Document de travail

« WEAK » TRENDS FOR INFERENCE AND FORECASTING IN FINITE SAMPLES

Tendances « faibles » pour inférer et prévoir en échantillons de taille finie

N° 2004-12
Décembre 2004

GUILLAUME CHEVILLON*
OFCE et Economics Department, Oxford University

***e-mail : guillaume.chevillon@sciences-po.fr**
Adresse : OFCE, 69 quai d'Orsay, 75340 Paris cedex 07, France.

I am grateful to David Hendry, Bent Nielsen, Mike Clements and Sophocles Mavroeidis for helpful comments and suggestions.

Observatoire Français des Conjonctures Économiques
69, Quai d'Orsay 75340 Paris Cedex 07
Tel : 01 44 18 54 00 Fax : 01 45 56 06 15
E-mail: ofce@ofce.sciences-po.fr Web: http://www.ofce.sciences-po.fr

'Weak' trends for inference and forecasting in finite samples

Abstract

This paper studies the small sample properties of processes which exhibit both a stochastic and a deterministic trend. Whereas for estimation, inference and forecasting purposes the latter asymptotically dominates the former, it is not so when only a finite number of observations is available and large non-linearities in the parameter estimators result. To analyze this dependence, we resort to local-asymptotics and present the concept of a 'weak' trend whose coefficient is of order $O(T^{-1/2})$, so that the deterministic trend is $O(T^{1/2})$ and the process $O_p(T^{1/2})$. In this framework, parameter estimates, unit-root test statistics and forecast errors are functions of 'drifting' Ornstein-Uhlenbeck processes. We derive a comparison of direct and iterated multi-step estimation and forecasting of a—potentially misspecified—random walk with drift, and show that we explain well the non-linearities exhibited in finite samples. Another main benefit of direct multi-step estimation stems from some different behaviors of the 'multi-step' unit-root and drift tests under the weak and strong (constant coefficient) trend frameworks which could lead to testing which framework is more relevant. A Monte Carlo analysis validates the local-asymptotics approximation to the distributions of finite sample biases and test statistics.

Keywords: Stochastic Trend, Deterministic Trend, Local Asymptotics, Multi-step Forecasting.
JEL *Classification*: C22, C52, C53.

Résumé

Cet article étudie les propriétés en échantillons de petite taille de processus présentant conjointement une tendance stochastique et une déterministe. Tandis que cette dernière domine asymptotiquement la précédente pour ce qui est de l'estimation, de l'inférence et de la prévision, ce n'est pas le cas en présence d'un nombre fini d'observations et que les estimateurs de paramètres présentent de fortes non-linéarités. Afin d'analyser cette dépendance, nous recourons à des méthodes localement asymptotiques et introduisons le concept de tendance "faible" dont le coefficient est d'ordre $O(T^{-1/2})$, ce qui rend la tendance déterministe d'ordre $O(T^{1/2})$ et le processus $O_p(T^{1/2})$. Dans ce cadre, les estimateurs des paramètres, les statistiques de tests de racine unitaire et les erreurs de prévision sont fonction de processus d'Ornstein-Uhlenbeck "avec dérive". Nous présentons une comparaison de méthodes d'estimation et de prévision multi-étapes directe et itérée d'une marche aléatoire avec dérive (potentiellement mal spécifiée). Nous parvenons ainsi à expliquer la non-linéarité rencontrée en échantillons de taille finie. Un autre bénéfice de l'estimation multi-étapes provient des comportements différents des statistiques de tests "multi-étapes" de racine unitaire et de dérive selon que le cadre de tendance faible ou forte (à coefficient constant) s'applique, ce qui pourrait mener à un test entre ces derniers. Une simulation de Monte-Carlo valide l'approche d'approximation localement asymptotique pour ce qui concerne les distributions des biais d'estimation et de statistiques de tests en échantillons finis.

Mots-clefs: Tendance Stochastique, Tendance Déterministe, Asymptote Locale, Prévision multi-étapes.
Codes **JEL**: C22, C52, C53.

Contents

1 Introduction

There has been sustained interest over the last twenty years in studying the properties of deterministic and stochastic trends. Determining how to best represent the trending behavior of some economic process is still of interest and not yet entirely settled, as Diebold and Senhadji (1996) and Phillips (2004) show. Although it might be difficult to distinguish between deterministic and stochastic trends in small samples, it is essential to discern which is most appropriate since their long-term properties are radically different. Yet, as Sampson (1991) showed, when the parameters are to be estimated, the consequences of either model—when well-specified—are similar for prediction—as long as the true model is being used as Clements and Hendry (2001) point out.

Chevillon and Hendry (2004) have shown that when estimating the parameters of an AR(1) model when the model is truly a random walk with a drift which is close to zero, and with some potential misspecified error serial correlation, the estimators can be badly biased in finite samples owing to the presence of both a stochastic and a deterministic trend. When analyzing the consequences thereof for multi-step forecasting, via comparing an iterated one-step ahead procedure to a direct multi-step technique, they showed that this setting was highly beneficial to the latter forecasting method.

Our aim here is to introduce the concept of a 'weak' trend to model these findings and to show that there is in fact small-sample ambivalence and a continuum between deterministic and pure stochastic trends when allowing the parameter of the deterministic trend to vanish asymptotically.

The data generating process that we use is, thus, nearly I(1) with a local-to-zero drift and is generated by the model

$$y_{t,T} = \tau_T + \rho_T y_{t-1,T} + \epsilon_t, \quad (t, T = 1, 2, ...) \tag{1}$$

where $y_{0,T} = y_0^*$ is any random variable whose distribution is fixed and independent of T, including a constant. Thus, $\{y_{t,T}\}$ formally constitutes a triangular array of the type $\{y_{t,T} : t = 1, ..., T; T = 1, 2...\}$. But this is not central to our discussion and we refer to the process generated by (1) as $\{y_t\}$. In order to assess the small sample properties of the estimators, we assume a local-to-zero drift, and a local-to-unity root so that:

$$\tau_T = \psi/T^k \text{ and } \rho_T = \exp(\phi/T), \tag{2}$$

where k, ψ and ϕ can potentially take any real (not necessarily non-zero) values. If $\psi \neq 0$ and $k > 0$, the intercept in (1) tends to zero as the sample size increases. The parameter ϕ in (2) can be treated as a noncentrality parameter as in Phillips (1987b), so that, depending on its sign, the process $\{y_t\}$, for finite T, is either stable over some stretches of the data ($\phi < 0$), or not ($\phi \geq 0$); moreover, it is deterministically trending if $\phi = 0$ and $\psi \neq 0$; difference stationary if $\phi = 0$ and $\psi = 0$ (for appropriate choice of y_0^*); or explosive if $\phi > 0$. The parameters of the process tend, however, to those of a random walk without drift as, when $T \to \infty$, $\tau_T \to 0$ and $\rho_T \to 1$. Notice that the rates of approach to the limiting values vary for the two parameters: $O_p\left(T^{-k}\right)$ for the intercept and $O_p\left(T^{-1}\right)$ for the slope. Depending on the values of the parameters (k, ψ, ϕ) different asymptotic distributions result.

