Econometrics Journal (2002), volume **5**, pp. 196–224.

Forecasting autoregressive time series in the presence of deterministic components

SERENA NG[†] AND TIMOTHY J. VOGELSANG[‡]

[†]*Department of Economics, Johns Hopkins University, 3400 N. Charles St. Baltimore, MD 21218, USA*
E-mail: serena.ng@jhu.edu

[‡]*Department of Economics, Cornell University Uris Hall, Ithaca, NY 14853-7601, USA*
E-mail: tjv2@cornell.edu

Received: February 2002

Summary This paper studies the error in forecasting an autoregressive process with a deterministic component. We show that when the data are strongly serially correlated, forecasts based on a model that detrends the data using OLS before estimating the autoregressive parameters are much less precise than those based on an autoregression that includes the deterministic components, and the asymptotic distribution of the forecast errors under the two-step procedure exhibits bimodality. We explore the conditions under which feasible GLS trend estimation can lead to forecast error reduction. The finite sample properties of OLS and feasible GLS forecasts are compared with forecasts based on unit root pretesting. The procedures are applied to 15 macroeconomic time series to obtain real time forecasts. Forecasts based on feasible GLS trend estimation tend to be more efficient than forecasts based on OLS trend estimation. A new finding is when a unit root pretest rejects non-stationarity, use of GLS yields smaller forecast errors than OLS. When the series to be forecasted is highly persistent, GLS trend estimation in conjunction with unit root pretests can lead to sharp reduction in forecast errors.

Keywords: *Forecasting, Trends, Unit root, GLS estimation.*

1. INTRODUCTION

An important use of econometric modeling is generating forecasts. If one is interested in forecasting a single economic time series, the starting point is often autoregressive models. Alternatively, one could base forecasts on structural models that incorporate economic theory. The usefulness of structural models is often measured by forecast precision compared to those of autoregressive models. Given the many uses of forecasts from autoregressive models, it seems sensible to construct these forecasts using the best methodology possible. We show in this paper that the way in which deterministic components (mean, trend) are estimated matters in important ways when the data are strongly serially correlated. In particular, we show that use of GLS detrending can improve forecasts compared to OLS when the errors are highly persistent.

Consider data generated by the trend plus noise model:

$$y_t = m_t + u_t, \tag{1}$$

$$u_t = \alpha u_{t-1} + e_t, \tag{2}$$

$$m_t = \delta_0 + \delta_1 t + \cdots + \delta_p t^p \tag{3}$$

where $e_t \sim i.i.d.(0, \sigma_e^2)$, and p is the order of the polynomial in time. We restrict attention to the empirically relevant cases of $p = 0$ and $p = 1$. We focus on AR(1) errors for the sake of simplicity as it enables derivation of asymptotic results and more importantly, it provides intuition as to how the trend estimates affect forecasts.[1]

An appealing feature of the trend plus noise model is that the unconditional mean of the series does not depend on the dynamic parameter α. In contrast, the unconditional mean of a process generated by $y_t = c + \alpha y_{t-1} + e_t$ depends on α. In particular, for $|\alpha| < 1$, $E(y_t) = c/(1 - \alpha)$, whereas for $\alpha = 1$, $E(y_t) = E(y_0) + ct$. Because we consider both stationary and unit root errors, it is important for clarity that the unconditional mean parameters, i.e. trend parameters, do not depend on α. In addition the trend plus noise model has been analyzed by Elliott *et al.* (1996), Canjels and Watson (1997), Phillips and Lee (1996) and Vogelsang (1998) to study the role of trend parameter estimation in inference. Some results from these papers are directly useful for analyzing the behavior of forecasts.

Assuming a quadratic loss function, the minimum mean squared error of the h-step ahead forecast of y_t, conditional upon lags of y_t, is

$$y_{t+h|t} = m_{t+h} + \alpha^h (y_t - m_t). \tag{4}$$

If we write the DGP as

$$y_t = \beta_0 + \beta_1 t + \alpha y_{t-1} + e_t, \tag{5}$$

where $\beta_0 = (1 - \alpha)\delta_0 + \alpha\delta_1$, $\beta_1 = (1 - \alpha)\delta_1$, the h step ahead forecast is

$$y_{t+h|t} = \sum_{i=0}^{h-1} \alpha^i (\beta_0 + \beta_1 (t + h - i)) + \alpha^h y_t. \tag{6}$$

If we know α and $\delta = (\delta_0, \delta_1)'$, the two parameterizations give exactly the same forecasts since one model can be reparameterized as the other exactly. However, α and δ are population parameters which we do not observe. In practice, we have three choices. First, we can jointly estimate the parameters by quasi-maximum likelihood. Second, we can first obtain estimates of δ by OLS or GLS, detrend the data, then estimate α by OLS, and ultimately use (4) to generate forecasts. Third, we can estimate β_0, β_1, and α simultaneously from (5) by OLS and then use (6) to make forecasts. In this paper, we focus on the latter two least squares method. We refer to these procedures as one-step and two-step procedures respectively.[2]

A quick review of textbooks reveals that, although (1) and (2) are often used instead of (5) to present the theory of optimal prediction,[3] the practical recommendation is not unanimous.

[1] We would expect to obtain similar results for more general ARMA models. Clearly, generalization of our results to ARMA models is worth considering in future work, though economic forecasting exercises tend to favor simple, low order, autoregressive models, see Stock and Watson (1998).

[2] Maximum likelihood and feasible GLS are equally efficient asymptotically in the strictly stationary framework. Estimation by GLS will be considered below.

[3] See, for example, Hamilton (1994, p. 81) and Box *et al.* (1994, p. 157). An exception is Clements and Hendry (1994).

For example, Pindyck and Rubinfeld (1998, p. 565) and Johnston and DiNardo (1997, p. 192, 232) used (1) and (2) to motivate the theory, but the examples appear to be based upon (5) (e.g. see Table 7.15 of Johnston and DiNardo (1997)). Examples considered in Diebold (1997), on the other hand, are based on an estimated trend function with a correction for serial correlation in the noise component (e.g. p. 231). This is consistent with assuming (4) as the forecasting model.

This paper is motivated by the fact that while $y_{t+h|t}$ is unique, its feasible counterpart is not. Depending on the treatment of the deterministic terms, the mean-squared forecast errors can be different. Efficient estimation of the trend coefficients and of α have separately received a great deal of attention in the literature.[4] The theme of this paper is that when the objective of the exercise is forecasting, estimation of these parameters can no longer be considered in isolation. This is especially important when the data are highly persistent.[5]

In this paper, we focus on two issues. Do the one- and two-step OLS forecasts differ in ways that should matter to practitioners? Does efficient estimation of the deterministic components improve forecasts? The answers to both of these questions are yes. Our results provide three useful recommendations for practitioners. First, if OLS is used to construct feasible forecasts, then one-step OLS usually leads to better forecasts than two-step OLS. Second, GLS estimation of the deterministic trend function, especially Prais–Winsten GLS, usually improves forecasts over OLS. Third, following unit root pretests (which are useful with highly persistent series) GLS forecasts should be used when the unit root null is rejected. Specifically, forecasts based on what is referred to as the PW_1 procedure below yield gains over the OLS procedures both in simulations and in empirical applications, and it is easy to implement.

The remainder of the paper is organized as follows. Theoretical and empirical properties of the forecast errors under least squares and GLS detrending are presented in Sections 2 and 3. In Section 4, we compare the forecasting procedures as applied to some common macroeconomic time series. Proofs are given in an appendix. We begin in the next section with forecasts based on OLS estimation of the trend function.

2. FORECASTS UNDER LEAST SQUARES ESTIMATION

Throughout our analysis, we assume that the data are generated by (1). Given $\{y_t\}_{t=1}^T$, we consider the one-step ahead forecast error given information at time T,

$$\begin{aligned} e_{T+1|T} &= y_{T+1} - \widehat{y}_{T+1|T} \\ &= y_{T+1} - y_{T+1|T} + y_{T+1|T} - \widehat{y}_{T+1|T} \\ &= e_{T+1} + \widehat{e}_{T+1|T}, \end{aligned}$$

where $e_{T+1} = y_{T+1} - y_{T+1|T}$ and $\widehat{e}_{T+1|T} = y_{T+1|T} - \widehat{y}_{T+1|T}$. The innovation, e_{T+1}, is unforecastable given information at time T and is beyond the control of the forecaster. A forecast is best in a mean-squared sense if $\widehat{y}_{T+1|T}$ is made as close to $y_{T+1|T}$ as measured by mean squared error. Throughout, we refer to $\widehat{e}_{T+1|T}$ as the forecast error.

[4] See Canjels and Watson (1997) and Vogelsang (1998) for inference on $\widehat{\delta}_1$ when u_t is highly persistent.

[5] Stock (1995, 1996, 1997) considers forecasting time series with a large autoregressive root in the absence of a deterministic component. Diebold and Kilian (2000) also consider forecasting highly persistent time series with a simple linear deterministic trend function, but they focus on the effects of unit root pretests rather than trend function estimation.

We consider two strategies, labelled OLS_1 and OLS_2 hereafter:

(1) OLS_1: Estimate (5) by OLS directly and then use (6) for forecasting.
(2) OLS_2: Estimate δ from (1) by OLS to obtain $\widehat{u}_t = y_t - \widehat{m}_t$. Then obtain $\widehat{\alpha}$ from (2) by OLS with u_t replaced by $\widehat{u}_t$. Forecasts are obtained from (4).

2.1. Finite sample properties

We first consider the finite sample properties of the forecast errors using Monte-Carlo experiments. Focusing on AR(1) processes, we generate data according to (1) and (2) for $\alpha = -0.4$, 0, 0.4, 0.8, 0.9, 0.95, 0.975, 0.99, 1.0, 1.01. We set $\delta_0 = \delta_1 = 0$ without loss of generality. The choice of the parameter set reflects the fact that many macro-economic time series are highly and positively autocorrelated. The errors are $N(0, 1)$ generated using the rndn() function in Gauss V3.27 with seed = 99. For this section, we assume that $u_1 = 0$. Results reported in Section 3.1 suggest that the rankings of OLS_1 and OLS_2 do not depend critically on this assumption on u_1.

We use $T = 100$ in the estimations to obtain up to $h = 10$ steps ahead forecasts. We use 10 000 replications to obtain the forecast errors $\widehat{e}_{T+h|T} = y_{T+h|T} - \widehat{y}_{T+h|T}$, and then evaluate the root mean-squared error (RMSE), $\sqrt{E(\widehat{e}^2_{T+h|T})}$, and the mean absolute error (MAE), $E(|\widehat{e}_{T+h|T}|)$. The MAE and the RMSE provide qualitatively similar information and only the RMSE will be reported.

Table 1(a) reports results for $h = 1$ (one step ahead forecasts). As benchmarks, we consider two infeasible forecasts: (i) OLS_2^{α} which assumes α is known and (ii) OLS_2^{δ} which assumes δ is known. From these, we see that when $p = 0$, the error in estimating α dominates the error in estimating δ_0. But when $p = 1$, the error in $\widehat{\delta}$ dominates unless $\alpha \geq 1.0$. The RMSE for OLS_2 is smaller than the sum of OLS_2^{α} and OLS_2^{δ}. The RMSE at $h = 10$ confirms that the error from estimating α vanishes when α is far away from one but increases (approximately) linearly with the forecast horizon when α is close to unity. However, the error in estimating δ does not vanish with the forecast horizon even when $\alpha = 0$ as the RMSE for OLS_2^{α} shows.

The RMSE for OLS_1 and OLS_2 when both parameters have to be estimated are quite similar when $\alpha < 1.0$ for $p = 0$ and when $\alpha < 0.8$ for $p = 1$. These similarities end as the error process becomes more persistent.[6] When $p = 0$, OLS_2 exhibits a sudden increase in RMSE and is sharply inferior to OLS_1 at $\alpha = 1$. When $p = 1$ the RMSE for OLS_1 is always smaller than OLS_2 when $\alpha \geq 0.8$. The difference is sometimes as large as 20% when α is close to unity. Results for $h = 10$ in Table 1(b) show a sharper contrast in the two sets of forecast errors. For a given procedure, the forecast errors are much larger when $p = 1$.

The finite sample simulations show that the method of trend estimation is by and large irrelevant when α is small. However, when $p = 0$, the forecast errors for OLS_1 exhibit a sharp decrease at $\alpha = 1$, while those of OLS_2 show a sharp increase. When $p = 1$, one step least squares also clearly dominates two steps least squares in terms of RMSE when the data are persistent. Thus, for empirically relevant cases when α is large, the method of trend estimation matters for forecasting. In the next subsection, we report an asymptotic analysis which provides some theoretical explanations for these simulation results.

[6] Sampson (1991) showed that under least squares trend estimation, the deterministic terms have a higher order effect on the forecast errors when u_t is non-stationary.

Table 1. (a) RMSE of OLS forecast errors: $T = 100$, $h = 1$.

α	OLS_1	OLS_2	OLS_2^α	OLS_2^δ	OLS_1	OLS_2	OLS_2^α	OLS_2^δ
	$p = 0$				$p = 1$			
−0.400	0.142	0.141	0.100	0.100	0.228	0.225	0.205	0.100
0.000	0.143	0.141	0.099	0.101	0.228	0.225	0.203	0.101
0.400	0.144	0.143	0.099	0.101	0.230	0.228	0.202	0.101
0.800	0.153	0.152	0.096	0.104	0.242	0.249	0.200	0.104
0.900	0.163	0.162	0.092	0.108	0.253	0.270	0.196	0.108
0.950	0.175	0.173	0.084	0.114	0.263	0.292	0.186	0.114
0.975	0.183	0.179	0.068	0.122	0.264	0.310	0.169	0.122
0.990	0.180	0.180	0.041	0.131	0.257	0.319	0.146	0.131
1.000	0.174	0.196	0.000	0.142	0.244	0.314	0.109	0.142
1.010	0.191	0.292	0.087	0.161	0.220	0.297	0.036	0.161

(b) RMSE of OLS forecast errors: $T = 100$, $h = 10$.

