Multi-Step Estimation for Forecasting

Michael P. Clements and David F. Hendry*
Department of Economics, University of Warwick
Nuffield College, Oxford.

March 31, 2008

Abstract

We delineate conditions which favour multi-step, or dynamic, estimation for multi-step forecasting. An analytical example shows how dynamic estimation (DE) may accomodate incorrectly-specified models as the forecast lead alters, improving forecast peformance for some mis-specifications. However, in correctly-specified models, reducing finite-sample biases does not justify DE. In a Monte Carlo forecasting study for integrated processes, estimating a unit root in the presence of a neglected negative moving-average error may favour DE, though other solutions exist to that scenario. A second Monte Carlo study obtains the estimator biases and explains these using asymptotic approximations.

1 Introduction

Minimizing multi-step (in-sample) criteria for estimating unknown parameters has a long pedigree, although the method does not seem to have been subject to many formal analyses (see e.g. Klein, 1971). Cox (1961) applied this idea to the exponentially-weighted moving average (EWMA) or integrated moving-average IMA(1,1) model, and Findley (1983) and Weiss (1991) among others considered multi-step estimation criteria for autoregressive (AR) models. The intuition is that when a model is not well specified, minimization of *1*-step errors need not deliver reliable forecasts at longer lead times, so estimation by minimizing the in-sample counterpart of the desired step-ahead horizon may yield better forecasts.

When models are mis-specified for the data generation process (DGP), mean-square forecast error (MSFE(h)) rankings can alter as the forecast horizon h increases. Indeed, a necessary condition for such a result in large samples is that the models under consideration are mis-specified.[1] This implication depends only on the DGP providing the correct conditional expectation, which is the minimum MSFE predictor. As it is not possible to prove that *1*-step estimation is optimal when models are mis-specified, multi-step, or dynamic, estimation (DE, also called 'adaptive forecasting': Tsay, 1993, and Lin and Tsay, 1995) could improve multi-period forecast accuracy.

Here, we investigate model mis-specifications which may sustain DE. Empirical studies have been used by some authors (e.g. Lin and Tsay, 1995) to gauge the practical usefulness of DE, from which it is difficult to deduce precisely which characteristics are responsible for the outcomes. We consider a

*Financial support from the UK Economic and Social Research Council under grant R000233447 is gratefully acknowledged by both authors. Thanks to Agustin Maravall and Grayham Mizon for helpful discussions. A longer version of this paper is available as Warwick Economic Research Paper No.447, University of Warwick.

[1]However, mis-specification is not sufficient for reversals of forecast-accuracy ranking to occur, since model selection based on MSFE is not invariant under linear transformations when $h > 1$: Clements and Hendry (1993).

simple analytic example to illustrate some of the issues involved, before presenting a Monte Carlo study to identify features which might favour DE for multi-period forecasting.

The basic formulation is as follows. Consider an h-period ahead forecast from a vector autoregression (VAR) for the n variables $\mathbf{x}_t$:

$$\mathbf{x}_t = \boldsymbol{\Upsilon}\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t \tag{1}$$

where we only specify that $\mathsf{E}[\boldsymbol{\epsilon}_t] = \mathbf{0}$. h-periods ahead from an end-of-sample point T:

$$\mathbf{x}_{T+h} = \boldsymbol{\Upsilon}^h \mathbf{x}_T + \sum_{i=0}^{h-1} \boldsymbol{\Upsilon}^i \boldsymbol{\epsilon}_{T+h-i}. \tag{2}$$

A corresponding forecast is made, either from a 'powered-up' *1*-step parameter estimator or using an *h*-step estimator. The *1*-step estimator is defined by:

$$\widehat{\boldsymbol{\Upsilon}} = \underset{\boldsymbol{\Upsilon}}{\text{argmin}} \left| \sum_{t=1}^{T} \left(\mathbf{x}_t - \boldsymbol{\Upsilon}\mathbf{x}_{t-1}\right)\left(\mathbf{x}_t - \boldsymbol{\Upsilon}\mathbf{x}_{t-1}\right)' \right| = \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_{t-1}' \left(\sum_{t=1}^{T} \mathbf{x}_{t-1}\mathbf{x}_{t-1}' \right)^{-1}. \tag{3}$$

with forecasts given by:

$$\hat{\mathbf{x}}_{T+h} = \widehat{\boldsymbol{\Upsilon}}^h \mathbf{x}_T \tag{4}$$

and with average conditional error:

$$\mathsf{E}\left[\mathbf{x}_{T+h} - \hat{\mathbf{x}}_{T+h} \mid \mathbf{x}_T\right] = \left(\boldsymbol{\Upsilon}^h - \mathsf{E}\left[\widehat{\boldsymbol{\Upsilon}}^h\right]\right)\mathbf{x}_T. \tag{5}$$

The *h*-step estimator is defined by:

$$\widetilde{\boldsymbol{\Upsilon}^h} = \underset{(\boldsymbol{\Upsilon}^h)}{\text{argmin}} \left| \sum_{t=1}^{T} \left(\mathbf{x}_t - \left(\boldsymbol{\Upsilon}^h\right)\mathbf{x}_{t-h}\right)\left(\mathbf{x}_t - \left(\boldsymbol{\Upsilon}^h\right)\mathbf{x}_{t-h}\right)' \right| = \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_{t-h}' \left(\sum_{t=1}^{T} \mathbf{x}_{t-h}\mathbf{x}_{t-h}' \right)^{-1} \tag{6}$$

so that using the *h*-step estimator of the 'powered-up' parameter directly:

$$\tilde{\mathbf{x}}_{T+h} = \widetilde{\boldsymbol{\Upsilon}^h}\mathbf{x}_T, \tag{7}$$

with average conditional error:

$$\mathsf{E}\left[\mathbf{x}_{T+h} - \tilde{\mathbf{x}}_{T+h} \mid \mathbf{x}_T\right] = \left(\boldsymbol{\Upsilon}^h - \mathsf{E}\left[\widetilde{\boldsymbol{\Upsilon}^h}\right]\right)\mathbf{x}_T. \tag{8}$$

The relative accuracy of the multi-step forecast procedure (7) compared to (4) is determined by that of the powered estimate versus the estimated power. When $\widehat{\boldsymbol{\Upsilon}}$ is badly biased for $\boldsymbol{\Upsilon}$, powered estimated values will deviate increasingly from the powered 'true' values, in which case, direct estimation of $\boldsymbol{\Upsilon}^h$ may have potential. Alternatively, when:

$$\mathsf{E}\left[\mathbf{x}_{T+1} \mid \mathbf{x}_T\right] = \boldsymbol{\Psi}\mathbf{x}_T,$$

but:

$$\mathsf{E}\left[\mathbf{x}_{T+h} \mid \mathbf{x}_T\right] \neq \boldsymbol{\Psi}^h\mathbf{x}_T,$$

powering up $\widehat{\boldsymbol{\Psi}}$ may prove a poor strategy, no matter how precisely it is estimated.

However, the issue is more subtle than this. First, in stationary processes, dynamic mis-specification *per se* is not sufficient to ensure poor multi-step forecasts since *1*-step estimation of $\boldsymbol{\Upsilon}$ in (1) yields the least-squares approximation and *h*-step forecasts converge on the unconditional expectation with $\boldsymbol{\Upsilon}^h$

tending to zero. Since increasing divergence from powered estimates (that did not accidentally capture unit roots) seems unlikely, we will focus on integrated processes.

Secondly, when $\boldsymbol{\epsilon}_t \sim \mathsf{IN}_n(\mathbf{0}, \boldsymbol{\Omega})$, finite-sample biases are unlikely to be large enough to offset the inefficiency of DE (see section 4). Thus, mis-specification seems required, so we consider the third possibility of unmodelled moving-average (MA) errors in the DGP, such that in (1):

$$\boldsymbol{\epsilon}_t = \boldsymbol{\theta}(L)\,\boldsymbol{\zeta}_t, \quad \text{where} \quad \boldsymbol{\zeta}_t \sim \mathsf{IN}_n(\mathbf{0}, \boldsymbol{\Sigma}),$$

L denotes the lag operator, and $\boldsymbol{\theta}_0 = \mathbf{I}_n$. When $\boldsymbol{\theta}(L) = \mathbf{I}_n + \boldsymbol{\theta}_1 L$:

$$\mathsf{E}[\mathbf{x}_{T+1} \mid \mathbf{x}_T] = \boldsymbol{\Upsilon}\mathbf{x}_T + \boldsymbol{\theta}_1 \mathsf{E}[\boldsymbol{\zeta}_T \mid \mathbf{x}_T] = (\boldsymbol{\Upsilon} + \boldsymbol{\theta}_1\boldsymbol{\Lambda})\,\mathbf{x}_T = \boldsymbol{\Psi}\mathbf{x}_T,$$

where $\mathsf{E}[\boldsymbol{\zeta}_t \mid \mathbf{x}_t] = \boldsymbol{\Lambda}\mathbf{x}_t$, but:

$$\mathsf{E}[\mathbf{x}_{T+h} \mid \mathbf{x}_T] = \boldsymbol{\Upsilon}^h\mathbf{x}_T + \boldsymbol{\Upsilon}^{h-1}\boldsymbol{\theta}_1 \mathsf{E}[\boldsymbol{\zeta}_T \mid \mathbf{x}_T] = \boldsymbol{\Upsilon}^{h-1}(\boldsymbol{\Upsilon} + \boldsymbol{\theta}_1\boldsymbol{\Lambda})\,\mathbf{x}_T = \boldsymbol{\Upsilon}^{h-1}\boldsymbol{\Psi}\mathbf{x}_T.$$

Defining:

$$\mathsf{E}[\mathbf{x}_{T+h} \mid \mathbf{x}_T] = \boldsymbol{\Psi}_h\mathbf{x}_T$$

then $\boldsymbol{\Psi}_h = \boldsymbol{\Psi}^h$ only if $\boldsymbol{\theta}(L) = \mathbf{I}_n$, in which case $\boldsymbol{\Psi} = \boldsymbol{\Upsilon}$. (5) and (8) remain intact upon replacing $\boldsymbol{\Upsilon}^h$ by $\boldsymbol{\Psi}_h$, but $\mathsf{E}[\widehat{\boldsymbol{\Psi}}^h] \neq \boldsymbol{\Psi}_h$. This analysis suggests it is worth examining whether the substantive biases due to the interaction between estimated unit roots and neglected negative MA errors discussed by (*inter alia*) Molinas (1986) and Schwert (1989) could justify DE over *1*-step estimation. Such transpires to be the case, but we also consider other solutions to that cause of mis-specification, including using *1*-step instrumental-variables (IV) estimators following Hall (1989) .

