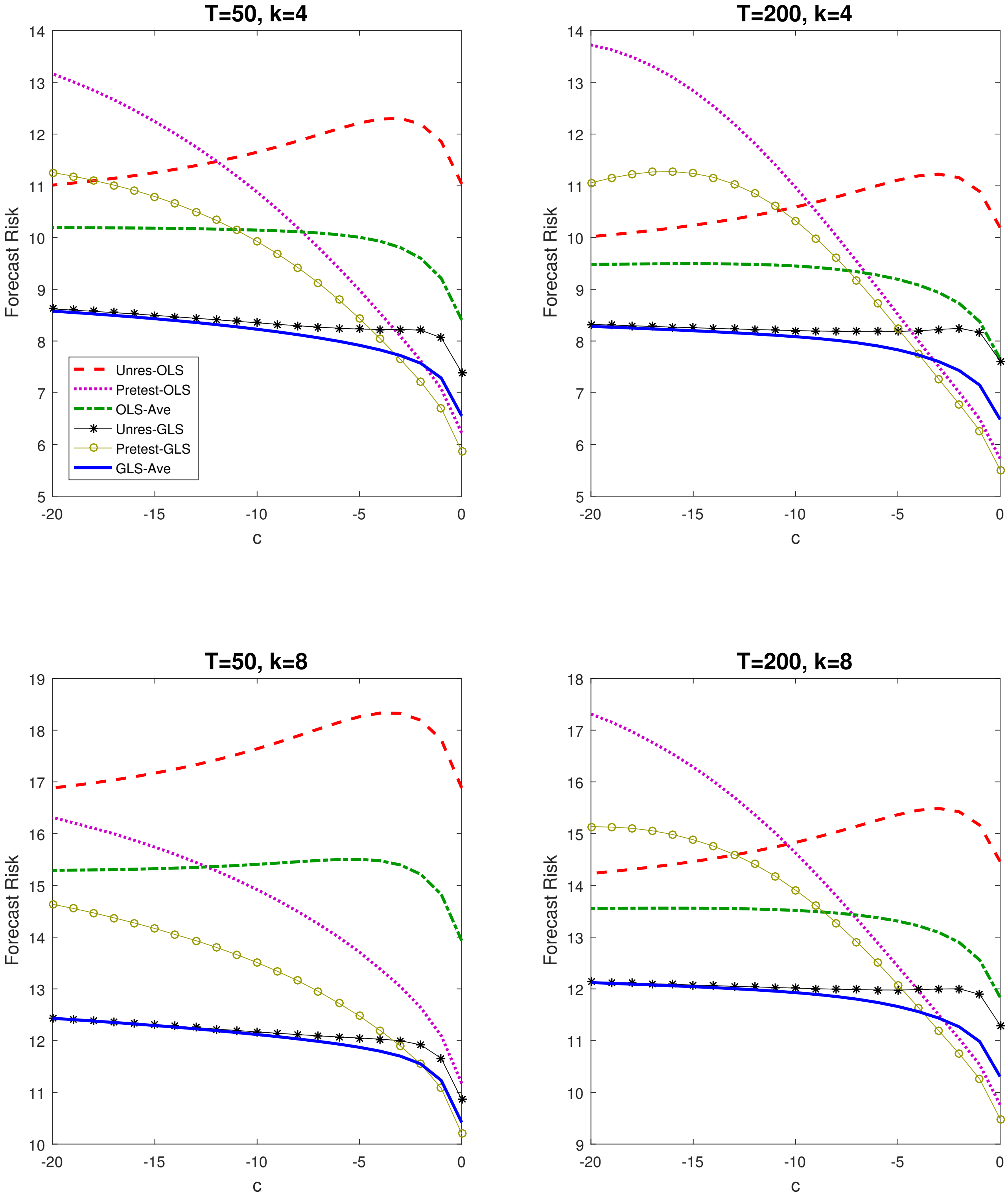

Figure 4b: Forecast risk of OLS averaging and GLS averaging, $p = 1$

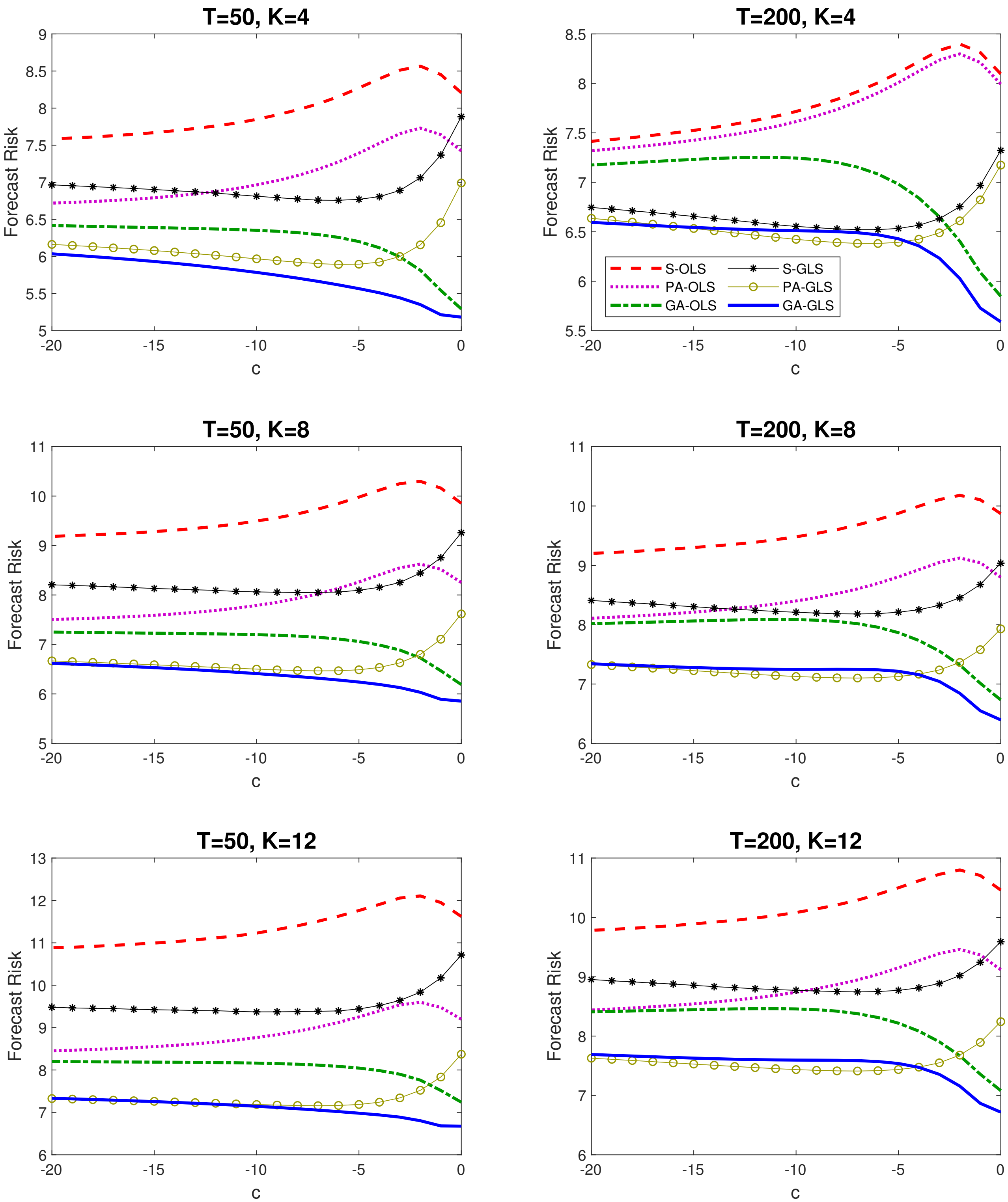

Figure 5a: Forecast risk of General OLS averaging and General GLS averaging, $p = 0$