This article studies the principal aspects of weakly trending processes in terms of estimation and forecasting. It is therefore organized as follows. We first present our motivations for this analysis in section 2. We then provides the framework of our analysis and defines the drifting Ornstein-Uhlenbeck process that we require subsequently. In section 4, we derive some asymptotic results concerning sample statistics and use them in §5 to establish the limiting distributions of the biases from estimation of a weakly drifting AR(1) process. Section 6 gives the distribution of specification tests under weakly trending properties and shows that the behavior of the statistics radically differ from the strong trend case. The consequences of our framework for forecasting are considered in §7. A Monte Carlo analysis of the validity of the weak trend approximation in small samples follows in section 8. An appendix presents the proofs of the main theorems and lemmata and recalls the principal variables used.

2 Multi-step forecasting motivations

In the course of comparing two methods for forecasting at varying horizon, namely iterating a one-step ahead forecast (which they call iterated multi-step, or IMS) and direct multi-step estimation and forecasting (DMS), Chevillon and Hendry (2004) notice that the biases of the parameter estimators of a random walk with drift exhibit non-linearity in small samples. For simplicity, consider the IMA(1,1), with $x_0 = 0$ and, where we allow some misspecification $\theta \in (-1, 1)$ in:

$$\begin{aligned} \mathsf{DGP} \quad : \quad & x_t = \tau + x_{t-1} + \epsilon_t, \\ & \epsilon_t = \zeta_t + \theta \zeta_{t-1}, \end{aligned} \tag{3}$$

and $\zeta_t \sim \mathsf{IN}[0, \sigma_\zeta^2]$. Re-write (3) as $x_{T+h} = x_T + h\tau + \sum_{i=0}^{h-1} \epsilon_{T+h-i}$, with corresponding forecasts from the two methods given by:

$$\mathsf{M}_{IMS} \quad : \quad \widehat{x}_{T+h} = \widehat{\rho}^{\{h\}} \widehat{\alpha} + \widehat{\rho}^h x_T, \quad \text{and} \tag{4}$$

$$\mathsf{M}_{DMS} \quad : \quad \widetilde{x}_{T+h} = \widetilde{\tau}_h + \widetilde{\rho}_h x_T, \tag{5}$$

where the estimators are obtained by ordinary least-squares (omitting misspecification for both methods and the additional residual autocorrelation for DMS). When $\tau = 0$ and the intercept is not estimated, Banerjee, Hendry, and Mizon (1996) have shown that $\widetilde{\rho}_h$ is asymptotically more (respectively less) accurate than $\widehat{\rho}^h$ if θ is negative (resp. positive). By contrast, the presence of a non-zero drift means that IMS and DMS estimators share the same asymptotic distribution: both $(T^{1/2}(\widetilde{\tau}_h - h\tau), T^{3/2}(\widetilde{\rho}_h - 1))'$ and $(T^{1/2}\left(\widehat{\rho}^{\{h\}}\widehat{\alpha} - h\tau\right), T^{3/2}\left(\widehat{\rho}^h - 1\right))'$ converge in law towards

$$\mathsf{N}_2\left[\mathbf{0}, h^2 \begin{pmatrix} 4 & -6/\tau \\ -6/\tau & 12/\tau^2 \end{pmatrix} (1+\theta)^2 \sigma_\zeta^2\right]. \tag{6}$$

The distributions differ in finite samples, though since the conditional moments are non-

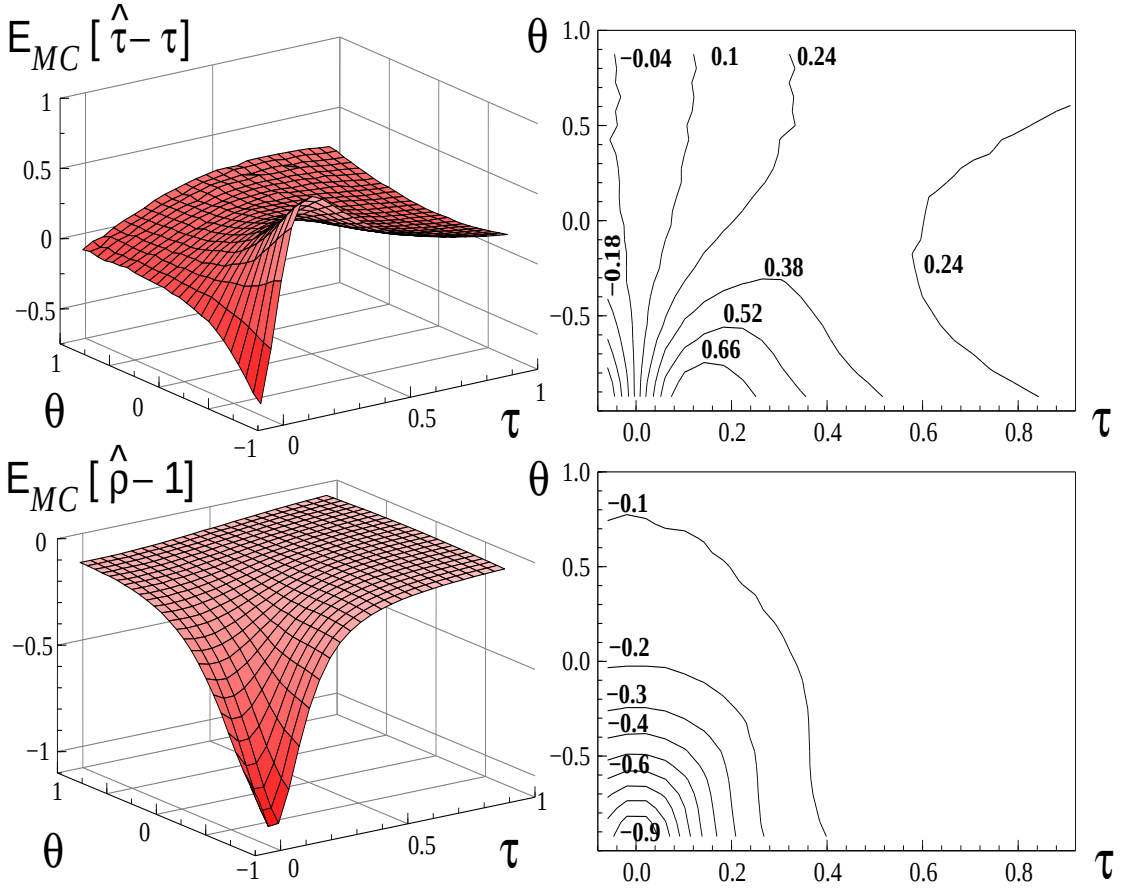

Figure 1: Monte Carlo estimates of the slope and intercept estimator biases for 1-step estimation for a sample of $T = 25$ observations, $10,000$ replications and varying drift and moving average coefficient. The right-hand-side panels (b and d) exhibit a set of contours for the panels on their left. The lines join points at the same altitude (z-axis) on the 3D graphs.