The plan of the paper is as follows. Section 2 describes the relationships between estimation criteria and forecast evaluation. Section 3 outlines a simplified version of the forecast-error taxonomy of Clements and Hendry (1995b, 1996) to delineate the roles of estimation and model mis-specification for *1*-step and DE. Section 4 shows that for correctly specified models, there is little finite-sample bias reduction from *2*-step estimation versus squaring the *1*-step estimator in a first-order AR. Section 5 reports the Monte Carlo forecasting study, showing that neglected moving-average errors have relatively benign effects on medium-term forecast performance using OLS, except in the presence of unit roots, parallelling the poor performance of unit-root tests in this situation. In section 6, we consider the IV estimator of Hall (1989) and a hybrid IV multi-step estimator. Their asymptotic distributions are derived for *h*-step forecasts, together with a Monte Carlo study. Section 7 concludes.

2 Estimation and evaluation criteria

Granger (1993) suggests that, at least asymptotically, the criterion used to evaluate forecasts (C_F) should also be used to estimate the parameters of the forecasting model (C_E). In the context of assessing multi-step forecasts, Weiss (1991) shows that (in large samples) parameter estimation based on minimizing squared in-sample *h*-step forecast errors is optimal if C_F is a squared-error loss function defined over *h*-step forecast errors. His proof assumes stationarity, and the Monte Carlo evidence of the small-sample performance of DE in stationary DGPs indicates that OLS is generally as good as DE. Stoica and Nehorai (1989) give similar asymptotic results for ARMA processes, and their Monte Carlo also shows small gains from DE (a possible exception being when the model is 'under-parameterized' for the DGP).[2] They conjecture that EWMA schemes may yield gains from DE because the processes to

[2] Based on simulating an ARMA(2,2) DGP and an AR(1) model.

be predicted are higher order than an IMA(1,1) model. More recently, Tsay (1993) and Lin and Tsay (1995) offer an upbeat assessment based on empirical studies. Weiss (1995) explores procedures which base estimation on the forecast criterion function for a variety of forecast criteria, such as asymmetric quadratic loss: also see Christoffersen and Diebold (1994).

We consider the gains (if any) from matching estimation and evaluation criteria, taking C_F as the MSFE loss function. Unfortunately, rankings by such C_F-measures, for a given C_E, may depend on the selected linear transformations of the C_F arguments (levels, differences, etc.), perhaps necessitating a different DE in each instance.

3 A partial taxonomy of forecast errors

Consider a forecast for a vector of variables $\mathbf{x}_t$ at horizon h from an estimated model M based on the conditional expectation with respect to the model where the information set is $\mathcal{I}_t$:[3]

$$\mathbf{x}_{M,T+h} = \mathsf{E}_{\mathsf{M}}\left[\mathbf{x}_{T+h} \mid \mathcal{I}_T\right],$$

Typically M will contain unknown parameters, $\mathbf{\Phi}$, which need to be estimated, so a complete description of the forecast generation process will require both the specification of the model and the estimation criterion C_E. Using estimates $\widehat{\mathbf{\Phi}}$, the operational forecast errors are:

$$\widehat{\boldsymbol{\nu}}_{T+h} = \mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{M,T+h}.$$

To contrast *1*-step and multi-step estimation, we consider the special case of the forecast-error taxonomy in Clements and Hendry (1996) where the DGP is constant with accurate initial conditions $\mathbf{x}_T$. The degree of integration of the data, and whether unit roots are imposed or estimated, also matters: see Clements and Hendry (1995a) and Lin and Tsay (1995).

The issues are most easily seen in a first-order univariate AR model with an intercept when the DGP is the ARMA(1,q):

$$\begin{aligned} x_{T+h} &= \tau + \Upsilon x_{T+h-1} + \nu_{T+h} \\ &= \sum_{i=0}^{h-1} \Upsilon^i \tau + \Upsilon^h x_T + \sum_{i=0}^{h-1} \Upsilon^i \nu_{T+h-i} \end{aligned} \tag{9}$$

where $\nu_t = \theta(L)\zeta_t$ and $\zeta_t \sim \mathsf{IN}(0, \sigma_\zeta^2)$. M is:

$$\mathsf{M}\text{: } x_t = \tau_p + \Upsilon_p x_{t-1} + v_t$$

so $\mathbf{\Phi} = (\tau_p : \Upsilon_p)$. We allow $\tau_p \neq \tau$ and $\Upsilon_p \neq \Upsilon$, because $\theta(L) \neq 1$ but v_t is treated as white noise. The *1*-step parameter estimates $(\hat{\tau}_p : \hat{\Upsilon}_p)$ yield the h-step ahead forecasts:

$$\hat{x}_{T+h} = \sum_{i=0}^{h-1} \hat{\Upsilon}_p^i \hat{\tau}_p + \hat{\Upsilon}_p^h x_T. \tag{10}$$

From (9) and (10), the *h*-step ahead forecast errors are:

$$\hat{\nu}_{T+h} = \sum_{i=0}^{h-1} \left[\Upsilon^i \tau - \hat{\Upsilon}_p^i \hat{\tau}_p\right] + \left(\Upsilon^h - \hat{\Upsilon}_p^h\right) x_T + \sum_{i=0}^{h-1} \Upsilon^i \nu_{T+h-i}. \tag{11}$$

[3] In some states of nature it may be desirable to 'correct' the conditional expectation: see Clements and Hendry (1994a, 1995b) and Hendry and Clements (1994b, 1994a).

Denote deviations between sample estimates and population parameters by $\delta_\tau = \hat{\tau}_p - \tau_p$ and $\delta_\Upsilon = \hat{\Upsilon}_p - \Upsilon_p$. We neglect powers of and interactions between δ_k and additional finite-sample biases in $\hat{\tau}_p$ and $\hat{\Upsilon}_p$ around their plims, using the approximation:

$$\hat{\Upsilon}_p^i = (\Upsilon_p + \delta_\Upsilon)^i \approx \Upsilon_p^i + i\Upsilon_p^{i-1}\delta_\Upsilon = \Upsilon_p^i + C_i,$$

to obtain:

$$\hat{\Upsilon}_p^i \hat{\tau}_p \approx \left(\Upsilon_p^i + C_i\right)(\tau_p + \delta_\tau) \approx \Upsilon_p^i \tau_p + C_i \tau_p + \Upsilon_p^i \delta_\tau,$$

so using $\sum_{i=0}^{h-1} \Upsilon^i = (1-\Upsilon^h)/(1-\Upsilon) = \Upsilon^{\{h\}}$ (similarly for Υ_p and C), the first term in (11) is:

$$\sum_{i=0}^{h-1}\left[\Upsilon^i(\tau - \tau_p) + \left(\Upsilon^i - \Upsilon_p^i\right)\tau_p - \left(C_i\tau_p + \Upsilon_p^i\delta_\tau\right)\right] = \Upsilon^{\{h\}}(\tau - \tau_p) + \left(\Upsilon^{\{h\}} - \Upsilon_p^{\{h\}}\right)\tau_p - \left(C^{\{h\}}\tau_p + \Upsilon_p^{\{h\}}\delta_\tau\right)$$

The term multiplied by x_T is approximated by:

$$\left(\Upsilon^h - \hat{\Upsilon}_p^h\right)x_T = \left(\Upsilon^h - \Upsilon_p^h\right)x_T + \left(\Upsilon_p^h - \hat{\Upsilon}_p^h\right)x_T \approx \left(\Upsilon^h - \Upsilon_p^h\right)x_T - C_h x_T.$$

This leads to the decomposition in table 1.

Table 1 Forecast-error taxonomy for *1*-step estimation.

$\hat{\nu}_{T+h}$	$\simeq$	$\left(\Upsilon^{\{h\}} - \Upsilon_p^{\{h\}}\right)\tau_p + \left(\Upsilon^h - \Upsilon_p^h\right)x_T$	slope mis-specification
	$+$	$\Upsilon^{\{h\}}(\tau - \tau_p)$	intercept mis-specification
	$-$	$C_i^{\{h\}}\tau_p - C_h x_T$	slope estimation
	$-$	$\Upsilon_p^{\{h\}}\delta_\tau$	intercept estimation
	$+$	$\sum_{i=0}^{h-1}\Upsilon^i \nu_{T+h-i}$	error accumulation.

Conditional on x_T, the first two rows only have bias effects, whereas the remaining rows affect forecast error variances.

Consider the alternative of estimating $\tau_h = \sum_{i=0}^{h-1}\Upsilon^i\tau = \Upsilon^{\{h\}}\tau$ and $\Upsilon_h = \Upsilon^h$ by DE in:

$$x_t = \tau_h + \Upsilon_h x_{t-h} + u_t,$$

and forecasting by:

$$\tilde{x}_{T+h} = \widetilde{\tau}_h + \widetilde{\Upsilon}_h x_T. \tag{12}$$

From (9) and (12), the *h*-step ahead forecast error is:

$$\tilde{\nu}_{T+h} = (\tau_h - \widetilde{\tau}_h) + \left(\Upsilon_h - \widetilde{\Upsilon}_h\right)x_T + \sum_{i=0}^{h-1}\Upsilon^i\nu_{T+h-i}. \tag{13}$$

Denote deviations between multi-step sample estimates and population parameters by $\lambda_\tau = \widetilde{\tau}_h - \tau_h^*$ and $\lambda_\Upsilon = \widetilde{\Upsilon}_h - \Upsilon_h^*$, and neglect powers of λ_k and finite-sample biases in $\widetilde{\tau}_h$ and $\widetilde{\Upsilon}_h$ around their plims τ_h^* and Υ_h^*. Then, the corresponding forecast-error taxonomy for DE is shown in table 2.

Only the last row is in common with table 1, and there are no interactions between slope and intercept as in table 1: any of the remaining terms could be larger or smaller in mean and/or variance depending upon the specific example.

The Monte Carlo in section 5 calculates the relative magnitudes of the mis-specification and estimation effects for both OLS and DE for a number of examples.

Table 2 Forecast-error taxonomy for multi-step estimation.

$\tilde{\nu}_{T+h}$	$\simeq$	$(\Upsilon_h - \Upsilon_h^*)\, x_T$	slope mis-specification
	$+$	$(\tau_h - \tau_h^*)$	intercept mis-specification
	$-$	$\lambda_\Upsilon x_T$	slope estimation
	$-$	λ_τ	intercept estimation
	$+$	$\sum_{i=0}^{h-1} \Upsilon^i \nu_{T+h-i}$	error accumulation.

