Forecasting with Difference-stationary and Trend-stationary Models

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

March 20, 2000

Abstract

Although difference-stationary (DS) and trend-stationary (TS) processes have been subject to considerable analysis, there are no direct comparisons for each being the data-generation process (DGP). We examine incorrect choice between these models for forecasting for both known and estimated parameters. Three sets of Monte Carlo simulations illustrate the analysis, to evaluate the biases in conventional standard errors when each model is mis-specified, compute the relative mean-square forecast errors of the two models for both DGPs, and investigate autocorrelated errors, so both models can better approximate the converse DGP. The outcomes are surprisingly different from established results.

Journal of Economic Literature classification: C32.

Key words: Difference stationary, trend stationary, forecastability.

*Financial support from the U.K. Economic and Social Research Council under grant L116251015 is gratefully acknowledged by both authors. We gratefully acknowledge helpful comments from two anonymous referees. Correspondence should be addressed to the first author, at: Department of Economics, University of Warwick, Coventry, CV4 7AL; email: M.P.Clements@Warwick.ac.uk.

1 Introduction

An important aspect of model selection when forecasting concerns the appropriate treatment of non-stationary variables. The two classes of non-stationary processes that we consider here are difference stationary (DS) and trend stationary (TS) processes.[1] The former contain stochastic trends, and are integrated of order one, I(1), so that differencing yields a stationary series. The latter are stationary about a deterministic function of time, here taken to be a linear trend. These two forms of non-stationarity have radically different implications for forecastability when the parameters of the processes are known: forecast-error variances grow linearly in the forecast horizon for the DS process, but are bounded for the TS process. This is unsurprising given that the unit root indefinitely accumulates previous disturbances, whereas in the TS process with known parameters, the conditional h-step ahead forecast error is simply that period's disturbance term. Uncertainty plays an add-on role in the TS process, but is integral to DS.

Sampson (1991) argued that allowing for parameter uncertainty leads to forecast-error variances which grow with the square of the forecast horizon for each of the DS and TS processes, so that asymptotically the two are indistinguishable in terms of their implications for forecastability. However, this result requires that the estimation sample size, T, remains fixed as the forecast horizon h goes to infinity. If T increases with h, no matter how slowly, then the forecast-error variance in a DS process will swamp that of the TS asymptotically, and the outcome will be similar to the known parameter case.

More importantly, it is artificial to compare predictability from a TS model when it is the data-generation process (DGP), with a DS model when it is the DGP. At best, the DGP is either a DS or a TS process in any particular instance, so the issue of interest is their relative predictability taking the DGP as each in turn. This question is of interest because empirically it may be difficult to distinguish between the two (see, e.g., Perron and Phillips, 1987, Rappoport and Reichlin, 1989, and Rothman, 1998).

In §2, we derive the forecast-error variances for DS and TS models when each in turn is the DGP, and compare the two, for both known and estimated parameters, confirming the results in Sampson (1991), but clarifying the conditions for their outcomes to be equivalent. Section 3 then compares the relative predictability from the two models in the empirically-relevant situations when the TS process is the DGP, and then when the DS process is the DGP, both with known and estimated parameters. Simulations in §4 illustrate the analytical results.

While the analytical calculations compare a first-order unit-root process and a deterministic process with white-noise errors, the results are unaffected qualitatively by allowing for dynamic generalizations of these processes. The uncertainties surrounding such elements impart terms of order $1/T$ to forecast variability, independently of the forecast horizon. These effects will not qualitatively influence the

[1] Other types of non-stationarities, such as structural breaks, are beyond the scope of this paper, but are likely to be one of the main causes of forecast failure when empirical econometric models are used in forecasting exercises: see, e.g., Clements and Hendry (1999). However, the TS model, by imposing a unit root, is the more robust forecasting model in the face of structural breaks for sequences of forecasts with a moving forecast origin: see Clements and Hendry (1996a), and the empirical evidence in Eitrheim, Husebø and Nymoen (1999).

results concerning the orders of magnitude of the TS and DS models for either the TS or DS DGPs. However, allowing for stationary ARMA error processes would enable the models to more closely mimic each other, affecting quantitative comparisons of predictability, so §5 reports some simulation results for DGPs with autocorrelated error processes. Section 6 concludes.

2 Predictability in difference-stationary and trend-stationary DGPs

2.1 Difference-stationary DGP and model

The simplest example of a trending $\mathsf{I}(1)$ process is ($y_0 = 0$):

$$y_t = y_{t-1} + \mu + \epsilon_t \;\text{ where }\; \epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]. \tag{1}$$

The h-step ahead forecast for known parameters, conditional on information available at time T is:

$$y_{DS,T+h} = \mu + y_{T+h-1} = \mu h + y_T. \tag{2}$$

Thus, the forecast is of a change in the variable from the forecast origin y_T determined by the local trend μh. The conditional multi-period forecast error is, therefore:

$$e_{DS,T+h} = y_{T+h} - y_{DS,T+h} = \mu h + y_T + \sum_{i=0}^{h-1} \epsilon_{T+h-i} - (\mu h + y_T) = \sum_{i=0}^{h-1} \epsilon_{T+h-i}. \tag{3}$$

As is well known, the cumulative error has a variance that increases at $O(h)$ in the horizon h:

$$\mathsf{V}\left[e_{T+h}\right] = h\sigma_\epsilon^2. \tag{4}$$

When the DGP is unknown, the h-step forecast error $\widehat{e}_{DS,T+h} = y_{T+h} - \widehat{y}_{DS,T+h}$ for the model:

$$y_t = \rho y_{t-1} + \delta + \epsilon_t,$$

where $\{\rho, \delta\}$ are estimated (denoted $\widehat{}$ for DS), is:

$$\begin{aligned} \widehat{e}_{DS,T+h} &= \mu h + y_T + \sum_{i=0}^{h-1} \epsilon_{T+h-i} - \left(\widehat{\delta} h + \widehat{\rho}^h y_T\right) \\ &= \left(\mu - \widehat{\delta}\right) h + \left(1 - \widehat{\rho}^h\right) y_T + \sum_{i=0}^{h-1} \epsilon_{T+h-i}. \end{aligned} \tag{5}$$

We neglect coefficient biases, treating such conditional forecasts as unbiased (see e.g., Clements and Hendry, 1998, ch. 5). Since the drift term μ in the $\mathsf{I}(1)$ process becomes the slope of a linear trend, for any given realization of the process, $\mu \neq \widehat{\delta}$ induces an error which is increasing in the forecast horizon. The case of a local-to-unity root (e.g., $\rho = 1 - k/T$ for small k) when making long-horizon forecasts is analyzed in Stock (1996), who shows that considerable forecast uncertainty will result.

Generally, the variances of multi-step forecast errors are difficult to derive for estimated parameters in $\mathsf{I}(1)$ processes, due to the non-standard nature of the estimator distributions. However, when the unit-root model is estimated unrestrictedly for non-zero μ, the estimator limiting distribution is normal (see

West, 1988): the estimate of ρ converges at a rate of $T^{\frac{3}{2}}$, (so its variance can be neglected), whereas $\mathsf{V}[\widehat{\mu}] = \sigma_\epsilon^2 T^{-1}$, emphasizing the importance of accurately estimating the local trend. Also, Clements and Hendry (1996b) show that when $\mu = 0$ and this is known, the distribution of $T(\widehat{\rho}^h - 1)$ is h times the Dickey–Fuller distribution; nevertheless, in their Monte Carlo, the resulting forecast errors are approximately normally distributed.

Here, we concentrate on the case in which ρ is correctly imposed at unity, which ensures $\delta = \mu$. Then, the forecast-error variance increases quadratically in the forecast horizon, h, for fixed T:

$$\mathsf{V}\left[\widehat{e}_{DS,T+h}\right] = h\left(\sigma_\epsilon^2 + h\mathsf{V}\left[\widehat{\mu}\right]\right) = h\sigma_\epsilon^2\left(1 + \frac{h}{T}\right). \tag{6}$$

Controlling the rate at which T and h go to infinity by (see Sampson, 1991):

$$T = Ah^\kappa \tag{7}$$

where $\kappa \geq 0$, then:

$$\mathsf{V}_{DS|DS} \equiv \mathsf{V}\left[\widehat{e}_{DS,T+h}\right] = h\sigma_\epsilon^2\left(1 + A^{-1}h^{1-\kappa}\right) \tag{8}$$

which is $O(h^2)$ for $\kappa = 0$; $O(h^{2-\kappa})$ for $0 < \kappa < 1$; and $O(h)$ for $\kappa \geq 1$ (see Sampson, 1991, eqn. 12). Throughout, the notation $\mathsf{V}_{x|y}$ indicates the conditional expected square error from using model x when the DGP is given by y. Here, the model and DGP are the same. We now compare $\mathsf{V}_{DS|DS}$ with that resulting when the DGP is trend stationary.