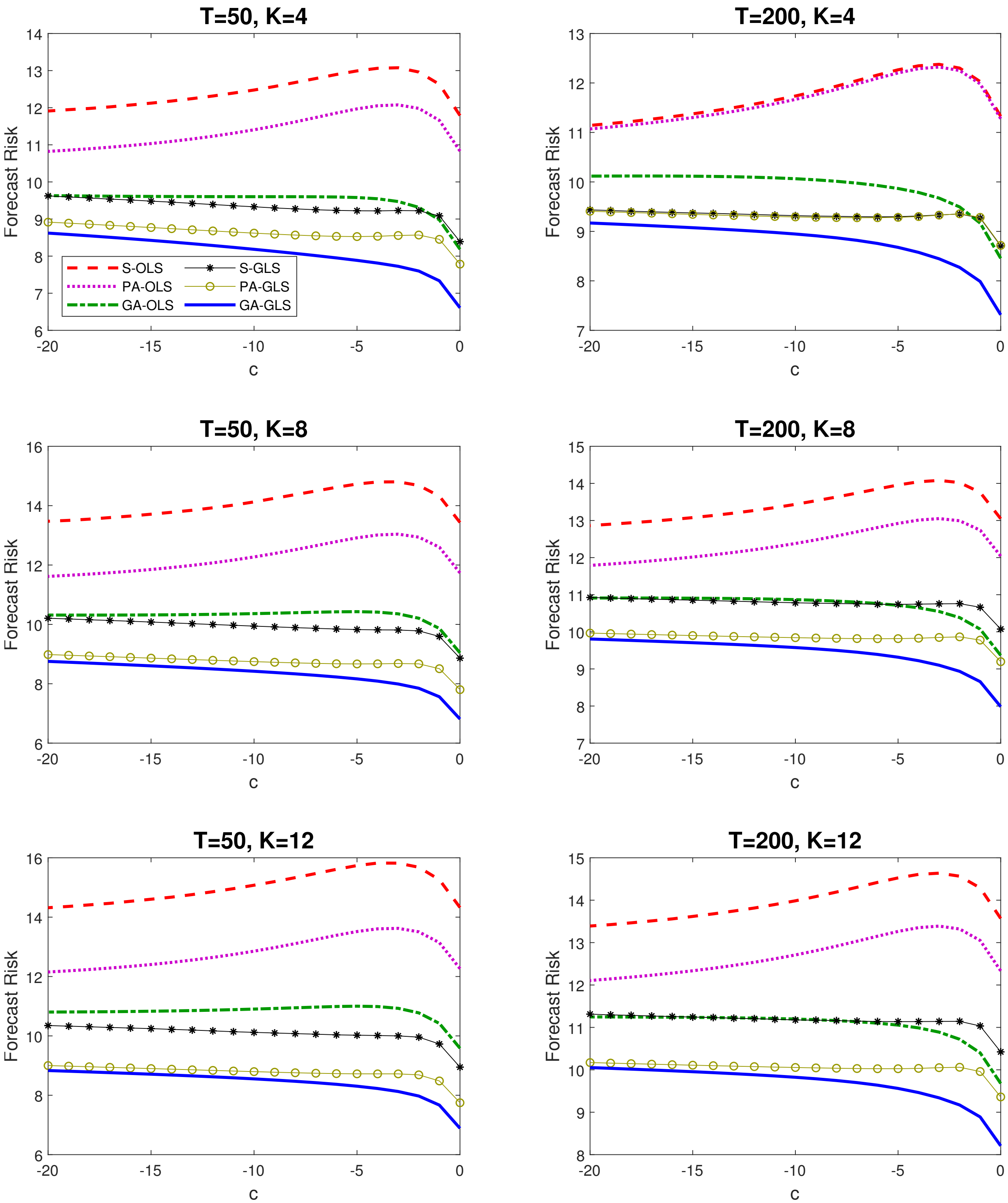

Figure 5b: Forecast risk of General OLS averaging and General GLS averaging, $p = 1$

Table 1: Percentage wins/losses of different forecasting methods

Method	S-GLS	PA-GLS	GA-GLS	PT-GLS	S-OLS	PA-OLS	GA-OLS	PT-OLS	ALL
S-GLS		3.25	4.88	45.53	80.49	27.64	18.70	43.90	2.44
PA-GLS	96.75		21.14	83.74	95.12	85.37	52.03	81.30	13.01
GA-GLS	95.12	78.86		93.50	100.00	92.68	73.98	93.50	**53.66**
PT-GLS	54.47	16.26	6.50		77.24	42.28	21.95	50.41	4.88
S-OLS	19.51	4.88	0.00	22.76		4.88	0.81	23.58	0.00
PA-OLS	72.36	14.63	7.32	57.72	95.12		6.50	59.35	0.81
GA-OLS	81.30	47.97	26.02	78.05	99.19	93.50		78.05	23.58
PT-OLS	56.10	18.70	6.50	49.59	76.42	40.65	21.95		1.63

Note: this table shows the percentage of the 123 series for which a method listed in a row outperforms a method in a column, include the other all in the last column.

Table 2: Relative MSFE of eight core macroeconomic time series

	Industrial production	Personal income	Mfg & trade sales	Nonag. employment	CPI	Consumption deflator	CPI exc. food	PPI
S-GLS	.962**	.989	.981	**.915***	.973	.995	.976	.983
PA-GLS	.961**	.958*	.967**	.919***	.958	.957**	.958*	.942**
GA-GLS	**.960****	**.950****	**.963****	.921***	**.952***	**.951****	**.955****	**.936****
PT-GLS	.974	.986	.974	.922***	.986	.982	.973	.979
S-OLS	.981	.995	.999	.946***	.979	1.005	.990	1.009
PA-OLS	.972**	.960	.983	.954***	.964	.961**	.961*	.948**
GA-OLS	.965**	.954*	.968**	.946***	.955*	.955**	.957**	.942**
PT-OLS	.960**	.983	.965**	.931***	.980	.993	.985	1.007

Note: *denotes 10%, **denotes 5%, and ***denotes 1% significance level for a two-sided Diebold and Mariano (1995) test. The benchmark is an unrestricted OLS estimation method with 12 lags.