α	OLS_1	OLS_2	OLS_2^α	OLS_2^δ	OLS_1	OLS_2	OLS_2^α	OLS_2^δ
	$p = 0$				$p = 1$			
−0.400	0.071	0.071	0.071	0.001	0.167	0.167	0.167	0.001
0.000	0.100	0.099	0.099	0.000	0.233	0.231	0.231	0.000
0.400	0.166	0.165	0.165	0.001	0.384	0.379	0.379	0.001
0.800	0.471	0.450	0.429	0.159	1.008	0.984	0.972	0.159
0.900	0.767	0.720	0.601	0.410	1.487	1.466	1.365	0.410
0.950	1.054	0.976	0.675	0.669	1.843	1.874	1.569	0.669
0.975	1.244	1.134	0.613	0.897	1.994	2.141	1.572	0.897
0.990	1.305	1.204	0.394	1.123	1.986	2.284	1.427	1.123
1.000	1.315	1.436	0.000	1.345	1.830	2.217	1.090	1.345
1.010	1.715	2.684	0.911	1.681	1.507	1.983	0.364	1.681

Note: OLS_1 and OLS_2 are forecasts based on (6) and (4) respectively. OLS_2^α refers to OLS_2 with α assumed known. OLS_2^δ refers to OLS_2 with δ assumed known.

2.2. Asymptotic properties of OLS forecasts

The goal of this section is to provide an asymptotic approximation for the sampling behavior of the forecast errors, and to use the asymptotic representations to provide theoretical explanations for why OLS_1 and OLS_2 give different forecasts. Because the difference between OLS_1 and OLS_2 occurs for $0.8 \le \alpha \le 1.01$, a useful asymptotic approximation is obtained by using a local-to-unity framework with non-centrality parameter c to characterize the DGP as:

$$\begin{aligned} y_t &= \delta_0 + \delta_1 t + u_t, \\ u_t &= \alpha_T u_{t-1} + e_t, \\ \alpha_T &= 1 + \frac{c}{T}, \end{aligned}$$

$$u_1 = \sum_{i=0}^{[\kappa T]} \alpha_T^i e_{1-i}, \tag{7}$$

where e_t is a martingale difference sequence with $2+d$ moments for some $d > 0$ and $E(e_t^2) = 1$, $[x]$ is the integer part of x, and $\kappa \in [0, 1]$. It can be shown that the forecast errors depend on $\widehat{\delta} - \delta$, which can be expressed in terms of stochastic processes that are invariant to δ (see the appendix). Because of this invariance to δ, without loss of generality we can set $\delta_0 = \delta_1 = 0$. For a given sample size, u_t is locally stationary when $c < 0$ and locally explosive when $c > 0$, but becomes an integrated process as the sample size increases to infinity.

The model of the initial condition follows Canjels and Watson (1997) and allows several relevant special cases. Using backward substitution we have

$$u_t = \sum_{i=0}^{t-2} \alpha_T^i e_{t-i} + \alpha_T^{t-1} u_1 = \sum_{i=0}^{t-2} \alpha_T^i e_{t-i} + \alpha_T^{t-1} \sum_{i=0}^{[\kappa T]} \alpha_T^i e_{1-i}.$$

Let $\Rightarrow$ denote weak convergence. Given the assumptions on e_t, a functional central limit theorem applies to both terms of u_t:

$$T^{-1/2} \sum_{i=0}^{[rT]-2} \alpha_T^i e_{t-i} \Rightarrow J_c(r), \qquad T^{-1/2} \sum_{i=0}^{[\kappa T]} \alpha_T^i e_{1-i} \Rightarrow J_c^-(\kappa),$$

where $J_c(r)$ and $J_c^-(r)$ are given by $dJ_c(r) = cJ_c(r) + dW(r)$ and $dJ_c^-(r) = cJ_c^-(r) + dW^-(r)$ respectively, with $W(r)$ and $W^-(r)$ being independent standard Wiener processes. Because $\lim_{T\to\infty} \alpha_T^{[rT]} = \lim_{T\to\infty}(1 + c/T)^{[rT]} = \exp(rc)$, it directly follows that

$$T^{-1/2} u_{[rT]} \Rightarrow J_c(r) + \exp(rc) J_c^-(\kappa) \equiv J_c^*(r).$$

When $\kappa = 0$, $u_1 = e_1$ and $T^{-1/2}u_1 = o_p(1)$. We adopt the standard convention that $J_c^-(0) = 0$ in which case $T^{-1/2}u_{[rT]} \Rightarrow J_c(r)$. Therefore, the $\kappa = 0$ case is asymptotically equivalent to setting $u_1 = 0$. When $\kappa > 0$, $T^{-1/2}u_1 \Rightarrow J_c^-(\kappa) = O_p(1)$. In this case, the initial condition is modeled as evolving from past unobserved observations going back $[\kappa T]$ time periods. Because we are modeling the errors as nearly $I(1)$, $u_1 = O_p(T^{1/2})$.

The asymptotic behavior of the forecast errors depends on demeaned and detrended variants of $J_c^*(r)$ which are the limits of OLS residuals, $\widehat{u}_t$, obtained from the regressions of u_t on 1 when $p = 0$, and on 1 and t when $p = 1$. Specifically we have

$$p = 0: \quad T^{-1/2}\widehat{u}_{[rT]} \Rightarrow \bar{J}_c^*(r) = J_c^*(r) - \int_0^1 J_c^*(s)ds,$$

$$p = 1: \quad T^{-1/2}\widehat{u}_{[rT]} \Rightarrow \widetilde{J}_c^*(r) = J_c^*(r) - (4 - 6r)\int_0^1 J_c^*(s)ds - (12r - 6)\int_0^1 sJ_c^*(s)ds.$$

If a detrending procedure can remove m_t without leaving asymptotic effects on the data, $\bar{J}_c^*(r)$ and $\widetilde{J}_c^*(r)$ would have been identically $J_c^*(r)$. Least squares detrending evidently leaves non-vanishing effects on the detrended data when errors are highly persistent. It is this observation that suggests GLS detrending can lead to more precise forecasts. Not surprisingly, the asymptotic distributions of the forecast errors all depend on the limiting distribution of $T(\widehat{\alpha} - \alpha)$. Let $\widehat{c} =$

$\lim_{T\to\infty} T(\widehat{\alpha}-1)$. Then, it directly follows that $T(\widehat{\alpha}-\alpha) = T(\widehat{\alpha}-1-c/T) \Rightarrow (\widehat{c}-c)$. When α is estimated by least squares, we more specifically have:

$$\widehat{c}-c = \frac{\int_0^1 B_c(r)dW(r)}{\int_0^1 B_c(r)^2 dr} \equiv \Phi(B_c(r), W)$$

where $B_c(r) = \bar{J}_c^*(r)$ when $p = 0$, and $B_c(r) = \widetilde{J}(r)_c^*$ when $p = 1$.

We use two theorems to summarize the asymptotic results.

Theorem 2.1. (OLS_1) *Let the data be generated by* (7). *Let p be the order of the polynomial in time. Let* $\widehat{e}_{T+1|T}$ *be obtained by one-step OLS estimation of* β *and* α.

- *For* $p = 0$, $\quad T^{1/2}\widehat{e}_{T+1|T} \Rightarrow W(1) - (\widehat{c}-c)\bar{J}_c^*(1)$.
- *For* $p = 1$, $\quad T^{1/2}\widehat{e}_{T+1|T} \Rightarrow \int_0^1 (2-6r)dW(r) - (\widehat{c}-c)\widetilde{J}_c^*(1)$.

Theorem 2.2. (OLS_2) *Let the data be generated by* (7). *Let p be the order of the polynomial in time. Let* $\widehat{e}_{T+1|T}$ *be obtained by two steps OLS estimation of* δ *and then* α.

- *For* $p = 0$, $\quad T^{1/2}\widehat{e}_{T+1|T} \Rightarrow c[J_c^*(1) - \bar{J}_c^*(1)] - (\widehat{c}-c)\bar{J}_c^*(1)$.
- *For* $p = 1$, $\quad T^{1/2}\widehat{e}_{T+1|T} \Rightarrow c[J_c^*(1) - \widetilde{J}_c^*(1)] + [\int_0^1 (6-12r)J_c^*(r)dr] - (\widehat{c}-c)\widetilde{J}_c^*(1)$.

The main implication of the two theorems is that forecasts based upon OLS_1 and OLS_2 are not asymptotically equivalent.[7] The difference is not due to estimation of α because OLS_1 and OLS_2 share the same dependence asymptotically on $\widehat{c}$. The difference between OLS_1 and OLS_2 arises from the differences in the estimation of the trend parameters. An interesting property of OLS_1 is that the sampling variability in the forecast errors from the trend estimates do not depend on c (i.e. α) while for OLS_2 there is a dependence on c. Therefore, forecasts based on OLS_2 are more sensitive to α, and this explains why the RMSE of OLS_2 forecasts showed large jumps (relative to OLS_1) for α close to one in the simulations.

Intuitively, OLS_2 is inferior because δ_0 is not identified when $c = 0$ and hence is not consistently estimable when $\alpha = 1$. By continuity, $\widehat{\delta}_0$ is imprecisely estimated in the vicinity of $c = 0$. In the local to unity framework, $T^{-1/2}(\widehat{\delta}_0 - \delta_0)$ is $O_p(1)$ and has a non-vanishing effect on OLS_2, and consequently, the forecast error is large. Theorem 2.2 confirms that this intuition also applies for $p = 1$. The forecast errors under OLS_2 are thus unstable and sensitive to c around $c = 0$. Theorems 2.1 and 2.2 illustrate in the most simple of models the care with which deterministic parameters need to be estimated when constructing forecasts.

The asymptotic results given by Theorems 2.1 and 2.2 can be used to understand the finite sample behavior of the conditional forecast errors and to shed light on the bias and the RMSE of the forecast errors as c varies. Five values of c are considered: $-15, -5, -2, 0, 1$. We approximate the limiting distributions by approximating $W(r)$ and $J_c(r)$ using partial sums of $i.i.d.$ $N(0, 1)$ errors with 500 steps in 10 000 simulations.[8] Given that the initial condition was set to zero in the simulations in the previous subsection, we focus on $\kappa = 0$ in which case $J_c^*(r) = J_c(r)$. We do not report asymptotic approximations for the case of $\kappa > 0$.

[7] Similar asymptotic results were also obtained for long horizon forecasts with horizon h satisfying $h/T \to \lambda \in (0, 1)$.

[8] The asymptotic MSE are in close agreement with the finite sample simulations for $T = 500$ which we did not report. In particular, the RMSE for $p = 0$ at $\alpha = 1$ are 0.078 and 0.089 respectively. The RMSE based on the asymptotic approximations are 0.0762 and 0.0877 respectively. For $p = 1$, the finite sample RMSE are 0.109 and 0.143. The asymptotic approximations are 0.108 and 0.142 respectively.

Densities of the asymptotic distributions are estimated using the Gaussian kernel with bandwidth selected as in Silverman (1986). These are plotted in Figures 1 and 2 for $p = 0$ and 1 respectively. The result that stands out is that the dispersion of OLS_2 is very sensitive to whether $c \geq 0$. This is consistent with our finite samples results. A notable aspect of Figure 1 is that the distribution of forecast errors for OLS_2 is bimodal when $p = 0$ and $c = 0$, i.e. when there is a unit root.[9] In this case, the forecast error under OLS_2 is $-(y_T - \widehat{m}_T)(\widehat{\alpha} - \alpha)$ in the limit. Since $\widehat{\alpha}$ is downward biased, the forecast errors will be positive if $y_T - m_T > 0$.[10] Figure 3 presents the limiting error distributions conditional on $y_T - m_T > 0$, and conditional median bias in the forecast errors is confirmed. Thus, when the deterministic terms have to be estimated, $\widehat{e}_{T+1|T}$ can also be expected to be positive if $y_T - \widehat{m}_T > 0$. The bimodality in Figure 1 arises because $y_T - \widehat{m}_T$ is unconditional and can take on positive or negative values. In other cases when the sign of the prediction error depends on terms other than the bias in $\widehat{\alpha}$, conditional median bias does not immediately imply bimodality in the forecast error distribution.

3. FORECASTS UNDER GLS ESTIMATION

The foregoing analysis shows that OLS_1 dominates OLS_2 when forecasting persistent time series. This raises the obvious question as to whether there exists a two-step procedure that can improve upon one-step least squares detrending. Two options come to mind. One possibility is to fix α to remove estimation of α. For highly persistent data, imposing a unit root (setting $\alpha = 1$) has been treated as a serious option. We will return to this subsequently. The other option is to reduce variability in the estimation of the deterministic trend parameters. Recall that the OLS estimate of δ_0 is inconsistent and in fact, there is no consistent estimate of δ_0 when errors are highly persistent in the local-to-unity framework. However, improved forecasts can be obtained if estimation of the trend parameters leaves no asymptotic effects on the detrended data. GLS estimation of δ is the obvious way to achieve this goal. In addition to improving forecasts through more precise trend parameters estimates, GLS potentially can improve forecasts by reducing sampling variability in estimates of α. This gain can be large when the data are highly persistent, as suggested by the findings of Elliott *et al.* (1996) concerning the dependence of the power of unit root tests on estimated trends. Thus, GLS estimation of δ has the potential to improve forecasts in two distinct ways.

The usefulness of GLS in forecasting was first analyzed by Goldberger (1962) who considered optimal prediction based upon the model $y_t = \delta' z_t + u_t$ but $E(uu') = \Omega$ is non-spherical. Goldberger showed that the best linear unbiased prediction can be obtained by quasi-differencing the data to obtain GLS estimates of δ, and then exploit the fact that if u_t is serially correlated, the relation between the u_{T+h} and u_T can be used to improve the forecast. When u_t is an AR(1) process with known parameter α, the one-step ahead optimal prediction reduces to $y_{T+1} = \widehat{\delta}' z_{T+1} + \alpha(y_T - \widehat{\delta}' z_T)$. This amounts to using (4) for forecasting with an efficient estimate of δ, assuming α is known.

[9]Abadir and Paruolo (1997) noted that kernel smoothers may not pick up discontinuities in the underlying densities. But the bimodality displayed in Figures 1 and 2 can be seen even from the histograms of the simulated distributions.

[10]Using Edgeworth expansions and assuming α is bounded away from the unit circle, Phillips (1979) showed that the exact distribution of $\widehat{e}_{T+h|T}$ will be skewed to the left if $y_T > 0$ once the dependence of $\widehat{\alpha}$ on y_T is taken into account. His result is for strictly stationary processes and applies in finite samples. Thus, the median bias conditional on $y_T > 0$ observed here does not arise for the same reason. Stock (1996) evaluated the forecasts conditional on $y_T > 0$ assuming m_t is known and confirms asymptotic median bias.