2.2 Trend-stationary DGP and model

The trend-stationary DGP is given by:

$$y_t = \phi + \gamma t + u_t \quad \text{where} \quad u_t \sim \mathsf{IN}[0, \sigma_u^2]. \tag{9}$$

The h-step ahead forecast for known parameters from (9), conditional on information available at time T is:

$$y_{TS,T+h} = \phi + \gamma\left(T + h\right), \tag{10}$$

with multi-period forecast error:

$$y_{T+h} - y_{TS,T+h} = u_{T+h}. \tag{11}$$

The conditional forecast-error variance is the variance of the disturbance term, $\mathsf{V}\left[u_{T+h}\right] = \sigma_u^2$. When parameters have to be estimated (denoted $\widetilde{}$ for TS), (10) becomes:

$$\widetilde{y}_{TS,T+h} = \widetilde{\phi} + \widetilde{\gamma}\left(T + h\right), \tag{12}$$

with the multi-period forecast error:

$$\widetilde{e}_{TS,T+h} = \left(\phi - \widetilde{\phi}\right) + \left(\gamma - \widetilde{\gamma}\right)\left(T + h\right) + u_{T+h}, \tag{13}$$

and error variance given by:

$$\mathsf{V}\left[\widetilde{e}_{TS,T+h}\right] = \mathsf{V}\left[\widetilde{\phi}\right] + (T+h)^2\, \mathsf{V}\left[\widetilde{\gamma}\right] + 2\,(T+h)\, \mathsf{C}\left[\widetilde{\phi}, \widetilde{\gamma}\right] + \sigma_u^2. \tag{14}$$

Direct calculation delivers $\mathsf{V}[\widetilde{\boldsymbol{\theta}}]$, where $\widetilde{\boldsymbol{\theta}} = [\widetilde{\phi} : \widetilde{\gamma}]'$, and substitution into (14) gives, on simplifying by approximating $(T+1) \simeq T$:

$$\begin{aligned} \mathsf{V}\left[\widetilde{e}_{TS,T+h}\right] &\simeq 4\sigma_u^2 T^{-1} - 12\,(T+h)\,\sigma_u^2 T^{-2} + 12\,(T+h)^2\,\sigma_u^2 T^{-3} + \sigma_u^2 \\ &= \sigma_u^2\left(1 + 4T^{-1} + 12hT^{-2} + 12h^2 T^{-3}\right). \end{aligned} \tag{15}$$

From (15), for fixed T, the forecast-error variance grows with the square of the forecast horizon. Using (7) to determine the behaviour of (15) as h and T go to infinity at different rates:

$$\mathsf{V}_{TS|TS} = \mathsf{V}\left[\widetilde{e}_{TS,T+h}\right] \simeq \sigma_u^2\left(1 + 4A^{-1}h^{-\kappa} + 12A^{-2}h^{1-2\kappa} + 12A^{-3}h^{2-3\kappa}\right). \tag{16}$$

Thus, (16) is $O(h^2)$ for $\kappa = 0$; $O(h^{2-3\kappa})$ for $0 < \kappa < \frac{2}{3}$; and $O(1)$ for $\kappa \geq \frac{2}{3}$ (see Sampson, 1991, eqn. 18). To more easily compare (15) for a trend-stationary DGP with (6) for a difference-stationary DGP when T is not assumed fixed, we calculate their ratio, and eliminate T using (7):

$$\frac{\mathsf{V}_{DS|DS}}{\mathsf{V}_{TS|TS}} \propto \frac{h + A^{-1}h^{2-\kappa}}{1 + 4A^{-1}h^{-\kappa} + 12A^{-2}h^{1-2\kappa} + 12A^{-3}h^{2-3\kappa}} \tag{17}$$

(dropping the multiplicative term $\sigma_\varepsilon^2/\sigma_u^2$). Thus, (17) reveals that $\forall\kappa \neq 0$, $\mathsf{V}_{DS|DS}/\mathsf{V}_{TS|TS} \rightarrow \infty$ as $h \rightarrow \infty$ (so $T \rightarrow \infty$ at the rate determined by (7)). The case $\kappa = 0$ corresponds to a fixed T, whence $\mathsf{V}_{DS|DS}/\mathsf{V}_{TS|TS} \rightarrow A^2/12$ as $h \rightarrow \infty$. When T grows as h increases, no matter how slowly (κ close to, but not equal to, zero), then $\mathsf{V}_{DS|DS}/\mathsf{V}_{TS|TS} \rightarrow \infty$, so the forecast-error variance of the DS process swamps that of the TS process in the limit, since for $0 \leq \kappa < 1$, $\mathsf{V}_{DS|DS}/\mathsf{V}_{TS|TS} \rightarrow A^2h^{2\kappa}/12$. Thus, only when T is fixed will the DS and TS DGPs be asymptotically indistinguishable, even allowing for parameter uncertainty in estimated models thereof.

However, such comparative findings cannot be used to interpret empirical evidence, since the DGP is at best one of the two processes DS or TS, whereas (17) compares the forecast-error variances for each being its own DGP. The relevant forecast-error variance ratio of DS and TS models applied to the same series will depend on which process generated the data. Since computer-reported conventional error-variance formulae are calculated assuming that the model is correctly specified for the process, and so correspond to either $\mathsf{V}_{DS|DS}$ or $\mathsf{V}_{TS|TS}$, one must be incorrectly reported when estimating both models, and the extent of such mistakes will become apparent below. In the next section, we derive the forecast-error variance ratios of the DS and TS models taking each as the DGP in turn.

3 Predictability in difference-stationary and trend-stationary models

Initially, we abstract from parameter-estimation uncertainty, then look at the extent to which the results change when the parameters have to be estimated.

3.1 Trend-stationary DGP using a DS model

When the process is trend stationary as in (9), the incorrect DS predictor $\mu h + y_T$, ignoring estimation uncertainty, yields the forecast error:

$$e_{DS,T+h} = y_{T+h} - y_{DS,T+h} = \phi + \gamma (T + h) + u_{T+h} - (\mu h + y_T) \quad (18)$$

with an expectation conditional on y_T of:

$$\mathsf{E}_{TS}\left[e_{DS,T+h} \mid y_T\right] = (\phi + \gamma T) - y_T + (\gamma - \mu) h = -u_T + (\gamma - \mu) h,$$

and an expected conditional squared error:

$$\mathsf{E}_{TS}\left[e^2_{DS,T+h} \mid y_T\right] = \sigma^2_u + \left(\mathsf{E}_{TS}\left[e_{DS,T+h} \mid y_T\right]\right)^2. \quad (19)$$

From appendix §7.2, the value of μ that minimizes the in-sample expected squared-error loss is $\mu = \gamma$. Treating μ as known at that value, and substituting into (19) gives:

$$\mathsf{V}_{DS|TS} = \mathsf{E}_{TS}\left[e^2_{DS,T+h} \mid y_T\right] = \sigma^2_u + u^2_T, \quad (20)$$

or unconditionally, $\mathsf{E}_{TS}[\mathsf{V}_{DS|TS}] = 2\sigma^2_u$, so that the relative loss to using the DS is:

$$\mathsf{E}_{TS}\left[\frac{\mathsf{V}_{DS|TS}}{\mathsf{V}_{TS|TS}}\right] = 2. \quad (21)$$

Thus, despite their radically-different behaviour when each is simultaneously assumed to be its own DGP, conditionally on a TS DGP, they differ only by the period-T squared disturbance, independently of h. Specifically, the DS model forecast-error variance, when the TS model is the DGP, is of the same order as the TS model variance ($O(1)$ in h).

Consider now the impact of parameter estimation uncertainty. First, we derive $\mathsf{V}[\widehat{\mu}]$ for the DS model given by (1) when (9) is the DGP, so $\widehat{\mu}$ is simply the mean of Δy_t:

$$\widehat{\mu} = T^{-1} \sum_{t=1}^{T} \Delta y_t = \frac{y_T - y_0}{T} \quad (22)$$

and substituting for y_T from (9):

$$\widehat{\mu} = \gamma + \frac{u_T - u_0}{T} \quad (23)$$

so that:

$$\mathsf{V}_{TS}\left[\widehat{\mu}\right] = \frac{2\sigma^2_u}{T^2}. \quad (24)$$

This expression for the variance of the estimated growth is an order of T smaller than an investigator would report on calculating the OLS formula $\widehat{\sigma}^2_u(\mathbf{x}'\mathbf{x})^{-1}$ (where $\mathbf{x}$ is a vector of ones), since in that case, from (1) and (9):

$$\epsilon_t = \gamma - \mu + \Delta u_t \quad (25)$$

which implies that:

$$\sigma_\epsilon^2 = \mathsf{E}_{TS}\left[\epsilon_t^2\right] = 2\sigma_u^2 \tag{26}$$

and since $(\mathbf{x}'\mathbf{x})^{-1} = T^{-1}$, then $\sigma_\epsilon^2(\mathbf{x}'\mathbf{x})^{-1} = 2\sigma_u^2 T^{-1}$, as against (24). Thus, the estimated growth rate would be far more precise than reported, greatly overstating forecast uncertainty. This bias derives from the differencing-induced negative residual autocorrelation.