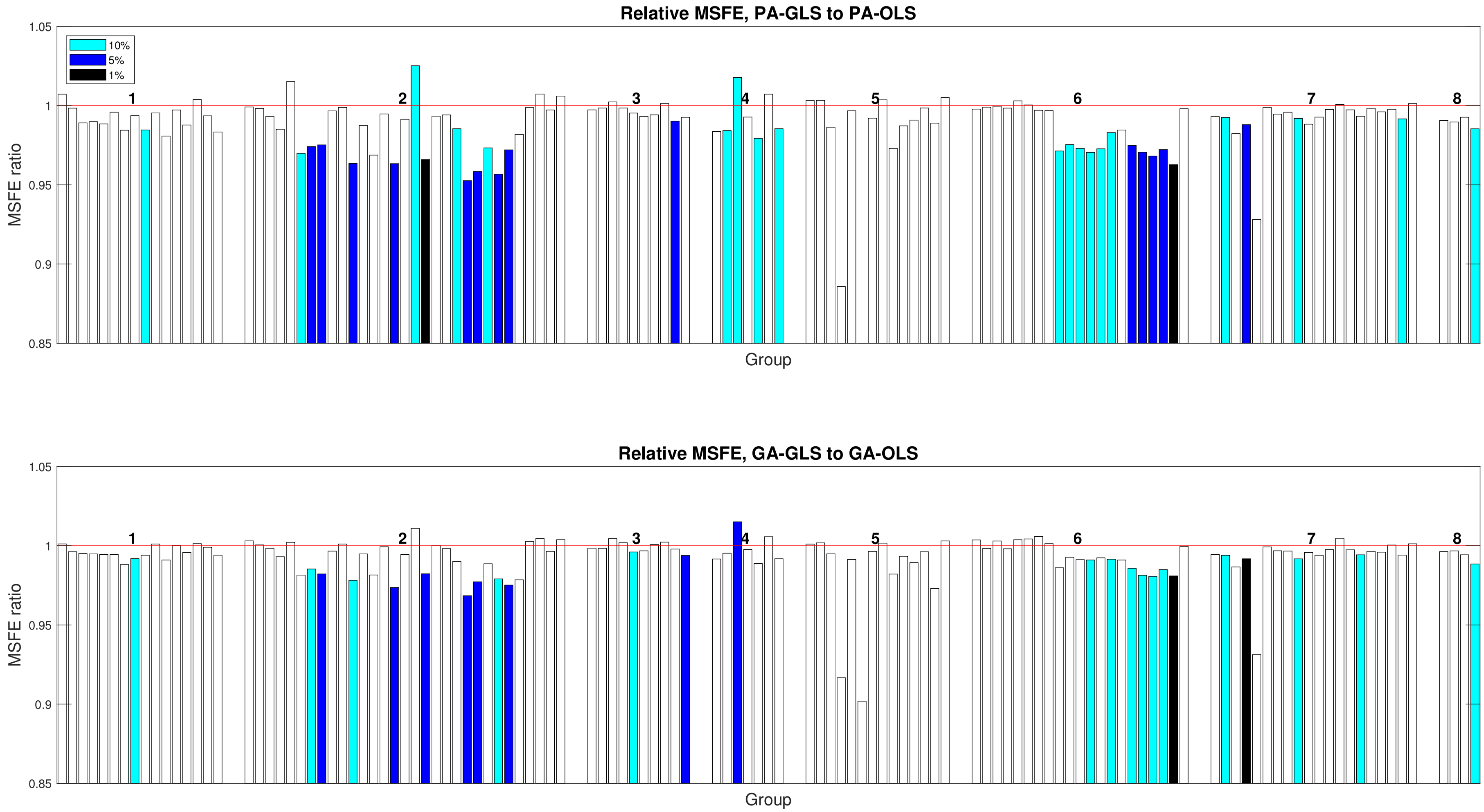

Figure 6: FRED-MD: forecast accuracy comparison between FGLS and OLS

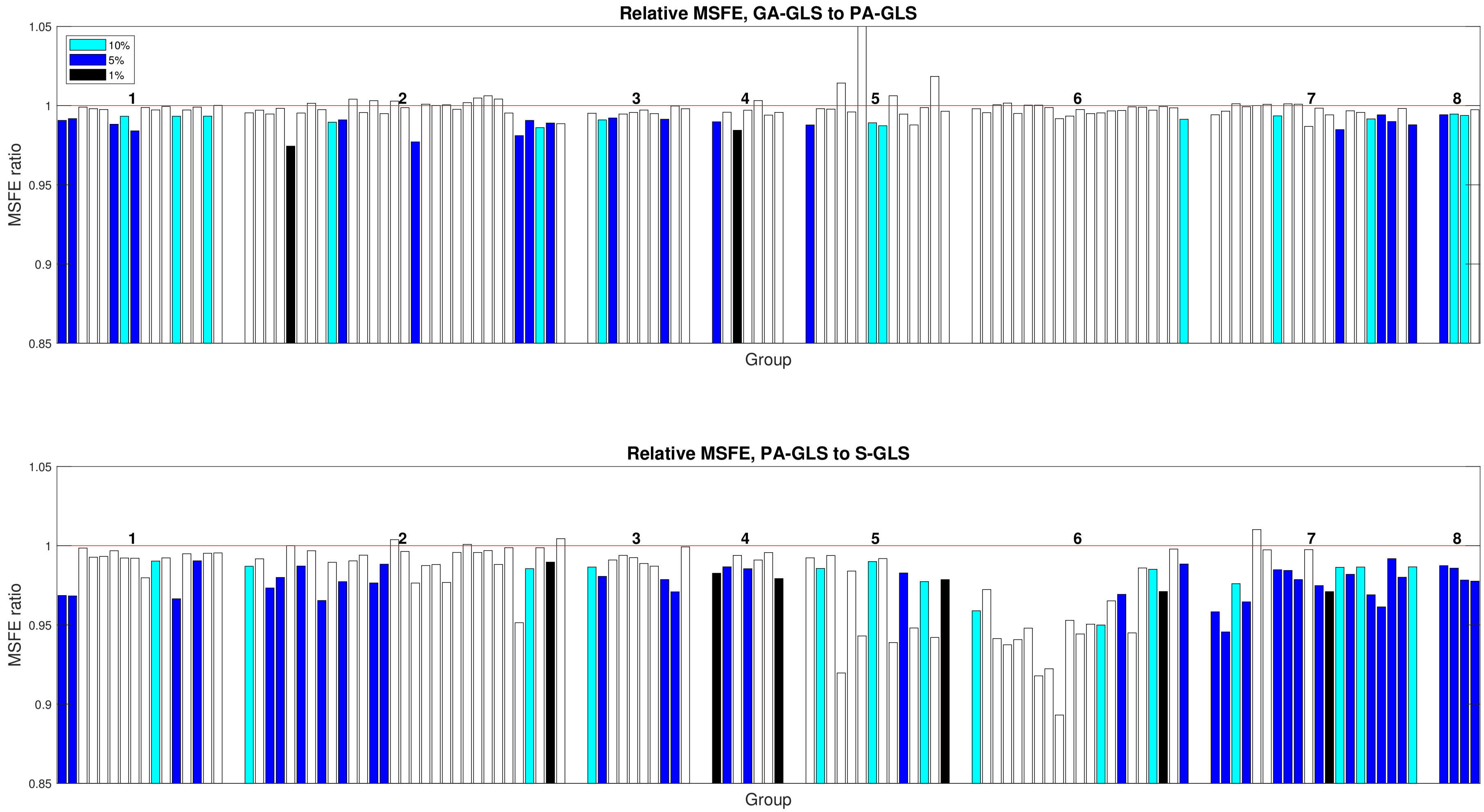

Figure 7: FRED-MD: forecast accuracy comparison between alternative methods of FGLS

Appendix A: Proofs

Let $W(.)$ denote a standard Brownian motion on $[0,1]$ and define the diffusion process: $dJ_c(r) = cJ_c(r) + dW(r)$. Define the demeaned and detrended versions of $J_c(.)$ as follows: $\bar{J}_c(r) = J_c(r) - \int_0^1 J_c(s)ds$, $\widetilde{J}_c(r) = J_c(r) - (4-6r)\int_0^1 J_c(s)ds - (12r-6)\int_0^1 sJ_c(s)ds$. Let $\beta = (\beta_0, \beta_1)'$, $z_t = (1,t)'$. For brevity, all proofs are provided only for the case $p = 1$. The proofs for $p = 0$ follow analogous arguments. We first state two lemmas that will be useful in developing the proofs of the results.

Lemma A.1 *Let $\dot{}$ and $\ddot{}$ denote the first stage and the second stage estimates of parameters in the unrestricted FGLS procedure. Under Assumptions A1-A2, as $T \to \infty$, we have*

(a) $$T(\dot{\alpha}-\alpha) \xrightarrow{d} \begin{cases} a\frac{\int_0^1 \widetilde{J}_c dW(r)}{\int_0^1 \widetilde{J}_c^{\,2} dr} & for\ p=1 \\ a\frac{\int_0^1 \bar{J}_c dW(r)}{\int_0^1 \bar{J}_c^{\,2} dr} & for\ p=0 \end{cases}$$

(b) $$\begin{cases} \frac{T^{\frac{1}{2}}}{\sigma}(\ddot{\beta}_1-\beta_1) \xrightarrow{d} a^{-1}\gamma_1 & for\ p=1 \\ \frac{T^{-\frac{1}{2}}}{\sigma}(\ddot{\beta}_0-\beta_0) \xrightarrow{d} 0 & for\ p=0,1 \end{cases}$$

where $\gamma_1 = (1 - a\dot{c} + \frac{1}{3}(a\dot{c})^2)^{-1}\int_0^1 (1-a\dot{c}s)d\dot{W}(s)$, $d\dot{W}(s) = dW(s) - (a\dot{c}-c)J_c(s)ds$.

(c) $$\frac{T^{-\frac{1}{2}}}{\sigma}\hat{u}_{[Tr]} \xrightarrow{d} \begin{cases} a^{-1}P(r) & for\ p=1 \\ a^{-1}J_c(r) & for\ p=0 \end{cases}$$

where $P(r) = J_c(r) - \gamma_1 r$.

(d) $$T(\ddot{\alpha}-\alpha) \xrightarrow{d} \begin{cases} a\frac{\int_0^1 P(r)dW(r) + \gamma_1\int_0^1 (cr-1)P(r)dr}{\int_0^1 P(r)^2 dr} & for\ p=1 \\ a\frac{\int_0^1 J_c dW(r)}{\int_0^1 J_c^2 dr} & for\ p=0 \end{cases}$$

(e) $\frac{T^{\frac{1}{2}}}{\sigma}(\ddot{\alpha}_1-\alpha_1, \cdots, \ddot{\alpha}_k-\alpha_k)' \xrightarrow{d} R \sim N(0, Q^{-1})$. $Q = E(L_tL_t')$. $L_t = (\Delta u_{t-1}, \cdots, \Delta u_{t-k})'$.

Lemma A.2 *Under Assumptions A1-A2, as $c \to -\infty$, we have*

(a) $\lim_{c\to-\infty} E[\int_0^1 crJ_c(r)dr]^2 = \frac{1}{3}$

(b) $\lim_{c\to-\infty} E[\int_0^1 crJ_c(r)dr \int_0^1 rdW(r)] = -\frac{1}{3}$

(c) $\lim_{c\to-\infty} E[\gamma_1^2 \int_0^1 (cr-1)^2 dr] = 1$

(d) $\lim_{c\to-\infty} E[(\ddot{c}-c)^2 \int_0^1 P(r)^2 dr] = 1$

(e) $\lim_{c\to-\infty} E[\gamma_1(\ddot{c}-c)\int_0^1 (cr-1)P(r)dr] = 0$

Proof of Lemma A.1: (a) From Lemma 1 of Hansen (1995), we have $\frac{u_{[Tr]}}{\sigma\sqrt{T}} \xrightarrow{d} a^{-1}J_c(r)$. Denoting $\widetilde{y}_t$ as the residual from regressing y_t on z_t, it follows that $\frac{\widetilde{y}_{[Tr]}}{\sigma\sqrt{T}} \xrightarrow{d} a^{-1}\widetilde{J}_c(r)$. By Frisch-Waugh-Lovell theorem and the independence of the estimates between the nonstationary and stationary components, we have

$$T(\dot{\alpha}-\alpha) = \frac{\sum_{t=1}^{T}\widetilde{y}_t e_t/T}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2/T^2} + o_p(1) \xrightarrow{d} \frac{\sigma^2 a^{-1}\int_0^1 \widetilde{J}_c dW(r)}{\sigma^2 a^{-2}\int_0^1 \widetilde{J}_c^2 dr} = a\frac{\int_0^1 \widetilde{J}_c dW(r)}{\int_0^1 \widetilde{J}_c^{\,2} dr} \tag{A.1}$$

(b) Denote the quasi-differenced error $\hat{v}_t = u_t - \dot{\alpha}u_{t-1}$, we have $\hat{v}_t = u_t - \dot{\alpha}u_{t-1} = (u_t - \alpha u_{t-1}) - (\dot{\alpha}-\alpha)u_{t-1}$. We first derive the limit of $\frac{1}{\sigma\sqrt{T}}\sum_{t=1}^{[rT]}\hat{v}_t$. Denoting $g(L) = \alpha_1 L + ... + \alpha_k L^k$, we have

$$\begin{aligned} u_t - \alpha u_{t-1} &= g(L)(u_t - u_{t-1}) + e_t = g(L)(u_t - \alpha u_{t-1} + \alpha u_{t-1} - u_{t-1}) + e_t \\ &= (1-g(L))^{-1}g(L)(\alpha-1)u_{t-1} + (1-g(L))^{-1}e_t \end{aligned}$$

By the Beveridge-Nelson decomposition,

$$\begin{aligned} &\frac{1}{\sigma\sqrt{T}}\sum_{t=1}^{[rT]}\hat{v}_t = \frac{1}{\sigma\sqrt{T}}\sum_{t=1}^{[rT]}[u_t - \alpha u_{t-1} - (\dot{\alpha}-\alpha)u_{t-1}] \\ &= \frac{1}{\sigma\sqrt{T}}\{(1-g(1))^{-1}g(1)(\alpha-1)\sum_{t=1}^{[rT]}u_{t-1} + (1-g(1))^{-1}\sum_{t=1}^{[rT]}e_t - (\dot{\alpha}-\alpha)\sum_{t=1}^{[rT]}u_{t-1}\} + o_p(1) \\ &\xrightarrow{d} a^{-1}\times(1-a)\times ac\times a^{-1}\int_0^r J_c(s)ds + a^{-1}W(r) - (\dot{c}-c)\int_0^r J_c(s)ds \\ &= a^{-1}[W(r) - (a\dot{c}-c)\int_0^r J_c(s)ds] := a^{-1}\dot{W}(r) \end{aligned} \tag{A.2}$$

From Theorem 5(b) in Canjels and Watson (1997),

$$\begin{aligned} \frac{\sqrt{T}}{\sigma}(\ddot{\beta}_1 - \beta_1) &= \frac{T^{-1/2}\sum_{t=1}^{T}\hat{v}_t(1-a\dot{c}\frac{t}{T})}{\sigma(1-a\ddot{c}+\frac{1}{3}(a\dot{c})^2)} + o_p(1) \\ &\xrightarrow{d} (1-a\dot{c}+\frac{1}{3}(a\dot{c})^2)^{-1}\int_0^1(1-a\dot{c}s)d\dot{W}(s) := a^{-1}\gamma_1 \end{aligned} \tag{A.3}$$

where $\gamma_1 = (1-a\dot{c}+\frac{1}{3}(a\dot{c})^2)^{-1}\int_0^1(1-a\dot{c}s)d\dot{W}(s)$, $d\dot{W}(s) = dW(s) - (a\dot{c}-c)J_c(s)ds$. The second result in (b) can be shown by a simple algebraic exercise using results from Canjels and Watson (1997) and is hence omitted.