4 Finite-sample behaviour under correct specification

An alternative reason for minimizing a *2*-step criterion when *2*-step forecasts are desired in dynamic models is the possible reduction in finite-sample bias that might result from doing so even when the models are correctly specified. As an illustration, suppose that the model is correctly specified for a stationary AR(1) with an intercept DGP. Then we obtain a finite sample bias for *1*-step estimation of the AR parameter, ρ:

$$\mathsf{E}\left[\hat{\rho}\right] \simeq \rho - \frac{1+3\rho}{T}. \tag{14}$$

The analysis of the bias in $\widetilde{\rho^2}$ (the *2*-step estimator) is close in form to that for the bias in the *1*-step estimator (see e.g. Hendry, 1984, for an exposition) and to $O(1/T)$ suggests that $\mathsf{E}[\widetilde{\rho^2}]$ is the square of (14). When $\rho = 0$, retaining terms of $O(1/T)$ only, we obtain:

$$\mathsf{E}\left[\widetilde{\rho^2} \mid \rho = 0\right] \simeq -\frac{1}{T},$$

as indicated in the figure. For OLS we obtain a bias of roughly equal magnitude but opposite sign:

$$\mathsf{E}\left[\hat{\rho}^2 \mid \rho = 0\right] = \left(\mathsf{E}\left[\hat{\rho} \mid \rho = 0\right]\right)^2 + \mathsf{V}\left[\hat{\rho} \mid \rho = 0\right] = \frac{1}{T^2} + \frac{1}{T}.$$

This result can be confirmed by Monte Carlo simulation using PcNaive (see Hendry, Neale and Ericsson, 1991). We selected $\rho = 0, 0.4, 0.8, 1.0$ and $T = 100$, and conducted the simulations recursively from $t = 10, \ldots, 100$ with 10,000 replications. Figure 1 plots $\mathsf{E}[\hat{\rho}^2 - \rho^2]$ (given by the dotted line) and $\mathsf{E}[\widetilde{\rho^2} - \rho^2]$ (the solid line) for the four values of ρ, where the horizontal axes are the estimation sample size t. The two approaches are similarly biased at large values of ρ and all sample sizes, so there is no gain in terms of finite-sample bias from direct *2*-step estimation. Biases are smaller at smaller ρ, but OLS still dominates, except perhaps at $\rho = 0$. The case of $\rho = 1$ is analysed in section 6.

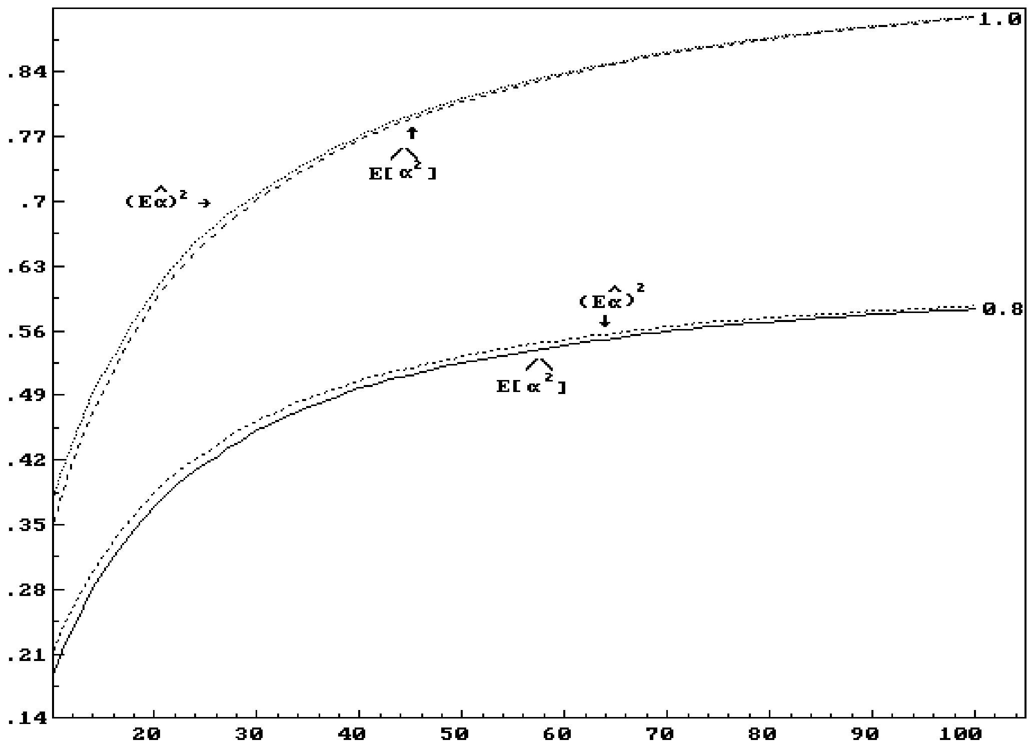

Figure 1: Monte Carlo biases for a correctly specified first-order autoregression

5 A Monte Carlo study of multi-step forecasting

On the basis of these results, we conjecture that neither sampling variability (where *1*-step should do best), nor (stationary) finite-sample bias are likely to indicate a role for DE. Alternatively, any divergence between $(\Psi_h - \Upsilon_p^h)$ on the one hand, versus $(\Psi_h - \Upsilon_h^*)$ on the other, might be important, particularly for integrated data where not all the terms are tending to zero in the forecast horizon. Recall Ψ_h is defined by $\mathsf{E}[x_{T+h} \mid x_T] = \Psi_h x_T$, and $\Psi_h = \Upsilon^h$ only when the DGP can be represented as an AR(1) process (ruling out MA errors). Thus, we designed the Monte Carlo to highlight potential biases when estimating unit-root processes yet neglecting negative moving-average errors.

The DGP is the 'nonseasonal Holt model' (see, among others, Ord, 1988, and Harvey, 1989), generating the data as the sum of unobserved components for the trend, intercept, and irregular elements:

$$x_t = \mu_t + \epsilon_t \tag{15}$$

$$\mu_t = \mu_{t-1} + \beta_t + \delta_{1t} \tag{16}$$

$$\beta_t = \beta_{t-1} + \delta_{2t}. \tag{17}$$

In (15)–(17), the disturbances ϵ_t, δ_{1t} and δ_{2t} are assumed to be normal, independent through time and of each other (at all lags and leads), with zero means, and variances σ_ϵ^2, $\sigma_{\delta_1}^2$, $\sigma_{\delta_2}^2$. The 'reduced form' implies the restricted ARIMA(0,2,2) model:

$$(1-L)^2 x_t = (1 - \theta_1 L - \theta_2 L^2) a_t \tag{18}$$

where the θ_i can be deduced by matching moments in (18) and (19), the latter obtained directly from (15)–(17) (see Harvey, 1989):

$$(1-L)^2 x_t = \epsilon_t + \delta_{1t} + \delta_{2t} - 2\epsilon_{t-1} - \delta_{1t-1} + \epsilon_{t-2}. \quad (19)$$

Zero restrictions on the variances σ^2_ϵ, $\sigma^2_{\delta_1}$, $\sigma^2_{\delta_2}$ enable data with different orders of integration to be generated from (15)–(17). These are summarised in Table 3.

Table 3 Monte Carlo Design.

	Parameter Restrictions	DGP	Description
I	None	See (18)	ARIMA(0,2,2)
	σ^2_ϵ $\sigma^2_{\delta_1}$ $\sigma^2_{\delta_2}$		
i	1 1 1		
ii	1 .5 1		
iii	1 1 .5		
II	$\sigma^2_{\delta_2} = 0, \quad \beta_t = \beta \neq 0$	$\Delta x_t = \beta + \delta_{1t} + \Delta\epsilon_t$	ARIMA(0,1,1) plus drift
III	$\sigma^2_{\delta_2} = 0, \quad \beta_t = \beta = 0$	$\Delta x_t = \delta_{1t} + \Delta\epsilon_t$	ARIMA(0,1,1)
IV	$\sigma^2_{\delta_2} = 0,$ $\beta_t = \beta \neq 0, \sigma^2_\epsilon = 0$	$\Delta x_t = \beta + \delta_{1t}$	ARIMA(0,1,0) plus drift
V	$\sigma^2_{\delta_2} = 0,$ $\beta_t = \beta = 0, \sigma^2_\epsilon = 0$	$\Delta x_t = \delta_{1t}$	ARIMA(0,1,0)
VI	$\sigma^2_{\delta_2} = \sigma^2_{\delta_1} = 0, \quad \beta \neq 0$	$x_t = \beta t + \epsilon_t$	Linear trend
VII	$\sigma^2_{\delta_2} = \sigma^2_{\delta_1} = 0, \quad \beta = 0$	$x_t = \epsilon_t$	White noise
VIII	$\sigma^2_{\delta_1} = 0, \quad \sigma^2_\epsilon = 0$	$(1-\tau_1 L)(1-\tau_2 L) x_t = \delta_{2t}$	
i	$\|\tau_1\| = \|\tau_2\| = 1$		ARIMA(0,2,0)
ii	$\|\tau_1\| = 1, \|\tau_2\| < 1$		ARIMA(1,1,0)
iii	$\|\tau_1\| < 1, \|\tau_2\| < 1$		ARIMA(2,0,0)
IX		$\Delta x_t = \tau_1 \Delta x_{t-1} + \delta_{2t} + \Delta\delta_{1t}$	
i	$\sigma^2_{\delta_1} = \sigma^2_{\delta_2} = 1, \sigma^2_\epsilon = 0,$ $\tau_1 = 0.5, \tau_2 = 1$		ARIMA(1,1,1)
ii	$\sigma^3_{\delta_1} = \sigma^2_{\delta_2} = 1, \sigma^2_\epsilon = 0,$ $\tau_1 = 0.1, \tau_2 = 1$		ARIMA(1,1,1)

We consider six AR forecasting models. M_1 is an AR(2) process for the levels, formulated in differences (i.e. with a unit root imposed):

$$\mathsf{M}_1\colon \Delta x_t = \alpha_0 + \alpha_1 \Delta x_{t-1} + \nu_t \quad (20)$$

and M^*_1 is the same model but with $\alpha_0 = 0$. M_2 is the same AR(2) model, but in levels without the unit root imposed, and M^*_2 is M_2 without a constant term. Finally, we consider an AR(1) model in levels, with a constant term (M_3), and without (M^*_3).