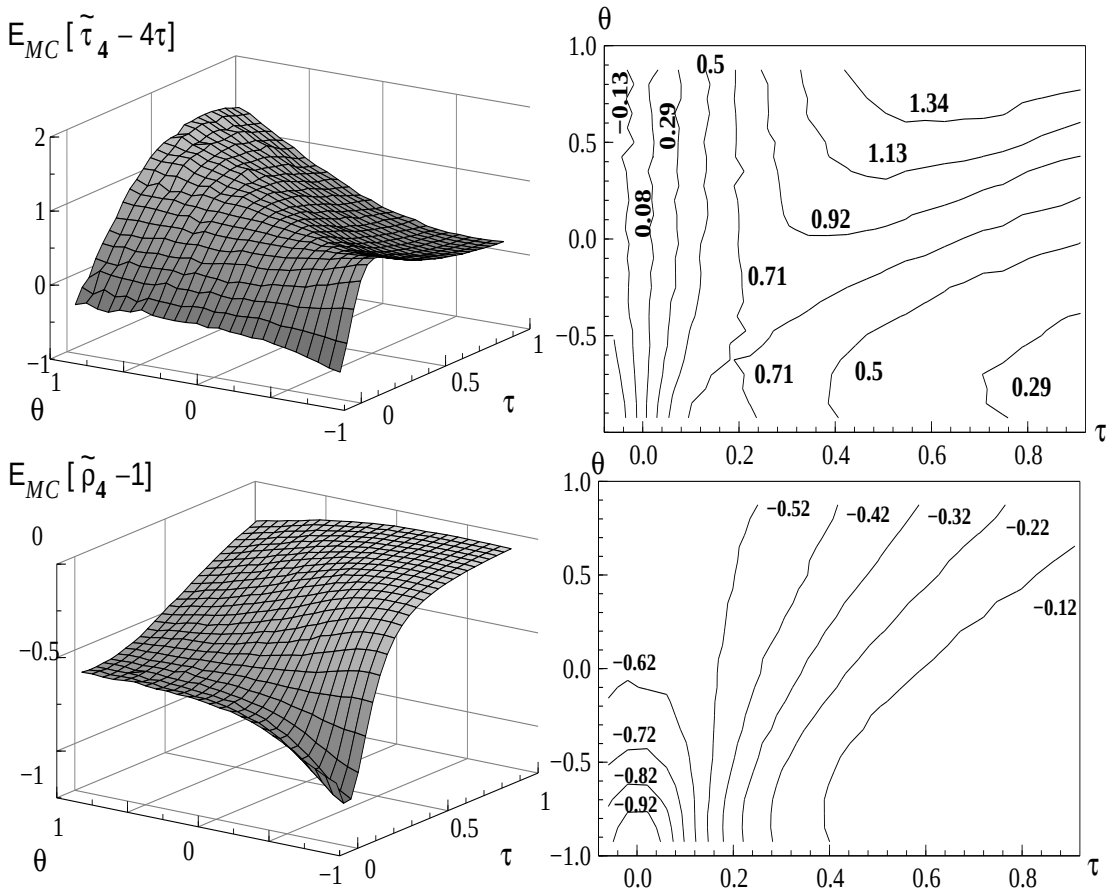

Figure 2: Monte Carlo estimates of the slope and intercept estimator biases for 4-step estimation for a sample of $T = 25$ observations, $10,000$ replications and varying drift and moving average coefficient. The right-hand-side panels (b and d) exhibit a set of contours for the panels on their left. The lines join points at the same altitude (z-axis) on the 3D graphs.

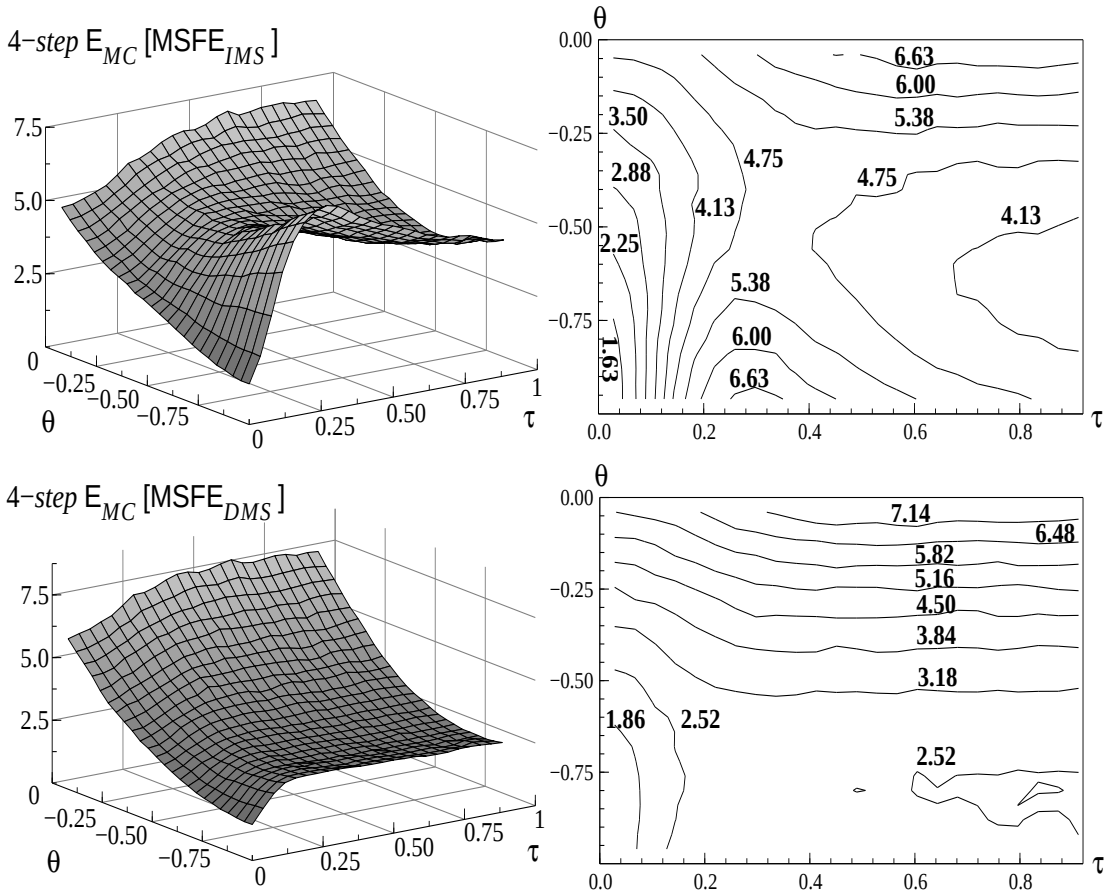

Figure 3: Monte Carlo estimates of the 4-step IMS and DMS MSFEs for a sample of $T = 25$ observations, $10,000$ replications and varying parameters.