(c) We have

$$\frac{1}{\sigma\sqrt{T}}\hat{u}_{[rT]} = \frac{1}{\sigma\sqrt{T}}(y_{[rT]} - \ddot{\beta}_0 - \ddot{\beta}_1[rT])$$
$$= \frac{1}{\sigma\sqrt{T}}u_{[rT]} - \frac{1}{\sigma\sqrt{T}}(\ddot{\beta}_0 - \beta_0) - \frac{\sqrt{T}}{\sigma}(\ddot{\beta}_1 - \beta_1)\frac{[rT]}{T}$$
$$\xrightarrow{d} a^{-1}J_c(r) - 0 - a^{-1}\gamma_1 r = a^{-1}(J_c(r) - \gamma_1 r) := a^{-1}P(r) \tag{A.4}$$

(d) Note that $\hat{u}_t = u_t - (\ddot{\beta}_0 - \beta_0) - (\ddot{\beta}_1 - \beta_1)t$, $\Delta\hat{u}_t = \Delta u_t - (\ddot{\beta}_1 - \beta_1)$. Defining $\ddot{\beta} = (\ddot{\beta}_0, \ddot{\beta}_1)'$, the effective error is

$$\xi_t = e_t - (u_t - \hat{u}_t) + \alpha(u_{t-1} - \hat{u}_{t-1}) + \alpha_1(\Delta u_{t-1} - \Delta\hat{u}_{t-1}) + \cdots + \alpha_k(\Delta u_{t-k} - \Delta\hat{u}_{t-k})$$
$$= e_t - z_t'(\ddot{\beta} - \beta) + (1 + \frac{ac}{T})z_{t-1}'(\ddot{\beta} - \beta) + \alpha_1(\ddot{\beta}_1 - \beta_1) + \cdots + \alpha_k(\ddot{\beta}_1 - \beta_1)$$
$$= e_t + \frac{ac}{T}(\ddot{\beta}_0 - \beta_0) + (\frac{t-1}{T}c - 1)a(\ddot{\beta}_1 - \beta_1) \tag{A.5}$$

which gives

$$T(\ddot{\alpha} - \alpha) = \frac{\sum_{t=1}^T \hat{u}_t\xi_t/T}{\sum_{t=1}^T \hat{u}_{t-1}^2/T^2} + o_p(1)$$
$$= \frac{\sum_{t=1}^T \hat{u}_t e_t/T + \sum_{t=1}^T \hat{u}_t \frac{ac}{T}(\ddot{\beta}_0 - \beta_0)/T + \sum_{t=1}^T \hat{u}_t(\frac{t-1}{T}c - 1)a(\ddot{\beta}_1 - \beta_1)/T}{\sum_{t=1}^T \hat{u}_{t-1}^2/T^2}$$
$$\xrightarrow{d} \frac{\sigma^2 a^{-1}\int_0^1 P(r)dW(r) + 0 + \sigma^2 a^{-1}\gamma_1\int_0^1(cr-1)P(r)dr}{\sigma^2 a^{-2}\int_0^1 P(r)^2 dr}$$
$$= a\frac{\int_0^1 P(r)dW(r) + \gamma_1\int_0^1(cr-1)P(r)dr}{\int_0^1 P(r)^2 dr} \tag{A.6}$$

thereby proving (d).

(e) From (d), $\Delta\hat{u}_t = \Delta u_t + O_p(T^{-1/2})$ so that $\hat{L}_t = L_t + O_p(T^{-1/2})$. Recalling the independence of the estimates between the nonstationary and stationary components, we have

$$\frac{T^{\frac{1}{2}}}{\sigma}(\ddot{\alpha}_1 - \alpha_1, \cdots, \ddot{\alpha}_k - \alpha_k)' = (\frac{1}{T}\sum_{t=1}^T \hat{L}_t\hat{L}_t')^{-1}(\frac{1}{\sigma\sqrt{T}}\sum_{t=1}^T \hat{L}_t e_t) + o_p(1)$$
$$= (\frac{1}{T}\sum_{t=1}^T L_t L_t')^{-1}(\frac{1}{\sigma\sqrt{T}}\sum_{t=1}^T L_t e_t) + o_p(1)$$
$$\xrightarrow{d} R \sim N(0, Q^{-1}). \tag{A.7}$$

where $Q = E(L_t L_t')$. ▲

Proof of Lemma A.2: (a) We have $J_c(r) = \int_0^r e^{c(r-s)} dW(s)$. It follows that

$$\lim_{c\to-\infty} E[\int_0^1 crJ_c(r)dr]^2 = \lim_{c\to-\infty} E[\int_0^1 cr \int_0^r e^{c(r-s)}dW(s)dr]^2$$

$$= \lim_{c\to-\infty} E[\int_0^1 \int_s^1 cre^{c(r-s)}drdW(s)]^2 = \lim_{c\to-\infty} E[\int_0^1 ((1-\frac{1}{c})e^{c(1-s)} - s + \frac{1}{c})dW(s)]^2$$

$$= \lim_{c\to-\infty} [(1-\frac{1}{c})^2 \frac{e^{2c}-1}{2c} + \frac{1}{3} + \frac{1}{c^2} + 2(1-\frac{1}{c})\frac{1}{c} - \frac{1}{c}] = \frac{1}{3} \tag{A.8}$$

(b) Similar to (a), we have

$$\lim_{c\to-\infty} E[\int_0^1 crJ_c(r)dr \int_0^1 rdW(r)] = \lim_{c\to-\infty} E[\int_0^1 cr\int_0^r e^{c(r-s)}dW(s)dr \int_0^1 rdW(r)]$$

$$= \lim_{c\to-\infty} E[\int_0^1 ((1-\frac{1}{c})e^{c(1-s)} - s + \frac{1}{c})dW(s) \int_0^1 rdW(r)] = \lim_{c\to-\infty} \int_0^1 ((1-\frac{1}{c})e^{c(1-s)} - s + \frac{1}{c})sds$$

$$= \lim_{c\to-\infty} [(1-\frac{1}{c})\frac{1}{c}(1-\frac{e^c-1}{c}) - \frac{1}{3} - \frac{1}{2c}] = -\frac{1}{3} \tag{A.9}$$

(c) From lemma A.1 (a) we know $\dot{c} - c = \frac{\int_0^1 \widetilde{J}_c dW(r)}{\int_0^1 \widetilde{J}_c{}^2 dr}$. As $c \to -\infty$, Phillips (1987) shows $\left(\int_0^1 J_c{}^2\right)^{-1} \int_0^1 J_c dW(r) = O_p(|c|^{1/2})$. Using techniques of Phillips (2014), we can easily verify this result also applies for the trend case, i.e., $\dot{c} - c = \left(\int_0^1 \widetilde{J}_c^2\right)^{-1} \int_0^1 \widetilde{J}_c dW(r) = O_p(|c|^{1/2})$, which implies $\dot{c}/c = 1 + O_p(|c|^{-1/2})$. Then it follows that

$$\lim_{c\to-\infty} E[\gamma_1^2 \int_0^1 (cr-1)^2 dr] = \lim_{c\to-\infty} (\frac{1}{3}c^2 - c + 1)E[(1 - a\dot{c} + \frac{1}{3}(a\dot{c})^2)^{-1} \int_0^1 (1-a\dot{c}s)d\dot{W}(s)]^2$$

$$= \lim_{c\to-\infty} (\frac{1}{3}c^2 - c + 1)E[(1 - a\dot{c} + \frac{1}{3}(a\dot{c})^2)^{-1} \int_0^1 (1-a\dot{c}s)dW(s) - \int_0^1 (1-a\dot{c}s)(a\dot{c}-c)J_c(s)ds]^2$$

$$= \lim_{c\to-\infty} \frac{1}{3}c^2 3^2 c^{-4} E[a^2c^2(\int_0^1 sdW(s) + \int_0^1 (1-a)cJ_c(s)ds)]^2 + O(|c|^{-1/2})$$

$$= \lim_{c\to-\infty} 3a^{-2}E[(\int_0^1 sdW(s) + \int_0^1 (1-a)cJ_c(s)ds)]^2 \tag{A.10}$$

With results (a) and (b) in hand, we have

$$\lim_{c\to-\infty} 3a^{-2}E[(\int_0^1 sdW(s) + \int_0^1 (1-a)cJ_c(s)ds)]^2$$

$$= \lim_{c\to-\infty} 3a^{-2}E[(\int_0^1 sdW(s))^2 + (\int_0^1 (1-a)cJ_c(s)ds)^2 + 2\int_0^1 sdW(s)\int_0^1 (1-a)cJ_c(s)ds]^2$$

$$= 3a^{-2}[\frac{1}{3} + \frac{1}{3}(1-a)^2 - 2(1-a)\frac{1}{3}] = 1 \tag{A.11}$$