The DS-model predictor when μ is estimated is $\widehat{y}_{DS,T+h} = \widehat{\mu}h + y_T$ with forecast error:

$$\begin{aligned}\widehat{e}_{DS,T+h} &= \phi + \gamma(T+h) + u_{T+h} - (\widehat{\mu}h + y_T) \\ &= (\gamma - \widehat{\mu})h + (u_{T+h} - u_T)\end{aligned} \tag{27}$$

and an expected squared forecast error of:

$$\begin{aligned}\mathsf{E}_{TS}\left[\mathsf{V}_{DS|TS}\right] &= h^2\mathsf{V}_{TS}[\widehat{\mu}] + 2\sigma_u^2 - 2\mathsf{E}_{TS}\left[u_T(\gamma - \widehat{\mu})h\right] \\ &= 2\sigma_u^2 + h^2 2\sigma_u^2 T^{-2} + 2h\sigma_u^2 T^{-1}\end{aligned} \tag{28}$$

where we have substituted from (23) and (24) to get the second line. Controlling the relative rate of increase of h and T as before ($T = Ah^\kappa$):

$$\mathsf{E}_{TS}\left[\frac{\mathsf{V}_{DS|TS}}{\mathsf{V}_{TS|TS}}\right] = \frac{2 + 2A^{-2}h^{2-2\kappa} + 2A^{-1}h^{1-\kappa}}{1 + 4A^{-1}h^{-\kappa} + 12A^{-2}h^{1-2\kappa} + 12A^{-3}h^{2-3\kappa}}. \tag{29}$$

For $\kappa = 0$, as $h \to \infty$:

$$\mathsf{E}_{TS}\left[\frac{\mathsf{V}_{DS|TS}}{\mathsf{V}_{TS|TS}}\right] \to \frac{A}{6} \tag{30}$$

so both forecast-error variances are of the same order in h, increasing in the square of the horizon.

For $1 > \kappa > 0$, the dominant term in h in the numerator is $2A^{-2}h^{2-2\kappa}$, and in the denominator is $12A^{-3}h^{2-3\kappa}$, so:

$$\mathsf{E}_{TS}\left[\frac{\mathsf{V}_{DS|TS}}{\mathsf{V}_{TS|TS}}\right] \simeq \frac{A}{6}h^k, \tag{31}$$

which goes to infinity in h at rate $O(h^\kappa)$.

For $\kappa > 1$, both the DS and TS forecast-error variances increase less rapidly than $O(h^0)$, and the ratio converges to 2, as in the 'known model' case in (21).

3.2 Difference-stationary DGP using a TS model

Suppose now that the DGP is the DS process, so that the DS model mean-squared forecast error (MSFE) is given by (4). The value of $\{\phi, \gamma\}$ that minimizes the in-sample prediction error for the TS model is $\{y_0, \mu\}$ (see §7.1), so the forecast error with known parameters is:

$$e_{TS,T+h} = \mu h + y_T + \sum_{i=0}^{h-1}\epsilon_{T+h-i} - \mu(T+h) - y_0 = y_T - y_0 - \mu T + \sum_{i=0}^{h-1}\epsilon_{T+h-i}. \tag{32}$$

Then, conditional on y_T:

$$\mathsf{V}_{TS|DS} = h\sigma_\epsilon^2 + (y_T - y_0 - \mu T)^2 \tag{33}$$

and the ratio relative to the DS-model variance is:

$$\frac{\mathsf{V}_{TS|DS}}{\mathsf{V}_{DS|DS}} = \frac{h\sigma_\epsilon^2 + (y_T - y_0 - \mu T)^2}{h\sigma_\epsilon^2}.$$

Taking expectations conditional on y_0:

$$\mathsf{E}_{DS}\left[\frac{\mathsf{V}_{TS|DS}}{\mathsf{V}_{DS|DS}}\right] = 1 + Th^{-1} = \frac{h+T}{h} \tag{34}$$

so that the forecast-error variances are of the same order in h ($O(h)$, and hence linear in the horizon). The TS model is penalized when h is small or T large.

Thus, the two models are indistinguishable in terms of their implications for predictability when we ignore estimation uncertainty, and derive the forecast-error variances assuming the DGP is either DS or TS. When the TS model is the DGP, the forecast-error variances of both models are $O(1)$, and when the DS model is the DGP, both are $O(h)$. There is qualitatively different behaviour dependent on which is the DGP, but not between the models of that DGP when parameters are known.

Next, we derive the forecast-error variances allowing for parameter estimation uncertainty with a DS DGP. The TS-model parameter estimates correspond to 'spurious detrending' – see Durlauf and Phillips (1988). Intermediate results used in the calculations are collected in the appendix, where equations (44) and (45) provide expressions for the estimates of the TS-model parameters $\{\phi, \gamma\}$ for the DS DGP, leading to the variances of the parameter estimates as:

$$\mathsf{V}_{DS}\left[\widetilde{\phi}\right] \simeq \frac{2\sigma_\epsilon^2 T}{15} \quad \text{and} \quad \mathsf{V}_{DS}\left[\widetilde{\gamma}\right] \simeq \frac{6\sigma_\epsilon^2}{5T} \tag{35}$$

(see (46) and (47)) ignoring terms of $O(T^{-2})$ or smaller.

Consider, instead, an investigator who used the conventional formula $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, where $\mathbf{X} = [\boldsymbol{\iota}\ \mathbf{t}]$ when $\boldsymbol{\iota}$ is a vector of ones and $\mathbf{t} = [1, 2, \ldots, T]'$. Since from (1) and (9):

$$u_t = y_0 - \phi + (\mu - \gamma)\, t + \sum_{i=0}^{t} \epsilon_i \tag{36}$$

the population variance of the TS-model disturbance will be heteroscedastic, $\sigma_{u,t}^2 = \sigma_\epsilon^2 t$. On average over the sample, $\overline{\sigma}_u^2 = T^{-1}\sum_{t=1}^{T}\sigma_\epsilon^2 t = \frac{1}{2}\sigma_\epsilon^2\,(T+1)$. The standard OLS formula yields approximate estimates for the parameter variances of $\widetilde{\phi}$ and $\widetilde{\gamma}$ of $2\sigma_\varepsilon^2$ and $6\sigma_\varepsilon^2 T^{-2}$, which compared to (35) are a factor of order T too small. Thus, the estimated growth and intercept would be far less precise than reported, by a factor of $\sqrt{T}$ for their standard errors, inducing serious under-estimation of the forecast uncertainty. This is exactly the opposite of the outcome using the DS model for a TS DGP.

The TS-model forecast error for estimated parameters becomes:

$$\begin{aligned}
\widetilde{e}_{TS,T+h} &= \mu h + y_T + \sum_{i=0}^{h-1}\epsilon_{T+h-i} - \widetilde{\phi} - \widetilde{\gamma}(T+h) \\
&= y_T - \widetilde{\gamma}T - \widetilde{\phi} - (\widetilde{\gamma} - \mu)h + \sum_{i=0}^{h-1}\epsilon_{T+h-i}
\end{aligned} \tag{37}$$

or after some algebra:

$$\mathsf{E}_{DS}\left[\mathsf{V}_{TS|DS}\right] = (T+h)\,\sigma_{\epsilon}^{2} + \mathsf{V}_{DS}\left[\widetilde{\phi}\right] + \mathsf{V}_{DS}\left[\widetilde{\gamma}\right](T+h)^{2} + 2\,(T+h)\,\mathsf{C}_{DS}\left[\widetilde{\phi},\widetilde{\gamma}\right]. \tag{38}$$