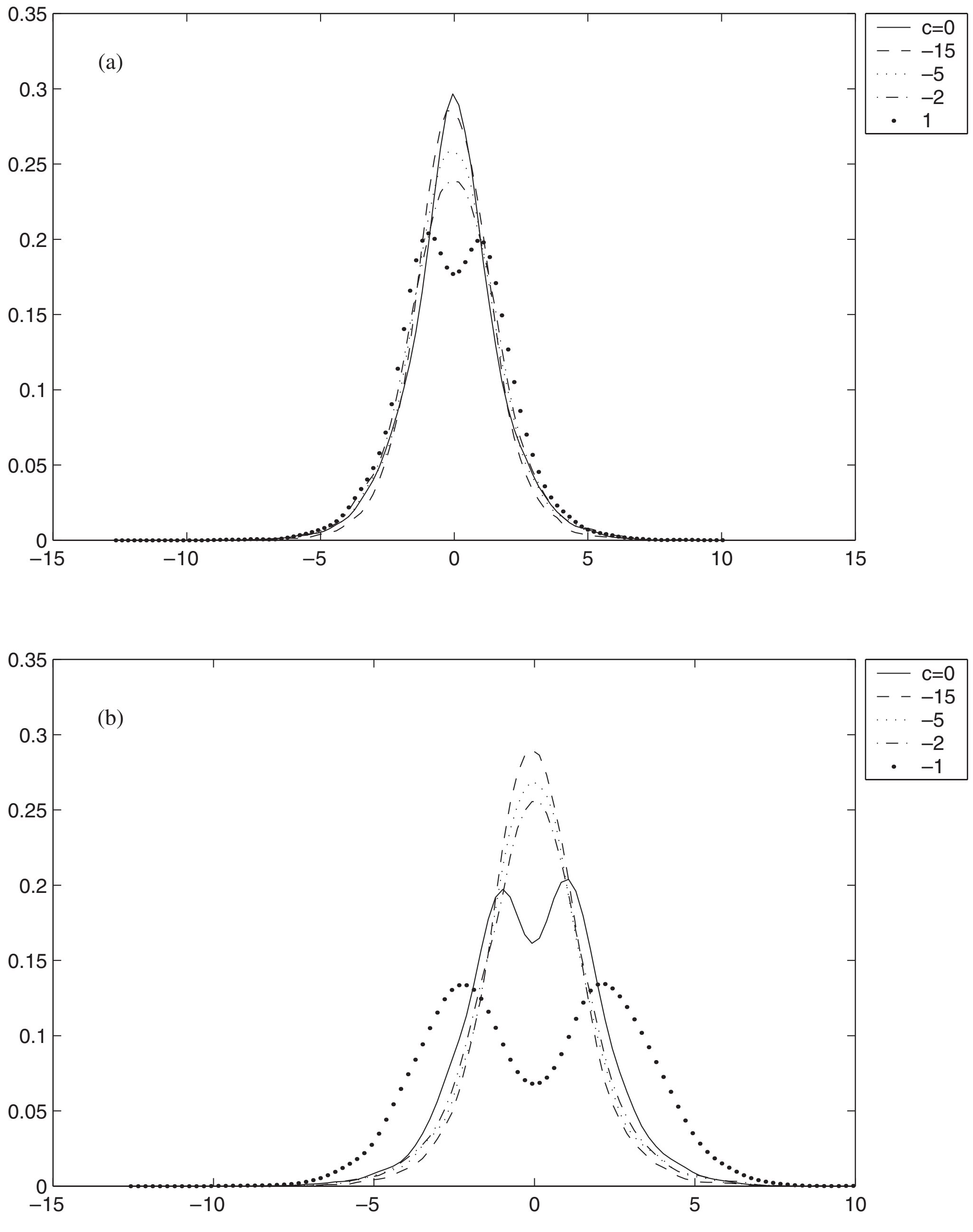

Figure 1. (a) OLS_1 $p = 0$. (b) OLS_2 $p = 0$.

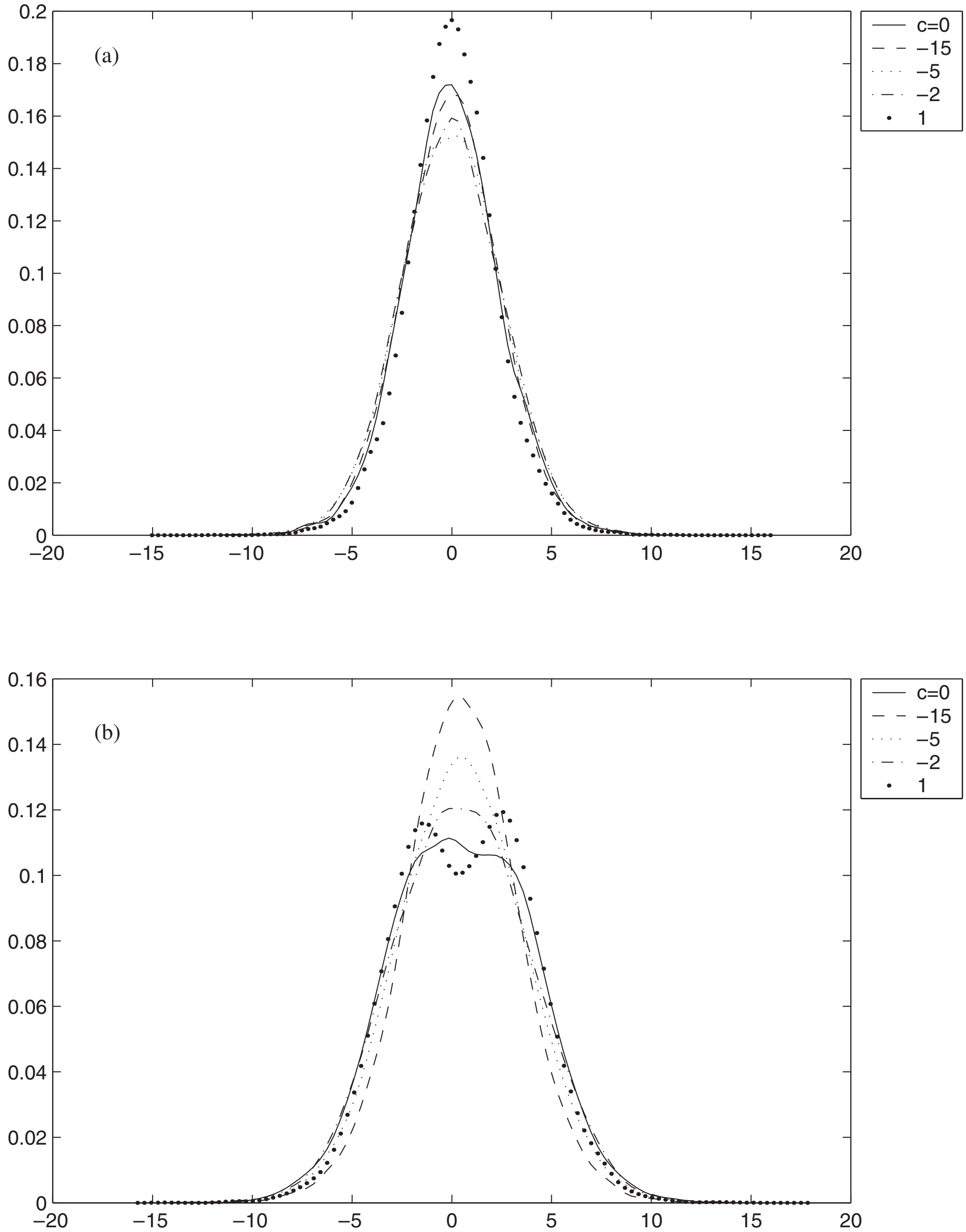

Figure 2. (a) OLS_1 $p = 1$. (b) OLS_2 $p = 1$.

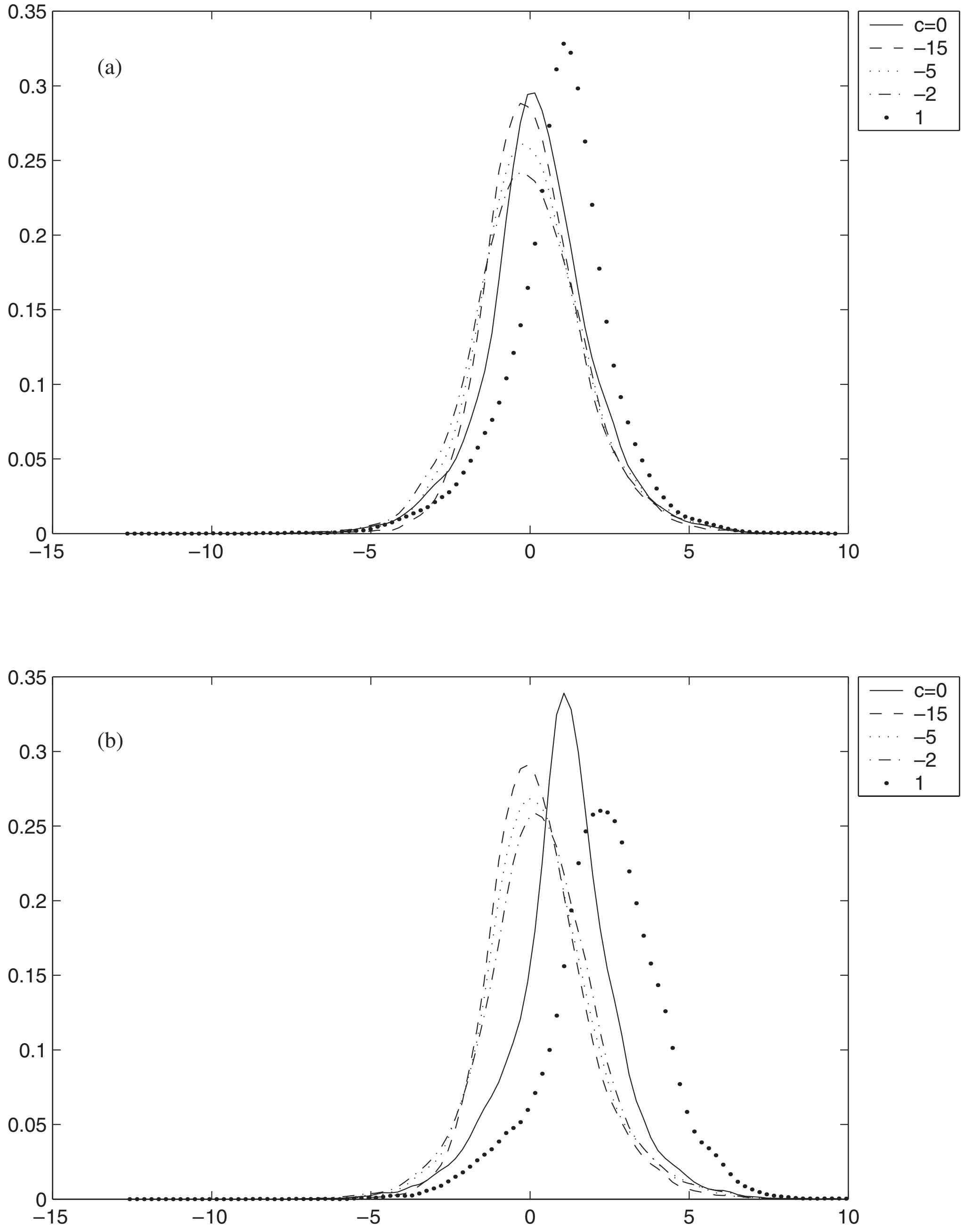

Figure 3. (a) OLS_1 $p = 0$: conditional on $y_T - m_T > 0$. (b) OLS_2 $p = 0$: conditional on $y_T - m_T > 0$.

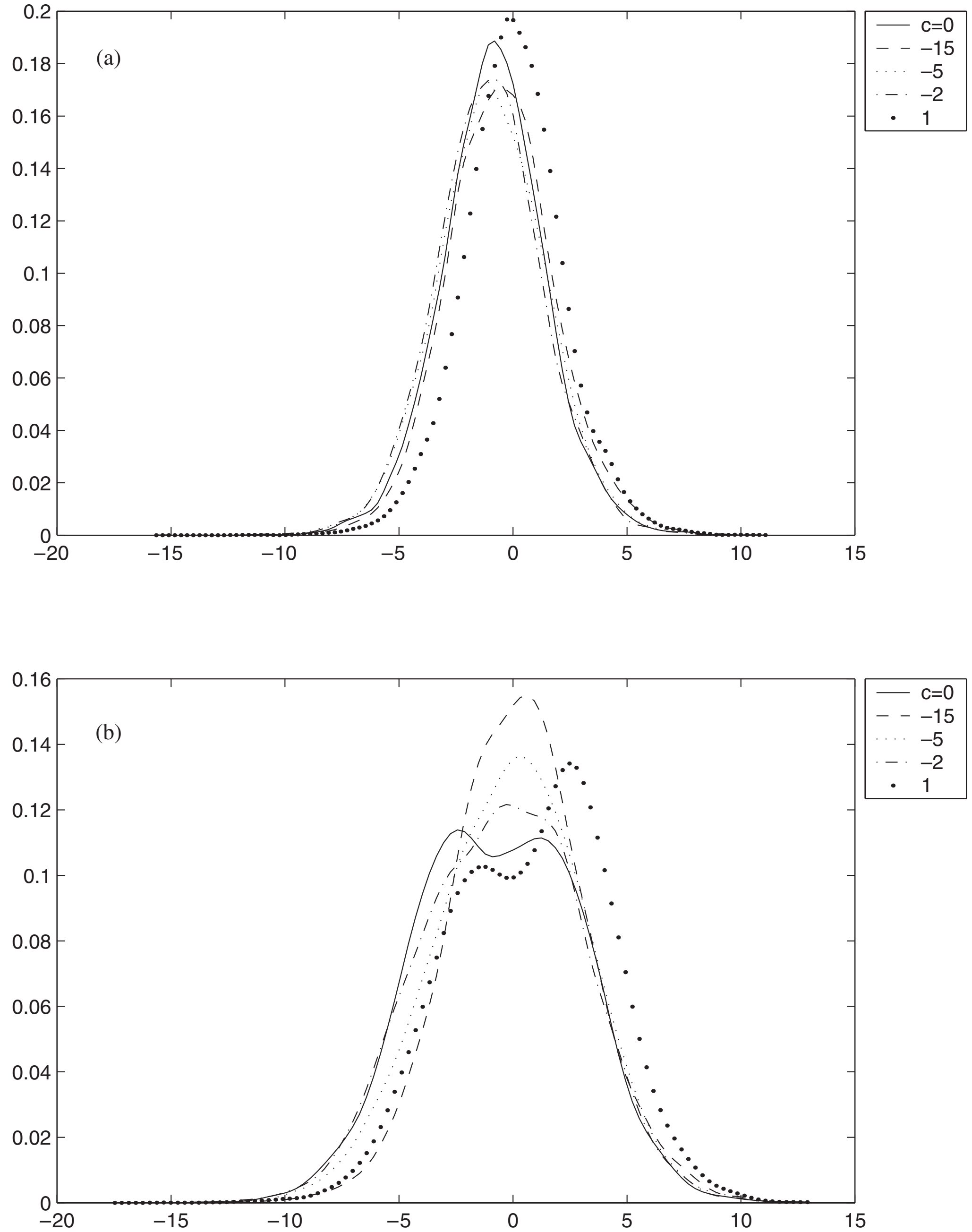

Figure 4. (a) OLS_1 $p = 1$: conditional on $y_T - m_T > 0$. (b) OLS_2 $p = 1$: conditional on $y_T - m_T > 0$.