(d) From the proof of Lemma A.2 (c), we know $\gamma_1 = O_p(|c|^{-1})$. Phillips (2014) shows as $c \to -\infty$, $\int_0^1 J_c(r)dr = O_p(|c|^{-1})$, $\int_0^1 rJ_c(r)dr = O_p(|c|^{-1})$, $\int_0^1 J_c(r)dW(r) = O_p(|c|^{-1/2})$. Recalling $P(r) = J_c(r) - \gamma_1 r$, it is easy to show $\int_0^1 P(r)dW(r) = O_p(|c|^{-1/2})$, $\int_0^1(cr - 1)P(r)dr = O_p(1)$. Then it follows $\gamma_1 \int_0^1(cr-1)P(r)dr = O_p(|c|^{-1})$, which is of smaller order than $\int_0^1 P(r)dW(r)$. From Phillips (1987), as $c \to -\infty$, $(-2c)\int_0^1 J_c(r)^2dr \xrightarrow{p} 1$, $(-2c)^{-1/2}\int_0^1 J_c(r)dW(r) \xrightarrow{d} N(0,1)$. It is easy to show that the two limits hold when $J_c(r)$ is replaced with $P(r)$. Thus, $\left(\int_0^1 P(r)^2dr\right)^{-1/2}\int_0^1 P(r)dW(r) \xrightarrow{d} N(0,1)$ as $c \to -\infty$. Then we have

$$\begin{aligned}\lim_{c\to-\infty} E[((\ddot{c}-c)^2\int_0^1 P(r)^2dr] &= \lim_{c\to-\infty}\left[E\frac{(\int_0^1 P(r)dW(r) + \gamma_1\int_0^1(cr-1)P(r)dr)^2}{(\int_0^1 P(r)^2dr)^2}\int_0^1 P(r)^2dr\right] \\ &= \lim_{c\to-\infty} E[\frac{\int_0^1 P(r)dW(r)}{(\int_0^1 P(r)^2dr)^{1/2}}]^2 + o(1) = 1 \qquad \text{(A.12)}\end{aligned}$$

(e) Using the results stated in the foregoing parts, it follows $\lim_{c\to-\infty} E[\gamma_1(\ddot{c}-c)\int_0^1(cr-1)P(r)dr] = \lim_{c\to-\infty} E[O_p(|c|^{-1})O_p(|c|^{1/2})O_p(1)] = \lim_{c\to-\infty} O(|c|^{-1/2}) = 0$. ▲

Proof of Theorem 1: (a) The forecast error can be expressed as

$$\begin{aligned}\frac{T^{\frac{1}{2}}}{\sigma}\widehat{e}_{[rT]} &= \frac{T^{\frac{1}{2}}}{\sigma}(\widehat{\mu}_{[rT]} - \mu_{[rT]}) \\ &= \frac{ac}{\sigma}T^{-1/2}(\widehat{u}_{[rT]} - u_{[rT]}) + \frac{T^{1/2}}{\sigma}(\ddot{\beta}_1 - \beta_1) + \frac{T}{\sigma}(\ddot{\alpha}-\alpha)T^{-1/2}\widehat{u}_{[rT]} \\ &\quad + \sum_{i=1}^{k}\frac{T^{1/2}}{\sigma}\alpha_i(\Delta\widehat{u}_{[rT]-i} - \Delta u_{[rT]-i}) + \sum_{i=1}^{k}\frac{T^{1/2}}{\sigma}(\ddot{\alpha}_i - \alpha_i)\Delta\widehat{u}_{[rT]-i} \\ &= \frac{ac}{\sigma}T^{-1/2}(\widehat{u}_{[rT]} - u_{[rT]}) + a\frac{T^{1/2}}{\sigma}(\ddot{\beta}_1 - \beta_1) + \frac{T}{\sigma}(\ddot{\alpha}-\alpha)T^{-1/2}\widehat{u}_{[rT]} \\ &\quad + \sum_{i=1}^{k}\frac{T^{1/2}}{\sigma}(\ddot{\alpha}_i - \alpha_i)\Delta\widehat{u}_{[rT]-i} \\ &= A_{[rT]} + B_{[rT]} \qquad \text{(A.13)}\end{aligned}$$

since $\Delta\widehat{u}_{[rT]-i} - \Delta u_{[rT]-i} = \beta_1 - \ddot{\beta}_1$. We have

$$\begin{aligned}
A_{[rT]} &= \frac{ac}{\sigma}T^{-1/2}(\widehat{u}_{[rT]} - u_{[rT]}) + a\frac{T^{1/2}}{\sigma}(\ddot{\beta}_1 - \beta_1) + \frac{T}{\sigma}(\ddot{\alpha} - \alpha)T^{-1/2}\widehat{u}_{[rT]} \\
&\xrightarrow{d} ac(a^{-1}P(r) - a^{-1}J_c(r)) + aa^{-1}\gamma_1 + a(\ddot{c} - c)a^{-1}P(r) \\
&= c(P(r) - J_c(r)) + \gamma_1 + (\ddot{c} - c)P(r) \\
&= \gamma_1(1 - cr) + (\ddot{c} - c)P(r) \triangleq U_1(c, a, r) \\
B_{[rT]} &= \sum_{i=1}^{k} \frac{T^{1/2}}{\sigma}(\ddot{\alpha}_i - \alpha_i)\Delta\widehat{u}_{[rT]-i} = -R^{'}\hat{L}_{[rT]} \qquad \text{(A.14)}
\end{aligned}$$

It follows that

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} A_t^2 &= \int_0^1 U_1(c, a, r)^2 dr \\
\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} B_t^2 &= \lim_{T\to\infty} R^{'}(\frac{1}{T}\sum_{t=1}^{T}\hat{L}_t\hat{L}_t^{'})R \to R^{'}QR \sim \chi_k^2 \\
\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} A_t B_t &= 0 \qquad \text{(A.15)}
\end{aligned}$$

For in-sample AMSE:

$$\begin{aligned}
m_1(c, a, 1, k) &= \lim_{T\to\infty} \frac{1}{\sigma^2}E\sum_{t=1}^{T}(\widehat{\mu}_t - \mu_t)^2 \\
&= \lim_{T\to\infty} E\frac{1}{T}\sum_{t=1}^{T}(A_t^2 + B_t^2 + 2A_tB_t) \\
&= E\left[\int_0^1 U_1(c, a, r)^2 dr\right] + k \qquad \text{(A.16)}
\end{aligned}$$

(b) As $c \to -\infty$,

$$\begin{aligned}
&\lim_{c\to-\infty} E\left[\int_0^1 U_1(c, a, r)^2 dr\right] \\
&= \lim_{c\to-\infty} E\left[\int_0^1 \{\gamma_1(1 - cr) + (\ddot{c} - c)P(r)\}^2 dr\right] \\
&= \lim_{c\to-\infty} E[\gamma_1^2\int_0^1 (cr - 1)^2 dr] + E[(\ddot{c} - c)^2\int_0^1 P(r)^2 dr] + 2E[\gamma_1(\ddot{c} - c)\int_0^1 (cr - 1)P(r)dr] \\
&= 1 + 1 + 2\cdot 0 = 2 \quad \text{[By Lemma A.2]} \qquad \text{(A.17)}
\end{aligned}$$

Then, as $c \to -\infty$, $(A.16)$ equals $2 + k$.

(c) For MSFE:

$$\begin{aligned} f_1(c,a,1,k) &= \lim_{T\to\infty} \frac{T}{\sigma^2} E(\widehat{\mu}_{T+1} - \mu_{T+1})^2 \\ &= \lim_{T\to\infty} E(A_{T+1}^2 + B_{T+1}^2 + 2A_{T+1}B_{T+1}) \\ &= E[U_1(c,a,1)^2] + E(R^{'}(\hat{L}_{T+1}\hat{L}_{T+1}^{'})R) + 0 \\ &= E[U_1(c,a,1)^2] + k \end{aligned} \tag{A.18}$$

since $E(R^{'}(\hat{L}_{T+1}\hat{L}^{'}_{T+1})R) = tr\left[E(\hat{L}_{T+1}\hat{L}^{'}_{T+1}RR^{'})\right] = E(R^{'}QR) = k$. ▲

Proof of Corollary 1: (a) First, note that the expression for $V_1(c,r)$ is derived from Hansen (2010) by transforming vector stochastic integrals to explicit Brownian motion processes. Following Lemma A.1 and Theorem 1 we have the restricted FGLS estimator as $V_1^{gls}(c,r) = J_c(1) - cJ_c(r)$, the same as Hansen's (2010) restricted OLS estimator (note that $J_c(1) = c\int_0^1 J_c(r)dr + W(1)$). So we simply drop the superscript gls to save notation. Using the definition of $V_1(c,r)$, we have $m_0(c,1) = E[\int_0^1 V_1(c,r)^2 dr] + k$, $f_0(c,1) = E[V_1(c,1)^2] + k$. For the averaging estimator, it follows that

$$\begin{aligned} & m_w(c,a,1,k) \\ &= \lim_{T\to\infty} \frac{1}{\sigma^2} \sum_{t=1}^{T} E(\hat{\mu}_t(w) - \mu_t)^2 \\ &= \lim_{T\to\infty} \frac{1}{\sigma^2} \sum_{t=1}^{T} E(w\hat{\mu}_t + (1-w)\widetilde{\mu}_t - \mu_t)^2 \\ &= \lim_{T\to\infty} \frac{1}{\sigma^2} \left[w^2 E\left\{\sum_{t=1}^{T}(\hat{\mu}_t - \mu_t)^2\right\} + (1-w)^2 E\left\{\sum_{t=1}^{T}(\widetilde{\mu}_t - \mu_t)^2\right\} + 2w(1-w)E\left\{\sum_{t=1}^{T}(\hat{\mu}_t - \mu_t)(\widetilde{\mu}_t - \mu_t)\right\}\right] \\ &= w^2\left[E\int_0^1 U_1(c,a,r)^2 dr + k\right] + (1-w)^2\left[E\int_0^1 V_1(c,r)^2 dr + k\right] + 2w(1-w)\left[E\int_0^1 U_1(c,a,r)V_1(c,r)dr\right. \\ &= w^2 m_1(c,a,1) + (1-w)^2 m_0(c,1) + 2w(1-w)m_{01}(c,a,1) + k \end{aligned}$$

(b) The MSFE $f_w(c,a,1,k)$ of the averaging estimator can be derived in a manner similar to that in (a) and is hence omitted.