Substituting for $\mathsf{V}_{DS}[\widetilde{\phi}] = \sigma_{\epsilon}^{2}\frac{2}{15}T$, $\mathsf{V}_{DS}[\widetilde{\gamma}] = \sigma_{\epsilon}^{2}\frac{6}{5}T^{-1}$, and using $\mathsf{E}_{DS}[\sum_{i=0}^{h-1}\epsilon_{T+h-i}(\widetilde{\phi}-y_0)] = 0$, $\mathsf{E}_{DS}[\sum_{i=0}^{h-1}\epsilon_{T+h-i}(\widetilde{\gamma}-\mu)] = \sigma_{\varepsilon}^{2}$, and $\mathsf{C}_{DS}[\widetilde{\phi},\widetilde{\gamma}] = -0.1\sigma_{\epsilon}^{2}$, from appendix equations (48)–(50):

$$\mathsf{E}_{DS}\left[\frac{\mathsf{V}_{TS|DS}}{\mathsf{V}_{DS|DS}}\right] \simeq \frac{\frac{6}{5}h + \frac{2}{15}Ah^{\kappa} + \frac{6}{5}h^{2-\kappa}A^{-1}}{h + A^{-1}h^{2-\kappa}}. \tag{39}$$

When $\kappa = 0$, the numerator and denominator are $O(h^2)$. For $0 < \kappa < 1$, the numerator and denominator are both $O(h^{2-\kappa})$. For $1 < \kappa$, the numerator is $O(h^{\kappa})$ and the denominator is $O(h)$, so the overall term is $O(h^{\kappa-1})$. These relative orders of magnitude are unaffected by $y_0 = 0$ when the constant term is omitted from the TS model.

Parameter estimation is actually advantageous for the TS model when the DGP is a DS process. If the pseudo-true values of the TS-model parameters are used ($\widetilde{\gamma} = \mu$, $\widetilde{\phi} = y_0$) instead of the estimated values, so that $\mathsf{V}_{DS}[\widetilde{\gamma}] = 0$ and $\mathsf{V}_{DS}[\widetilde{\phi}] = 0$, the performance of the TS model deteriorates. Taking the expectation of (33), we have for the pseudo-true parameters that:

$$\mathsf{E}_{DS}\left[\mathsf{V}_{TS|DS} \mid \widetilde{\gamma} = \mu, \widetilde{\phi} = y_0\right] = (T+h)\,\sigma_{\epsilon}^{2} \tag{40}$$

compared to the numerator of (39). When $h = 1$ and $\kappa \rightarrow \infty$, so that T becomes large, then $\mathsf{E}_{DS}[\mathsf{V}_{TS|DS}|\widetilde{\gamma} = \mu, \widetilde{\phi} = y_0]/\mathsf{E}_{DS}[\mathsf{V}_{TS|DS}] \rightarrow 15/2$. Intuitively, this occurs because the only information from the sample realization used by the TS model with known parameters is the value of y_0 (i.e., $\phi = y_0$), while the TS model with estimated parameters uses information up to period T via the estimates $\{\widetilde{\gamma}, \widetilde{\phi}\}$. Although $\mathsf{E}_{DS}[y_{T+1}] = \phi + \mu\,(T+1)$, on any sample realization, y_{T+1} can vary considerably from its expectation because $\{y_t\}$ is $\mathsf{I}(1)$. Forecasts based on using $\widetilde{\phi}$ and $\widetilde{\gamma}$ are more accurate since they capture part of the deviation of the sample path from the population path, and so commence closer to the forecast origin.

4 Monte Carlo

Two sets of Monte Carlo simulations illustrate the preceding analysis.[2] In the first (using *PcNaive*), artificial TS and DS DGPs with growth rates of 2.5% per period, and error standard deviations of 5%, generated data for $T = 100$ from which both DS and TS models were estimated. The aim was to evaluate the biases in estimating the standard errors of the growth coefficients from the conventional formulae, as shown in §3. This yielded table 1.

When the model coincides with the DGP – for both DGPs – the mean Monte Carlo standard errors (MCSEs) are close to the corresponding standard deviations computed from the sampling distributions

[2]Computations were performed using GiveWin, PcFiml 9, and PcNaive (see Doornik and Hendry, 1996, 1999, 1998) and the Gauss programming language, Aptech Systems, Inc., Washington.

Table 1 Monte Carlo and theory standard error and standard deviation comparisons ($\times 100$).

	MCSE	MCSD	SE	SD
	TS DGP			
TS model	0.0172	0.0172	-	-
DS model	0.7081	0.0721	0.707	0.0707
	DS DGP			
DS model	0.4956	0.4678	-	-
TS model	0.0430	0.5146	0.0543	0.622

(MCSDs), but as anticipated, they are out by the factors of $\sqrt{T}$ in opposite directions when the models are inappropriate. The uncertainty in the DS growth estimate when TS generates the data is genuinely larger than that of the TS model, but the two models have similar uncertainty when DS is the DGP. Moreover, numerical evaluation of the theory formulae delivers closely similar results, as the last two columns of table 1 show.

A second set of simulation experiments (using GAUSS) illustrate the formulae for the relative MSFEs of the TS and DS models given by (21), (34), (29), and (39).[3] The DS and TS DGPs were 'calibrated' on the log of UK Net National Income series over 1870–1993 (data from Friedman and Schwartz, 1982, and Attfield, Demery and Duck, 1995). Neither model is a congruent representation of that data, but the estimates of the parameters so obtained (including the error variances) are taken as 'typical' values for the DGPs; the latter are then assumed to have normal, independently-distributed disturbances with the estimated error variances. For the simulated data for the DS DGP, the first-period value is always set equal to the 1870 historical value. We consider estimation sample sizes from 20 to 200 in steps of 1: for each of these, we generate 1 to 200 step-ahead forecasts. For the DS model, the single parameter μ is estimated, and for the TS, the pair $\{\phi, \gamma\}$. We also consider the 'known model' case, for which the parameters are replaced by their pseudo-true values. Variations in the results across sample sizes due to the Monte Carlo are minimized by using common random numbers. Thus, for each of the $10,000$ replications, the first n values of the simulated sample of length n are identical to those of a sample of size $n+k$ ($k>0$).

The results are summarized in figure 1, which displays ratios of MSFEs for each pair of values $\{T, h\}$, where the denominator is always the MSFE of the model corresponding to the DGP. Panels *a* and *c* show ratios of MSFEs of DS to TS models when the DGP is a TS process, for known and estimated parameters respectively; panels *b* and *d* show ratios of MSFEs of TS to DS models when the DGP is a DS process, when parameters are known and estimated respectively. Thus, panels *a* and *b* confirm formulae (21) and (34), respectively. When parameters are known, and for the TS DGP, using

[3] Numerical evaluation of these formulae is straightforward, but would not have provided an independent check of the analytic calculations and approximations.

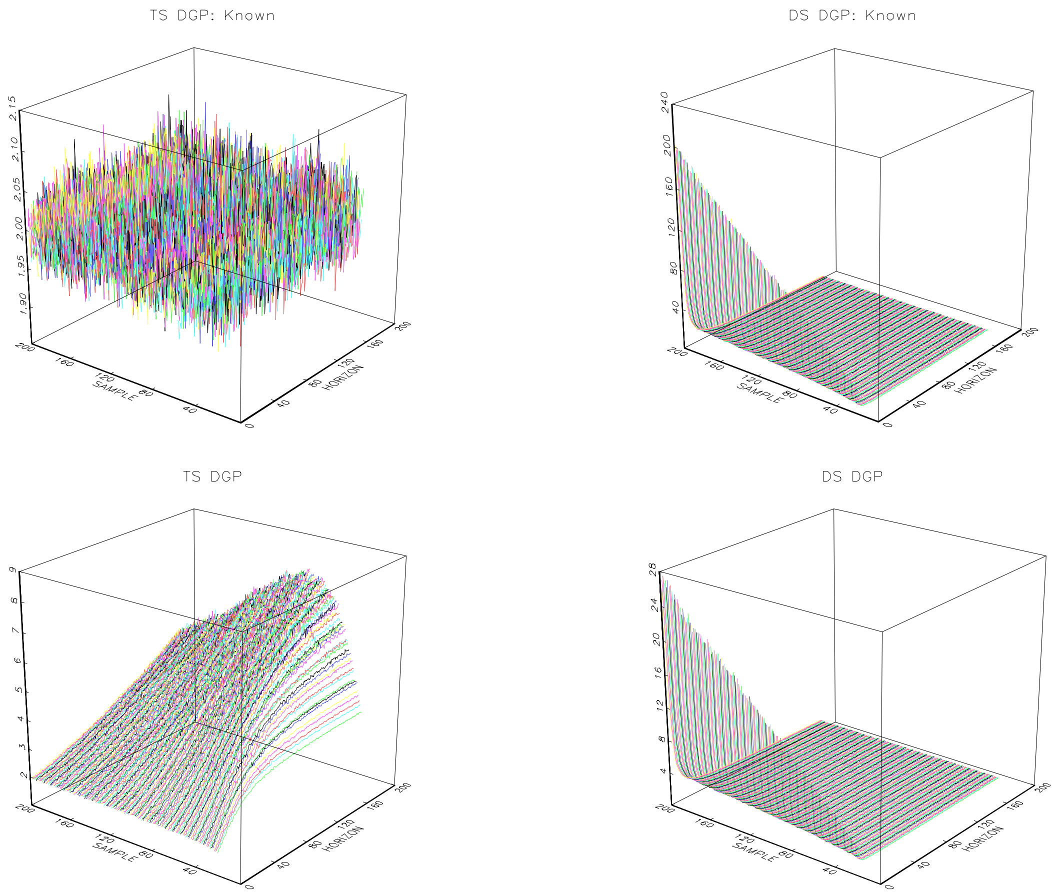

Figure 1   Ratios of MSFEs for the TS and DS models.

the wrong model doubles the MSFE, independently of the sample size and forecast horizon, in line with the formula. For the DS DGP, with known parameters, we find equal accuracy of the two models for large h, as the formula suggests, and also that the TS model fares relatively worse as the sampe size increases.