Trend estimation by GLS requires quasi-differencing the data at some α. We consider the Prais–Winsten (PW) transformation which includes information from the first observation, and the Cochrane–Orcutt (CO) transformation which drops information of the first observation. Specifically, given α and $z_t = (1, t)'$, the quasi-differenced data y_t^+ and z_t^+, are constructed as follows:

- PW: for $t = 2, \dots T$, $y_t^+ = y_t - \alpha y_{t-1}$, $z_t^+ = z_t - \alpha z_{t-1}$, with $y_1^+ = y_1$ and $z_1^+ = z_1$,
- CO: for $t = 2, \dots T$, $y_t^+ = y_t - \alpha y_{t-1}$, $z_t^+ = z_t - \alpha z_{t-1}$.

Then, $\widetilde{\delta} = (z^{+\prime} z^+)^{-1}(z^{+\prime} y^+)$ is the GLS estimate of δ, and $\widetilde{u}_t = y_t - \widetilde{\delta}' z_t$ is the GLS detrended data. The treatment of the initial condition under the PW follows Canjels and Watson (1997), Elliott *et al.* (1996) and Phillips and Lee (1996), who also favor quasi-differencing when the data are highly persistent.[11]

Goldberger assumed α is known, but in practice, this too has to be estimated. The method used to estimate α will inevitably affect the distribution of the trend parameter estimates and thus the forecast errors. We let $\dot{\alpha}$ denote a generic estimator of α used for obtaining feasible GLS estimates of the trend parameters. Canjels and Watson (1997) showed that the limiting behavior of the trend parameter estimates depends on $\dot{\alpha}$ in a complicated way. To make general observations about the relative merits of the different detrending methods in the asymptotic analysis, we define $\dot{c} = \lim_{T\to\infty} T(\dot{\alpha} - 1)$. Then, it directly follows that $T(\dot{\alpha} - \alpha) \Rightarrow (\dot{c} - c)$. If $\dot{\alpha}$ were obtained by (5) using OLS, for example, $\dot{c} - c = \Phi(B_c(r), W)$, where again $B_c(r) = \bar{J}_c^*(r)$ when an intercept is included, and $B_c(r) = \widetilde{J}_c^*(r)$ when a time trend is also included in the regression. We also define $d\dot{W}(r) = dW(r) - (\dot{c} - c)J_c^*(r)$ which is the asymptotic analog of the quasi-differenced errors where the effect of quasi-differencing is captured by $(\dot{c} - c)J_c^*(r)$ (see the appendix for details).

As in the case of OLS forecasts, the behavior of the GLS detrended errors plays an important role in the asymptotic behavior of the GLS forecasts. Under CO detrending,

$$p = 0: \qquad T^{-1/2}\widetilde{u}_{[rT]} \Rightarrow J_c^*(r) + \dot{c}^{-1}\int_0^1 d\dot{W}(s),$$

$$\begin{aligned} p = 1: \qquad T^{-1/2}\widetilde{u}_{[rT]} \Rightarrow {} & J_c^*(r) \\ & - \dot{c}^{-2}\int_0^1 (6 - 4\dot{c} - 12s + 6s\dot{c})d\dot{W}(s) - r\dot{c}^{-1}\int_0^1 (6 - 12s)d\dot{W}(s) \equiv C(r). \end{aligned}$$

Under PW detrending and with $\dot{\theta} = \left(1 - \dot{c} + \frac{1}{3}\dot{c}^2\right)^{-1}$,

$$p = 0: \qquad T^{-1/2}\widetilde{u}_{[rT]} \Rightarrow J_c^*(r) - J_c^-(\kappa),$$

$$\begin{aligned} p = 1: \qquad T^{-1/2}\widetilde{u}_{[rT]} \Rightarrow {} & J_c^*(r) - J_c^-(\kappa) - \\ & r\dot{\theta}\left(\int_0^1 (1 - \dot{c}s)d\dot{W}(s) + (1 - \tfrac{1}{2}\dot{c})J_c^-(\kappa)\right) \equiv P(r). \end{aligned}$$

Recall that detrending leaves no asymptotic effects on the data only when $T^{-1/2}\widetilde{u}_{[rT]} \Rightarrow J_c^*(r)$. Clearly, the CO does not achieve this goal. Under PW, this will also be the case except when $p = 0$ and $\kappa = 0$ (i.e. $u_1 = O_p(1)$), in which case, $J_c^-(0) = 0$. Thus when the data are persistent or α is being estimated, the efficiency of GLS forecasts cannot be presumed.

[11] Canjels and Watson (1997) referred to this as conditional GLS (their CC). We label this as PW only because it retains information in the first observation, in the same spirit as the Prais–Winsten transformation.

Note that the estimator used for α when constructing the feasible GLS estimates of the trend parameters does not necessarily have to be the same as the estimator of α used in the forecasting equation (4). To permit this level of generality, we continue to use $\dot{\alpha}$ to generically denote the estimate of α used for feasible GLS estimation of the trend parameters. We use $\ddot{\alpha}$ to generically denote the estimate of α used in (4), and we define $\ddot{c} = \lim_{T\to\infty} T(\ddot{\alpha} - 1)$ so that $T(\ddot{\alpha} - \alpha) \Rightarrow (\ddot{c} - c)$. In many cases, $\ddot{\alpha}$ is constructed from a regression of $\widetilde{u}_t$ on $\widetilde{u}_{t-1}$, where $\widetilde{u}_t$ is feasible GLS detrended data using $\dot{\alpha}$. It is only in the special case when α is not re-estimated that $\dot{\alpha}$ and $\ddot{\alpha}$ coincide. The limiting distributions of feasible GLS forecast errors are now summarized in the following two theorems.

Theorem 3.1. (CO) *Let the data be generated by* (7)*. Let p be the order of the polynomial in time. Let $\widetilde{e}_{T+1|T}$ be the forecast error obtained with the trend parameters estimated by CO.*

- *For $p = 0$,* $T^{1/2}\widetilde{e}_{T+1|T} \Rightarrow -c\dot{c}^{-1}\int_0^1 d\dot{W}(s) - (\ddot{c} - c)(J_c^*(1) + \dot{c}^{-1}\int_0^1 d\dot{W}(s)).$
- *For $p = 1$,* $T^{1/2}\widetilde{e}_{T+1|T} \Rightarrow c(J_c^*(1) - C(1)) - \dot{c}^{-1}\int_0^1 (6 - 12s) d\dot{W}(s) - (\ddot{c} - c)C(1).$

Theorem 3.2. (PW) *Let the data be generated by* (7)*. Let p be the order of the polynomial in time. Let $\widetilde{e}_{T+1|T}$ be the forecast error obtained with the trend parameters estimated by PW.*

- *For $p = 0$,* $T^{1/2}\widetilde{e}_{T+1|T} \Rightarrow cJ_c^-(\kappa) - (\ddot{c} - c)(J_c^*(1) - J_c^-(\kappa)).$
- *For $p = 1$,* $T^{1/2}\widetilde{e}_{T+1|T} \Rightarrow c(J_c^*(1) - P(1)) - \dot{\theta}\big(\int_0^1 (1 - \dot{c}s) d\dot{W}(s) + \dot{c}\big(1 - \tfrac{1}{2}\dot{c}\big) J_c^-(\kappa)\big) - (\ddot{c} - c)P(1).$

The theorems show that PW and CO do not generate equivalent forecasts and they both differ from the OLS forecasts. Even in the special case of a unit root with $c = 0$, the asymptotic expressions depend on both $\dot{c}$ and $\ddot{c}$ in highly nonlinear ways. However, two observations can be made. First, because the CO forecast error distributions depend on $\dot{c}^{-1}$ and $\dot{c}$ can take on values close to zero, the CO forecasts will likely be subject to large errors (see Canjels and Watson (1997) for a similar result). Second, the forecast error distributions depend on the initial condition, a result that parallels that of Elliott (1999) for the power of unit root tests.

Even though the trend parameters estimated by GLS are more efficient than those estimated by OLS, there is only one case when GLS leads unambiguously to more precise forecasts than OLS.

Lemma 3.1. *When $p = 0$, $u_1 = O_p(1)$ and $\ddot{\alpha}$ is constructed using PW–GLS demeaned data, then*

$$T^{1/2}\widetilde{e}_{T+1|T} \Rightarrow -\Phi(J_c(r), W)J_c(1).$$

When $p = 0$ and $u_1 = O_p(1)$, Theorem 3.2 implies that $T^{1/2}\widetilde{e}_{T+1|T} \Rightarrow -(\ddot{c} - c)J_c(1)$. The forecast error distribution now depends on trend estimation only to the extent that estimation of the autoregressive parameter in the forecasting equation is based on demeaned data. But Elliott *et al.* (1996) showed if $\ddot{\alpha}$ were constructed using PW–GLS demeaned data, the distribution of $\ddot{c} - c$ is the same as if the mean were known. It follows that if $\ddot{\alpha}$ were constructed using PW–GLS demeaned data, the effects of trend estimation on the forecast error can be completely removed. But note that this result requires efficient estimation of both δ and α, i.e. iterated GLS. For example, if $\ddot{\alpha}$ were constructed using OLS (rather than PW–GLS) data, then $\ddot{c} - c = \Phi(\bar{J}_c, W)$, $T^{1/2}\widetilde{e}_{T+1|T} \Rightarrow -\Phi(\bar{J}_c, W)J_c(1)$, and the result of the lemma does not follow.

In all other cases, GLS estimation of δ has non-vanishing effects on the forecast error distribution. Although one cannot unambiguously rank the different detrending methods for forecasting, results of Canjels and Watson (1997) suggest that GLS estimation of the trend parameters is more precise than OLS estimation, and that the PW provides more precise estimates of δ_1 than CO. One might therefore expect GLS detrending to provide more efficient forecasts than OLS detrending, and in particular, that PW might generate smaller forecast errors than the CO. Simulation results in the next subsection generally support these orderings.

3.1. Finite sample properties of feasible GLS forecasts

In this section, we consider six GLS estimators. Feasible GLS forecasts using n iterations for $QD = PW$ or CO are constructed using the following steps:

(1) Obtain, $\dot{\alpha}$, an initial estimate of α by OLS_1.
(2) Transform y_t and z_t by QD to obtain y_t^+ and z_t^+. Then compute $\widetilde{\delta} = (z^{+\prime}z^+)^{-1}(z^{+\prime}y^+)$ and $\widetilde{m}_t = \widetilde{\delta}' z_t$. Construct the forecast using (4), $y_{t+h|t} = \widetilde{m}_{t+h} + \dot{\alpha}^h(y_t - \widetilde{m}_t)$.
(3) If $n = 0$, stop. If $n = 1$, then re-estimate α from (1) with u_t replaced by $\widetilde{u}_t$ to give $\ddot{\alpha}$ and obtain $y_{t+h|t} = \widetilde{m}_{t+h} + \ddot{\alpha}^h(y_t - \widetilde{m}_t)$. If $n = 1$, stop. For $n > 1$, quasi-difference the data at $\ddot{\alpha}$, re-estimate δ, repeat the estimation of α using the newly detrended data and repeat until the change in $\ddot{\alpha}$ between iterations is small.

The objective of iterative estimation ($n > 0$) is to bring the estimates of α and δ closer to being jointly optimal. For models with no lagged dependent variable, the feasible GLS estimates are as efficient as the maximum likelihood estimates asymptotically. Setting $n = \infty$ provides a rough approximation to the forecast errors when the maximum likelihood estimator is used. The use of OLS (the generic estimator) in step 1 is based on Rao and Griliches (1969) who find that estimating α from (5) by OLS is more efficient than estimating it from an autoregression in least squares detrended data or by nonlinear least squares when α is positive.[12] In practice, $\dot{\alpha}$ could exceed unity, but quasi-differencing is valid only if $|\dot{\alpha}| < 1$. This problem is circumvented in the simulations as follows. If an initial $\dot{\alpha}$ exceeds one, it is reset to one prior to PW quasi-differencing. Our theoretical results show that the distribution of the CO detrended data depends on $\dot{c}^{-1}$ which does not exist when $\dot{c} = 0$. Numerical problems were indeed encountered if we allow $\dot{\alpha}$ to be unity. Therefore under CO, we set the upper bound of $\dot{\alpha}$ to 0.995.

The simulations are performed using four sets of assumptions on u_1:

- Assumption A: $u_1 = e_1$ is fixed.
- Assumption B: $u_1 = e_1 \sim N(0, 1)$.
- Assumption C: $u_1 \sim N(0, 1/(1 - \alpha^2))$ for $|\alpha| < 1$.
- Assumption D: $u_1 = \sum_{j=0}^{[\kappa T]} \alpha^j e_{1-j}$, $\kappa > 0$.