Minimizing the in-sample *h*-step errors is non-linear in the parameters of the original models, but can be approximated by a linear projection (and thus by OLS).[4] For example, M_1 in (20) implies *h*-step errors ($\nu_{h,t}$) of the form:

$$\Delta x_t = \alpha_0 \sum_{i=0}^{h-1} \alpha_1^i + \alpha_1^h \Delta x_{t-h} + \sum_{i=0}^{h-1} \alpha_1^i \nu_{t-i} = \alpha_{0,h} + \alpha_{1,h} \Delta x_{t-h} + \nu_{h,t}. \qquad (21)$$

We estimate $\boldsymbol{\alpha} = [\alpha_0 : \alpha_1]'$ by minimizing the sums of squares of $\nu_{h,t}$ with respect to the parameters:

$$\underset{\boldsymbol{\alpha}}{\text{argmin}} \sum_{t=h+1}^{T} \left(\Delta x_t - \alpha_0 \sum_{i=0}^{h-1} \alpha_1^i - \alpha_1^h \Delta x_{t-h} \right)^2 \qquad (22)$$

yielding a function which is non-linear in the parameters α_0 and α_1. The same outcome results by projecting Δx_t on a constant and Δx_{t-h} by OLS, defining $\boldsymbol{\alpha} = [\alpha_{0,h} : \alpha_{1,h}]'$ as in (21) since for this model, the form of (22) does not impose any restrictions on $[\alpha_{0,h} : \alpha_{1,h}]$ for *h*-step estimation. However, the parameters of the *1*-period model $(\alpha_0 : \alpha_1)$ are not necessarily uniquely identifiable from the OLS projection (e.g. consider $h = 2$).

In all cases we use OLS to estimate the models, and calculate *1*- to *h*-step MSFEs for *1*-step estimation but only *h*-period MSFEs for *h*-step estimation.

5.1 Analysis of results

There are two basic sets of results: (a) Table 6 reports MSFEs for predicting Δx_t using models M_1 and M_1^*; and (b) Table 7 shows MSFEs for predicting x_t using models M_2, M_2^* and M_3, M_3^* (the AR(2) and AR(1) models).

Tables 6 and 7 are summarised in tables 4 and 5 using response surfaces to highlight the conditions which favoured DE over OLS. The dependent variable is the log of the ratio of the MSFE for DE to that for OLS: in table 4 for *2*-steps ahead, and in table 5 for *4*-step ahead forecasts. Four regressions are reported in each table. The first pools results from table 6 (MSFEs for predicting Δx_t using M_1, labelled experiments 1–14) and table 7 (MSFEs for predicting x_t using model M_2, labelled experiments 15–28, and using M_3, labelled experiments 29–42). The other regressions are specific to the results for a particular model.

The explanatory variables are a constant, a dummy which is unity if at least one unit root is estimated, and zero otherwise (Dur), a dummy which is unity if there is a moving-average term, and zero otherwise (Dma), an interaction dummy for these two effects (Dur$\times$Dma), and finally, Dcanc, which is unity for DGP (IXi), M_2 and M_3 when the MA root is approximately cancelled by an AR root.

The tables report coefficient estimates and standard errors (in brackets). Dcanc is zero for all observations for the first sub-sample, and Dma and Dur$\times$Dma are collinear for the second and third, so the former is omitted. In no case are the constant, Dur and Dma significantly different from zero: in general, therefore, there is little gain (loss) from DE, and either estimating unit roots or neglecting MA components, in the absence of the other, does not alter this.

The conjunction of the two (signalled by Dur$\times$Dma), on the other hand, significantly favours DE, particularly at *4*-steps (table 5), except for the third sub-sample where the model is under-parameterised for many of the DGPs. Finally, the coefficient on Dcanc is always approximately equal magnitude and

[4] While linear projection is not in general equivalent to non-linear minimization, the two coincide here for multi-step estimation, but would not do so for models with explanatory variables where there are restrictions on the parameters.

opposite sign to Dur×Dma, so that there is no gain to DE when the MA term is effectively cancelled by an AR root (see section 5.1.3).

We now discuss the contributions to forecast errors of stochastic uncertainty, model mis-specification, and estimation uncertainty.

Table 4 Response surfaces for log(DE/OLS) for *2*-step forecasts.

	All experiments	Exp. 1-14	Exp. 15-28	Exp. 29-42
Constant	-0.011	0.003	-0.025	-0.015
	(0.017)	(0.015)	(0.038)	(0.017)
Dur	0.017	-0.003	0.032	0.021
	(0.024)	(0.034)	(0.051)	(0.023)
Dma	-0.013	-0.026		
	(0.027)	(0.020)		
Dur×Dma	-0.065	-0.121	-0.102	-0.019
	(0.035)	(0.040)	(0.043)	(0.019)
Dcanc	0.078		0.101	0.018
	(0.040)		(0.072)	(0.032)
R^2	0.301	0.829	0.400	0.123

Table 5 Response surfaces for log(DE/OLS) for *4*-step forecasts.

	All experiments	Exp. 1-14	Exp. 15-28	Exp. 29-42
Constant	-0.028	0.007	-0.029	-0.072
	(0.040)	(0.027)	(0.090)	(0.055)
Dur	0.047	0.001	0.050	0.094
	(0.058)	(0.060)	(0.119)	(0.072)
Dma	0.038	0.004		
	(0.065)	(0.035)		
Dur×Dma	-0.233	-0.355	-0.248	-0.062
	(0.084)	(0.071)	(0.101)	(0.061)
Dcanc	0.192		0.246	0.054
	(0.094)		(0.168)	(0.102)
R^2	0.347	0.911	0.431	0.169

5.1.1 Stochastic uncertainty

The uncertainty from accumulating future disturbances in the DGP sets a baseline level which cannot be improved upon by stochastic models. This source of forecast uncertainty, as measured by MSFE, is $\mathsf{O}(1)$ in the forecast horizon for $\mathsf{I}(0)$ processes, and $\mathsf{O}(h)$ for $\mathsf{I}(1)$ processes (see, e.g. Engle and Yoo, 1987, and Clements and Hendry, 1994b), as is reflected in our results. For example, from the first column of MSFEs in Table 7, the MSFE for DGP (VIIIi) is $\mathsf{O}(h^2)$ in the forecast horizon, since x_t is $\mathsf{I}(2)$; for DGP (VIIIii), it is $\mathsf{O}(h)$ in the forecast horizon, since x_t is $\mathsf{I}(1)$; and for DGP (VIIIiii), it is $\mathsf{O}(1)$ since x_t is $\mathsf{I}(0)$. Table 6 conveys the same information for Δx_t where the orders of integration and their rates of increase are reduced by unity.

Although the effects of this source of uncertainty can be obtained analytically, the figures in the tables are Monte Carlo estimates calculated numerically by simulating forecasts from (15)–(17) with the disturbances set to zero over the forecast period. Such estimates reflect Monte Carlo variability, but this is small for $10,000$ replications. For example, when the process is a linear trend or white noise (DGPs (VI,VII)), the theoretical conditional forecast error variances for predicting Δx_t are 1 for $h = 1$, and 2 for $h > 1$ compared with Monte Carlo estimates for the *1*- to *4*-step horizons of $[1, 1.99, 2.04, 2.01]$: see Table 6.

5.1.2 Model mis-specification

To gauge the impact of model mis-specification in the absence of parameter-estimation uncertainty, we generated forecasts using the pseudo-true values of the models' parameters under the DGP (referred to as $[\tau_p : \Upsilon_p]$ for *1*-step, and $[\tau_h^* : \Upsilon_h^*]$ for *h*-step, in section 3). These values are derived analytically. We computed this source of uncertainty including the stochastic element from the future disturbances, so these forecasts can be compared against the actual realizations.

Table 6 reveals the main points. When the model is correctly specified, the 'true model' and 'control' (for estimation uncertainty) columns are identical (DGPs (IV,V,VIIIi&ii) in Table 6). Failure to impose a valid unit root does not constitute model mis-specification (Table 6, DGP (VIIIi)), but imposing an invalid unit root does (Table 6, DGP (VIIIiii)) as over-parameterisation is not a form of mis-specification, but under-parameterisation is.

Abstracting from DGP (I), the impact of model mis-specification is largest at *1*-step horizons and is typically small at longer horizons (see the '*1*-step control' in Table 6, for DGPs (II,III,VI,VII)). An exception is where the model is mis-specified in terms of omitting a constant (eg. M_1^* for DGP (II)). In DGP (VIIIiii), the model incorrectly imposes a unit root for a stationary $\mathrm{AR}(2)$ process, so the impact of model mis-specification is more persistent in the forecast horizon.

For some DGPs (namely VI,VII,IXii), there is a marked improvement at $h = 2$ from the DE control ('*h*-step control': Υ_h^* and τ_h^*) relative to the *1*-period control. This occurs when DE provides a better approximation to the *2*-step ahead forecast function than simply iterating out the *1*-step forecast function. For DGP (VII), for example, Δx_{T+h} is unpredictable for $h > 1$. While the pseudo-true value Υ_2^* implies a DE forecast of zero, OLS is based on powering-up (the non-zero) first-order autocorrelation.

Comparing results for DGPs (VIIIi) and (I) suggests that neglecting the MA error is responsible for the large increase in forecast uncertainty relative to that inherent in predicting an $\mathsf{I}(2)$ process. The *1*-step and multi-step controls for DGP (I) have all had the slope parameter set to unity (the constant, where estimated, is zero). This is only correct asymptotically, but is a reasonable approximation unless the MA roots are close to minus one. DGP (IXi&ii) Table 6 has only a first order MA and a stationary root, compared to the second-order MA and unit root in DGP (I). The impact of model mis-specification is largest at a *1*-step horizon, and is more persistent than for DGPs (II,III), due to the interaction of the omitted MA error with the autoregressive dynamics. Nevertheless, that the AR dynamics are stationary implies that the impact of the MA mis-specification dies out in the horizon: this is not the case when the AR dynamics have a unit root as in DGP (I). The AR component in DGP (IXi) relative to DGP (VII), reduces the gains from the multi-step control characteristic of processes with MAs.

5.1.3 Estimation uncertainty: *1*-step and multi-step criteria

The impact of *1*-step estimation uncertainty is generally fairly small apart from DGPs (I,VIIIi), Table 6, when a unit root is being estimated. DE is significantly better for DGP (I), when a unit root is being estimated in conjunction with omitted MA errors, but not in the absence of MA mis-specification as in DGP (VIIIi). An omitted MA error in the absence of implicit unit-root estimation (Table 6, DGP (IXi&ii)) leads to only a small increase in overall forecast uncertainty above that due to model mis-specification, and no significant benefit from multi-step estimation. In summary, unit-root AR dynamics together with omitted MA errors seem necessary here for gains from multi-step over *1*-step estimation.