There is also a close correspondence between the theoretical formulae and the simulation results when the model parameters are estimated. Consider first the TS DGP in fig. 1*c*. For all but the shortest horizons, the ratio first increases as T grows, then tends back to 2 (for $\kappa > 1$ the ratio converges on 2), with maxima along $(12 - h)\, T^3 + 16hT^2 + 6h^3 = 0$ (e.g., $T = 70$ at $h = 200$): this explains the apparent 'quadratic' shape for large h.

Next, from (31), when $\kappa = 0.5$ (say) (so $T = A\sqrt{h}$), then $\mathsf{V}_{DS|TS}/\mathsf{V}_{TS|TS} \rightarrow \sqrt{h}$ for $A = 6$. This relation holds, e.g., from $(T = 42,\ h = 49)$ to $(T = 84,\ h = 196)$, and explains the less-than-linear increase in the ratio.

For $\kappa = 1$, $T/h = A$ is constant, and if we consider $A = 1$, so $h = T$, (a diagonal across the surface in the cube), then from (29), the ratio increases linearly in h, according to:

$$\mathsf{E}_{TS}\left[\frac{\mathsf{V}_{DS|TS}}{\mathsf{V}_{TS|TS}}\right] = \frac{6}{1+28h^{-1}}.$$

Figure 1*d* (for the DS DGP) is consistent with (39). For large T, say $T = 200$, and $h = 1$, the formula predicts a ratio of around 28, which is also evident from the figure. For large h, ($\kappa = 0$), the ratio of the TS to DS model MSFE asymptotes to $6/5$.

5 TS and DS processes with ARMA errors

The above analytical results are derived for DS and TS DGPs given by $y_t = y_{t-1} + \mu + \epsilon_t$ and $y_t = \phi + \gamma t + u_t$ with white-noise disturbances. The models have the same forms, although the derivations allowed for the autocorrelation induced when the wrong model is fitted. The analysis suggests that allowing for dynamic generalizations of these simple processes will not qualitatively affect the results. Consider (29), which summarizes the relative performance of the two models for the TS DGP. While the variance of the DS-model drift $\widehat{\mu}$ is $1/T^2$, so is estimated an order T more accurately than that of any additional stationary variables in the model, it enters the formula for the forecast-error variability multiplied by h^2, and is thus the dominant parameter-estimation uncertainty effect. Allowing more general DS and TS DGPs and estimating the corresponding models would add order $1/T$ terms to the numerator and denominator of (29), but the 'large h' results would be unaltered for all κ. Similarly, inspection of (39) for the DS DGP indicates that the qualitative results would be unaltered by adding order $1/T$ stationary-regressor estimation effects to either the numerator or denominator. Nevertheless, in this section, we investigate by simulation more general error processes, so the DS and TS models better approximate the converse DGPs.

For the TS DGP, the disturbance u_t becomes a first-order autoregression (AR(1)):

$$y_t = \phi + \gamma t + u_t, \quad u_t = \rho u_{t-1} + a_t, \;\; a_t \sim \mathsf{IN}\left[0, \sigma_a^2\right] \tag{41}$$

with $0 \leq \rho < 1$. Then $\rho = 0$ in (41) replicates our original formulation. But as ρ increases towards (but less than) unity, the degree of persistence in y_t increases, and the DS model can more successfully approximate the TS DGP. This can be seen by rewriting (41) as:

$$\begin{aligned} y_t &= \rho y_{t-1} + \phi(1-\rho) + \rho\gamma + \gamma(1-\rho)t + a_t \\ &= \rho y_{t-1} + \phi^* + \gamma^* t + a_t \end{aligned} \tag{42}$$

so that $\phi^* \rightarrow \gamma$ and $\gamma^* \rightarrow 0$ as $\rho \rightarrow 1$. The DS model of this DGP is unchanged, but the TS model is augmented by a lagged value of y. We simulate data from (42) with $\sigma_a^2 = (1-\rho^2)\sigma_u^2$ so that the variance of u_t in (41) does not depend on ρ. In the simulations, $\left\{\phi, \gamma, \sigma_u^2\right\}$ are calibrated as in §4, and $\rho = \{0, 0.3, 0.6, 0.9\}$.

For the DS DGP, the disturbance follows a first-order moving average (MA(1)):

$$y_t = y_{t-1} + \mu + \epsilon_t, \quad \epsilon_t = b_t - \psi b_{t-1}, \;\; b_t \sim \mathsf{IN}\left[0, \sigma_b^2\right]$$

where $0 \leq \psi < 1$. Then $\psi = 0$ delivers the original formulation, and the TS model can more accurately characterize the DS model as ψ approaches 1. The TS model estimated for this DS DGP is unchanged, while the DS model is the correctly specified ARIMA$(0, 1, 1)$ with drift. We set $\sigma_b^2 = \sigma_\epsilon^2 / \left(1 + \psi^2\right)$ so that the variance of ϵ_t does not depend on ψ. The DGP is calibrated as in §4, and $\psi = \{0, 0.3, 0.6, 0.9\}$. All results are for $10{,}000$ replications.

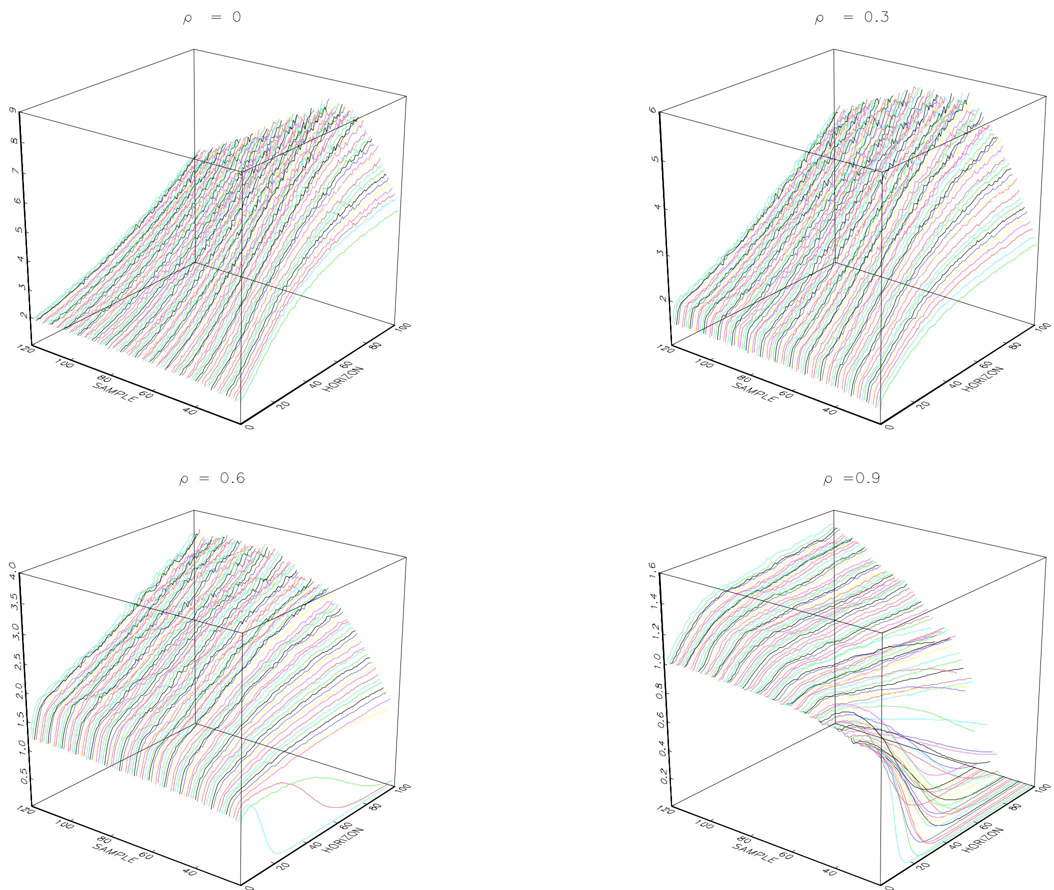

Figure 2 Simulation ratios of MSFEs for TS DGP with AR disturbance.