Under A, we condition u_1 to a constant. Under B, $u_1 = O_p(1)$ and does not depend on unknown parameters. Under C, u_1 depends on α. Elliott (1999) showed that unit root tests based on GLS detrending are farther away from the asymptotic power envelope under C than B. Canjels and

[12] Using the regression model $y_t = \delta' z_t + u_t$ where u_t is AR(1) with parameter α strictly bounded away from the unit circle and z_t does not include a constant, Rao and Griliches (1969) showed, via Monte-Carlo experiments, that GLS estimation of δ in conjunction with an initial estimate of α obtained from (5) is desirable for the mean-squared-error of $\widehat{\delta}$ when $|\alpha| > 0.3$.

Watson (1997) found, under D, that the efficiency of estimating δ_1 by PW–GLS is reduced when $\kappa > 0$. Assumption B is a special case of D with $\kappa = 0$. In the local asymptotic framework, u_1 is $O_p(T^{1/2})$ under both Assumptions C and D. Simulations are performed under the same experimental design described earlier, except that under Assumption C, only cases with $|\alpha| < 1$ are evaluated. Canjels and Watson (1997) found that for small values of κ, the PW performs well. Here, we report results for $\kappa = 1$, which is considerably large, to put the PW to a challenge.

The results with $h = 1$ and $T = 100$ are reported in columns 3 through 8 of Table 2 for $p = 0$ and likewise in Table 3 for $p = 1$. The forecast errors are smaller, as expected, as the sample size increases. When $\alpha < 0.8$, the gain from GLS estimation over the two OLS procedures is small, irrespective of the assumption on u_1. This is perhaps to be expected since the asymptotic equivalence of OLS and GLS detrending follows from the classic result of Grenander and Rosenblatt (1957) when u_t is stationary. However, as persistence in u_t increases, there are notable differences.

Abadir and Hadri (2000) noted that the bias in the parameters estimated from a model without deterministic components can increase with the sample size if u_1 is relatively large (e.g. exceeding 32). To see if the mean-squared forecast errors exhibit non-monotonicity when deterministic terms are present, Tables 4 and 5 report results for $T = 50$ while Tables 6 and 7 report results for $T = 250$ for $h = 10$. Qualitatively, the results are not sensitive to the sample size. Compared with the results for different sample sizes, we find no evidence of non-monotonicity, as the forecast root-mean-squared errors fall with the sample size roughly at rate $\sqrt{T}$.

For $p = 0$, first notice that PW_0 displays a sharp increase in RMSE around $\alpha = 1$ just like OLS_2 and PW_0 gives less precise forecasts than either OLS forecast. This is the case whether we condition u_1 to zero or let it be drawn from the unconditional distribution. This shows that GLS estimation of δ alone will not always reduce forecast errors. However, PW_1 and PW_∞, greatly improves forecast precision over PW_0 and OLS. This matches the intuition that efficient forecasts depend on efficient estimation of *both* the trend and the slope parameters. With CO on the other hand, iteration does not make much difference and CO is usually dominated by PW_1 and PW_∞. Neither PW_1 nor PW_∞ dominate the other with the best forecast depending on α and u_1. This dependence on the initial condition is predicted by theory but is problematic in practice because the assumption on u_1 cannot be validated. However, when the data are mildly persistent, PW_1 is similar to OLS_1, when the data are moderately persistent, PW_1 outperforms PW_∞, and when the data are extremely persistent, PW_1 dominates OLS_1 and is second best to PW_∞. It is perhaps the best feasible GLS forecast when $p = 0$.

Results for $p = 1$ are reported in Table 3. Because the contribution of $\widehat{\delta}$ to the forecast error is large (as can be seen from OLS_2^α in Table 1(a)), the reduction in forecast error due to efficient estimation of trends is also more substantial. The results in Table 3 show that irrespective of the assumption on the initial condition, the forecast errors are smallest with PW_∞. Even at the one period horizon, the error reduction is 30% over OLS_2. From a RMSE point of view, the choice among the feasible GLS forecasts is clear when $p = 1$.

We also report results for two forecasts based on pretesting for a unit root. Setting $\widehat{\alpha} = 1$ will generate the best one-step ahead forecast if there is indeed a unit root, and in such a case, even long horizon forecasts can be shown to be consistent. Of course, if the unit root is falsely imposed, the forecast precision can suffer. But one can expect forecast error reduction if we impose a unit root for α close to but not identically one. Campbell and Perron (1991) presented some simulation evidence in this regard for $p = 0$, and Diebold and Kilian (2000) considered

Table 2. (a) RMSE of GLS and UP forecast errors: $p = 0$, $T = 100$, $h = 1$, $u_1 = e_1 = 0$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.143	0.141	0.143	0.142	0.143	0.142	0.143	0.142	0.148	0.148
0.400	0.144	0.143	0.144	0.143	0.144	0.143	0.144	0.143	0.146	0.147
0.800	0.153	0.152	0.153	0.146	0.153	0.146	0.153	0.147	0.151	0.157
0.900	0.163	0.162	0.163	0.145	0.163	0.144	0.163	0.146	0.175	0.187
0.950	0.175	0.173	0.174	0.145	0.175	0.141	0.175	0.139	0.174	0.182
0.975	0.183	0.179	0.179	0.159	0.180	0.143	0.180	0.140	0.142	0.146
0.990	0.180	0.180	0.173	0.205	0.174	0.152	0.175	0.145	0.102	0.105
1.000	0.174	0.196	0.168	0.287	0.163	0.165	0.164	0.153	0.068	0.070

(b) RMSE of GLS and UP forecast errors: $p = 0$, $T = 100$, $h = 1$, $u_1 = e_1 \sim N(0, 1)$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.141	0.140	0.141	0.141	0.141	0.141	0.141	0.141	0.150	0.151
0.400	0.144	0.143	0.144	0.143	0.144	0.143	0.144	0.143	0.150	0.150
0.800	0.152	0.151	0.152	0.149	0.152	0.150	0.152	0.153	0.159	0.161
0.900	0.161	0.161	0.161	0.150	0.161	0.149	0.161	0.153	0.180	0.187
0.950	0.174	0.173	0.173	0.152	0.174	0.145	0.174	0.144	0.172	0.179
0.975	0.182	0.180	0.179	0.164	0.180	0.145	0.180	0.141	0.140	0.144
0.990	0.180	0.181	0.173	0.208	0.175	0.153	0.176	0.145	0.101	0.104
1.000	0.173	0.196	0.167	0.289	0.162	0.165	0.163	0.153	0.067	0.068

(c) RMSE of GLS and UP forecast errors: $p = 0$, $T = 100$, $h = 1$, $u_1 \sim$ $(0, 1/(1-\alpha^2)$ $|\alpha| < 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.147	0.146	0.147	0.146	0.147	0.146	0.147	0.146	0.143	0.143
0.400	0.148	0.147	0.148	0.147	0.148	0.147	0.148	0.147	0.144	0.146
0.800	0.154	0.154	0.154	0.155	0.154	0.155	0.154	0.165	0.142	0.155
0.900	0.162	0.163	0.162	0.173	0.162	0.164	0.162	0.177	0.164	0.182
0.950	0.172	0.174	0.172	0.198	0.172	0.165	0.172	0.165	0.172	0.183
0.975	0.180	0.184	0.176	0.225	0.178	0.167	0.178	0.160	0.141	0.146
0.990	0.181	0.190	0.173	0.257	0.175	0.168	0.175	0.158	0.102	0.105

(d) RMSE of GLS and UP forecast errors: $p = 0$, $T = 100$, $h = 1$, $u_1 = \sum_{j=0}^{\kappa T} \alpha^j e_{1-j}$, $\kappa = 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.147	0.145	0.147	0.146	0.147	0.146	0.147	0.146	0.156	0.157
0.400	0.148	0.147	0.148	0.147	0.148	0.148	0.148	0.148	0.155	0.155
0.800	0.153	0.153	0.153	0.155	0.153	0.156	0.153	0.166	0.177	0.176
0.900	0.161	0.161	0.161	0.172	0.161	0.163	0.161	0.175	0.198	0.199
0.950	0.171	0.173	0.170	0.196	0.171	0.165	0.171	0.165	0.175	0.176
0.975	0.179	0.182	0.175	0.222	0.176	0.165	0.177	0.160	0.136	0.137
0.990	0.179	0.187	0.171	0.248	0.172	0.165	0.173	0.156	0.100	0.101
1.000	0.171	0.194	0.165	0.289	0.160	0.164	0.161	0.153	0.069	0.069

Note: OLS_1 and OLS_2 are forecasts based on (6) and (4) respectively. CO_n (Cochrane–Orcutt) and PW_n (Prais–Winsten) are forecasts based on GLS estimation of the trend function, with estimation of α iterated n times. UP_{PW_1} is the forecast based on a unit root pretest where PW_1 is used if a unit root is rejected. UP_{OLS_1} is the forecast based on a unit root pretest where OLS_1 is used if a unit root is rejected.

Table 3. (a) RMSE of GLS and UP forecast errors: $p = 1$, $T = 100$, $h = 1$, $u_1 = e_1 = 0$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.228	0.225	0.228	0.227	0.228	0.227	0.228	0.227	0.227	0.228
0.400	0.230	0.228	0.230	0.227	0.230	0.227	0.230	0.227	0.227	0.230
0.800	0.242	0.249	0.242	0.227	0.242	0.226	0.242	0.226	0.251	0.262
0.900	0.253	0.270	0.253	0.231	0.253	0.225	0.253	0.223	0.254	0.259
0.950	0.263	0.292	0.280	0.245	0.263	0.227	0.263	0.222	0.209	0.210
0.975	0.264	0.310	0.298	0.265	0.264	0.232	0.264	0.221	0.175	0.176
0.990	0.257	0.319	0.312	0.279	0.257	0.233	0.257	0.218	0.149	0.150
1.000	0.244	0.314	0.332	0.274	0.242	0.222	0.242	0.204	0.123	0.123

(b) RMSE of GLS and UP forecast errors: $p = 1$, $T = 100$, $h = 1$, $u_1 = e_1 \sim N(0, 1)$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.226	0.223	0.226	0.225	0.226	0.225	0.226	0.225	0.225	0.226
0.400	0.229	0.228	0.229	0.227	0.229	0.227	0.229	0.227	0.228	0.230
0.800	0.241	0.249	0.241	0.230	0.241	0.229	0.241	0.230	0.259	0.266
0.900	0.252	0.270	0.252	0.234	0.252	0.227	0.252	0.225	0.254	0.259
0.950	0.262	0.293	0.262	0.247	0.262	0.228	0.262	0.222	0.209	0.211
0.975	0.264	0.311	0.279	0.266	0.264	0.232	0.264	0.221	0.174	0.175
0.990	0.258	0.321	0.319	0.280	0.257	0.234	0.257	0.219	0.150	0.151
1.000	0.244	0.316	0.344	0.275	0.243	0.223	0.243	0.205	0.125	0.124

(c) RMSE of GLS and UP forecast errors: $p = 1$, $T = 100$, $h = 1$, $u_1 \sim (0, 1/(1 - \alpha^2)\ |\alpha| < 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.230	0.227	0.230	0.229	0.230	0.229	0.230	0.229	0.226	0.228
0.400	0.232	0.231	0.232	0.230	0.232	0.230	0.232	0.230	0.226	0.231
0.800	0.241	0.252	0.241	0.240	0.241	0.237	0.241	0.240	0.243	0.258
0.900	0.253	0.276	0.253	0.253	0.253	0.240	0.253	0.238	0.253	0.259
0.950	0.264	0.302	0.273	0.266	0.264	0.240	0.264	0.233	0.209	0.211
0.975	0.266	0.318	0.304	0.277	0.266	0.240	0.266	0.228	0.174	0.175
0.990	0.260	0.324	0.336	0.281	0.259	0.236	0.259	0.221	0.150	0.150

(d) RMSE of GLS and UP forecast errors: $p = 1$, $T = 100$, $h = 1$, $u_1 = \sum_{j=0}^{\kappa T} \alpha^j e_{1-j}$, $\kappa = 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.229	0.227	0.229	0.229	0.229	0.229	0.229	0.229	0.229	0.230
0.400	0.231	0.230	0.231	0.230	0.231	0.230	0.231	0.230	0.230	0.232
0.800	0.241	0.251	0.241	0.238	0.241	0.236	0.241	0.239	0.268	0.272
0.900	0.253	0.275	0.253	0.251	0.253	0.238	0.253	0.237	0.258	0.260
0.950	0.264	0.302	0.273	0.266	0.264	0.240	0.264	0.232	0.211	0.212
0.975	0.267	0.319	0.306	0.278	0.266	0.241	0.266	0.228	0.177	0.177
0.990	0.261	0.325	0.329	0.281	0.260	0.236	0.260	0.221	0.152	0.152
1.000	0.247	0.317	0.349	0.272	0.245	0.223	0.245	0.206	0.123	0.123

Note: OLS_1 and OLS_2 are forecasts based on (6) and (4) respectively. CO_n (Cochrane–Orcutt) and PW_n (Prais–Winsten) are forecasts based on GLS estimation of the trend function, with estimation of α iterated n times. UP_{PW_1} is the forecast based on a unit root pretest where PW_1 is used if a unit root is rejected. UP_{OLS_1} is the forecast based on a unit root pretest where OLS_1 is used if a unit root is rejected.