constant, owing to the presence of a stochastic trend and an omitted moving-average component:

$$\begin{aligned} \mathsf{E}\left[x_{t+h} \mid x_t\right] &= \tau\left(h - \frac{\theta}{\frac{1}{t} + \frac{t-1}{t}\left(1+\theta\right)^2}\right) + \left(1 + \frac{\theta}{1+\left(1+\theta\right)^2\left(t-1\right)}\right) x_t, \qquad (7) \\ \mathsf{V}\left[x_{t+h} \mid x_t\right] &= \sigma_\zeta^2\left(1+\theta^2+\left(1+\theta\right)^2\left(h-1\right) - \frac{\theta^2}{1+\left(1+\theta\right)^2\left(t-1\right)}\right). \end{aligned}$$

The interaction between the stochastic and deterministic trends in small samples—where their influences have more similar magnitudes—therefore affects estimation. Given that the estimated parameters correspond to (7), we expect that, when θ is negative, one should under-estimate the unit-root and over-estimate the intercept. The negative asymptotic covariance of the biases in (6) reinforces this. Such a misestimation converts the intercept from a 'drift' term to an 'equilibrium mean' of the (pseudo-) stationary estimated process. The behaviors of the estimators for the two methods are therefore non-linear and non-monotonic in the parameters of the DGP and the horizon. In such a setting, DMS is more robust to unmodeled residual autocorrelation, as in Hall (1989).

Chevillon and Hendry (2004) proceed to a Monte Carlo simulation for a sample size of $T = 25$ observations and we reproduce here their graphs of the estimator biases, respectively figures[1] 1 and 2 for IMS and DMS estimation. We notice that the intercept estimation bias is non linear in the value of the drift and that, in the presence of omitted negative serial correlation of the residuals, its value achieves a local maximum for some value of the drift.In turn, these translate into the Monte Carlo means of the 4-step ahead unconditional Mean Square Forecast Errors as in figure 3 for the two models and response surfaces for the parameters.

First, in panels b and d, for non-zero θ and τ, DMS entails a lower mean square forecast error (MSFE) than IMS. The striking feature is that MSFE is generally increasing in θ for DMS, which means that a more misspecified model will forecast better, but this is not true for IMS when the drift is greater than 0.1, as the IMS MSFE surface is saddle shaped: it is increasing in θ for $\theta \geq -0.5$, decreasing elsewhere; increasing in τ for value smaller than about 0.3, but decreasing for higher values. It is this behavior which we wish to model via local asymptotics in this paper. But prior to this, we need to define our framework.

3 Preliminary theory

3.1 A weak trend

In our analysis, we assume throughout that $\{\epsilon_t\}_1^\infty$ in (1) is an innovation sequence which satisfies the following conditions (Phillips, 1987b):

[1] Figure panels are referred to as a to d left to right, top to bottom. Graphs were produced using GiveWin and the Ox programming language.

(i) $\quad \mathsf{E}[\epsilon_t] = 0 \quad$ for all t;

(ii) $\quad \sup_t \mathsf{E}|\epsilon_t|^{\beta+\eta} < \infty \quad$ for some $\beta > 2$ and $\eta > 0$;

(iii) $\quad \sigma^2 = \lim_{T\to\infty} T^{-1}\mathsf{E}\left[u_T^2\right]$ exists and $\sigma^2 > 0$, $\quad$ where $u_T = \sum_{j=1}^{T} \epsilon_j$; and

(iv) $\quad \{\epsilon_t\}$ is strongly mixing with mixing coefficients α_m such that $\sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty$.

Conditions (i)–(iv) are quite standard and reasonably weak. The innovation sequence $\{\epsilon_t\}$ is therefore heterogeneously distributed and weakly dependent over time. Thus many finite-order ARMA models, under general conditions, are possible. In particular, we will study the case of MA(1) errors:

$$\epsilon_t = \zeta_t + \theta\zeta_{t-1}, \tag{8}$$

and $\zeta_t \sim \mathsf{N}(0, \sigma_\zeta^2)$. Equation (1) can be re-written for $y_{t,T}$ as:

$$\begin{aligned} y_{t,T} &= \rho_T^t y_0 + \sum_{i=0}^{t-1} \rho_T^i (\tau_T + \epsilon_{t-i}), \\ &= \rho_T^t y_0 + \tau_T \sum_{i=0}^{t-1} \rho_T^i + \sum_{i=0}^{t-1} \rho_T^i \epsilon_{t-i}, \end{aligned}$$

and, letting $\lambda = e^{\phi}$,

$$\begin{aligned} &\text{if } \phi \neq 0: \quad y_{t,T} = \lambda^{t/T} y_0 + \psi T^{-k} \frac{1-\lambda^{t/T}}{1-\lambda^{1/T}} + \textstyle\sum_{i=0}^{t-1} \lambda^{i/T} \epsilon_{t-i}, \\ &\text{if } \phi = 0: \quad y_{t,T} = y_0 + \psi t T^{-k} + \textstyle\sum_{i=0}^{t-1} \epsilon_{t-i}. \end{aligned} \tag{9}$$

Notice that the distribution of $y_{t,T}$ is varying continuously for $\phi \to 0$. Let $[w]$ denotes the integer part of w for any real scalar w. Define, then, X_T in $D[0,1]$, the space of real-valued functions on the interval $[0,1]$ which are right continuous and have finite left limits (cadlag):[2]

$$\forall r \in [0,1], \quad X_T(r) = T^{-1/2} u_{[Tr]} \Rightarrow \sigma W(r), \quad \text{as } T \to \infty \tag{10}$$

where '$\Rightarrow$' denotes weak convergence of the associated probability measure, and $W(r)$ is a standard Brownian motion on $C[0,1]$ (the subspace of $D[0,1]$ of continuous functions); and, when $\{\epsilon_t\}$ is weakly stationary with covariance function series $\left\{\gamma_i^{(\epsilon)}\right\}_{i=1}^{\infty}$, the variance σ^2 is given by:

$$\sigma^2 = \lim_{T\to\infty} T^{-1}\mathsf{E}\left[u_T^2\right] = \lim_{T\to\infty} T^{-1}\left(\sum_{t=1}^{T} \mathsf{E}\left[\epsilon_t^2\right] + 2\sum_{t=1}^{T}\sum_{i=1}^{t-1} \mathsf{E}\left[\epsilon_t \epsilon_{t-i}\right]\right) = \sigma_\epsilon^2 + 2\sum_{i=1}^{\infty} \gamma_i^{(\epsilon)}. \tag{11}$$

Notice that $\lim_{T\to\infty} T^{-1}\mathsf{E}[u_T u_{T-i}] = \sigma^2$ for fixed i, and, when (8) is satisfied, $\sigma^2 = (1+\theta)^2 \sigma_\zeta^2$. We can now formally introduce the concept which constitutes the focus of this paper.