(c) We have

$$
\begin{aligned}
\lim_{c\to 0} m_{01}(c,a,1) &= \lim_{c\to 0} E\int_0^1 U_1(c,a,r)V_1(c,r)dr \\
&= \lim_{c\to 0} E\int_0^1 [\gamma_1(1-cr)+(\ddot{c}-c)P(r)][W(1)-c\bar{J}_c(r)]dr \\
&= \lim_{c\to 0} E[\gamma_1 W(1)] + \lim_{c\to 0} E\left[(\ddot{c}\int_0^1 P(r)dr)W(1)\right] \qquad (A.19)\\
&= 1+0=1
\end{aligned}
$$

The first term in $(A.19)$ is $E[E\{\gamma_1 W(1)|\widetilde{J}_0(.)\}] = E[(1-a\dot{c}+\frac{1}{3}(a\dot{c})^2)^{-1}\{(1-a\dot{c})E[W(1)^2|\widetilde{J}_0(.)]+(a\dot{c})^2/3\}] = 1$ since $E[W(1)|\widetilde{J}_0(.)] = 0$. That the second term is zero follows from the facts that $\ddot{c} = [2\int_0^1 P(r)^2dr]^{-1}[P(1)^2-1]$, $E[P(r)W(1)] = 0$, $r \in [0,1]$ and the law of iterated expectations. ▲

Proof of Corollary 2: The proof is straightforward following the proof of Corollary 1 and is hence omitted.

Proof of Theorem 2: We have

$$
\begin{aligned}
&\lim_{T\to\infty} E\frac{M_0(c,a,1,k)-T\sigma^2}{\sigma^2} \\
&= \lim_{T\to\infty} E[\frac{1}{\sigma^2}\sum_{t=1}^T(e_t^2-\sigma^2)+\frac{1}{\sigma^2}\sum_{t=1}^T(\widetilde{\mu}_t-\mu_t)^2+\frac{2\widehat{\sigma}^2}{\sigma^2}(m_{01}(c,a,1)+k)-\frac{2}{\sigma^2}\sum_{t=1}^T e_t(\widetilde{\mu}_t-\mu_t)] \\
&= 0+m_0(c,1)+k+2(m_{01}(c,a,1)+k)-\lim_{T\to\infty} E\frac{2}{\sigma^2}\sum_{t=1}^T e_t(\widetilde{\mu}_t-\mu_t) \qquad (A.20)
\end{aligned}
$$

The last term is -2 times

$$
\lim_{T\to\infty} E\frac{1}{\sigma^2}\sum_{t=1}^T e_t(\widetilde{\mu}_t-\mu_t) = E\int_0^1[-c\bar{J}_c(r)+W(1)]dW(r)+E\chi_k^2 \qquad (A.21)
$$

As $c \to 0$, we have $E\int_0^1[-c\bar{J}_c(r)+W(1)]dW(r) \to EW(1)^2 = 1$, so the last term amounts to -2 times $\lim_{c\to 0}[m_{01}(c,a,1)+k]$ so that the limit of $(A.20)$ is $\lim_{c\to 0}[m_0(c,1)+k]$.

For the unrestricted case,

$$
\begin{aligned}
&\lim_{T\to\infty} E\frac{M_1(c,a,1,k)-T\sigma^2}{\sigma^2} \\
&= \lim_{T\to\infty} E[\frac{1}{\sigma^2}\sum_{t=1}^T(e_t^2-\sigma^2)+\frac{1}{\sigma^2}\sum_{t=1}^T(\hat{\mu}_t-\mu_t)^2+\frac{2\widehat{\sigma}^2}{\sigma^2}(m_1(c,a,1)+k)-\frac{2}{\sigma^2}\sum_{t=1}^T e_t(\hat{\mu}_t-\mu_t)] \\
&= 0+m_1(c,a,1)+k+2(m_1(c,a,1)+k)-\lim_{T\to\infty} E\frac{2}{\sigma^2}\sum_{t=1}^T e_t(\hat{\mu}_t-\mu_t) \qquad (A.22)
\end{aligned}
$$

The last term is -2 times

$$\lim_{T\to\infty} E\frac{1}{\sigma^2}\sum_{t=1}^{T} e_t(\hat{\mu}_t-\mu_t) = E\int_0^1 [\gamma_1(1-cr)+(\ddot{c}-c)P(r)]dW(r) + E\chi_k^2 \tag{A.23}$$

Using Lemma A.2, we have

$$\lim_{c\to-\infty} E\int_0^1 \gamma_1(1-cr)dW(r) = \lim_{c\to-\infty} E[(1-a\dot{c}+\frac{1}{3}(a\dot{c})^2)^{-1}\int_0^1(1-a\dot{c}s)d\dot{W}(s)\int_0^1(1-cr)dW(r)]$$

$$= \lim_{c\to-\infty} E[(1-a\dot{c}+\frac{1}{3}(a\dot{c})^2)^{-1}(\int_0^1(1-a\dot{c}s)dW(s)\int_0^1(1-cr)dW(r)$$

$$+\int_0^1(1-a\dot{c}s)(c-a\dot{c})ds\int_0^1(1-cr)dW(r))] = 3a^{-2}(\frac{1}{3}a-\frac{1}{3}a(1-a)) = 1 \tag{A.24}$$

$$\lim_{c\to-\infty} E\int_0^1(\ddot{c}-c)P(r)dW(r) = \lim_{c\to-\infty} E[\frac{(\int_0^1 P(r)dW(r))^2}{\int_0^1 P(r)^2dr} + o_p(1)] = 1 \tag{A.25}$$

Substituting $(A.24)$ and $(A.25)$ in $(A.23)$ establishes that the limit of $(A.22)$ equals $\lim_{c\to-\infty}[m_1(c,a,1)+k]$. ▲

Proof of Theorem 3: To prove this result, we need to derive the explicit forms of m_{1K}^{ols} and m_{0K}^{ols}. We first consider the case where $l\le k$. Let $H_t=(\Delta y_{t-1},...,\Delta y_{t-K})'$ and for $i\le j$, $H_{[i,j],t}=(\Delta y_{t-i},...,\Delta y_{t-j})'$, $\alpha_{[i,j]}=(\alpha_i,...,\alpha_j)'$. Let $x_t=(t,y_{t-1})'$. Define the orthogonalized series $H_t^*[x_t^*]$ as the residuals from regressing $H_t[x_t]$ on a constant. Let

$$\Sigma = E(H_t^*H_t^{*'}) = \begin{bmatrix} \Sigma 11_{[l\times l]} & \Sigma 12_{[l\times(K-l)]} \\ \Sigma 21_{[(K-l)\times K]} & \Sigma 22_{[(K-l)\times(K-l)]} \end{bmatrix}_{[K\times K]}$$

For unrestricted estimation, we reformulate the regression as

$$\Delta y_t = \theta_0 + x_t^{*'}\theta_1 + H_{[1,l],t}^{*'}\alpha_{[1,l]} + \epsilon_t^* \tag{A.26}$$

where the effective error is $\epsilon_t^* = H_{[l+1,k],t}^{*'}\alpha_{[l+1,k]}+e_t$, with $H_{[i,j],t}^*$ defined analogously to $H_{[i,j],t}$. θ_0 and θ_1 are functions of the true parameters; specifically, $\theta_1=(-\beta_1(\alpha-1),\alpha-1)'$. From this regression, it follows that

$$\begin{aligned}
T^{1/2}(\breve{\alpha}_{[1,l]}-\alpha_{[1,l]}) &= (\frac{1}{T}\sum_{t=1}^{T}H_{[1,l],t}^*H_{[1,l],t}^{*'})^{-1}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H_{[1,l],t}^*\epsilon_t^*)+o_p(1) \\
&= (\frac{1}{T}\sum_{t=1}^{T}H_{[1,l],t}^*H_{[1,l],t}^{*'})^{-1}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H_{[1,l],t}^*e_t) \\
&\quad + (\frac{1}{T}\sum_{t=1}^{T}H_{[1,l],t}^*H_{[1,l],t}^{*'})^{-1}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H_{[1,l],t}^*H_{[l+1,k],t}^{*'}\alpha_{[l+1,k]})+o_p(1) \\
&\xrightarrow{d} N(0,\sigma^2\Sigma_{11}^{-1})+\Sigma_{11}^{-1}\Sigma_{12}\alpha_{[l+1,K]}
\end{aligned} \tag{A.27}$$

with $\alpha_{[l+1,K]} = (\alpha_{l+1}, ..., \alpha_k, \alpha_{k+1}, ..., \alpha_K)'$, where $(\alpha_{k+1}, ..., \alpha_K)' = (0, ..., 0)'$. We can write

$$\breve{\mu}_t(l) - \mu_t = (\breve{\theta}_0 - \theta_0) + x_t^{*'}(\breve{\theta}_1 - \theta_1) + H_{[1,l],t}^{*'}(\breve{\alpha}_{[1,l]} - \alpha_{[1,l]}) - H_{[l+1,k],t}^{*'}\alpha_{[l+1,k]} \qquad (A.28)$$

We now calculate the cross product of the misspecified unrestricted estimator with the estimator from the largest unrestricted model. Denoting $\breve{\theta}_0^K$, $\breve{\theta}_1^K$, $\breve{\alpha}_{[1,l]}^K$, and $\breve{\alpha}_{[l+1,K]}^K$ as the estimates from the largest model, we have:

$$\breve{\mu}_t(K) - \mu_t = (\breve{\theta}_0^K - \theta_0) + x_t^{*'}(\breve{\theta}_1^K - \theta_1) + H_{[1,l],t}^{*'}(\breve{\alpha}_{[1,l]}^K - \alpha_{[1,l]}) + H_{[l+1,K],t}^{*'}(\breve{\alpha}_{[l+1,K]}^K - \alpha_{[l+1,K]}) \quad (A.29)$$