Table 4. (a) RMSE of GLS and UP forecast errors: $p = 0$, $T = 50$, $h = 1$, $u_1 = e_1 = 0$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.200	0.197	0.200	0.200	0.200	0.200	0.200	0.200	0.211	0.211
0.400	0.204	0.202	0.204	0.202	0.204	0.202	0.204	0.202	0.207	0.209
0.800	0.224	0.221	0.224	0.208	0.224	0.209	0.224	0.213	0.237	0.247
0.900	0.243	0.239	0.241	0.210	0.242	0.210	0.242	0.212	0.241	0.251
0.950	0.256	0.250	0.250	0.220	0.252	0.214	0.252	0.213	0.203	0.209
0.975	0.256	0.253	0.253	0.249	0.249	0.225	0.250	0.219	0.162	0.165
0.990	0.250	0.259	0.261	0.302	0.240	0.238	0.240	0.224	0.123	0.124
1.000	0.246	0.278	0.290	0.364	0.231	0.249	0.232	0.231	0.104	0.104

(b) RMSE of GLS and UP forecast errors: $p = 0$, $T = 50$, $h = 1$, $u_1 = e_1 \sim N(0, 1)$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.201	0.198	0.201	0.200	0.201	0.201	0.201	0.201	0.213	0.213
0.400	0.206	0.203	0.206	0.204	0.206	0.204	0.206	0.204	0.214	0.216
0.800	0.223	0.221	0.223	0.212	0.223	0.213	0.223	0.222	0.246	0.252
0.900	0.241	0.238	0.240	0.217	0.241	0.214	0.241	0.218	0.240	0.247
0.950	0.254	0.250	0.249	0.226	0.252	0.215	0.252	0.214	0.200	0.205
0.975	0.254	0.253	0.253	0.251	0.248	0.222	0.249	0.216	0.159	0.162
0.990	0.248	0.258	0.260	0.301	0.238	0.233	0.238	0.219	0.124	0.126
1.000	0.242	0.275	0.285	0.362	0.229	0.243	0.229	0.223	0.103	0.103

(c) RMSE of GLS and UP forecast errors: $p = 0$, $T = 50$, $h = 1$, $u_1 \sim N(0, 1/(1 - \alpha^2)$ $|\alpha| < 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.201	0.198	0.201	0.200	0.201	0.200	0.201	0.200	0.200	0.202
0.400	0.205	0.202	0.205	0.203	0.205	0.204	0.205	0.204	0.202	0.208
0.800	0.220	0.219	0.220	0.220	0.220	0.221	0.220	0.237	0.222	0.244
0.900	0.235	0.236	0.234	0.244	0.234	0.229	0.234	0.237	0.239	0.255
0.950	0.249	0.253	0.245	0.276	0.246	0.238	0.246	0.233	0.203	0.211
0.975	0.255	0.266	0.261	0.308	0.249	0.244	0.249	0.234	0.163	0.167
0.990	0.253	0.272	0.272	0.336	0.243	0.246	0.243	0.231	0.126	0.127

(d) RMSE of GLS and UP forecast errors: $p = 0$, $T = 50$, $h = 1$, $u_1 = \sum_{j=0}^{\kappa T} \alpha^j e_{1-j}$, $\kappa = 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.202	0.198	0.202	0.201	0.202	0.201	0.202	0.201	0.215	0.216
0.400	0.205	0.202	0.205	0.203	0.205	0.204	0.205	0.204	0.214	0.216
0.800	0.222	0.221	0.222	0.221	0.222	0.220	0.222	0.235	0.257	0.259
0.900	0.239	0.240	0.238	0.246	0.239	0.231	0.239	0.238	0.242	0.244
0.950	0.253	0.255	0.247	0.276	0.249	0.239	0.249	0.234	0.193	0.195
0.975	0.256	0.264	0.259	0.303	0.248	0.242	0.248	0.231	0.156	0.157
0.990	0.251	0.267	0.271	0.323	0.241	0.240	0.241	0.226	0.127	0.128
1.000	0.241	0.273	0.284	0.359	0.228	0.242	0.228	0.224	0.103	0.103

Note: OLS_1 and OLS_2 are forecasts based on (6) and (4) respectively. CO_n (Cochrane–Orcutt) and PW_n (Prais–Winsten) are forecasts based on GLS estimation of the trend function, with estimation of α iterated n times. UP_{PW_1} is the forecast based on a unit root pretest where PW_1 is used if a unit root is rejected. UP_{OLS_1} is the forecast based on a unit root pretest where OLS_1 is used if a unit root is rejected.

Table 5. (a) RMSE of GLS and UP forecast errors: $p = 1$, $T = 50$, $h = 1$, $u_1 = e_1 = 0$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.326	0.318	0.326	0.323	0.326	0.323	0.326	0.323	0.325	0.328
0.400	0.332	0.329	0.332	0.326	0.332	0.326	0.332	0.326	0.333	0.339
0.800	0.358	0.374	0.358	0.331	0.358	0.327	0.358	0.326	0.360	0.367
0.900	0.375	0.409	0.718	0.342	0.374	0.330	0.374	0.324	0.300	0.303
0.950	0.378	0.433	1.338	0.360	0.376	0.335	0.376	0.322	0.252	0.253
0.975	0.370	0.443	1.693	0.371	0.368	0.335	0.368	0.318	0.224	0.225
0.990	0.357	0.441	1.749	0.369	0.355	0.328	0.355	0.308	0.202	0.203
1.000	0.342	0.432	1.634	0.359	0.341	0.315	0.341	0.294	0.180	0.181

(b) RMSE of GLS and UP forecast errors: $p = 1$, $T = 50$, $h = 1$, $u_1 = e_1 \sim N(0, 1)$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.327	0.318	0.327	0.323	0.327	0.323	0.327	0.323	0.325	0.328
0.400	0.333	0.329	0.333	0.326	0.333	0.326	0.333	0.326	0.334	0.340
0.800	0.356	0.372	0.356	0.332	0.356	0.328	0.356	0.327	0.358	0.364
0.900	0.371	0.406	0.420	0.342	0.371	0.328	0.371	0.322	0.296	0.299
0.950	0.372	0.429	1.060	0.359	0.371	0.331	0.371	0.318	0.248	0.250
0.975	0.364	0.440	1.517	0.371	0.362	0.334	0.362	0.317	0.222	0.223
0.990	0.352	0.441	1.607	0.372	0.350	0.329	0.350	0.310	0.203	0.203
1.000	0.339	0.433	1.467	0.363	0.337	0.318	0.337	0.297	0.182	0.181

(c) RMSE of GLS and UP forecast errors: $p = 1$, $T = 50$, $h = 1$, $u_1 \sim N(0, 1/(1-\alpha^2)$ $|\alpha| < 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.327	0.318	0.327	0.323	0.327	0.323	0.327	0.323	0.322	0.327
0.400	0.332	0.327	0.332	0.325	0.332	0.325	0.332	0.325	0.324	0.336
0.800	0.357	0.376	0.357	0.341	0.357	0.335	0.357	0.336	0.356	0.367
0.900	0.372	0.412	0.521	0.358	0.372	0.340	0.372	0.332	0.300	0.304
0.950	0.373	0.434	1.169	0.370	0.372	0.341	0.372	0.327	0.253	0.254
0.975	0.363	0.442	1.544	0.375	0.362	0.337	0.362	0.320	0.224	0.224
0.990	0.352	0.440	1.625	0.372	0.350	0.329	0.350	0.309	0.203	0.203

(d) RMSE of GLS and UP forecast errors: $p = 1$, $T = 50$, $h = 1$, $u_1 = \sum_{j=0}^{\kappa T} \alpha^j e_{1-j}$, $\kappa = 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.327	0.319	0.327	0.324	0.327	0.324	0.327	0.324	0.327	0.330
0.400	0.332	0.328	0.332	0.325	0.332	0.326	0.332	0.326	0.336	0.342
0.800	0.356	0.376	0.356	0.342	0.356	0.336	0.356	0.337	0.357	0.362
0.900	0.372	0.412	0.718	0.359	0.371	0.340	0.371	0.333	0.297	0.299
0.950	0.372	0.434	1.181	0.371	0.371	0.340	0.371	0.326	0.249	0.251
0.975	0.363	0.441	1.527	0.374	0.361	0.336	0.361	0.318	0.221	0.223
0.990	0.351	0.440	1.613	0.372	0.350	0.329	0.350	0.309	0.200	0.201
1.000	0.339	0.433	1.581	0.363	0.338	0.318	0.338	0.296	0.178	0.178

Note: OLS_1 and OLS_2 are forecasts based on (6) and (4) respectively. CO_n (Cochrane–Orcutt) and PW_n (Prais–Winsten) are forecasts based on GLS estimation of the trend function, with estimation of α iterated n times. UP_{PW_1} is the forecast based on a unit root pretest where PW_1 is used if a unit root is rejected. UP_{OLS_1} is the forecast based on a unit root pretest where OLS_1 is used if a unit root is rejected.

Table 6. (a) RMSE of GLS and UP forecast errors: $p = 0$, $T = 250$, $h = 10$, $u_1 = e_1 = 0$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.064	0.063	0.064	0.063	0.064	0.063	0.064	0.063	0.063	0.064
0.400	0.106	0.106	0.106	0.105	0.106	0.105	0.106	0.105	0.105	0.106
0.800	0.295	0.290	0.295	0.283	0.295	0.283	0.295	0.283	0.283	0.295
0.900	0.478	0.467	0.478	0.424	0.478	0.427	0.478	0.431	0.430	0.479
0.950	0.662	0.647	0.662	0.520	0.662	0.525	0.662	0.529	0.710	0.796
0.975	0.825	0.805	0.823	0.611	0.823	0.596	0.824	0.594	0.939	0.989
0.990	0.967	0.933	0.949	0.864	0.947	0.709	0.953	0.697	0.765	0.783
1.000	0.975	1.099	0.889	1.870	0.889	0.928	0.900	0.888	0.316	0.323

(b) RMSE of GLS and UP forecast errors: $p = 0$, $T = 250$, $h = 10$, $u_1 = e_1 \sim N(0, 1)$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.064	0.064
0.400	0.106	0.105	0.106	0.105	0.106	0.105	0.106	0.105	0.107	0.108
0.800	0.296	0.292	0.296	0.290	0.296	0.290	0.296	0.291	0.296	0.302
0.900	0.482	0.471	0.482	0.451	0.482	0.455	0.482	0.471	0.486	0.509
0.950	0.663	0.648	0.663	0.560	0.663	0.562	0.663	0.578	0.766	0.824
0.975	0.821	0.798	0.821	0.640	0.821	0.616	0.821	0.616	0.952	0.993
0.990	0.955	0.917	0.936	0.878	0.933	0.709	0.940	0.697	0.765	0.784
1.000	0.964	1.090	0.876	1.852	0.876	0.927	0.887	0.891	0.317	0.321

(c) RMSE of GLS and UP forecast errors: $p = 0$, $T = 250$, $h = 10$, $u_1 \sim (0, 1/N(1 - \alpha^2)\ |\alpha| < 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.064
0.400	0.106	0.105	0.106	0.105	0.106	0.105	0.106	0.105	0.105	0.106
0.800	0.297	0.293	0.297	0.305	0.297	0.306	0.297	0.311	0.272	0.296
0.900	0.481	0.474	0.481	0.559	0.481	0.551	0.481	0.652	0.388	0.477
0.950	0.662	0.653	0.662	0.866	0.662	0.758	0.662	0.837	0.632	0.766
0.975	0.820	0.810	0.820	1.120	0.820	0.836	0.820	0.854	0.926	0.987
0.990	0.961	0.955	0.933	1.388	0.931	0.893	0.940	0.878	0.765	0.784

(d) RMSE of GLS and UP forecast errors: $p = 0$, $T = 250$, $h = 10$, $u_1 = \sum_{j=0}^{\kappa T} \alpha^j e_{1-j}, \kappa = 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.068	0.068
0.400	0.106	0.105	0.106	0.105	0.106	0.105	0.106	0.105	0.108	0.109
0.800	0.297	0.293	0.297	0.305	0.297	0.306	0.297	0.311	0.348	0.341
0.900	0.481	0.474	0.481	0.559	0.481	0.551	0.481	0.652	0.688	0.656
0.950	0.662	0.653	0.662	0.866	0.662	0.758	0.662	0.837	0.984	0.966
0.975	0.820	0.810	0.820	1.120	0.820	0.836	0.820	0.854	1.003	1.004
0.990	0.961	0.955	0.933	1.388	0.931	0.893	0.940	0.878	0.760	0.766
1.000	0.968	1.094	0.878	1.855	0.877	0.928	0.889	0.892	0.318	0.325

Note: OLS_1 and OLS_2 are forecasts based on (6) and (4) respectively. CO_n (Cochrane–Orcutt) and PW_n (Prais–Winsten) are forecasts based on GLS estimation of the trend function, with estimation of α iterated n times. UP_{PW_1} is the forecast based on a unit root pretest where PW_1 is used if a unit root is rejected. UP_{OLS_1} is the forecast based on a unit root pretest where OLS_1 is used if a unit root is rejected.

Table 7. (a) RMSE of GLS and UP forecast errors: $p = 1$, $T = 250$, $h = 10$, $u_1 = e_1 = 0$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133	0.133
0.400	0.222	0.221	0.222	0.220	0.222	0.220	0.222	0.220	0.220	0.222
0.800	0.593	0.589	0.593	0.570	0.593	0.570	0.593	0.570	0.570	0.593
0.900	0.889	0.895	0.889	0.812	0.889	0.813	0.889	0.813	0.878	0.939
0.950	1.145	1.187	1.145	1.011	1.145	0.995	1.145	0.992	1.218	1.250
0.975	1.328	1.436	1.329	1.206	1.327	1.125	1.327	1.112	1.070	1.079
0.990	1.427	1.658	1.436	1.465	1.421	1.245	1.421	1.195	0.777	0.780
1.000	1.341	1.726	1.360	1.567	1.326	1.217	1.326	1.118	0.345	0.350

(b) RMSE of GLS and UP forecast errors: $p = 1$, $T = 250$, $h = 10$, $u_1 = e_1 \sim N(0, 1)$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.135	0.134	0.135	0.134	0.135	0.134	0.135	0.134	0.134	0.135
0.400	0.224	0.223	0.224	0.223	0.224	0.223	0.224	0.223	0.223	0.224
0.800	0.599	0.598	0.599	0.586	0.599	0.586	0.599	0.587	0.586	0.599
0.900	0.899	0.911	0.899	0.844	0.899	0.842	0.899	0.845	0.907	0.952
0.950	1.158	1.208	1.158	1.042	1.158	1.021	1.158	1.020	1.224	1.253
0.975	1.346	1.459	1.346	1.224	1.346	1.143	1.346	1.130	1.080	1.090
0.990	1.447	1.669	1.449	1.456	1.444	1.245	1.444	1.196	0.781	0.785
1.000	1.337	1.702	1.351	1.536	1.326	1.197	1.326	1.104	0.341	0.344

(c) RMSE of GLS and UP forecast errors: $p = 1$, $T = 250$, $h = 10$, $u_1 \sim N(0, 1/(1 - \alpha^2)$ $|\alpha| < 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.133	0.134
0.400	0.224	0.223	0.224	0.223	0.224	0.223	0.224	0.223	0.219	0.222
0.800	0.601	0.600	0.601	0.604	0.601	0.604	0.601	0.610	0.554	0.594
0.900	0.902	0.916	0.902	0.925	0.902	0.912	0.902	0.935	0.835	0.928
0.950	1.159	1.221	1.159	1.191	1.159	1.121	1.159	1.126	1.207	1.251
0.975	1.349	1.489	1.349	1.386	1.349	1.237	1.349	1.214	1.067	1.079
0.990	1.459	1.704	1.462	1.542	1.456	1.300	1.456	1.243	0.774	0.779

(d) RMSE of GLS and UP forecast errors: $p = 1$, $T = 250$, $h = 10$, $u_1 = \sum_{j=0}^{\kappa T} \alpha^j e_{1-j}$, $\kappa = 1$.