The results for the TS DGP are presented in fig. 2. Panel *a*, where $\rho = 0$, corresponds to panel *c* of fig. 1. For $\rho = 0.3$, (panel *b*) the shape of the surface is similar, but the vertical axis indicates that

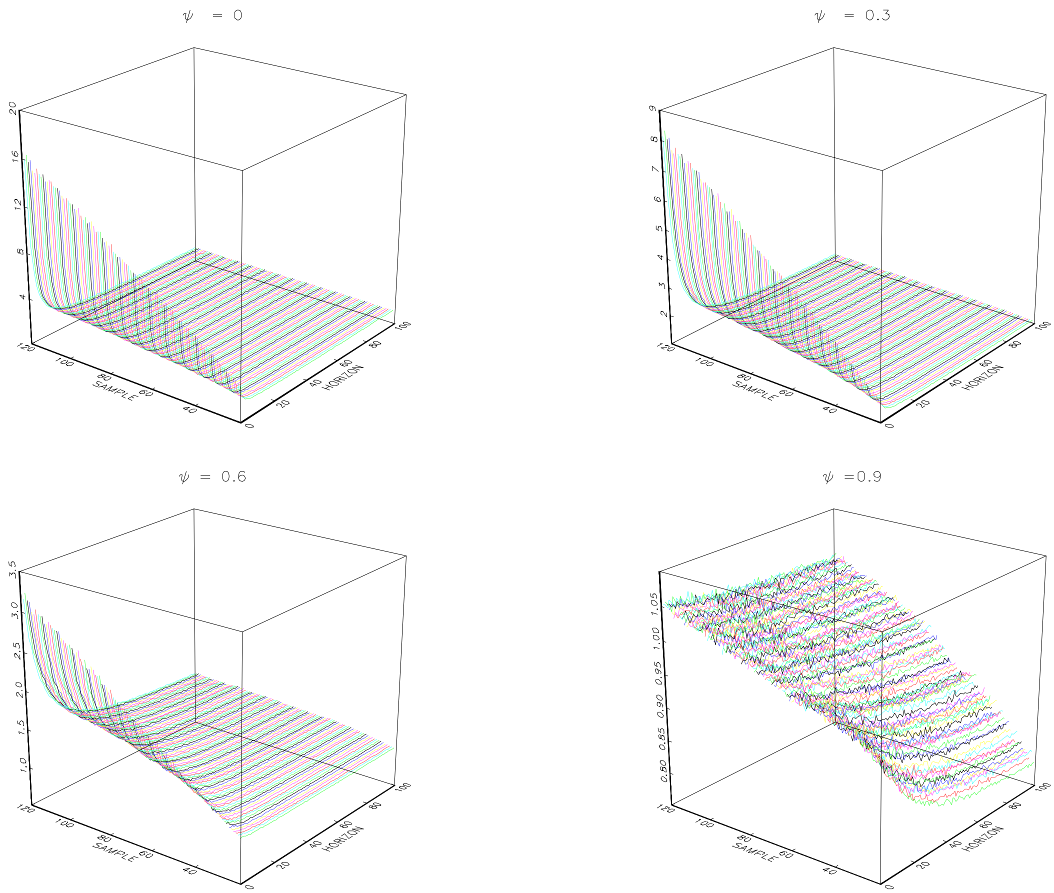

Figure 3 Simulation ratios of MSFEs for DS DGP with MA disturbance.

the DS model is more competitive. For $\rho = 0.9$, the DS model is a good approximation to the TS DGP, particularly for small T and h. Indeed, for $T < 50$ the one-parameter DS model yields more accurate forecasts than the three-parameter TS model. The results for the DS DGP are presented in fig. 3, and indicate that for $\psi = 0.9$, the TS model is more accurate when $T < 100$.

Thus, as we alter the DGP in the direction of the 'other model', for 'small T' the DGP model may yield less accurate forecasts. In many empirical settings, both DS and TS models may appear to fit a series almost equally well, mirroring the difficulties that unit-roots tests have in distinguishing the null from nearby stationary alternatives, such as trend stationarity. For the UK NNI data series, fig. 4 shows 100-, 75-, 50-, and 25-step ahead forecasts from both the simple DS and TS models with conventionally-computed 95% confidence intervals. The prediction intervals for TS are generally too narrow, while those of the DS model are too wide. Overall, TS forecasts are more accurate than the DS forecasts for the two longer horizons, but less accurate for the shorter.

Figure 5 provides a final way of comparing the forecast performance of the two models. Rather than look at a sequence of multi-step forecasts for four different origins, we move the origin through the sample, calculating sequences of multi-step forecasts where the models are re-estimated at each point. The initial origin is at $T = 50$, and we generate around seventy 1-step forecast errors, down to one 70-step forecast error (requiring caution in interpreting the longer-horizon outcomes). For each of these, we calculate MSFEs, which are plotted against the forecast lengths (shown on the x-axis). The four models are the simple TS and DS models, the TS model with a lagged-dependent variable, and the DS model with an MA term. No one model is best at all horizons, and it is hard to interpret the findings in terms of the simulation results, because the MSFE for a given horizon is of necessity calculated by varying the sample size. Nevertheless, the simple TS model is the least accurate up to 50-steps ahead, as would be the case if the actual process was either unit root or stationary with autoregressive dynamics. The TS model with an AR error (TSar, in the figure) is competitive with the DS models, consistent with this explanation. At longer horizons, the TS models are more accurate than the DS models, as would be the case if the process did not contain a unit root, but note that the longer horizon MSFEs are based on relatively few observations.

To find some general rules that might guide model choice in practice, Diebold and Kilian (1999) study by Monte Carlo whether unit-root pre-testing produces more accurate forecasts. They find positively in favour of pre-testing versus routinely differencing, suggesting success in selecting the more useful representations for forecasting, at least in the absence of structural breaks.

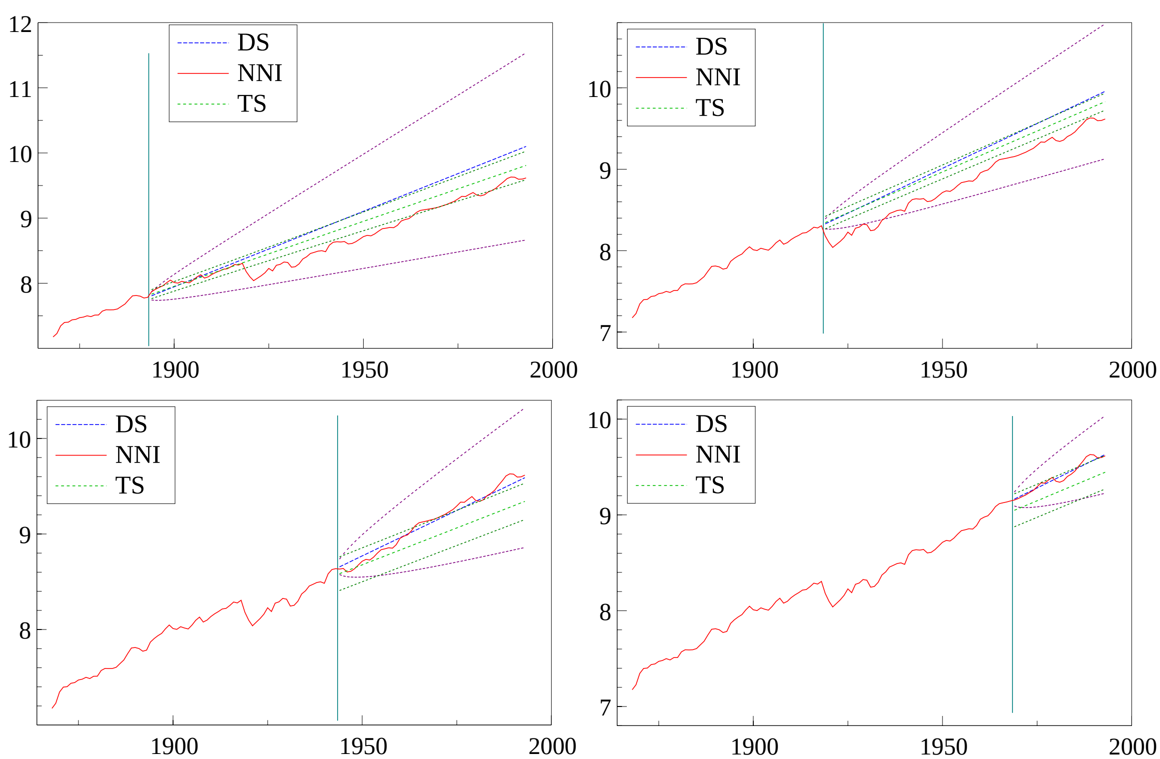

Figure 4 TS and DS forecasts for the historical data.