[2] $D = D[0,1]$ is endowed with the uniform metric

$$\begin{aligned} ||\cdot|| : \quad & D \to \mathbb{R} \\ & f \to \sup_r |f| \end{aligned}$$

Definition 1 *A time series $\{y_t\}$ which is generated by (1) and (2), and where $\{\epsilon_t\}$ satisfies (i)–(iv), is said to exhibit a 'weak' drift or (deterministic) trend if $\psi \neq 0$ and $k > 0$. When $\psi = 0$, $\{y_t\}$ is near-integrated ($\phi \neq 0$) or integrated ($\phi = 0$) as in Phillips (1987b). By contrast, a non-weak drift is said to be strong, and this is the case for $\psi \neq 0$ and $k \leq 0$.*

The terminology that we use here corresponds to the usage popularized by Staiger and Stock (1997) in the context of 'weak instruments'. The concepts of integratedness and of a near-integrated process follow, respectively, Box and Jenkins (1976) and Phillips (1987b). It should be noted that when $\psi \neq 0$ and $k > 0$, the weak trend includes cases when the process is either strongly autoregressive (even stationary), for $\phi < 0$, or mildly explosive, when $\phi > 0$, in finite samples.

3.2 A useful functional

To derive the results concerning the asymptotic properties of weakly trending processes, we need to define the functional $K_{\psi,\phi}(r)$:

$$K_{\psi,\phi}(r) = \psi f_\phi(r) + \sigma \int_0^r e^{\phi(r-s)} dW(s), \quad \text{for } r \in [0,1],$$

where we use the continuous deterministic functional $f_{(\cdot)} : \mathbb{R} \to C[0,1]$, such that for $\phi \in \mathbb{R}\backslash\{0\}$:[3]

$$f_\phi(\cdot) : r \to \frac{e^{\phi r} - 1}{\phi},$$

and $f_0(r) = r$. Notice the uniform continuity of the function $(\psi, \phi) \to K_{\psi,\phi}$ in $(\cdot, 0)$, since

$$\lim_{\phi \to 0} \left|\left| \mathsf{E}\left[(K_{\psi,\phi} - K_{\psi,0})^2 \right] \right|\right| = 0.$$

Here, $K_{\psi,\phi}(r)$ is a Gaussian process for fixed r and

$$K_{\psi,\phi}(r) \sim \mathsf{N}\left(\psi f_\phi(r), \sigma^2 f_{2\phi}(r)\right), \quad \forall r \in [0,1],$$

where '$\sim$' means equality in distribution. $K_{\psi,\phi}(r)$ follows the linear stochastic differential equation with white noise:

$$dK_{\psi,\phi}(r) = [\psi + \phi K_{\psi,\phi}(r)]\, dr + \sigma dW(r), \tag{12}$$

and initial condition $K_{\psi,\phi}(0) = 0$, so that, in the case where $\psi = 0$, it reduces to an Ornstein-Uhlenbeck process, satisfying the stochastic differential equation:

$$dK_{0,\phi}(r) = \phi K_{0,\phi}(r)\, dr + \sigma dW(r),$$

and initial condition $K_{0,\phi}(0) = 0$. For a fixed $r \in [0,1]$, the process $K_{0,\phi}(r)$ is normally distributed with mean zero and variance $\sigma^2 [\exp(2\phi r) - 1] / (2\phi)$. The presence of a non-zero trend implies that:

$$K_{\psi,\phi}(r) = \psi f_\phi(r) + K_{0,\phi}(r). \tag{13}$$

[3] We will straightforwardly extend this definition below to allow for $r > 1$ in forecasting.

The case $\phi = 0$ is slightly different since it, then, implies

$$K_{0,0}(r) = \sigma W(r),$$

but

$$K_{\psi,0}(r) = \psi r + \sigma W(r).$$

When $\psi = 0$, the theory matches that of Phillips (1987b). Notice that (12) could allow one to define $K_{\psi,\phi}(r)$ as an *Ornstein-Uhlenbeck process with drift* (or OU-d).

Remark 1 *Note that the formula (12) leads to the following properties (using Itô's lemma):*

$$dK^2_{\psi,\phi}(r) = \left(2K_{\psi,\phi}(r)\left[\psi + \phi K_{\psi,\phi}(r)\right] + \sigma^2\right)dr + 2\sigma K_{\psi,\phi}(r)\,dW(r), \text{ and} \tag{14}$$

$$\left\{K_{\psi,\phi}(1)\right\}^2 = \sigma^2 + 2\int_0^1 \left[\psi K_{\psi,\phi}(r) + \phi K^2_{\psi,\phi}(r)\right]dr \tag{15}$$

$$+2\sigma\int_0^1 K_{\psi,\phi}(r)\,dW(r), \tag{16}$$

which will prove useful.

4 Asymptotics for weakly trending processes

The first step is to find the asymptotic distribution of some population statistics in terms of the functional $K_{\psi,\phi}(r)$ as it was defined above. The first results that we need correspond to the order of magnitude of the intercept, namely k in (2). The following lemma provides the asymptotic distribution of the process.

Lemma 1 *If $\{y_t\}$ is a time series generated by (1) and (2), then as $T \to \infty$*

$$Y_T(r) = T^{-p}y_{[Tr],T} \Rightarrow \begin{cases} K_{0,\phi}(r) & \text{if } p = 1/2, \text{ and either } k > \frac{1}{2}, \text{ or } \psi = 0, \\ K_{\psi,\phi}(r) & \text{if } p = k = 1/2, \text{ and } \psi \neq 0, \\ \psi f_\phi(r) & \text{if } p > 1/2, \text{ and } k = 1 - p, \end{cases}$$

where, in the other cases, $T^{-p}y_{[Tr],T}$ diverges.