α	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}
0.000	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135
0.400	0.224	0.223	0.224	0.223	0.224	0.223	0.224	0.223	0.223	0.224
0.800	0.601	0.600	0.601	0.604	0.601	0.604	0.601	0.610	0.614	0.611
0.900	0.902	0.916	0.902	0.925	0.902	0.912	0.902	0.935	1.013	1.013
0.950	1.159	1.221	1.159	1.191	1.159	1.121	1.159	1.126	1.258	1.263
0.975	1.349	1.489	1.349	1.386	1.349	1.237	1.349	1.214	1.080	1.081
0.990	1.459	1.704	1.462	1.542	1.456	1.300	1.456	1.243	0.783	0.784
1.000	1.335	1.698	1.349	1.537	1.325	1.197	1.325	1.105	0.345	0.344

Note: OLS_1 and OLS_2 are forecasts based on (6) and (4) respectively. CO_n (Cochrane–Orcutt) and PW_n (Prais–Winsten) are forecasts based on GLS estimation of the trend function, with estimation of α iterated n times. UP_{PW_1} is the forecast based on a unit root pretest where PW_1 is used if a unit root is rejected. UP_{OLS_1} is the forecast based on a unit root pretest where OLS_1 is used if a unit root is rejected.

the case of $p = 1$.[13] Stock and Watson (1998) considered the usefulness of unit root pretests in empirical applications. However, they forecast using OLS_1 when the unit root hypothesis is rejected. In light of the efficiency of GLS over the OLS, we use PW_1 under the alterative of stationarity. The PW_1 is used because it has desirable properties both when $p = 0$ and $p = 1$, and also because it is easy to implement. Specifically, we use the DF–GLS test (based on the PW transformation) with one lag to test for a unit root. If we cannot reject a unit root and $p = 0$, $\widehat{y}_{T+1|T} = y_T$. If $p = 1$, the mean of the first differenced series is estimated. Denoting this by $\overline{\Delta y}$, then $\widehat{y}_{T+1|T} = y_T + \overline{\Delta y}$. If a unit root is rejected and a PW_1 forecast is obtained, the procedure is labelled UP_{PW_1} below. If a unit root is rejected and OLS_1 is used (as in Stock and Watson), we refer to the procedure as UP_{OLS_1}.

The UP forecast errors are given in the last two columns of Tables 2 to 7. If the unit root test always rejects correctly, the RMSE for $\alpha < 1$ would have coincided with PW_1 or OLS_1. This apparently is not the case and reflects the fact that power of the unit root test is less than one. The increase in RMSE from falsely imposing a unit root is larger when $p = 0$. Furthermore, forecasts based on unit root pretests can sometimes be worse than without unit root pretesting (see line 5 of Table 2(a)). Nonetheless, the reductions in forecast errors are quite substantial in many of the cases when α is very close to or at unity. This arises not just because variability in $\widehat{\alpha}$ is suppressed, but also because first differencing bypasses the need to estimate δ_0, the key source of variability with any two-step procedure. Irrespective of the assumption on u_1, UP_{PW_1} has smaller RMSE than UP_{OLS_1}, reflecting the improved efficiency of PW_1 over OLS_1. For both UP procedures, the trade-offs involved are clear: large reduction in RMSE when the data are persistent versus an increase in RSME when the largest autoregressive root is far from unity.

An overview of the alternatives to OLS_1 (the preferred OLS forecast) is as follows. The two UP procedures usually yield the minimum RMSE when α is very close to one. The problem, of course, is that 'close' depends on the data in question. Of the GLS forecasts, PW_∞ performs very well when $p = 1$, and the PW_1 also yields significant improvements over the OLS procedures. For $p = 0$, the results are sensitive to u_1 and the PW_1 is more robust than PW_∞. Feasible GLS based on PW_1 and PW_∞ with or without a unit root pretest dominates OLS and should be used in practice.

4. EMPIRICAL EXAMPLES

In this section, we take the procedures to 15 US macroeconomic time series. These are GDP, investment, exports, imports, final sales, personal income, employee compensation, M2 growth rate, unemployment rate, 3 month, 1 year, and 10 year yield on treasury bills, FED funds rate, inflation in the GDP deflator and the CPI. Except for variables already in rates, the logarithm of the data is used. Inflation in the CPI is calculated as the change in the price index between the last month of two consecutive quarters. All data span the sample 1960:1–1998:4 and are taken from FRED.[14] Throughout, we use $k = 4$ lags in the forecasting model. Stock and Watson (1998) found little to gain from using data dependent rules for selecting the lag length in forecasting exercises. Four lags are also used in the unit root tests. We assume a linear time trend for the seven National Account series. Although the unit root test is performed each time the sample

[13] Diebold and Kilian (2000) found that pretesting is better than always setting $\widehat{\alpha} = 1$ and is often better than always using the OLS estimate of α.

[14] The web site address is `http://www.stls.frb.org/fred`.

is extended, we only keep track of unit root test results for the sample as a whole. Except for investment, the unit root hypothesis cannot be rejected in the full sample for the first seven series. For the remaining variables, we use $p = 0$. The DFGLS rejects a unit root in M2 growth, unemployment rate and CPI inflation.

Since the preceding analysis assumes $k = 1$, a discussion on quasi-differencing when $k > 1$ is in order. We continue to obtain $\dot{\alpha}_i$, $i = 1, \ldots, k$ from (5) with additional lags of y_t added to the regression. We experimented with two possibilities. One option is to quasi-difference with $\dot{\alpha} = \sum_{i=1}^{k} \dot{\alpha}_i$. The alternative option is to let $x_t^+ = x_t - \sum_{i=1}^{k} \dot{\alpha}_i x_{t-i}$ for $t = k+1, \ldots, T$ $(x = y, z)$. For the CO, we lose the first k observations but no further modification is required. For the PW, we additionally assume $x_i^+ = x_i - \sum_{j=1}^{i} \dot{\alpha}_j x_{i-j}$ for $i = 1, \ldots, k$ $(x = y, z)$. The forecasts are then based on four lags of the quasi-transformed data. Based on our limited experimentation, both approaches give very similar forecast errors and we only report results based on the first procedure. That is, quasi-differencing using the sum of the autoregressive parameters.

Our results are based on 100 real time, one period ahead forecasts. Specifically, the first forecast is based on estimation up to 1973:4. The sample is then extended by one period, the models re-estimated, and a new forecast is obtained. Because we do not know the data generating process for the observed data, the forecast errors reflect not only parameter uncertainty, but also potential model misspecification. Procedures sensitive to model misspecification may have larger errors than are found in the simulations when the forecasting model is correctly specified. We also carried out formal tests for the equality of the MSE of the forecasts using the tests proposed by West (1996). For each series we tested the hypothesis, denoted H_A, that all forecasts (excluding the UP forecasts) yield equivalent MSE (on average). The tests are carried out by testing whether the sample means of the 100 real time MSEs are consistent with equal population means of the underlying MSE processes. Therefore, we are essentially testing equality of the means of vectors of time series consisting of the real time forecast mean square errors. West (1996) showed that Wald statistics for testing equality of the means constructed using serial correlation robust standard errors have asymptotic chi-square distributions. We computed serial correlation robust standard errors using spectral density kernel methods with the quadratic spectral kernel and the bandwidth chosen using the data-dependent method recommended by Andrews (1991) using the VAR(1) plug-in method.

Our results are summarized in terms of the average RMSE and are reported in Table 8. We group the series according to whether $p = 1$ or $p = 0$. For $p = 1$ we see that PW_∞ gives the best forecast in four of the seven cases. Surprisingly (given the simulation results in the previous section), CO_0 and CO_1 give the best forecasts in three cases. However, iterating CO in those cases makes the forecasts worse. The OLS forecasts are never the best although OLS_1 is often much better than OLS_2. Unit root pretesting often improves the forecasts which is not surprising given that six of the series appear to be $I(1)$. Because the UP forecasts are the same when a unit root is not rejected, UP_{PW_1} and UP_{OLS_1} usually have the same RMSE. The last column gives p-values for the test of equality of forecasts. We see that, with the exception of the investment series, we strongly reject the null that all the forecasts are equally precise. This suggests that the dominance of GLS over OLS can be taken seriously.

For $p = 0$, less crisp comparisons of forecasts can be made. Except for the GDP deflator series, the tests for equality of forecasts cannot be rejected. The null of equal forecasts for the GDP deflator series is rejected because the CO_0 forecast is much worse than any of the other forecasts. For the series for which a unit root can be rejected (m2sl, unrate, cpiaus) the RMSE are essentially identical across forecasts. This is to be expected given that the simulations in the previous section showed that the method of trend estimation is largely irrelevant for $I(0)$ series.

Table 8. Empirical examples: average RMSE of 100 real time, one period ahead forecasts.

Series	OLS_1	OLS_2	CO_0	PW_0	CO_1	PW_1	CO_∞	PW_∞	UP_{PW_1}	UP_{OLS_1}	H_A
					$p = 1$						
gdpc92-$I(1)$	0.280	0.292	0.279	0.294	0.279	0.286	0.287	0.285	0.278	0.278	0.000
gpdic92-$I(0)$	1.587	1.605	1.578	1.599	1.579	1.592	1.581	1.591	1.592	1.587	0.149
expgsc92-$I(1)$	0.196	0.204	0.195	0.208	0.193	0.184	0.191	0.182	0.181	0.181	0.000
impgsc92-$I(1)$	1.198	1.192	1.195	1.152	1.181	1.145	1.178	1.125	1.123	1.123	0.002
finslc92-$I(1)$	1.011	1.033	0.999	1.008	0.995	0.974	0.995	0.957	0.870	0.870	0.000
dpic92-$I(1)$	0.356	0.365	0.351	0.356	0.349	0.350	0.352	0.347	0.342	0.342	0.000
wascur-$I(1)$	0.239	0.247	0.237	0.247	0.237	0.242	0.241	0.241	0.238	0.238	0.000
					$p = 0$						
m2sl-$I(0)$	3.002	2.993	2.992	2.991	2.991	2.989	2.991	2.989	2.989	3.002	0.249
unrate-$I(0)$	0.353	0.353	0.353	0.354	0.353	0.354	0.358	0.355	0.354	0.353	0.524
tb3ma-$I(1)$	1.591	1.605	1.491	1.601	1.573	1.583	1.577	1.569	1.549	1.549	0.681
gs1-$I(1)$	1.474	1.476	1.366	1.475	1.447	1.469	1.458	1.468	1.452	1.452	0.860
gs10-$I(1)$	0.864	0.857	0.911	0.863	0.854	0.863	0.855	0.864	0.856	0.856	0.555
fed-$I(1)$	1.987	1.985	1.957	1.991	1.976	2.009	1.974	2.000	1.934	1.934	0.360
gdpdef-$I(1)$	1.156	1.131	1.213	1.155	1.121	1.148	1.117	1.151	1.115	1.115	0.000
cpiaucs-$I(0)$	2.221	2.206	2.206	2.207	2.211	2.217	2.199	2.218	2.217	2.221	0.139

Note: The $I(0)/I(1)$ after each series name indicates whether a unit root can be rejected in the errors of the full series, using the DFGLS of Elliott *et al.* (1996). Columns OLS_1 to UP_{OLS_1} are the averaged RMSE over 100 continuously updated forecasts. The column labeled H_A reports asymptotic p-values for the joint hypothesis that mean square errors of all the forecasts (not including UP_{PW_1} and UP_{OLS_1}) are the same (on average). The series names are those used by FRED. They are deciphered as follows: gdpc92 = gross domestic product, gdpic92 = investment, expgsc92 = exports, impgsc92 = imports, finslc92 = final sales, dpic92 = personal income, wascur = employee compensation, m2sl = M2 growth rate, unrate = unemployment rate, tb3ma = 3 month t-bill yield, gs1 = one year t-bill yield, gs10 = 10 year t-bill yield, fed = FED funds rate, gdpdef = GDP deflator based inflation, cpiaucs = CPI based inflation.

Of the five series for which a unit root cannot be rejected, GLS gives better forecasts than OLS in four cases. UP_{PW_1} performs slightly better than UP_{OLS_1} again supporting the recommendation that GLS be used when a unit root is rejected.

5. CONCLUSION

In this paper, we focused on the role played by trend function estimation when forecasting autoregressive time series. We showed that the forecast errors based upon one-step OLS trend estimation and two-step OLS trend estimation have rather different empirical and theoretical properties when the autoregressive root is large. One-step OLS clearly dominates two-step OLS in terms of forecast precision. We then showed that efficient estimation of deterministic trend parameters by GLS may improve forecasts over OLS. Specifically, finite sample simulations and empirical applications show that iterative GLS, especially Prais–Winsten, yields smaller forecast errors than one-step OLS when applied to series with highly persistent errors. Prais–Winsten GLS without iteration is not recommended because this does not lead to jointly optimal trend and autoregressive parameter estimation and can give forecasts inferior to OLS. Iterative Prais–Winsten GLS is preferred over iterative Cochrane–Orcutt GLS because the latter tends to be unreliable for highly persistent time series. In practice, we find one iteration to yield satisfactory and robust results. We also confirmed in this paper that unit root pretests can improve forecast accuracy when the errors have a root close to unity (but, unit root pretests can reduce forecast accuracy for persistent but stationary errors). Whether or not a practitioner chooses to use a unit root pretest, estimation of the trend and the autoregressive parameters by one iteration of Prais–Winsten GLS is recommended when constructing forecasts of autoregressive time series.

ACKNOWLEDGEMENTS

This paper was presented at Penn State University, the University of Maryland, the University of Montreal, and the 2000 Meeting of the Econometric Society in Boston. The authors would like to thank the seminar participants and Ken West for constructive comments. We would also like to thank an anonymous referee and the editor for comments on an earlier draft. Tim Vogelsang thanks the Center for Analytic Economics at Cornell University.

REFERENCES

Abadir, K. M. and K. Hadri (2000). Is more information a good thing? Bias nonmonotonicity in stochastic difference equations. *Bulletin of Economic Research 52*, 91–100.