In Table 7, DGP (VIII), the AR(2) model is correctly specified, so that the *1*- and *h*-step estimation columns measure the impact of estimation uncertainty, and estimating the unit roots (two for VIIIi, one for VIIIii) significantly inflates the forecast-error variances, but there is no gain to DE.

1-step estimation is also better here than DE when there are unit roots and the model is under-parameterised (Table 7, DGPs (VIIIi&ii) and AR(1) model), although there is little between the two when the model is under-parameterised and unit roots are not being implicitly estimated (Table 7, DGP (VIIIiii), AR(1) model).

Further evidence that DE appears to be most advantageous when unit roots are being estimated and there are omitted MA components is provided by Table 7, DGPs (II,III), the AR(1) model. However, merely over-parameterizing the autoregressive part of the model largely removes the gains from DE and yields a lower forecast-error variance: the extra autoregression corrects the missing MA error. Table 7 DGP (I), AR(2) model shows gains from DE when MA components are present and a *double* unit root is being estimated (but not when the AR component is under-parameterized).

At first sight, Table 7, DGP (IXi) is an anomaly since the DGP is an ARIMA$(1,1,1)$, and the models are AR(2) and AR(1) in levels. Thus, a unit root is being estimated with a neglected MA error, but *1*-step estimation is preferred. The reason is that the second autoregressive root approximately cancels the MA term so that the DGP is nearly ARIMA$(0,1,0)$. The implied value of θ is:

$$\theta = \frac{\sqrt{(q^2+4q)}-2-q}{2} \quad \text{where} \quad q = \frac{\sigma^2_{\delta_2}}{\sigma^2_{\delta_1}},$$

(see Harvey, 1993). For DGP (IXi), $\theta \simeq -0.382$, so that the AR and MA polynomials are $(1-L)(1-0.5L)$ and $(1-0.382L)$, leading to near cancellation of the MA root with the second AR root.

For DGP (IXii), $\theta \simeq -0.605$ but $\tau_1 = 0.1$, ensuring that the process is ARIMA$(1,1,1)$, and delivering the expected gains from DE for both the AR(2) and under-parameterised AR(1) model (see Table 7).

5.1.4 Explaining other researchers' findings

Our results help explain some other researchers' findings. First, Stoica and Nehorai (1989) find that in the ARMA$(2,2)$ DGP given by:

$$y_t - 0.98y_{t-2} = e_t - 0.87e_{t-1} - 0.775e_{t-2},$$

DE yields superior forecasts to OLS when the forecasting model is an AR(1). They interpret this as being due to the model being under-parameterized and only poorly approximating the true process. Consistent with this, the gains to DE disappear for an AR(6) model. However, the process has a near-unit root and negative MA errors, precisely the circumstances under which DE performs well. If we set the AR coefficient to -0.90 instead of -0.98, and re-run their simulations, the gains to DE disappear, suggesting that the relevant feature is not under-parameterization *per se*. However, over-parameterization is likely to lessen the gains to DE (see section 5.1.3).

Secondly, Weiss (1991) considers a stationary second-order autoregressive-distributed lag process with a strongly exogenous $\mathsf{I}(0)$ variable. Various mis-specified models are estimated by OLS and DE, and forecast performances are compared. Our analysis indicates little gain to DE in these circumstances, and indeed Weiss finds the differences between DE and OLS MSFEs are small.

6 Unit roots and neglected moving-average errors

We have established that neglecting MA errors in unit-root processes provides a rationale for DE at short horizons due to 'model mis-specification effects' (Table 6, DGPs VI,VII,IXii). In this section, we derive the asymptotic distributions of the estimators, and perform a Monte Carlo to compare the empirical distributions of the unit-root estimates from OLS and DE to check whether the divergence between parameter estimates accounts for the forecast differences. The Monte Carlo examines a range of values of the MA parameter, namely $\theta = \{-0.9, -0.5, -0.1, 0, 0.1, 0.5, 0.9\}$ and sample sizes $T = \{25, 50, 100\}$, using 50000 replications. We include the IV estimator suggested by Hall (1989) in the comparison, and a fourth estimator, obtained by applying IV to DE (IVDE), which is motivated below. The DGP is:

$$\begin{aligned} y_t &= y_{t-1} + u_t \\ u_t &= \epsilon_t + \theta\epsilon_{t-1} \\ \epsilon_t &\sim \mathsf{IN}\left(0, \sigma_\epsilon^2\right), \end{aligned}$$

where $\sigma_\epsilon^2 = 1$ and the *h*-step forecast function is:

$$\hat{y}_{T+h} = \tilde{\rho}_{(h)} y_T,$$

where $\tilde{\rho}_{(h)}$ is alternatively $(\hat{\rho}_{0\text{LS}})^h$, $\hat{\rho}_{\text{DE}_h}$, $(\hat{\rho}_{\text{IV}})^h$, and $\hat{\rho}_{\text{IVDE}_h}$ for OLS, DE, IV and IVDE. The first two are defined by the scalar versions of (3) to the power h and (6) at h (rather than 2) lags, and the last two by:

$$\begin{aligned} \text{IV: } (\hat{\rho}_{\text{IV}})^h &= \left(\frac{\sum y_t y_{t-2}}{\sum y_{t-1} y_{t-2}}\right)^h \\ \text{IVDE: } \hat{\rho}_{\text{IVDE}_h} &= \frac{\sum y_t y_{t-h-1}}{\sum y_{t-h} y_{t-h-1}}. \end{aligned}$$

Hall (1989) shows that the IV estimator of ρ in $y_t = \rho y_{t-1} + e_t$ has the Dickey–Fuller (DF) distribution when an instrument dated y_{t-k}, $k > 1$, is used (or more generally, when u_t follows an MA(q) process, for $k > q$). $\hat{\rho}_{0\text{LS}}$ will not have the DF distribution because of the bias induced by the correlation between y_{t-1} and u_t. IV is valid because the unit root in the process implies that y_t is correlated with y_s for all s, but y_s ($s < t$) will not be correlated with u_t for $t - s$ sufficiently large (depending on the order of the MA), and thus are valid instruments. Below, we derive limiting distributions for any *h* of the four estimators to examine the impact of estimation on multi-step forecasting for a first-order neglected MA error process.

First, for the IV estimator at $h = 1$:

$$T\left(\hat{\rho}_{\text{IV}} - 1\right) = \left(T^{-2}\sum_{t=1}^{T} y_{t-1}y_{t-2}\right)^{-1} T^{-1}\sum_{t=1}^{T} y_{t-2}u_t. \tag{23}$$

Using the usual notation that:

$$S_t = \sum_{s=1}^{t} u_s,$$

and:

$$\sigma^2 = \lim_{T\to\infty} T^{-1}\mathsf{E}\left(S_T^2\right), \quad \sigma_u^2 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} u_t^2$$

then:

$$\sigma^2 = (1+\theta)^2 \sigma_\epsilon^2 \quad \text{and} \quad \sigma_u^2 = \left(1+\theta^2\right)\sigma_\epsilon^2.$$

The denominator in (23) is:

$$T^{-2}\sum_{t=1}^{T} y_{t-1}y_{t-2} \Rightarrow \sigma^2 \int_0^1 W(r)^2 \,\mathrm{d}r \tag{24}$$

where $\Rightarrow$ denotes weak convergence (see e.g. Banerjee, Dolado, Galbraith and Hendry, 1993). Since:

$$T^{-1}\sum_{t=1}^{T} y_{t-1}u_t \Rightarrow \frac{\sigma^2}{2}\left[W(1)^2 - 1\right] + \frac{\sigma^2 - \sigma_u^2}{2} \tag{25}$$

and:

$$T^{-1}\sum_{t=1}^{T} u_t u_{t-1} \Rightarrow \theta\sigma_\epsilon^2.$$

the numerator:

$$T^{-1}\sum_{t=1}^{T} y_{t-2}u_t = T^{-1}\sum_{t=1}^{T} y_{t-1}u_t - T^{-1}\sum_{t=1}^{T} u_t u_{t-1} \Rightarrow \frac{\sigma^2}{2}\left(W(1)^2 - 1\right) \tag{26}$$

and hence:

$$T(\hat{\rho}_{\mathrm{IV}} - 1) \Rightarrow \frac{1}{2}\left(W(1)^2 - 1\right)\left(\int_0^1 W(r)^2 \,\mathrm{d}r\right)^{-1} \tag{27}$$

which is the Dickey–Fuller distribution (see Dickey and Fuller, 1979, 1981) for testing for a unit root in a univariate process, independently of the residual autocorrelation. The numerator in (27) is (see Fuller, 1976):

$$\int_0^1 W(r)\,\mathrm{d}W(r) = \frac{1}{2}\left(W(1)^2 - 1\right) \sim \frac{1}{2}\left(\chi^2(1) - 1\right)$$

and as $\mathsf{P}(\chi^2(1) \leq 1) \simeq 0.7$, this imparts a negative shift to the distribution.

The outcome in (27) contrasts with:

$$T(\hat{\rho}_{\mathrm{OLS}} - 1) = \left(T^{-2}\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} T^{-1}\sum_{t=1}^{T} y_{t-1}u_t. \tag{28}$$

The denominator converges to the same limit as (24), and the numerator is given by (25). Hence:

$$T(\hat{\rho}_{\mathrm{OLS}} - 1) \Rightarrow \left(\int_0^1 W(r)^2 \,\mathrm{d}r\right)^{-1}\left[\frac{1}{2}\left(W(1)^2 - 1\right) + \frac{\theta}{(1+\theta)^2}\right]. \tag{29}$$

Consequently, the appropriately normalized distribution is non-central; in particular, when $\theta < 0$, the leftward shift of the limiting distribution is exacerbated. For $\theta > -0.2$, the non-centrality is minor (≤ -0.3), but when $-1 < \theta < -0.73$, the non-centrality exceeds 10, and for $\theta < -0.9$, it is very large (≥ -90).

To compare the performance of the estimators for multi-step forecasting, we need the asymptotic distributions of $(\hat{\rho}_{\mathrm{OLS}})^h$ and $(\hat{\rho}_{\mathrm{IV}})^h$ (the appendix provides details). For OLS:

$$T\left((\hat{\rho}_{\mathrm{OLS}})^h - 1\right) \Rightarrow \left(\int_0^1 W(r)^2 \,\mathrm{d}r\right)^{-1} h\left[\frac{1}{2}\left(W(1)^2 - 1\right) + \frac{\theta}{(1+\theta)^2}\right] \tag{30}$$

indicating a bias h times as large as for $\hat{\rho}_{\text{OLS}}$. However, a better approximation in finite samples takes account of lower-order biases:

$$T\left(\left(\hat{\rho}_{\text{OLS}}\right)^{h}-1\right)\Rightarrow\left[\sum_{i=0}^{h-1}\left(1+T^{-1}B\right)^{i}\right]B\simeq h\left[1+\frac{(h-1)}{2T}B\right]B \tag{31}$$

where B is (29). Thus, the increase in the bias for finite T should be less than h times that of the *1*-step.