Proof. From (9), for $r \in [0,1]$,

$$y_{[Tr],T} = e^{\phi[Tr]/T}y_0 + \left[\sum\nolimits_{i=0}^{[Tr]-1} e^{\phi i/T}\right]\psi/T^k + \sum\nolimits_{i=0}^{[Tr]-1} e^{\phi i/T}\epsilon_{[Tr]-i}.$$

Thus:

$$\begin{aligned} T^{-p}y_{[Tr],T} &= T^{-(p+k)}\psi\left[\sum\nolimits_{i=0}^{[Tr]-1} e^{\phi i/T}\right] + T^{-p}\sum\nolimits_{i=1}^{[Tr]} e^{\phi([Tr]-i)/T}\epsilon_i, \\ &= T^{1-(p+k)}\psi f_\phi(r)\left(1 + O\left(T^{-1}\right)\right) + T^{-(p-1/2)}\left[T^{-1/2}\sum\nolimits_{i=1}^{[Tr]} e^{\phi([Tr]-i)/T}\epsilon_i\right], \end{aligned}$$

where $T^{-1/2}\sum_{i=1}^{[Tr]} e^{\phi([Tr]-i)/T}\epsilon_i \Rightarrow K_{0,\phi}(r)$ (Phillips, 1987b). Hence the different cases. ■

Notice that the case $p > 1/2$ and $k = 1-p$ includes $p = 1$ and $k = 0$, which corresponds to a non-degenerate linear trend, whose results are well-known. This framework allows for negative values of k, and the series may exhibit, for instance, a quadratic trend if $p = 2$ and $k = -1$. Since our purpose is to model the interaction between the deterministic trend and the unit root, we focus, in the rest of this paper, on the case $k = 1/2$, for which $T^{-1/2}y_{[Tr],T} \Rightarrow K_{\psi,\phi}(r)$.

Remark 2 *Notice the triangular aspect of our framework, where the distribution of $y_t|y_{t-1}$ depends on the sample size. Avoiding this issue, when we assume a unit-root, could lead to the following DGP:*

$$y_t = \tau_t + y_{t-1} + \epsilon_t, \quad (t = 1, 2, ...) \tag{17a}$$

$$\tau_t = \psi/\sqrt{t}, \tag{17b}$$

in which case

$$T^{-1/2}y_{[Tr]} \Rightarrow Z_\psi(r) = 2\psi\sqrt{r} + \sigma W(r),$$

and we notice that the deterministic components in $K_{\psi,0}(r)$ and $Z_\psi(r)$ are of different orders of magnitude with respect to r. The intercept that is observed for a sample of T observations is the average $\overline{\tau}_T = T^{-1}\sum\tau_t$, such that $\overline{\tau}_T/\left(2\psi T^{-1/2}\right) \to 1$ as $T \to \infty$.

We can now state our result concerning the sample statistics in terms of the functional defined above.

Lemma 2 *If $\{y_t\}$ is a near random–walk process with weak drift generated by (1) and (2), where $k = 1/2$, then as $T \to \infty$,*

(a) $T^{-1/2}y_{[Tr],T} \Rightarrow K_{\psi,\phi}(r)$;
(b) $T^{-3/2}\sum y_{t,T} \Rightarrow \int_0^1 K_{\psi,\phi}(r)\,dr$;
(c) $T^{-2}\sum y_{t,T}^2 \Rightarrow \int_0^1 K_{\psi,\phi}^2(r)\,dr$;
(d) $T^{-1}\sum y_{t-1,T}\epsilon_t \Rightarrow \sigma\int_0^1 K_{\psi,\phi}(r)\,dW(r) + \frac{1}{2}\left(\sigma^2 - \sigma_\epsilon^2\right)$;

with $\sigma_\epsilon^2 = \lim T^{-1}\sum \mathsf{E}\left(\epsilon_t^2\right)$. Joint weak convergence of (a)–(d) also applies.

Similar results hold for Remark (2) and $Z_\psi(r)$, replacing respectively (1)–(2) and $K_{\psi,\phi}(r)$.

Proof. It can be seen in Phillips (1987b) in the case $\psi = 0$; the proof in the non-zero case can be seen as a special case of the multi-step moments analyzed below. ■

Notice that, when (8) is satisfied, $\sigma_\epsilon^2 = \left(1+\theta^2\right)\sigma_\zeta^2$ and $\frac{1}{2}\left(\sigma^2 - \sigma_\epsilon^2\right) = \theta$. The results can be used to approximate the sample moments of weakly trending non-stationary time series. Since $K_{\psi,\phi}(r)$ is Gaussian, it is easy to show that:

$$\int_0^1 K_{\psi,\phi}(r)\,dr \sim \mathsf{N}\left(\psi\int_0^1 f_\phi(r)\,dr, \sigma^2\frac{\left(e^\phi - 1\right)^2\left(1 - e^{-2\phi}\right)}{2\phi^3}\right),$$

where $\int_0^1 f_\phi(r)\,dr = \left(e^\phi - \phi - 1\right)/\phi^2$ for $\phi \neq 0$ and (notice the continuity) $\int_0^1 f_0(r)\,dr = 1/2$. We, now, have at our disposal the elements necessary to establish the asymptotic properties derived from the estimation, inference and forecasting of a weakly trending process.

5 Estimation

5.1 OLS estimator biases

We derive in this subsection the asymptotic biases that result from the estimation of the misspecified first-order autoregressive model AR(1) with a fixed starting value $y_{0,T}$ and a fixed intercept; it is, potentially wrongly, assumed that the disturbances (ϵ_t) are weakly stationary and uncorrelated across time in:

$$y_{t,T} = \tau + \rho y_{t-1,T} + \epsilon_t, \quad (t = 1, 2, ..., T).$$

The estimation method used is ordinary least-squares (OLS) over a sample of size T. The corresponding biases, under the weak trend and local unit-root case as in (2), are given by:

$$\begin{bmatrix} \sqrt{T}\left(\widehat{\tau}_T - \tau_T\right) \\ T\left(\widehat{\rho}_T - \rho_T\right) \end{bmatrix} = \begin{bmatrix} 1 & T^{-3/2}\sum_{t=1}^T y_{t-1,T} \\ T^{-3/2}\sum_{t=1}^T y_{t-1,T} & T^{-2}\sum_{t=1}^T y_{t-1,T}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\sum_{t=1}^T \epsilon_t \\ T^{-1}\sum_{t=1}^T y_{t-1,T}\epsilon_t \end{bmatrix},$$

and the estimators are therefore consistent. Their limiting distribution is implied by lemma (2) and the continuous mapping theorem as:

$$\begin{bmatrix} \sqrt{T}\left(\widehat{\tau}_T - \tau_T\right) \\ T\left(\widehat{\rho}_T - \rho_T\right) \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \dfrac{\sigma W(1)\int_0^1 K_{\psi,\phi}^2(r)\,dr - \sigma\int_0^1 K_{\psi,\phi}(r)\,dr\left(\int_0^1 K_{\psi,\phi}(r)\,dW(r) + \frac{1}{2}\left(\sigma^2 - \sigma_\epsilon^2\right)\right)}{\int_0^1 K_{\psi,\phi}^2(r)\,dr - \left(\int_0^1 K_{\psi,\phi}(r)\,dr\right)^2} \\ \dfrac{\sigma\int_0^1 K_{\psi,\phi}(r)\,dW(r) - \sigma W(1)\int_0^1 K_{\psi,\phi}(r)\,dr + \frac{1}{2}\left(\sigma^2 - \sigma_\epsilon^2\right)}{\int_0^1 K_{\psi,\phi}^2(r)\,dr - \left(\int_0^1 K_{\psi,\phi}(r)\,dr\right)^2} \end{bmatrix}, \tag{18}$$

and, since $\rho_T = \exp(\phi/T) = 1 + \phi/T + O_p\left(T^{-2}\right)$, non-centrality of the slope estimator implies that:

$$\left\{T\left(\widehat{\rho}_T - 1\right) - T\left(\widehat{\rho}_T - \rho_T\right)\right\} \quad \Rightarrow \quad \phi, \tag{19}$$

which parallels $$\left\{\sqrt{T}\widehat{\tau}_T - \sqrt{T}\left(\widehat{\tau}_T - \tau_T\right)\right\} \quad = \quad \psi. \tag{20}$$