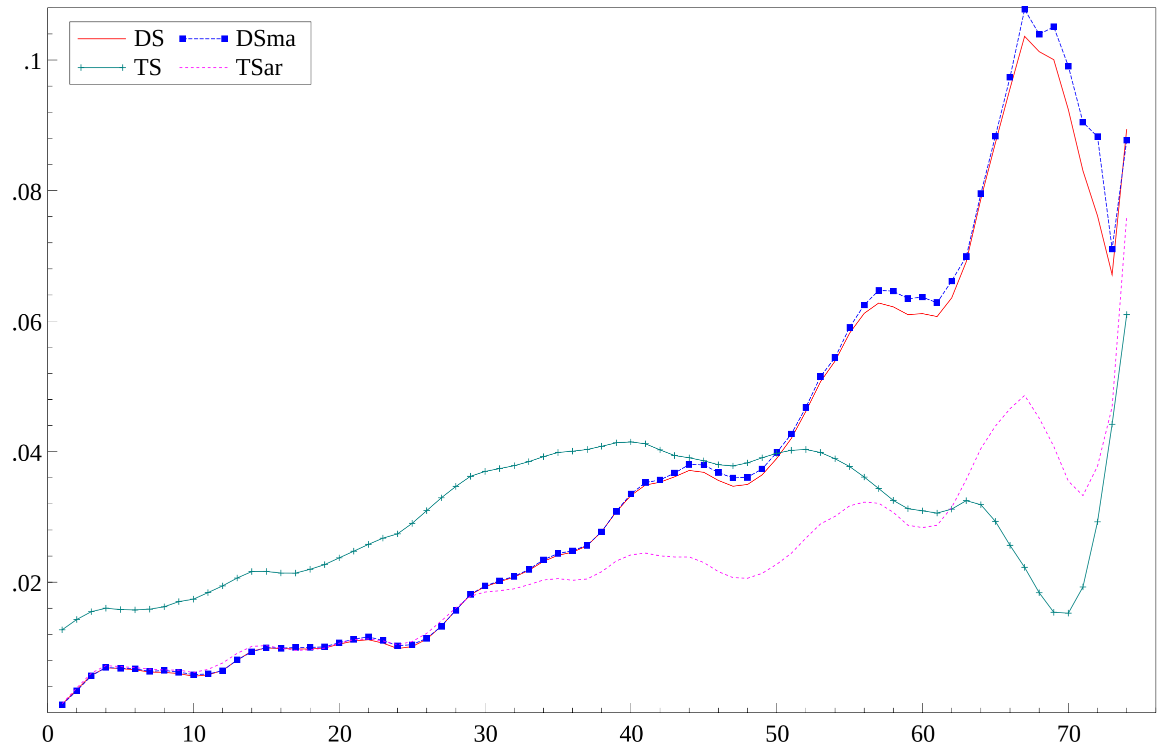

Figure 5 Empirical TS and DS model MSFEs for UK NNI data.

6 Conclusions

DS and TS processes have markedly different implications for forecasting when the properties of each model are derived treating it as the DGP. Allowing parameter-estimation uncertainty, forecast-error variances grow with the square of the forecast horizon for each model, assuming that the estimation sample, T, remains fixed as the forecast horizon h goes to infinity. If T increases with h, no matter how slowly, then the known-parameter case prevails, and the forecast-error variance of a DS process swamps that of the TS asymptotically.

However, a more meaningful approach is to compare the predictability of the two models when the DGP is either DS or TS, so that the other model is mis-specified, since in practice, only one model will (at best) approximate the DGP. In this setting, in the absence of parameter-estimation uncertainty, the two models are indistinguishable in terms of their implications for predictability. When the TS model is the DGP, the forecast-error variances of both models are $O(1)$, and when the DS model is the DGP, both are $O(h)$. There is qualitatively different behaviour dependent on which is the DGP, but not between the models of that DGP when parameters are known. A richer pattern of results emerges under parameter-estimation uncertainty. For the TS DGP, both models' forecast-error variances increase in the square of the horizon for fixed T ($\kappa = 0$), the DS/TS variance ratio goes to infinity as T increases – but less quickly than h ($0 < \kappa \leq 1$) – and for faster rates of increase of T, the ratio converges to 2. For the DS DGP, both the TS and DS models' variances are of the same order, $O(h^{2-\kappa})$, for $0 \leq \kappa \leq 1$. Only when

T increases at a faster rate than h does the order of the TS model variance exceed that of the DS model. The Monte Carlo simulations corroborated these results.

A third Monte Carlo simulation suggested that such results still held qualitatively when the corresponding models allowed for residual autocorrelation, but by making them closer substitutes, there was naturally less quantitative difference between these alternative representations.

References

Attfield, C. L. F., Demery, D., and Duck, N. W. (1995). Estimating the UK demand for money function: A test of two approaches. Mimeo, Economics Department, University of Bristol.

Clements, M. P., and Hendry, D. F. (1996a). Intercept corrections and structural change. *Journal of Applied Econometrics*, **11**, 475–494.

Clements, M. P., and Hendry, D. F. (1996b). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics*, **58**, 657–684.

Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press. The Marshall Lectures on Economic Forecasting.

Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-Stationary Economic Time Series*. Cambridge, Mass.: MIT Press. The Zeuthen Lectures on Economic Forecasting.

Diebold, F. X., and Kilian, L. (1999). Unit root tests are useful for selecting forecasting models. mimeo, Department of Economics, University of Pennsylvania.

Doornik, J. A., and Hendry, D. F. (1996). *GiveWin: An Interactive Empirical Modelling Program*. London: Timberlake Consultants Press.

Doornik, J. A., and Hendry, D. F. (1998). Monte Carlo simulation using PcNaive for Windows. Unpublished typescript, Nuffield College, University of Oxford.

Doornik, J. A., and Hendry, D. F. (1999). *Modelling Dynamic Systems using PcFiml 9 for Windows*. London: Timberlake Consultants Press.

Durlauf, S. N., and Phillips, P. C. B. (1988). Trends versus random walks in time series analysis. *Econometrica*, **56**, 1333–1354.

Eitrheim, Ø., Husebø, T. A., and Nymoen, R. (1999). Equilibrium-correction versus differencing in macroeconometric forecasting. *Economic Modelling*, **16**, 515–544.

Friedman, M., and Schwartz, A. J. (1982). *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975*. Chicago: University of Chicago Press.

Perron, P., and Phillips, P. C. B. (1987). Does GNP have a unit root? – A re-evaluation. *Economics Letters*, **23**, 139–145.

Rappoport, P., and Reichlin, L. (1989). Segmented trends and non-stationary time series. *Economic Journal*, **99**, 168–177.

Rothman, P. (1998). Forecasting asymmetric unemployment rates. *The Review of Economics and Statistics*, **80**, 164–168.

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.

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

West, K. D. (1988). Asymptotic normality when regressors have a unit root. *Econometrica*, **56**, 1397–1417.

7 Appendix

7.1 Pseudo-true values $\{\phi, \gamma\}$ in a TS model for a DS DGP

The values of $\{\phi, \gamma\}$ that minimize the in-sample prediction error of the TS model are obtained by minimizing:

$$\begin{aligned}\mathsf{E}_{DS}\sum_{t=1}^{T}\left[y_t-(\phi+\gamma t)\right]^2 &= \mathsf{E}_{DS}\sum_{t=1}^{T}\left[y_0+\mu t+\sum_{s=1}^{t}\epsilon_s-(\phi+\gamma t)\right]^2 \\ &= (\mu-\gamma)^2\frac{1}{6}T(T+1)(2T+1)+T\phi(\phi-2y_0) \\ &\quad+\left(\sigma_\epsilon^2-2(\phi-y_0)(\mu-\gamma)\right)\frac{1}{2}T(T+1).\end{aligned}$$

The first-order conditions $\frac{\partial(\cdot)}{\partial\gamma}=0$ for a minimum yield:

$$0=-2(\mu-\gamma)\frac{1}{6}T(T+1)(2T+1)+2(\phi-y_0)\frac{1}{2}T(T+1),$$

so $\phi-y_0=\frac{1}{3}(\mu-\gamma)(2T+1)$; and $\frac{\partial(\cdot)}{\partial\phi}=0$:

$$0=2(\phi-y_0)T-2(\mu-\gamma)\frac{1}{2}T(T+1),$$

so $\phi-y_0=\frac{1}{2}(\mu-\gamma)(T+1)$, with solution $\{\phi=y_0,\gamma=\mu\}$.