Here $\sqrt{T}(\breve{\alpha}_{[1,K]}^K - \alpha_{[1,K]}) = \sqrt{T}(\breve{\alpha}_{[1,l]}^{K'} - \alpha_{[1,l]}', \breve{\alpha}_{[l+1,K]}^{K'} - \alpha_{[l+1,K]}')' \xrightarrow{d} N(0, \sigma^2\Sigma^{-1})$. Let $H_{[1,l]}^* = (H_{[1,l],l+2}^*, ..., H_{[1,l],T}^*)'$, $H^* = (H_{l+2}^*, ..., H_T^*)'$. For $m_{1K}^{ols}(c, \delta, 1, l)$, we calculate

$$\begin{aligned}
m_{1K}^{ols}(c, \delta, 1, l) &= E[\lim_{T\to\infty} \frac{1}{\sigma^2} \sum_{t=1}^{T} (\breve{\mu}_t(l) - \mu_t)(\breve{\mu}_t(K) - \mu_t)] \\
&= E[\lim_{T\to\infty} \frac{T}{\sigma^2} (\breve{\theta}_0 - \theta_0)'(\breve{\theta}_0^K - \theta_0) + \frac{1}{\sigma^2}(\breve{\theta}_1 - \theta_1)' \sum_{t=1}^{T} x_t^* x_t^{*'}(\breve{\theta}_1^K - \theta_1) \\
&\quad + \frac{1}{\sigma^2}(\breve{\alpha}_{[1,l]} - \alpha_{[1,l]})' \sum_{t=1}^{T} H_{[1,l],t}^* H_{[1,l],t}^{*'}(\breve{\alpha}_{[1,l]}^K - \alpha_{[1,l]}) \\
&\quad + \frac{1}{\sigma^2}(\breve{\alpha}_{[1,l]} - \alpha_{[1,l]})' \sum_{t=1}^{T} H_{[1,l],t}^* H_{[l+1,K],t}^{*'}(\breve{\alpha}_{[l+1,K]}^K - \alpha_{[l+1,K]}) \\
&\quad - \frac{1}{\sigma^2}\alpha_{[l+1,k]}' \sum_{t=1}^{T} H_{[l+1,k],t}^* H_{[1,l],t}^{*'}(\breve{\alpha}_{[1,l]}^K - \alpha_{[1,l]}) \\
&\quad - \frac{1}{\sigma^2}\alpha_{[l+1,k]}' \sum_{t=1}^{T} H_{[l+1,k],t}^* H_{[l+1,K],t}^{*'}(\breve{\alpha}_{[l+1,K]}^K - \alpha_{[l+1,K]}) + o_p(1)] \\
&= 1 + E(F_{1c}) + E \lim_{T\to\infty} [\frac{1}{\sigma^2}(\breve{\alpha}_{[1,l]} - \alpha_{[1,l]})' \sum_{t=1}^{T} H_{[1,l],t}^* H_t^{*'}(\breve{\alpha}_{[1,K]}^K - \alpha_{[1,K]}) \\
&\quad - \frac{1}{\sigma^2}\alpha_{[l+1,k]}' \sum_{t=1}^{T} H_{[l+1,k],t}^* H_t^{*'}(\breve{\alpha}_{[1,K]}^K - \alpha_{[1,K]})] \\
&= E(F_{1c}) + 1 + \lim_{T\to\infty} tr(H_{[1,l]}^*(H_{[1,l]}^{*'}H_{[1,l]}^*)^{-1}H_{[1,l]}^{*'}H^*(H^{*'}H^*)^{-1}H^{*'}) + 0 \\
&= E(F_{1c}) + 1 + l \qquad (A.30)
\end{aligned}$$

where $F_{1c} = \lim_{T\to\infty} \frac{1}{\sigma^2}(\breve{\theta}_1 - \theta_1)' \sum_{t=1}^{T} x_t^* x_t^{*'}(\breve{\theta}_1^K - \theta_1) = O_p(1)$ with $E(F_{1c}) \xrightarrow{p} 2$, as $c \to -\infty$.[see equations (15) and (36) in Hansen, 2010]. The last two equalities in $(A.30)$ hold since

$$
\begin{aligned}
&E \lim_{T\to\infty}[\frac{1}{\sigma^2}(\breve{\alpha}_{[1,l]}-\alpha_{[1,l]})' \sum_{t=1}^{T} H^*_{[1,l],t}H_t^{*'}(\breve{\alpha}^K_{[1,K]}-\alpha_{[1,K]}) - \frac{1}{\sigma^2}\alpha'_{[l+1,k]}\sum_{t=1}^{T} H^*_{[l+1,k],t}H_t^{*'}(\breve{\alpha}^K_{[1,K]}-\alpha_{[1,K]})] \\
&= E[\lim_{T\to\infty}\frac{1}{\sigma^2}(\sum_{t=1}^{T} H^*_{[1,l],t}e_t)'(\sum_{t=1}^{T} H^*_{[1,l],t}H^{*'}_{[1,l],t})^{-1}\sum_{t=1}^{T} H^*_{[1,l],t}H_t^{*'}(\sum_{t=1}^{T} H_t^*H_t^{*'})^{-1}\sum_{t=1}^{T} H_t^*e_t \\
&+ \lim_{T\to\infty}\frac{1}{\sigma^2}(\sum_{t=1}^{T} H^*_{[1,l],t}H^{*'}_{[l+1,k],t}\alpha_{[l+1,k]})'(\sum_{t=1}^{T} H^*_{[1,l],t}H^{*'}_{[1,l],t})^{-1}\sum_{t=1}^{T} H^*_{[1,l],t}H_t^{*'}(\sum_{t=1}^{T} H_t^*H_t^{*'})^{-1}\sum_{t=1}^{T} H_t^*e_t \\
&- \lim_{T\to\infty}\frac{1}{\sigma^2}\alpha_{[l+1,k]}\sum_{t=1}^{T} H^*_{[l+1,k],t}H_t^{*'}(\sum_{t=1}^{T} H_t^*H_t^{*'})^{-1}\sum_{t=1}^{T} H_t^*e_t + o_p(1)] \\
&= \lim_{T\to\infty} tr(H^*_{[1,l]}(H^{*'}_{[1,l]}H^*_{[1,l]})^{-1}H^{*'}_{[1,l]}H^*(H^{*'}H^*)^{-1}H^{*'}) + 0 - 0 \\
&= l \qquad\qquad \text{(A.31)}
\end{aligned}
$$

using the properties of a projection matrix. Hence $(A.30)$ reduces to $2+1+l$ as $c\to-\infty$.

For the restricted model, we can write

$$
\widetilde{\mu}_t(l)-\mu_t = (\widetilde{\theta}_0-\theta_0) + H^{*'}_{[1,l],t}(\widetilde{\alpha}_{[1,l]}-\alpha_{[1,l]}) - H^{*'}_{[l+1,k],t}\alpha_{[l+1,k]} - \frac{ac}{T}y^*_{t-1} \qquad \text{(A.32)}
$$

Then we calculate

$$
\begin{aligned}
m^{ols}_{0K}(c,\delta,1,l) &= E\lim_{T\to\infty}\frac{1}{\sigma^2}\sum_{t=1}^{T}(\widetilde{\mu}_t(l)-\mu_t)(\breve{\mu}_t(K)-\mu_t) \\
&= E\lim_{T\to\infty}[\frac{T}{\sigma^2}(\widetilde{\theta}_0-\theta_0)'(\breve{\theta}^K_0-\theta_0) - \frac{ac}{T\sigma^2}\sum_{t=1}^{T} y^*_{t-1}x_t^{*'}(\breve{\theta}^K_1-\theta_1) \\
&+ \frac{1}{\sigma^2}(\widetilde{\alpha}_{[1,l]}-\alpha_{[1,l]})'\sum_{t=1}^{T} H^*_{[1,l],t}H^{*'}_{[1,l],t}(\widetilde{\alpha}^K_{[1,l]}-\alpha_{[1,l]}) \\
&+ \frac{1}{\sigma^2}(\widetilde{\alpha}_{[1,l]}-\alpha_{[1,l]})'\sum_{t=1}^{T} H^*_{[1,l],t}H^{*'}_{[l+1,K],t}(\widetilde{\alpha}^K_{[1+1,K]}-\alpha_{[l+1,K]}) \\
&- \frac{1}{\sigma^2}\alpha'_{[l+1,k]}\sum_{t=1}^{T} H^*_{[l+1,k],t}H^{*'}_{[1,l],t}(\widetilde{\alpha}^K_{[1,l]}-\alpha_{[1,l]}) \\
&- \frac{1}{\sigma^2}\alpha'_{[l+1,k]}\sum_{t=1}^{T} H^*_{[l+1,k],t}H^{*'}_{[l+1,K],t}(\widetilde{\alpha}^K_{[l+1,K]}-\alpha_{[l+1,K]}) + o_p(1)] \\
&= E(F_{01c}) + 1 + l \qquad\qquad \text{(A.33)}
\end{aligned}
$$

where $F_{01c} = \lim_{T\to\infty} -\frac{ac}{T\sigma^2}\sum_{t=1}^{T} y^*_{t-1}x_t^{*'}(\breve{\theta}^K_1-\theta_1) = O_p(1)$. It follows $F_{01c} \xrightarrow{p} 0$ as $c\to 0$. So $(A.33)$ reduces to $1+l$ as $c\to 0$.

We next consider the case where $l > k$, and show the results for m_{1K}^{ols} and m_{0K}^{ols} remain the same. For unrestricted estimation, similar to $(A.26)$, we reformulate the regression as

$$\Delta y_t = \theta_0 + x_t^{*'}\theta_1 + H_{[1,l],t}^{*'}\alpha_{[1,l]} + \epsilon_t^* \tag{A.34}$$

The effective error is $\epsilon_t^* = -H_{[k+1,l],t}^{*'}\alpha_{[k+1,l]} + e_t$, where $\alpha_{[1,l]} = (\alpha_1, ..., , \alpha_k, ...\alpha_l)'$ are the parameters corresponding to the selected lags and $\alpha_{[k+1,l]} = (\alpha_{k+1}, ..., \alpha_l)'$ are the parameters corresponding to the over-specified lags. Note that the true parameters $\alpha_{[k+1,l]} = (\alpha_{k+1}, ..., \alpha_l)' = (0, ..., 0)'$. In this regression, it follows that $T^{1/2}(\breve{\alpha}_{[1,l]} - \alpha_{[1,l]}) \to N(0, \sigma^2\Sigma_{11}^{-1})$, which is different from $(A.27)$. Nevertheless, the subsequent calculations are exactly the same as in $(A.30 - A.31)$, so the result remains the same. The same conclusion applies to the restricted counterpart.