The parameters ϕ and ψ, thus imply some non-centrality in the limit of, respectively, the unit-root and the intercept estimators. Notice that the presence of a weak drift implies that both the stochastic and deterministic trends are of the same asymptotic orders of magnitude. The unit-root estimator is thus super-consistent and the corresponding bias is of order $O_p\left(T^{-1}\right)$ and not $O_p\left(T^{-3/2}\right)$ as in the presence of a strong trend. Indeed, in the latter case—$k = 0$ and $\tau_T = \tau^*$ in

(2)—OLS estimation leads to:

$$\begin{bmatrix} T^{1/2}(\widetilde{\tau}^* - \tau^*) \\ T^{3/2}(\widetilde{\rho}^* - 1) \end{bmatrix} \underset{T\to\infty}{\overset{L}{\to}} \mathsf{N}\left(0, \sigma^2 \begin{bmatrix} 4 & -6/\tau^* \\ -6/\tau^* & 12/\tau^{*2} \end{bmatrix}\right). \tag{21}$$

The consistency of the intercept estimator is more straightforward since both the weak drift and its estimator tend towards zero; yet, it remains possible to characterize its bias. When $\psi = 0$, (18) simplifies to the results in Phillips (1987b) concerning near-integratedness.

5.2 Powered-up one-step ahead estimators

To derive the implied distributions of the nonlinear combination of the OLS estimators which provide the unconditional expectation of $y_{t,T}$ in terms of $y_{t-h,T}$ for $h > 1$. This can be written as in the multi-step parameterization in the model:

$$y_{t,T} = \tau_{h,T} + \rho_{h,T} y_{t-h,T} + w_{h,t}, \quad (t, T = h, h+1, ...), \text{ for } h \geq 1, \tag{22}$$

with $\mathsf{E}[w_{h,t}] = 0$, but $\mathsf{E}[w_{h,t}|y_{t-h,T}]$ may be nonzero, in which case:

$$\begin{aligned} \mathsf{E}[y_{t,T}] &= \tau_{h,T} + \rho_{h,T}\mathsf{E}[y_{t-h,T}], \\ \text{but } \mathsf{E}[y_{t,T}|y_{t-h,T}] &\neq \tau_{h,T} + \rho_{h,T} y_{t-h,T}. \end{aligned}$$

We assume that the data generating process is given by (1), which implies that y_t is given by:

$$y_t = \left(\sum_{i=0}^{h-1} \rho_T^i\right) \tau_T + \rho_T^h y_{t-h} + \sum_{j=0}^{h-1} \rho_T^j \epsilon_{t-j}.$$

Define $\rho_T^{\{h\}} = \sum_{i=0}^{h-1} \rho_T^i$ and similarly for $\widehat{\rho}_T^{\{h\}}$. Let, then, $\widehat{\tau}_{\{h\},T} = \widehat{\rho}_T^{\{h\}}\widehat{\tau}_T$. Thus, in (22): $\tau_{h,T} = \rho_T^{\{h\}}\tau_T$, $\rho_{h,T} = \rho_T^h$ and $w_{h,t} = \sum_{j=0}^{h-1} e^{j\phi/T}\epsilon_{t-j}$; using the delta method:

$$\frac{\widehat{\tau}_{\{h\},T} - \tau_{h,T}}{\widehat{\tau}_T - \tau_T} \Rightarrow h \quad \text{and} \quad \frac{\widehat{\rho}_T^h - \rho_T^h}{\widehat{\rho}_T - \rho_T} \Rightarrow h. \tag{23}$$

5.3 Direct multi-step estimators

We now wish to compare the powered-up one-step ahead estimators to those obtained by direct in-sample minimization of the multi-step disturbances $w_{h,t}$, whose variance we define as:

$$\sigma^2_{w_h} = \lim_{T\to\infty} T^{-1}\mathsf{E}\left[\sum_{t=h}^{T} w_{h,t}^2\right] = h\sigma_\epsilon^2 + 2\sum_{i=1}^{h-1}(h-i)\gamma_i^{(\epsilon)}, \tag{24}$$

and let:

$$T^{-1/2} v_{h,[Tr]} = T^{-1/2}\sum_{i=h}^{[Tr]} w_{h,i} = T^{-1/2}\sum_{j=0}^{h-1} e^{j\phi/T} L^j u_{[Tr]},$$

where L is the lag operator such that $L^j u_{[Tr]-j}$. The asymptotic multi-step variance of $\{v_{h,T}\}$ is:

$$\sigma_h^2 = \lim_{T\to\infty} T^{-1}\mathsf{E}\left[v_{h,T}^2\right] = h^2\sigma^2,$$

so that, in terms of variance, $\{v_{h,T}\}$ behaves as h times the near stochastic trend $\{u_t\}$. Then, simply, for $r \in [0,1]$,

$$X_{h,T}(r) = T^{-1/2} v_{h,[Tr]} \Rightarrow \sigma_h W(r) = h\sigma W(r), \quad \text{as } T \to \infty$$

and we derive the equivalent of lemma 2 for the multi-step statistics.

Theorem 1 *If $\{y_t\}$ is a near random-walk process with weak drift generated by (1) and (2), then, for all fixed integer $h \in [1,T)$, as $T \to \infty$,*

$$\begin{array}{ll} (a_h) & T^{-1/2} y_{[Tr],T} \Rightarrow K_{\psi,\phi}(r) \\ (b_h) & T^{-3/2} \sum_{t=h}^{T} y_{t,T} \Rightarrow \int_0^1 K_{\psi,\phi}(r)\, dr; \\ (c_h) & T^{-2} \sum_{t=h}^{T} y_{t,T}^2 \Rightarrow \int_0^1 K_{\psi,\phi}^2(r)\, dr; \\ (d_h) & T^{-1} \sum_{t=h}^{T} y_{t-h,T} w_{h,t} \Rightarrow h\sigma \int_0^1 K_{\psi,\phi}(r)\, dW(r) + \frac{1}{2}\left[h\sigma^2 - \sigma_{w_h}^2\right]. \end{array}$$

Joint convergence of (a_h)–(d_h) also holds. (Proof in appendix).