For IV, the appendix shows that:

$$T\left(\left(\hat{\rho}_{\text{IV}}\right)^{h}-1\right)\Rightarrow\left(\int_{0}^{1}W\left(r\right)^{2}\mathrm{d}r\right)^{-1}\frac{h}{2}\left(W\left(1\right)^{2}-1\right)\simeq\left[1+\frac{(h-1)}{2T}C\right]C \tag{32}$$

taking account of lower-order biases as for OLS, where C is given by (27) and replaces B from (29).

The appendix also derives the asymptotic bias for the *h*-step DE:

$$T\left(\hat{\rho}_{\text{DE}_h}-1\right)\Rightarrow\left(\int_{0}^{1}W\left(r\right)^{2}\mathrm{d}r\right)^{-1}h\left[\ \tfrac{1}{2}\ \left(W\left(1\right)^{2}-1\right)+\frac{\theta}{h\left(1+\theta\right)^{2}}\right]. \tag{33}$$

The IVDE estimator is motivated by noting that in the DE estimator the regressor is correlated with the disturbance term, but when u_t is $\text{MA}(1)$, y_{t-h-1} is a suitable instrument. Thus, for the *h*-step IVDE:

$$T\left(\hat{\rho}_{\text{IVDE}_h}-1\right)=\frac{T^{-1}\sum_{t=1}^{T}y_{t-h-1}\sum_{s=0}^{h-1}u_{t-s}}{T^{-2}\sum_{t=1}^{T}y_{t-h}y_{t-h-1}}. \tag{34}$$

Similar arguments to those set out in the appendix B for the DE estimator show that:

$$T\left(\hat{\rho}_{\text{IVDE}_h}-1\right)\Rightarrow\left(\int_{0}^{1}W\left(r\right)^{2}\mathrm{d}r\right)^{-1}\frac{h}{2}\left[\left(W\left(1\right)^{2}-1\right)\right],$$

which is the same as the *h*-step IV limiting distribution (32).

Thus, for OLS, the leftward non-centrality due to $\theta<0$ increases in *h*, whereas for DE it does not, and their limiting distributions coincide if $\theta=0$, suggesting no gain to DE in that case (as borne out by our Monte Carlo). IV is better still, and asymptotically the distributions of IV and IVDE coincide.

6.1 A Monte Carlo study of the estimators

The results of some of the simulations are illustrated graphically in figure 2 which shows the histograms of OLS and DE for 2 and 4-steps ahead forecasts ($\left(\hat{\rho}_{\text{OLS}}\right)^{h}$ and $\hat{\rho}_{\text{DE}_h}$, $h=2,4$) for $\theta=\{-0.9,-0.5,0,0.9\}$ at $T=100$. Although the IV methods (IV and IVDE) have no moments of any order for any value of θ, this does not prevent apparently sensible values of the mean and variance being obtained by simulation for some values of θ (see Sargan, 1982, and Hendry, 1990). Moreover, moments would exist if more than one lag were used as an instrument (eg. $y_{t-k},\ldots,y_{t-k-p}$, where $k>q,p>0$).

A striking feature of the results, in line with the asymptotic bias formulae, is the marked increase in bias for OLS as h increases, when there is a negative moving-average error, and the correspondingly much smaller increase in bias for DE. In the presence of negative MAs, OLS and DE are median-biased for both $h=2$ and $h=4$, their means and medians being roughly similar otherwise.

As expected, Hall's IV estimator is approximately median unbiased for all θ, as is IVDE, but for $\theta\geq 0$, IV generally has a smaller Monte Carlo variance than IVDE (results not reported). The inter-quartile ranges ($25\%-75\%$) for the two IV estimators are similar except when $\theta=-0.9$.

Figure 3 plots the histograms of the resulting 2- and 4-step forecast errors for OLS and DE corresponding to the estimates plotted in figure 2. These are visually much better behaved, due to the relatively dominant role played by error accumulation.

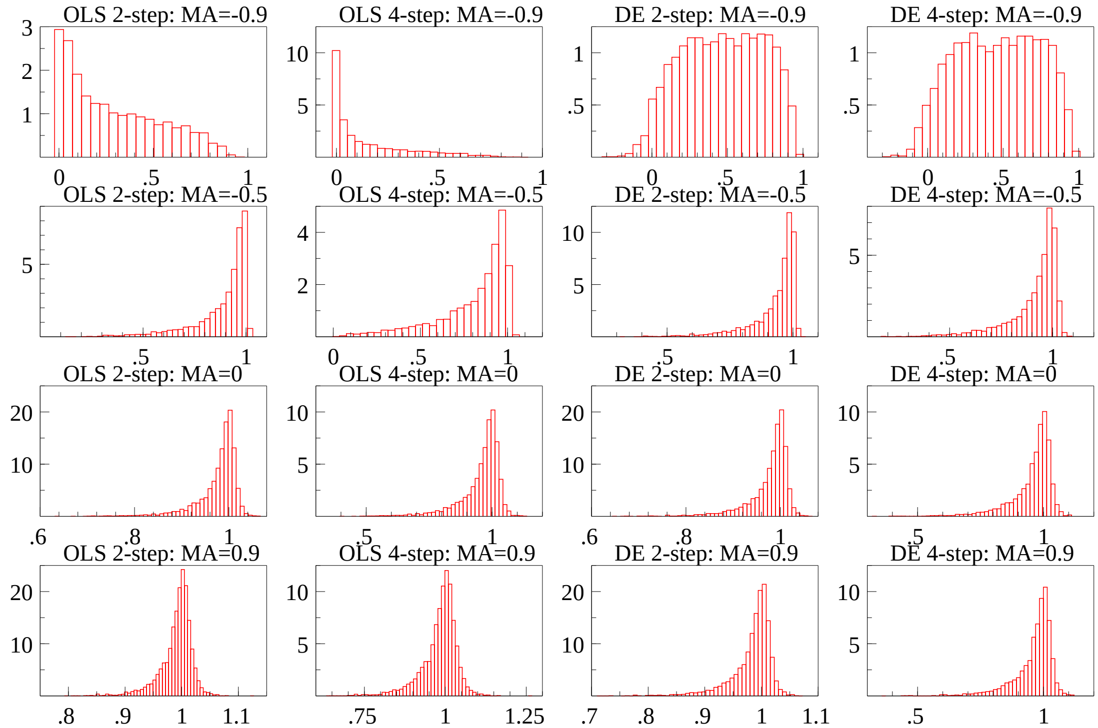

Figure 2: Histograms of OLS and DE at powers of 2 and 4

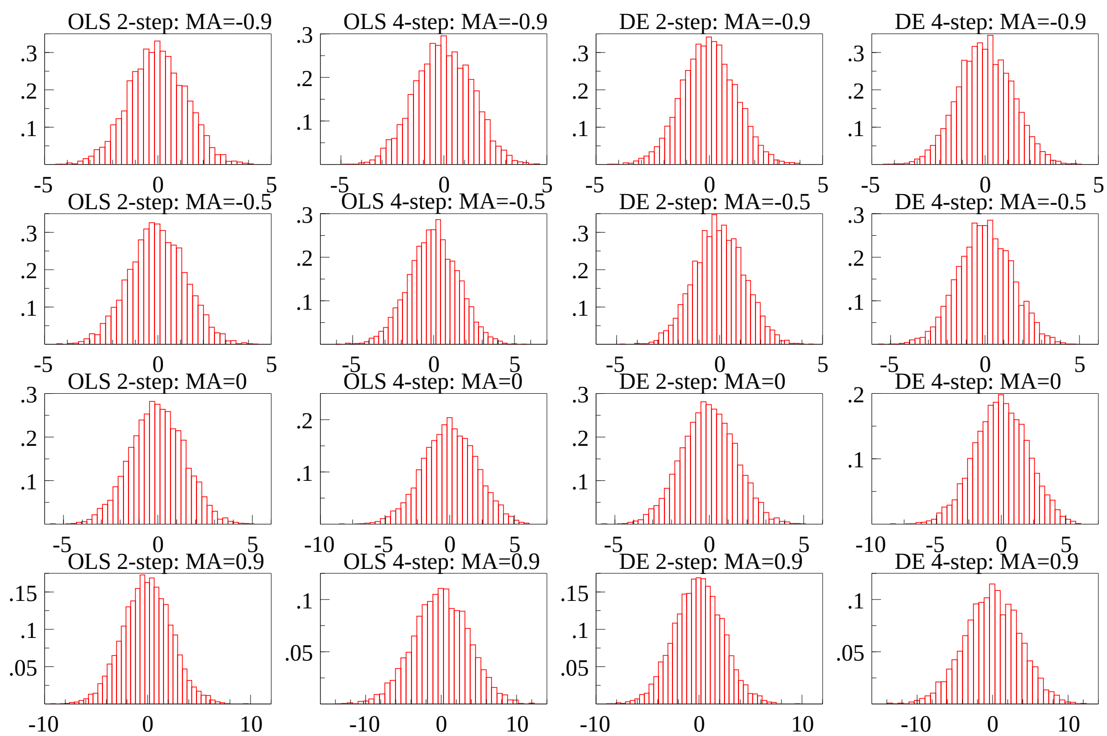

Figure 3: Histograms of the OLS and DE 2 and 4-step forecast errors

Table 8 compares the values of $(\hat{\rho}_{\text{OLS}})^h$ for $h = 2, 4$ predicted by (30) and (31) (taking account of lower-order biases) with the means of the Monte Carlo.[5] The asymptotic formulae use estimates of B (see (29)) from the same set of replications. The table illustrates the usefulness of the asymptotic formulae for OLS for $T = 100$, and $\theta = -0.9, -0.5$. The rows (30) are calculated as $hBT^{-1}+1$, where $B = T(\hat{\rho}_{\text{OLS}} - 1)$, and the rows (31) as $BT^{-1}(\sum_{i=0}^{h-1}(1+T^{-1}B)^i)+1$. Allowing for lower-order biases is important for the accuracy of the formulae when $\theta = -0.9$.