Now we prove the unbiasedness property. We elaborate on the steps to prove the result for the case $l \leq k$, with similar steps applicable to the case $l > k$ with the same conclusion. Firstly, for the restricted case

$$\begin{aligned}
&E \lim_{T\to\infty} \frac{M_0^{ols}(c, \delta, 1, l) - T\sigma^2}{\sigma^2} \\
&= E \lim_{T\to\infty} [\frac{1}{\sigma^2}\sum_{t=1}^{T}(e_t^2 - \sigma^2) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(\widetilde{\mu}_t(l) - \mu_t)^2 + \frac{2\breve{\sigma}_K^2}{\sigma^2}(m_{0K}^{ols}(c, \delta, 1, l)) - \frac{2}{\sigma^2}\sum_{t=1}^{T} e_t(\widetilde{\mu}_t(l) - \mu_t)] \\
&= 0 + m_0^{ols}(c, \delta, 1, l) + 2m_{0K}^{ols}(c, \delta, 1, l) - E \lim_{T\to\infty} \frac{2}{\sigma^2}\sum_{t=1}^{T} e_t(\widetilde{\mu}_t(l) - \mu_t) \qquad \text{(A.35)}
\end{aligned}$$

The last term is -2 times

$$\begin{aligned}
&E \lim_{T\to\infty} \frac{1}{\sigma^2}\sum_{t=1}^{T} e_t(\widetilde{\mu}_t(l) - \mu_t) \\
&= E \lim_{T\to\infty} [\frac{1}{\sigma^2}\sum_{t=1}^{T} e_t(\widetilde{\theta}_0 - \theta_0) - \frac{ac}{T\sigma^2}\sum_{t=1}^{T} e_t y_{t-1}^* \\
&\quad + \frac{1}{\sigma^2}\sum_{t=1}^{T} e_t H_{[1,l],t}^{*'}(\widetilde{\alpha}_{[1,l]} - \alpha_{[1,l]}) - \frac{1}{\sigma^2}\sum_{t=1}^{T} e_t H_{[l+1,k],t}^{*'}\alpha_{[l+1,k]}] \\
&= 1 + E(F_{01c}) + E \lim_{T\to\infty} \frac{1}{\sigma^2}[\sum_{t=1}^{T} e_t H_{[1,l],t}^{*'}(\frac{1}{T}\sum_{t=1}^{T} H_{[1,l],t}^{*} H_{[1,l],t}^{*'})^{-1}(\frac{1}{T}H_{[1,l],t}^{*}e_t)] + 0 \\
&= 1 + E(F_{01c}) + l \qquad \text{(A.36)}
\end{aligned}$$

which is $m_{0K}^{ols}(c, \delta, 1, l)$. Note that here we have $-\frac{ac}{T\sigma^2}\sum_{t=1}^{T} e_t y_{t-1}^* \xrightarrow{d} F_{01c}$. To show $-\frac{ac}{T\sigma^2}\sum_{t=1}^{T} e_t y_{t-1}^*$ and $-\frac{ac}{T\sigma^2}\sum_{t=1}^{T} y_{t-1}^* x_t^{*'}(\breve{\theta}_1^K - \theta_1)$ follow the same limit F_{01c}, notice that $y_{t-1}^* = Sx_t^*$, where

$S = [0, 1]$. We have

$$\lim_{T\to\infty} -\frac{ac}{T\sigma^2}\sum_{t=1}^{T} y^*_{t-1}x^{*'}_t(\breve{\theta}^K_1 - \theta_1) - (-\frac{ac}{T\sigma^2}\sum_{t=1}^{T} e_t y^*_{t-1})$$
$$= \lim_{T\to\infty}\frac{ac}{T\sigma^2}[\sum_{t=1}^{T} Sx^*_t x^{*'}_t \sum_{t=1}^{T}(x^*_t x^{*'}_t)^{-1}\sum_{t=1}^{T} x^*_t e_t - \sum_{t=1}^{T} Sx^*_t e_t]$$
$$= \lim_{T\to\infty}\frac{ac}{T\sigma^2}[S\sum_{t=1}^{T} x^*_t e_t - S\sum_{t=1}^{T} x^*_t e_t] = 0 \qquad \text{(A.37)}$$

Then, adding the terms in $(A.35)$ yields the final result $m^{ols}_0(c, \delta, 1, l)$.

For the unrestricted case,

$$E\lim_{T\to\infty}\frac{M^{ols}_1(c, \delta, 1, l) - T\sigma^2}{\sigma^2}$$
$$= E\lim_{T\to\infty}[\frac{1}{\sigma^2}\sum_{t=1}^{T}(e^2_t - \sigma^2) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(\breve{\mu}_t(l) - \mu_t)^2 + \frac{2\breve{\sigma}^2_K}{\sigma^2}(m^{ols}_{1K}(c, \delta, 1, l)) - \frac{2}{\sigma^2}\sum_{t=1}^{T} e_t(\breve{\mu}_t(l) - \mu_t)]$$
$$= 0 + m^{ols}_1(c, \delta, 1, l) + 2m^{ols}_{1K}(c, \delta, 1, l) - E\lim_{T\to\infty}\frac{2}{\sigma^2}\sum_{t=1}^{T} e_t(\breve{\mu}_t(l) - \mu_t) \qquad \text{(A.38)}$$

The last term is -2 times

$$E\lim_{T\to\infty}\frac{1}{\sigma^2}\sum_{t=1}^{T} e_t(\breve{\mu}_t(l) - \mu_t)$$
$$= E\lim_{T\to\infty}[\frac{1}{\sigma^2}\sum_{t=1}^{T} e_t(\breve{\theta}_0 - \theta_0) + \frac{1}{\sigma^2}\sum_{t=1}^{T} e_t x^{*'}_t(\breve{\theta}_1 - \theta_1)$$
$$+ \frac{1}{\sigma^2}\sum_{t=1}^{T} e_t H^{*'}_{[1,l],t}(\breve{\alpha}_{[1,l]} - \alpha_{[1,l]}) - \frac{1}{\sigma^2}\sum_{t=1}^{T} e_t H^{*'}_{[l+1,k],t}\alpha_{[l+1,k]}]$$
$$= 1 + E(F_{1c}) + l \qquad \text{(A.39)}$$

which is $m^{ols}_{1K}(c, \delta, 1, l)$. Hence, adding the terms in $(A.38)$, we obtain the final result $m^{ols}_1(c, \delta, 1, l)$.▲

Proof of Theorem 4: This result is proved in Theorem 3; see $(A.30$-$A.33)$.

Proof of Theorem 5: We follow the steps as in the proof of Theorem 3. First we derive the explicit forms of m^{gls}_{1K} and m^{gls}_{0K}. For $l < k$ (the misspecified case), Lemma A.1 (a)-(d) still holds, and compared to (e) of Lemma A.1 now we have

$$T^{1/2}(\ddot{\alpha}_{[1,l]} - \alpha_{[1,l]}) \xrightarrow{d} N(0, \sigma^2 Q^{-1}_{11}) + Q^{-1}_{11}Q_{12}\alpha_{[l+1,K]} \qquad \text{(A.40)}$$

where

$$Q = E(L_t L_t^{'}) = \begin{bmatrix} Q_{11_{[l \times l]}} & Q_{12_{[l \times (K-l)]}} \\ Q_{21_{[(K-l) \times K]}} & Q_{22_{[(K-l) \times (K-l)]}} \end{bmatrix}$$

For any $i \leq j$, define $L_{[i,j],t} = (\Delta u_{t-i}, ..., \Delta u_{t-j})^{'}$. Following the steps in proving Theorem 1, the forecast error from the misspecified FGLS model can be expressed as

$$\begin{aligned} \frac{T^{\frac{1}{2}}}{\sigma} \widehat{e}_{[rT]} &= \frac{T^{\frac{1}{2}}}{\sigma} (\widehat{\mu}_{[rT]} - \mu_{[rT]}) \\ &= A_{[rT]} + \dot{B}_{[rT]} + C_{[rT]} \end{aligned} \quad \text{(A.41)}$$

where $A_{[rT]}$ is defined as in $(A.14)$ and

$$\begin{aligned} \dot{B}_{[rT]} &= \sum_{i=1}^{l} \frac{T^{1/2}}{\sigma} (\ddot{\alpha}_i - \alpha_i) \Delta \widehat{u}_{[rT]-i} = \frac{T^{1/2}}{\sigma} (\ddot{\alpha}_{[1,l]} - \alpha_{[1,l]})^{'} \hat{L}_{[1,l],[rT]} \\ C_{[rT]} &= - \sum_{i=l+1}^{K} \frac{T^{1/2}}{\sigma} \alpha_i \Delta u_{[rT]-i} = - \frac{T^{1/2}}{\sigma} \alpha_{[l+1,K]}^{'} L_{[l+1,K],[rT]} \end{aligned} \quad \text{(A.42)}$$

Following the results of Theorem 1, Corollary 1 and proof of Theorem 3, the cross products $m_{1K}^{gls}(c, a, \delta, 1, l)$, $m_{0K}^{gls}(c, a, \delta, 1, l)$ can be easily derived:

$$\begin{aligned} \lim_{c \to -\infty} m_{1K}^{gls}(c, a, \delta, 1, l) &= 1 + 1 + l \\ \lim_{c \to 0} m_{0K}^{gls}(c, a, \delta, 1, l) &= 1 + l \end{aligned} \quad \text{(A.43)}$$

which also hold for $l > k$. The subsequent unbiasedness property can be established in a manner similar to the proof of Theorem 3 and is hence omitted.▲

Proof of Theorem 6: This result is proved in Theorem 5.