7.2 Pseudo-true value of γ in a DS model for a TS DGP

The in-sample (1 to T) residual sum of squares of the DS predictor for the TS DGP (known parameters) is:

$$\sum_{t=1}^{T}\left[\phi+\gamma t+u_t-(\mu+y_{t-1})\right]^2=\sum_{t=1}^{T}(\gamma-\mu+\Delta u_t)^2 \qquad (43)$$

so:

$$\mathsf{E}_{TS}\sum_{t=1}^{T}\left[\phi+\gamma t+u_t-(\mu-y_{t-1})\right]^2=T\left(\gamma-\mu\right)^2+2T\sigma_u^2,$$

which is minimized over μ by setting $\mu=\gamma$.

7.3 OLS estimates of a TS model parameters for a DS DGP

After some algebra, the OLS estimates of $\widetilde{\phi}$ and $\widetilde{\gamma}$ in (9), when (1) is the DGP, are:

$$\widetilde{\gamma}-\mu\simeq 12T^{-3}\sum_{t=1}^{T}tS_t-6T^{-2}\sum_{t=1}^{T}S_t, \tag{44}$$

and:

$$\widetilde{\phi}-y_0\simeq 4T^{-1}\sum_{t=1}^{T}S_t-6T^{-2}\sum_{t=1}^{T}tS_t, \tag{45}$$

corresponding to 'spurious de-trending' – see Durlauf and Phillips (1988). $S_j=\sum_{s=1}^{j}\epsilon_s$ is the partial sum of the DS DGP disturbance, and the approximations are obtained from $T\simeq T+1$.

The following derivation evaluates one of the summations necessary to calculate the variances of these parameters.

7.3.1 Derivation of $\mathsf{E}_{DS}\left[\left(\sum_{t=1}^{T}tS_t\right)^2\right]$

Since $\sum_{t=1}^{T}tS_t=\sum_{j=1}^{T}\sum_{t=j}^{T}t\epsilon_j$:

$$\mathsf{E}_{DS}\left[\left(\sum_{j=1}^{T}\sum_{t=j}^{T}t\epsilon_j\right)^2\right]=\sigma_\epsilon^2\sum_{j=1}^{T}\left(\sum_{t=j}^{T}t\right)^2,$$

with:

$$\sum_{t=j}^{T}t=\sum_{t=1}^{T}t-\sum_{t=1}^{j-1}t=\frac{1}{2}T\left(T+1\right)-\frac{1}{2}\left(j-1\right)j,$$

$$\left(\sum_{t=j}^{T}t\right)^2=\frac{1}{4}T^2\left(T+1\right)^2+\frac{1}{4}\left(j-1\right)^2j^2-\frac{1}{2}T\left(T+1\right)j\left(j-1\right),$$

and so:

$$\sigma_\epsilon^2\sum_{j=1}^{T}\left(\sum_{t=j}^{T}t\right)^2\simeq\frac{2}{15}\sigma_\epsilon^2T^5.$$

$\mathsf{E}_{DS}\left[\left(\sum_{t=1}^{T}S_t\right)^2\right]$ and $\mathsf{E}_{DS}\left[\left(\sum_{t=1}^{T}S_t\right)\left(\sum_{t=1}^{T}tS_t\right)\right]$ can be derived similarly, and are approximately $\sigma_\epsilon^2\frac{T^3}{3}$ and $\frac{5}{24}\sigma_\epsilon^2T^4$, respectively.

7.3.2 Derivation of $\mathsf{V}_{DS}\left[\widetilde{\phi}\right]$

From (45):

$$\mathsf{V}_{DS}\left[\widetilde{\phi}\right]=\mathsf{E}_{DS}\left[\left(4T^{-1}\sum_{t=1}^{T}S_t-6T^{-2}\sum_{t=1}^{T}tS_t\right)^2\right],$$

and substituting the expressions derived above for the expectations of terms involving partial sums:

$$\mathsf{V}_{DS}\left[\widetilde{\phi}\right]\simeq\frac{16}{T^2}\left(\sigma_\epsilon^2\frac{T^3}{3}\right)-\frac{48}{T^3}\left(\frac{5}{24}\sigma_\epsilon^2T^4\right)+\frac{36}{T^4}\left(\frac{2}{15}\sigma_\epsilon^2T^5\right)\simeq\sigma_\epsilon^2T\frac{2}{15}. \quad (46)$$

7.3.3 Derivation of $\mathsf{V}_{DS}\left[\widetilde{\gamma}\right]$

From (44):

$$\mathsf{V}_{DS}\left[\widetilde{\gamma}\right]=\mathsf{E}_{DS}\left[(\widetilde{\gamma}-\mu)^2\right]=\mathsf{E}_{DS}\left[\left(12T^{-3}\sum_{t=1}^{T}tS_t-6T^{-2}\sum_{t=1}^{T}S_t\right)^2\right],$$

and substituting the expressions derived above for the expectations of terms involving partial sums:

$$\mathsf{V}_{DS}\left[\widetilde{\gamma}\right]\simeq\frac{144}{T^6}\left(\frac{2}{15}\sigma_\epsilon^2T^5\right)+\frac{36}{T^4}\left(\frac{1}{3}\sigma_\epsilon^2T^3\right)-\frac{144}{T^5}\left(\frac{5}{24}\sigma_\epsilon^2T^4\right)\simeq\frac{6}{5}\sigma_\epsilon^2T^{-1}. \quad (47)$$

7.3.4 Derivation of $\mathsf{E}_{DS}\left[S_T\widetilde{\phi}\right]$

$$\begin{aligned}\mathsf{E}_{DS}\left[S_T\widetilde{\phi}\right] &= \mathsf{E}_{DS}\left[4T^{-1}S_T\sum_{t=1}^{T}S_t-6T^{-2}S_T\sum_{t=1}^{T}tS_t\right] \quad (48)\\ &\simeq 4T^{-1}\sigma_\epsilon^2\left[T^2+T-\frac{T^2}{2}\right]-\sigma_\epsilon^2\frac{6}{T^2}\frac{T^3}{3}\simeq\sigma_\epsilon^2\left[2T-2T\right]=0,\end{aligned}$$

which uses $\sum_{s=1}^{T}\sum_{t=s}^{T}t=\sum_{s=1}^{T}[\frac{1}{2}T\left(T+1\right)-\frac{1}{2}\left(s-1\right)s]\simeq\frac{1}{3}T^3$.

7.3.5 Derivation of $\mathsf{E}_{DS}\left[S_T\left(\widetilde{\gamma}-\gamma\right)\right]$

$$\mathsf{E}_{DS}\left[S_T\left(\widetilde{\gamma}-\gamma\right)\right]=\mathsf{E}_{DS}\left[\frac{12S_T\sum_{t=1}^{T}tS_t-6\left(T+1\right)S_T\sum_{t=1}^{T}S_t}{T\left(T+1\right)\left(T-1\right)}\right]\simeq\sigma_\epsilon^2\left[\frac{12\frac{T^3}{3}-6\frac{T^3}{2}}{T^3}\right]=\sigma_\epsilon^2. \quad (49)$$

7.3.6 Derivation of $\mathsf{C}_{DS}\left[\widetilde{\phi},\widetilde{\gamma}\right]$

$$\begin{aligned}\mathsf{C}_{DS}\left[\widetilde{\phi},\widetilde{\gamma}\right] &= \mathsf{E}_{DS}\left[\left(\frac{12\sum_{t=1}^{T}tS_t-6\left(T+1\right)\sum_{t=1}^{T}S_t}{T\left(T+1\right)\left(T-1\right)}\right)\left(4T^{-1}\sum_{t=1}^{T}S_t-6T^{-2}\sum_{t=1}^{T}tS_t\right)\right]\\ &\simeq \sigma_\varepsilon^2\left[\frac{48}{T^4}\frac{5}{24}T^4-\frac{72}{T^5}\frac{2}{15}T^5-\frac{24}{T^3}\frac{T^3}{3}+\frac{36}{T^4}\frac{5}{24}T^4\right]=-0.1\sigma_\varepsilon^2. \quad (50)\end{aligned}$$