7 Conclusions

Model mis-specification is necessary but not sufficient to justify DE. Multi-step estimation criteria can lead to different parameter estimators from *1*-step, and can even alter the implicit model class. Conversely, a switch in the implicit model class implies model mis-specification. In stationary processes, DE, approximated by OLS projection, maintains the pseudo-true values of AR parameters when the step-ahead error horizons being optimized over are larger than the degree of the MA process, and results in enhanced forecast accuracy. However, the gains will typically fade rapidly in the forecast horizon. For example, if an AR(1) model is used to approximate an MA(1) process the pseudo-true value of the AR parameter for *2*-step projection is zero, yielding the optimal *2*-step forecast of zero: the *1*-step AR parameter will be non-zero, but will rapidly approach zero as it is powered up.

When the process contains unit roots, forecast success depends upon how accurately these are estimated. Here, large negative MA errors are pernicious since they exacerbate the downward bias of OLS. A separate Monte Carlo assessed the usefulness of DE in these circumstances, and the results bore out the hypothesis that the improved forecast accuracy from DE, when unit roots are estimated in the presence of neglected negative MA errors, stems from better estimates of the unit roots. The properties of two other estimators were also explored.

Few economic time series seem likely to exhibit negative moving-average error autocorrelation of the size liable to cause really serious problems for OLS. However, some economic variables may be $\mathsf{I}(2)$, represented as $\mathsf{I}(1)$ in differences with large negative moving averages (as in our DGP I), resulting in poor forecasts of growth rates. In such a state of nature, DE may be beneficial but there are sensible alternatives. The analysis in section 6 indicates that IV may be a better solution. Indeed, the limiting distributions of IVDE and IV coincide so that once the correlation between the regressor and the error have been taken care of by IV there are no additional gains to DE. A drawback of DE is that it is not invariant to linear transformations since it is dependent on the exact criterion selected for minimization, so very different decisions could result for levels versus first differences, say.

Finally, a focus of this paper has been when $\mathsf{E}[\mathbf{x}_{T+1} \mid \mathbf{x}_T] = \mathbf{\Psi}\mathbf{x}_T$, say, but $\mathsf{E}[\mathbf{x}_{T+h} \mid \mathbf{x}_T] \neq \mathbf{\Psi}^h\mathbf{x}_T$ because of neglected moving-average terms. Other causes may include non-linearity in the generating process, for example, an ARMA model of a fractionally integrated process (see Tiao and Tsay, 1994), and outliers (see Peña, 1994). However, in each instance we suspect reasonable alternatives exist, for example, MAD for outliers.

[5] The figures and Monte Carlo estimates in table 8 were based on 5000 replications.

References

Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. F. (1993). *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Christoffersen, P. F., and Diebold, F. X. (1994). Optimal prediction under asymmetric loss. Discussion paper, Department of Economics, University of Pennsylvania.

Clements, M. P., and Hendry, D. F. (1993). Model selection when forecasting. Discussion paper, Institute of Economics and Statistics, Oxford.

Clements, M. P., and Hendry, D. F. (1994a). Intercept corrections and structural change. *Journal of Applied Econometrics*. Forthcoming.

Clements, M. P., and Hendry, D. F. (1994b). Towards a theory of economic forecasting. In Hargreaves, C. (ed.), *Non-stationary Time-series Analyses and Cointegration*, pp. 9–52. Oxford: Oxford University Press.

Clements, M. P., and Hendry, D. F. (1995a). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127–146.

Clements, M. P., and Hendry, D. F. (1995b). Macro-economic forecasting and modelling. *Economic Journal*, **105**, 1001–1013.

Clements, M. P., and Hendry, D. F. (1996). Forecasting in macro-economics. In Cox, D. R., Hinkley, D. V., and Barndorff-Nielsen, O. E. (eds.), *Time Series Models In econometrics, finance and other fields*, pp. 101–141. London: Chapman and Hall.

Cox, D. R. (1961). Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society B*, **23**, 414–422.

Dickey, D. A., and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**, 427–431.

Dickey, D. A., and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, **49**, 1057–1072.

Engle, R. F., and Yoo, B. S. (1987). Forecasting and testing in co-integrated systems. *Journal of Econometrics*, **35**, 143–159.

Findley, D. F. (1983). On the use of multiple models for multi-period forecasting. *ASA Proc. Bus. Econ. Sec.*, 528–531.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. New York: John Wiley & Sons.

Granger, C. W. J. (1993). Comment on 'On the limitations of comparing mean squared forecast errors', by M.P. Clements and D.F. Hendry. *Journal of Forecasting*, **12**, 617–637.

Hall, A. (1989). Testing for a unit root in the presence of moving average errors. *Biometrika*, **76**, 49–56.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A. C. (1993). *Time Series Models* 2nd (first edition 1981) edn. Hemel Hempstead: Harvester Wheatsheaf.

Hendry, D. F. (1984). Monte Carlo experimentation in econometrics. In Griliches, Z., and Intriligator, M. D. (eds.), *Handbook of Econometrics*, Vol. 2–3, Ch. 16. Amsterdam: North-Holland.

Hendry, D. F. (1990). Using PC-NAIVE in teaching econometrics. *Oxford Bulletin of Economics and Statistics*, **53**, 199–223.

Hendry, D. F., and Clements, M. P. (1994a). Can econometrics improve economic forecasting?. *Swiss Journal of Economics and Statistics*, **130**, 267–298.

Hendry, D. F., and Clements, M. P. (1994b). On a theory of intercept corrections in macro-economic forecasting. In Holly, S. (ed.), *Money, Inflation and Employment: Essays in Honour of James Ball*, pp. 160–182. Aldershot: Edward Elgar.

Hendry, D. F., Neale, A. J., and Ericsson, N. R. (1991). *PC-NAIVE, An Interactive Program for Monte Carlo Experimentation in Econometrics. Version 6.0*. Oxford: Institute of Economics and Statistics, University of Oxford.

Klein, L. R. (1971). *An Essay on the Theory of Economic Prediction*. Chicago: Markham Publishing Company.

Lin, J.-L., and Tsay, R. S. (1995). Co-integration and Forecasting. mimeo, Institute of Economics, Academia Sinica, Taipei, Taiwan.

Molinas, C. (1986). A note on spurious regressions with integrated moving average errors. *Oxford Bulletin of Economics and Statistics*, **48**, 279–282.

Ord, J. K. (1988). Future developments in forecasting. the time series connexion. *International Journal of Forecasting*, **4**, 389–401.

Peña, D. (1994). Second-generation Time series Models: A Comment on 'Some Advances in Non-linear and Adaptive Modelling in Time-series', by tiao and tsay. *Journal of Forecasting*, **13**, 133–140.

Sargan, J. D. (1982). On Monte Carlo estimates of moments that are infinite. In Basmann, R. L., and Rhodes, G. F. (eds.), *Advances in Econometrics: A Research Annual*, Vol. 1, pp. 267–299. Greenwich, Connecticut: Jai Press Inc.

Schwert, G. W. (1989). Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics*, **7**, 147–159.

Stoica, P., and Nehorai, A. (1989). On multistep prediction error methods for time series models. *Journal of Forecasting*, **8**, 357–368.

Tiao, G. C., and Tsay, R. S. (1994). Some Advances in Non-linear and Adaptive Modelling in Time-series. *Journal of Forecasting*, **13**, 109–131.

Tsay, R. S. (1993). Comment: Adaptive Forecasting. *Journal of Business and Economic Statistics*, **11**, 140–142.

Weiss, A. A. (1991). Multi-step estimation and forecasting in dynamic models. *Journal of Econometrics*, **48**, 135–149.

Weiss, A. A. (1995). Estimating time series models using the relevant cost function. Paper presented at the NBER/NSF/Journal of Applied Econometrics Forecasting Seminar, Department of Economics, University of Southern California.

8 Appendix

To derive the asymptotic distribution of $(\hat{\rho}_{\text{OLS}})^h$ when $\rho = 1$, first consider $(\hat{\rho}_{\text{OLS}})^2$.

$$(\hat{\rho}_{\text{OLS}})^2 = \left[1 + T^{-1}\left(T^{-1}\sum_{t=1}^{T} u_t y_{t-1}\right)\left(T^{-2}\sum_{t=1}^{T} y_{t-1}^2\right)^{-1}\right]^2 = \left(1 + T^{-1}\overline{B}\right)^2,$$

Then:

$$T\left((\hat{\rho}_{\text{OLS}})^2 - 1\right) = \left(2 + T^{-1}\overline{B}\right)\overline{B} \Rightarrow \left(2 + T^{-1}B\right)B$$

where B is given by (29). A direct generalization yields (31):

$$T\left((\hat{\rho}_{\text{OLS}})^h - 1\right) \Rightarrow \left(\sum_{i=0}^{h-1}\left(1 + T^{-1}B\right)^i\right)B.$$

The derivation of the asymptotic distribution of $(\hat{\rho}_{\text{IV}})^h$ follows in a similar vein. For $(\hat{\rho}_{\text{IV}})^2$:

$$(\hat{\rho}_{\text{IV}})^2 = \left[1 + T^{-1}\left(T^{-1}\sum_{t=1}^{T} u_t y_{t-2}\right)\left(T^{-2}\sum_{t=1}^{T} y_{t-1} y_{t-2}\right)^{-1}\right]^2 = \left(1 + T^{-1}\overline{C}\right)^2,$$

so:

$$T\left((\hat{\rho}_{\text{IV}})^2 - 1\right) = \left(2 + T^{-1}\overline{C}\right)\overline{C} \Rightarrow \left(2 + T^{-1}C\right)C$$

where C is given by (27). Thus (32) results:

$$T\left((\hat{\rho}_{\text{IV}})^h - 1\right) \Rightarrow \left(\sum_{i=0}^{h-1}\left(1 + T^{-1}C\right)^i\right)C.$$

For the *h*-step DE:

$$T\left(\hat{\rho}_{\text{DE}_h} - 1\right) = \frac{T^{-1}\sum_{t=1}^{T} y_{t-h}\left(\sum_{i=0}^{h-1} u_{t-i}\right)}{T^{-2}\sum_{t=1}^{T} y_{t-h}^2}. \tag{35}$$

The denominator is again (24). For the numerator, the various terms are:

$$T^{-1}\sum_{t=1}^{T} y_{t-h} u_{t-h+1} \Rightarrow \frac{\sigma^2}{2}\left[W(1)^2 - 1\right] + \frac{\sigma^2 - \sigma_u^2}{2},$$

whereas for $j = 2, \ldots, h$:

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} y_{t-h} u_{t-h+j} &= T^{-1}\sum_{t=1}^{T} y_{t-h+1} u_{t-h+j} - T^{-1}\sum_{t=1}^{T} u_{t-h+1} u_{t-h+j} \\ &\Rightarrow \frac{\sigma^2}{2}\left[W(1)^2 - 1\right] + \frac{\sigma^2 - \sigma_u^2}{2} - \theta\sigma_\epsilon^2; \end{aligned}$$

so that:

$$T^{-1}\sum_{t=1}^{T} y_{t-h}\sum_{i=0}^{h-1} u_{t-i} \Rightarrow \sigma^2\left[\frac{h}{2}\left(W(1)^2 - 1\right) + h\frac{\sigma^2 - \sigma_u^2}{2\sigma^2} - (h-1)\,\theta\frac{\sigma_\epsilon^2}{\sigma^2}\right] \tag{36}$$

and hence we obtain equation (33).