Abadir, K. M. and P. Paruolo (1997). Two mixed normal densities from cointegrating analysis. *Econometrica 65*, 617–80.

Andrews, D. W. K. (1991). Heteroskedastic and autocorrelation consistent matrix estimation. *Econometrica 59*, 817–54.

Box, G. E. P., G. M. Jenkins and G. C. Reinsel (1994). *Time Series Analysis: Forecasting and Control*. Englewood, NJ: Prentice-Hall.

Campbell, J. Y. and P. Perron (1991). Pitfalls and opportunities: what macroeconomists should know about unit roots. *NBER Macroeconomic Annual*, vol. 6, pp. 141–201. MIT Press.

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

Clements, M. P. and D. Hendry (1994). Towards a theory of economic forecasting. In C. P. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*. Oxford University Press.

Diebold, F. X. (1997). *Elements of Forecasting*. Cincinatti, Ohio: South Western Publishing.

Diebold, F. X. and L. Kilian (2000). Unit root tests are useful for selecting forecasting models. *Journal of Business and Economic Statistics 18*, 265–73.

Elliott, G. (1999). Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *International Economic Review 40*, 767–83.

Elliott, G., T. J. Rothenberg and J. H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica 64*, 813–36.

Goldberger, A. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association 57*, 369–75.

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

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Johnston, J. and J. Dinardo (1997). *Econometric Methods*. New York: Mcgraw Hill.
Phillips, P. (1979). The sampling distribution of forecasts from a first order autoregression. *Journal of Econometrics 9*, 241–61.
Phillips, P. and C. Lee (1996). Efficiency gains from quasi-differencing under non-stationarity. Cowles Foundation Working Paper.
Pindyck, R. S. and D. Rubinfeld (1998). *Econometric Models and Economic Forecasts*. Mcgraw Hill.
Rao, P. and Z. Griliches (1969). Small sample properties of several two stage regression methods in the context of autocorrelated errors. *Journal of the American Statistical Association 64*, 251–72.
Sampson, M. (1991). The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend stationary time-series models. *Journal of Applied Econometrics 6*, 67–76.
Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
Stock, J. H. (1995). Point forecasts and prediction intervals for long horizon forecasts. Mimeo, Kennedy School of Government, Harvard University.
Stock, J. H. (1996). Var, error-correction and pretest forecasts at long horizon. *Oxford Bulletin of Economics and Statistics 58*, 685–701.
Stock, J. H. (1997). Cointegration, long run comovementes and long horizon forecasting. *Advances in Econometrics: Proceedings of the Seventh World Congress of the Econometric Society*. Cambridge University Press.
Stock, J. H. and M. W. Watson (1998). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. Mimeo, Harvard University.
Vogelsang, T. J. (1998). Trend function hypothesis testing in the presence of serial correlation parameters. *Econometrica 65*, 123–48.
West, K. (1996). Asymptotic inference about predictability ability. *Econometrica 64*, 1067–84.

APPENDIX

The following lemma provides asymptotic limits that are used in the derivation of the limiting distribution of the forecast errors. The proof of the lemma is straightforward and hence omitted.

Lemma 5.1. *When* $p = 0$, $T^{-1/2}(\widehat{\delta}_0 - \delta_0) \Rightarrow \int_0^1 J_c^*(r)dr$. *When* $p = 1$, $T^{-1/2}(\widehat{\delta}_0 - \delta_0) \Rightarrow \int_0^1 (4 - 6r) J_c^*(r)dr$, *and* $T^{1/2}(\widehat{\delta}_1 - \delta_1) \Rightarrow \int_0^1 (12r - 6) J_c^*(r)dr$.

Proof of Theorems 2.1 and 2.2. We begin with OLS_2 when $p = 0$.

$$\begin{aligned}
\widehat{e}_{T+1|T} &= (\alpha - \widehat{\alpha})(u_T + \delta_0 - \widehat{\delta}_0) - cT^{-1}(\delta_0 - \widehat{\delta}_0), \\
T^{1/2}\widehat{e}_{T+1|T} &= T(\alpha - \widehat{\alpha})(T^{-1/2}u_T + T^{-1/2}(\delta_0 - \widehat{\delta}_0)) - cT^{-1/2}(\delta_0 - \widehat{\delta}_0), \\
&= cT^{-1/2}(\widehat{\delta}_0 - \delta_0) - T(\widehat{\alpha} - \alpha)T^{-1/2}\widehat{u}_T, \\
&\Rightarrow c[J_c^*(1) - \bar{J}_c^*(1)] - \Phi(\bar{J}_c^*, W)\bar{J}_c^*(1).
\end{aligned}$$

When $p = 1$, $\widehat{m}_t = \widehat{\delta}_0 + \widehat{\delta}_1 t$ and therefore $(1 - \widehat{\alpha}L)(m_{T+1} - \widehat{m}_{T+1}) = (1 - \widehat{\alpha})(m_T - \widehat{m}_T) + (\delta_1 - \widehat{\delta}_1)$. It follows that

$$\begin{aligned}
\widehat{e}_{T+1|T} &= (1 - \widehat{\alpha})(m_T - \widehat{m}_T) + (\delta_1 - \widehat{\delta}_1) + (\alpha - \widehat{\alpha})u_T, \\
&= (\alpha - \widehat{\alpha})(m_T - \widehat{m}_T + u_T) + (\delta_1 - \widehat{\delta}_1) - cT^{-1}(\widehat{u}_T - u_T), \\
T^{1/2}\widehat{e}_{T+1|T} &= cT^{-1/2}(u_T - \widehat{u}_T) - T^{1/2}(\widehat{\delta}_1 - \delta_1) - T(\widehat{\alpha} - \alpha)T^{-1/2}\widehat{u}_T, \\
&\Rightarrow c[J_c^*(1) - \widetilde{J}_c^*(1)] - \int_0^1 (12r - 6)J_c^*(r)dr - \Phi(\widetilde{J}_c^*(r), W)\widetilde{J}_c^*(1).
\end{aligned}$$

Under OLS_1, let $\beta = (\beta_0\ \beta_1)'$. Then

$$\begin{aligned}\widehat{y}_{T+1|T} &= \widehat{\beta}_0 + \widehat{\beta}_1(T+1)\widehat{\alpha}y_T,\\ \widehat{e}_{T+1|T} &= (\beta_0 - \widehat{\beta}_0) + (\beta_1 - \widehat{\beta}_1)(T+1) + (\alpha - \widehat{\alpha})(m_T + u_T),\\ &= -[1,\ T+1](\beta - \widehat{\beta}) + (\alpha - \widehat{\alpha})(m_T + u_T).\end{aligned}$$

We first show that the forecast error is invariant to the true values of δ_0 and δ_1. By partitioned regression, recall that $z_t = (1, t)$, and let $y_{-1} = \{y_0, y_1, \ldots, y_{T-1}\}$. Also let D be a $T \times 2$ matrix with 0 in the first column and 1 in the first column. Then

$$\begin{aligned}\widehat{\beta} - \beta &= (z'z)^{-1}z'e - (z'z)^{-1}z'y_{-1}(\widehat{\alpha} - \alpha),\\ &= (z'z)^{-1}z'e - (z'z)^{-1}z'(z\delta - D\delta + u_{-1})(\widehat{\alpha} - \alpha),\\ &= (z'z)^{-1}z'e - (\delta_0 - \delta_1,\ \ \delta_1)'(\widehat{\alpha} - \alpha) - (z'z)^{-1}z'u_{-1}(\widehat{\alpha} - \alpha).\end{aligned}$$

Substituting this result into the expression for $\widehat{e}_{T+1|T}$, we have

$$\begin{aligned}\widehat{e}_{T+1|T} &= -[1, T+1][(z'z)^{-1}z'e - (z'z)^{-1}z'u_{-1}(\widehat{\alpha} - \alpha)]\\ &\quad + (\delta_0 - \delta_1 + \delta_1 T + \delta_1)(\widehat{\alpha} - \alpha) + (\alpha - \widehat{\alpha})(m_T + u_T),\\ &= -[1, T+1][(z'z)^{-1}z'e - (z'z)^{-1}z'u_{-1}(\widehat{\alpha} - \alpha)] + (\alpha - \widehat{\alpha})u_T,\end{aligned}$$

which does not depend on δ. Therefore, without loss of generality, we let $\delta = 0$ so that $y_T = u_T$. Consider the artificial regression $x_t = A_0 + A_1 t + e_t$ where e_t is white noise. Then $\widehat{\beta} - \beta$ and $\widehat{e}_{T+1|T}$ simplify to

$$\begin{aligned}\widehat{\beta} - \beta &= (z'z)^{-1}z'e - (z'z)^{-1}z'u_{-1}(\widehat{\alpha} - \alpha),\\ &\equiv \begin{bmatrix}[c]c\widehat{A}_0 - A_0\\ \widehat{A}_1 - A_1\end{bmatrix} - \begin{bmatrix}[c]c\widehat{\delta}_0 - \delta_0\\ \widehat{\delta}_1 - \delta_1\end{bmatrix}(\widehat{\alpha} - \alpha),\\ \widehat{e}_{T+1|T} &= (\beta_0 - \widehat{\beta}_0) + (\beta_1 - \widehat{\beta}_1)(T+1) + (\alpha - \widehat{\alpha})u_T.\end{aligned}$$

Therefore,

$$\begin{aligned}\widehat{e}_{T+1|T} &= (\delta_0 - \widehat{\delta}_0) - (\widehat{A}_0 - A_0) + (\delta_1 - \widehat{\delta}_1)(\alpha - \widehat{\alpha})(T+1) - (\widehat{A}_1 - A_1)(T+1)\\ &\quad +(\alpha - \widehat{\alpha})u_T,\\ &= (\alpha - \widehat{\alpha})(m_T - \widehat{m}_T + u_T) + (\delta_1 - \widehat{\delta}_1)(\alpha - \widehat{\alpha}) - (\widehat{A}_0 - A_0)\\ &\quad -(\widehat{A}_1 - A_1)(T+1),\\ &= (\alpha - \widehat{\alpha})\widehat{u}_T + (\delta_1 - \widehat{\delta}_1)(\alpha - \widehat{\alpha}) - (\widehat{A}_0 - A_0) - (\widehat{A}_1 - A_1)(T+1),\\ T^{1/2}\widehat{e}_{T+1|T} &= T(\alpha - \widehat{\alpha})T^{-1/2}\widehat{u}_T - T^{1/2}(\widehat{A}_0 - A_0) - T^{3/2}(\widehat{A}_1 - A_1) + o_p(1)\\ &\Rightarrow -\Phi(\widetilde{J}_c^*, W)\widetilde{J}_c^*(1) - \int_0^1 (6r - 2)dW(r).\end{aligned}$$

For $p = 0$, the last term A_1 does not exist and $A_0 = T^{-1}\sum_{t=1}^T e_t$. Thus,

$$\begin{aligned}T^{1/2}\widehat{e}_{T+1|T} &= T(\alpha - \widehat{\alpha})\widehat{u}_T - T^{-1/2}\sum_{t=1}^{T} e_t\\ &\Rightarrow -\Phi(\bar{J}_c^*, W)\bar{J}_c^*(1) - W(1).\end{aligned}$$

□

GLS Detrending

Proof of Theorems 3.1 and 3.2. Recall that $\widetilde{\delta}$ denotes the feasible GLS estimates of δ. Using the algebraic results from OLS_2 we have for CO and PW:

$$p = 0: \qquad T^{1/2}\widetilde{e}_{T+1|T} = cT^{-1/2}(u_T - \widetilde{u}_T) - T(\ddot{\alpha} - \alpha)T^{-1/2}\widetilde{u}_T,$$
$$p = 1: \qquad T^{1/2}\widetilde{e}_{T+1|T} = cT^{-1/2}(u_T - \widetilde{u}_T) - T^{1/2}(\widetilde{\delta}_1 - \delta_1) - T(\ddot{\alpha} - \alpha)T^{-1/2}\widetilde{u}_T,$$

where $T^{-1/2}\widetilde{u}_T = T^{-1/2}u_T - T^{-1/2}(\widetilde{\delta}_0 - \delta_0)$ for $p = 0$ and $T^{-1/2}\widetilde{u}_T = T^{-1/2}u_T - T^{-1/2}(\widetilde{\delta}_0 - \delta_0) - T^{1/2}(\widetilde{\delta}_1 - \delta_1)$ for $p = 1$. Given that $T^{-1/2}u_T \Rightarrow J_c^*(1)$ and $T(\ddot{\alpha} - \alpha) \Rightarrow (\ddot{c} - c)$, all that remains to be established are the limits of $T^{-1/2}(\widetilde{\delta}_0 - \delta_0)$ and $T^{1/2}(\widetilde{\delta}_1 - \delta_1)$ for feasible CO and PW detrending.

The results for $T^{1/2}(\widetilde{\delta}_1 - \delta_1)$ for $p = 1$ can be found in Canjels and Watson (1997):

$$CO: T^{1/2}(\widetilde{\delta}_1 - \delta_1) \Rightarrow \dot{c}^{-1}\int_0^1 (6 - 12s)d\dot{W}(s),$$
$$PW: T^{1/2}(\widetilde{\delta}_1 - \delta_1) \Rightarrow \dot{\theta}\left(\int_0^1 (1 - \dot{c}s)d\dot{W}(s) + \dot{c}(1 - \tfrac{1}{2}\dot{c})J_c^-(\kappa)\right).$$

Using the limiting results in the appendix of Canjels and Watson (1997), it is a simple algebraic exercise to show that for $p = 1$:

$$CO: T^{1/2}(\widetilde{\delta}_0 - \delta_0) \Rightarrow \dot{c}^{-2}\int_0^1 (6 - 4\dot{c} - 12s + 6s\dot{c})d\dot{W}(s),$$
$$PW: T^{1/2}(\widetilde{\delta}_0 - \delta_0) \Rightarrow J_c^-(\kappa),$$

and for $p = 0$:

$$CO: T^{1/2}(\widetilde{\delta}_0 - \delta_0) \Rightarrow -\dot{c}^{-1}\int_0^1 d\dot{W}(s),$$
$$PW: T^{1/2}(\widetilde{\delta}_0 - \delta_0) \Rightarrow J_c^-(\kappa).$$

The final expressions in the theorems then follow from algebraic simplification. □