Table 6 AR(2) model imposing a unit root. MSFE differences.

		Constant				No constant			
DGP	True Model	1-step Control	1-step Est.	h-step Control	h-step Est.	1-step Control	1-step Est.	h-step Control	h-step Est.
Ii	3.08	8.97	8.82	-	-	8.97	8.88	-	-
	5.03	8.02	10.76	8.02	8.00	8.02	9.07	8.02	8.06
	6.06	9.09	14.23	9.09	9.21	9.09	11.26	9.09	9.17
	6.93	9.96	17.05	9.96	10.12	9.96	13.37	9.96	10.05
Iii	2.30	7.47	7.46	-	-	7.47	7.45	-	-
	4.25	6.45	8.91	6.45	6.53	6.45	7.32	6.45	6.51
	5.32	7.58	12.13	7.58	7.80	7.58	9.35	7.58	7.70
	6.21	8.46	14.77	8.46	8.77	8.46	11.23	8.46	8.61
Iiii	2.30	8.19	6.85	-	-	8.19	7.59	-	-
	3.52	6.53	7.75	6.53	5.71	6.53	7.69	6.53	6.26
	3.80	6.82	8.99	6.82	6.03	6.82	9.15	6.82	6.50
	3.97	6.99	9.78	6.99	6.15	6.99	10.78	6.99	6.66
II	2.02	2.72	2.78	-	-	3.12	3.15	-	-
	3.00	3.04	3.07	3.02	3.06	3.24	3.25	3.24	3.28
	3.02	3.03	3.04	3.04	3.09	3.30	3.31	3.27	3.32
	2.99	2.99	3.00	3.01	3.04	3.25	3.25	3.23	3.27
III	2.02	2.72	2.78	-	-	2.72	2.75	-	-
	3.00	3.04	3.07	3.00	3.06	3.04	3.06	3.00	3.05
	3.02	3.03	3.04	3.02	3.09	3.03	3.03	3.02	3.08
	2.99	2.99	3.00	2.99	3.04	2.99	2.99	2.99	3.03
IV	1.03	1.03	1.06	-	-	1.23	1.25	-	-
	0.98	0.98	0.99	1.04	1.00	1.21	1.21	1.18	1.19
	0.98	0.98	0.99	1.05	1.00	1.25	1.24	1.19	1.21
	1.01	1.01	1.02	1.07	1.03	1.26	1.26	1.21	1.22
V	1.03	1.03	1.06	-	-	1.03	1.05	-	-
	0.98	0.98	0.99	0.98	1.00	0.98	0.98	0.98	0.99
	0.98	0.98	0.99	0.98	1.00	0.98	0.98	0.98	0.99
	1.01	1.01	1.02	1.01	1.03	1.01	1.01	1.01	1.02
VI	1.00	1.51	1.54	-	-	2.00	2.01	-	-
	1.99	2.11	2.12	2.02	2.02	2.22	2.22	2.22	2.24
	2.04	2.06	2.07	2.06	2.08	2.31	2.31	2.27	2.30
	2.01	2.02	2.03	2.04	2.05	2.27	2.27	2.25	2.28
VII	1.00	1.51	1.54	-	-	1.51	1.53	-	-
	1.99	2.11	2.12	1.99	2.02	2.11	2.12	1.99	2.02
	2.04	2.06	2.07	2.04	2.08	2.06	2.07	2.04	2.08
	2.01	2.02	2.03	2.01	2.05	2.02	2.03	2.01	2.05
VIIIi	1.02	1.02	1.05	-	-	1.02	1.03	-	-
	1.99	1.99	2.12	1.99	2.13	1.99	2.06	1.99	2.06
	2.98	2.98	3.22	2.98	3.28	2.98	3.12	2.98	3.13
	3.96	3.96	4.33	3.96	4.48	3.96	4.19	3.96	4.22
VIIIii	1.02	1.02	1.04	-	-	1.02	1.03	-	-
	1.23	1.23	1.26	1.23	1.28	1.23	1.24	1.23	1.25
	1.28	1.28	1.30	1.28	1.32	1.28	1.28	1.28	1.30
	1.32	1.32	1.35	1.32	1.38	1.32	1.33	1.32	1.35
VIIIiii	1.02	1.17	1.20	-	-	1.17	1.19	-	-
	0.98	1.16	1.18	1.19	1.17	1.16	1.17	1.19	1.15
	1.02	1.14	1.16	1.17	1.14	1.14	1.15	1.17	1.12
	1.14	1.20	1.21	1.21	1.20	1.20	1.20	1.21	1.19
IXi	2.08	2.68	2.74	-	-	2.68	2.71	-	-
	2.52	2.70	2.73	2.69	2.76	2.70	2.70	2.69	2.72
	2.56	2.60	2.63	2.60	2.66	2.60	2.60	2.60	2.63
	2.67	2.69	2.72	2.69	2.74	2.69	2.69	2.69	2.71

Table 7 AR(2) model and AR(1) model. MSFE levels.

AR(2)		Constant		No const.		**AR(1)**	Constant		No const.	
DGP	True	1-step	h-step	1-step	h-step		1-step	h-step	1-step	h-step
Ii	3.08	8.79	-	9.16	-		21.28	-	50.32	-
	8.18	29.68	24.83	30.01	25.49		78.78	81.46	196.44	199.53
	18.36	73.35	54.99	74.77	55.32		179.10	191.74	452.40	466.89
	35.10	144.83	102.94	151.46	101.10		323.94	358.90	823.78	863.86
Iii	2.30	7.47	-	7.71	-		20.43	-	49.51	-
	6.60	25.28	20.79	24.89	21.05		76.96	79.63	194.97	198.06
	15.97	63.62	46.60	62.51	46.15		176.05	188.64	450.02	464.51
	32.03	127.75	88.67	127.88	85.51		319.94	354.77	820.98	861.06
Iiii	2.30	6.33	-	7.39	-		7.71	-	15.00	-
	4.33	17.64	15.48	22.29	18.50		23.01	23.62	52.10	52.85
	7.68	38.59	31.38	52.67	37.49		49.32	52.40	117.19	120.79
	12.56	69.41	54.60	101.26	64.61		86.44	95.15	210.57	220.62
II	2.02	2.86	-	2.84	-		3.14	-	3.10	-
	3.00	4.01	4.00	3.96	3.97		4.38	4.24	4.21	4.19
	4.02	5.35	5.31	5.21	5.23		5.87	5.52	5.46	5.42
	4.97	6.72	6.65	6.49	6.52		7.43	6.82	6.72	6.67
III	2.02	2.83	-	2.77	-		3.09	-	3.03	-
	3.00	3.92	3.91	3.78	3.78		4.50	4.13	4.10	4.01
	4.02	5.16	5.08	4.93	4.91		6.13	5.29	5.39	5.13
	4.97	6.35	6.29	6.01	5.99		7.75	6.45	6.60	6.18
IV	1.03	1.09	-	1.09	-		1.07	-	1.07	-
	2.04	2.22	2.24	2.22	2.24		2.20	2.20	2.20	2.20
	3.01	3.43	3.47	3.40	3.43		3.40	3.42	3.38	3.39
	3.96	4.70	4.79	4.66	4.72		4.67	4.71	4.65	4.66
V	1.03	1.08	-	1.06	-		1.06	-	1.04	-
	2.04	2.16	2.19	2.11	2.12		2.14	2.15	2.09	2.09
	3.01	3.24	3.31	3.16	3.19		3.20	3.26	3.13	3.14
	3.96	4.35	4.51	4.21	4.27		4.29	4.44	4.17	4.21
VI	1.00	1.59	-	1.61	-		2.10	-	2.08	-
	1.00	1.74	1.59	1.78	1.69		2.19	2.07	2.14	2.12
	1.02	1.73	1.62	1.89	1.85		2.39	2.12	2.32	2.28
	1.00	1.81	1.59	2.06	2.01		2.57	2.10	2.46	2.41
VII	1.00	1.04	-	1.02	-		1.03	-	1.02	-
	1.00	1.02	1.03	1.01	1.02		1.01	1.02	1.00	1.01
	1.02	1.03	1.06	1.02	1.04		1.03	1.05	1.02	1.03
	1.00	1.01	1.04	1.00	1.03		1.01	1.02	1.00	1.01
VIIIi	1.02	1.08	-	1.06	-		18.02	-	46.85	-
	5.03	5.59	5.70	5.42	5.46		74.28	77.01	192.87	196.00
	14.04	16.16	16.99	15.58	15.86		172.40	185.10	446.86	461.41
	29.91	35.53	38.89	34.15	35.30		316.30	351.19	818.40	858.53
VIIIii	1.02	1.06	-	1.04	-		1.40	-	1.38	-
	3.26	3.51	3.54	3.40	3.42		4.26	4.38	4.18	4.21
	6.30	6.95	7.10	6.67	6.75		7.89	8.29	7.73	7.85
	9.74	10.95	11.36	10.44	10.64		11.91	12.78	11.67	11.91
VIIIiii	1.02	1.05	-	1.04	-		1.11	-	1.09	-
	1.99	2.10	2.11	2.05	2.06		2.14	2.15	2.10	2.10
	2.53	2.67	2.71	2.60	2.62		2.72	2.72	2.67	2.64
	2.78	2.94	3.00	2.85	2.89		3.01	2.98	2.94	2.87
IXi	2.08	2.80	-	2.75	-		2.81	-	2.77	-
	4.60	6.45	6.50	6.26	6.30		6.46	6.55	6.30	6.33
	7.69	10.53	10.76	10.13	10.26		10.48	10.81	10.19	10.28
	11.03	14.72	15.34	14.15	14.43		14.58	15.36	14.21	14.41
IXii	10.41	14.60	-	14.57	-		16.41	-	16.96	-

Table 8 Estimates of E(rho) by OLS for T=100..

		$\theta = -0.9$	$\theta = -0.5$
$j = 2$	MC estimate	0.288	0.890
	(30)	-0.066	0.880
	(31)	0.218	0.884
$j = 4$	MC estimate	0.145	0.810
	(30)	-1.132	0.760
	(31)	0.048	0.781