Remark 3 *Theorem 1 relies strongly on the assumption that h is fixed since if it were not and, for instance $h = cT^{\alpha}$ for some $\alpha > 0$, then*

$$h^{-1}T^{-1} \sum_{t=h}^{T} y_{t-h,T} w_{h,t} \Rightarrow \sigma \int_0^1 K_{\psi,\phi}(r)\, dW(r) + \frac{1}{2}\sigma^2.$$

Notice that, when the error process is covariance stationary,

$$\sigma_{w_h}^2 = h\sigma_\epsilon^2 + 2\sum_{i=1}^{h-1} (h-i)\,\gamma_i^{(\epsilon)},$$

and we can re-write (d_h) from Theorem 1, as:

$$T^{-1} \sum_{t=h}^{T} y_{t-h,T} w_{h,t} \Rightarrow h\left\{\sigma \int_0^1 K_{\psi,\phi}(r)\, dW(r) + \frac{1}{2}\left[\sigma^2 - \sigma_\epsilon^2\right]\right\} - \sum_{i=1}^{h-1} (h-i)\,\gamma_i^{(\epsilon)},$$

i.e.

$$T^{-1}\left[\sum_{t=h}^{T} y_{t-h,T} w_{h,t} - h\sum_{t=1}^{T} y_{t-1,T}\epsilon_t\right] \Rightarrow -\sum_{i=1}^{h-1} (h-i)\,\gamma_i^{(\epsilon)}. \tag{25}$$

This latter term (25) shows that, if the process $\{\epsilon_t\}$ exhibits some serial correlation, then IMS and DMS estimations have different asymptotic properties. This result confirms that of Weiss (1991) who showed that when the estimated model is misspecified and exhibits omitted error serial correlation, there can exist cases when it is preferable to use multi-step estimation. When (8) is satisfied, the expressions simplify and $\sigma_{w_h}^2 = h\sigma^2 - 2\theta\sigma_\zeta^2$ and $\left(h\sigma^2 - \sigma_{w_h}^2\right)/2 = \theta$.

We can now derive the limiting distributions of the estimators and, in OLS estimation of (22), without modelling the error process, the multi-step biases become:

$$\begin{aligned}&\begin{bmatrix} \sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right) \\ T\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right) \end{bmatrix} \\ &\Rightarrow \begin{bmatrix} 1 & \int_0^1 K_{\psi,\phi}(r)\,dr \\ \int_0^1 K_{\psi,\phi}(r)\,dr & \int_0^1 K^2_{\psi,\phi}(r)\,dr \end{bmatrix}^{-1} \\ &\times \begin{bmatrix} h\sigma W(1) \\ h\left\{\sigma\int_0^1 K_{\psi,\phi}(r)\,dW(r)+\frac{1}{2}\left[\sigma^2-\sigma^2_\epsilon\right]\right\}-\sum_{i=1}^{h-1}(h-i)\,\gamma_i^{(\epsilon)} \end{bmatrix}.\end{aligned} \tag{26}$$

And we notice that $\rho_{h,T}=\exp(h\phi/T)=1+h\phi/T+O_p\left(T^{-2}\right)$, so that:

$$\left\{T\left(\widetilde{\rho}_{h,T}-1\right)-T\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right)\right\}\Rightarrow h\phi,$$

and the noncentrality of the multi-step estimator shifts with the horizon h. So does, in fact, the powered-up estimator, and we now compare them both.

5.4 Comparison of estimation techniques

If we compare both the powered-up (IMS) and direct (DMS) estimators, we notice that their biases differ asymptotically by only:

$$\begin{aligned}&\begin{bmatrix} \sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right) \\ T\left(\widetilde{\rho}_{h,T}-1\right) \end{bmatrix}-\begin{bmatrix} \sqrt{T}\left(\widehat{\tau}^{\{h\}}_T-\tau_{h,T}\right) \\ T\left(\widehat{\rho}^h_T-1\right) \end{bmatrix} \\ &\Rightarrow \frac{\left[\sum_{i=1}^{h-1}(h-i)\,\gamma_i^{(\epsilon)}\right]}{\int_0^1 K^2_{\psi,\phi}(r)\,dr-\left(\int_0^1 K_{\psi,\phi}(r)\,dr\right)^2}\begin{bmatrix} \int_0^1 K_{\psi,\phi}(r)\,dr \\ -1 \end{bmatrix}\end{aligned} \tag{27}$$

The interesting feature here is that the asymptotic efficiency gain—or loss—from using multi-step estimation depends on the first h terms of the autocovariance function of $\{\epsilon_t\}$, and not of those beyond. The larger the horizon, the more weight is accorded to the first autocovariance terms. Since this is the only dependence on h in (27), the natural consequence is that, if there exists h^*, such that $\gamma_i^{(\epsilon)}=0$, for $i\geq h^*$, whichever method dominates at horizon h^* will do so increasingly as the horizon grows thereafter. Notice, moreover, that

$$\int_0^1 K^2_{\psi,\phi}(r)\,dr-\left(\int_0^1 K_{\psi,\phi}(r)\,dr\right)^2=\int_0^1\left(K_{\psi,\phi}(r)-\int_0^1 K_{\psi,\phi}(s)\,ds\right)^2 dr\overset{a.s.}{>}0,$$

and, thus, is strictly positive with probability 1. Estimation of the unit-root is, therefore, such that the sign of the asymptotic distribution of $T\left(\widetilde{\rho}_{h,T}-\widehat{\rho}^h_T\right)$ is opposite to that of $\left[\sum_{i=1}^{h-1}(h-i)\,\gamma_i^{(\epsilon)}\right]$. For instance, given that unit-roots are more often under-estimated than the converse, negative residual autocorrelation favours multi-step estimation. Since the denominator of (27) is non-constant, and is not independent of $\int_0^1 K_{\psi,\phi}(r)\,dr$, we cannot easily predict the sign of the difference

in the intercept biases from the two estimation methods. Yet, as the sign of $\mathsf{E}\left[\int_0^1 K_{\psi,\phi}(r)\,dr\right]$ is that of ψ, assuming (i) that:

$$\mathsf{E}\left[\left(\int_0^1 K_{\psi,\phi}^2(r)\,dr - \left(\int_0^1 K_{\psi,\phi}(r)\,dr\right)^2\right)^{-1}\int_0^1 K_{\psi,\phi}(r)\,dr\right]$$

and

$$\mathsf{E}\left[\left(\int_0^1 K_{\psi,\phi}^2(r)\,dr - \left(\int_0^1 K_{\psi,\phi}(r)\,dr\right)^2\right)\right]\mathsf{E}\left[\int_0^1 K_{\psi,\phi}(r)\,dr\right]$$

have the same sign and (ii) that the intercept is over-estimated (in absolute value) on average—which is a natural assumption if the slope if underestimated—then again (27) implies that negative serial correlation of the error process benefits the multi-step estimator. As it is quite complicated to prove our assumptions analytically, we will use a Monte Carlo experiment in section 8.

5.5 Omitted moving averages

Now in the case from equation (8), $\sigma = 1+\theta$ and:

$$\begin{aligned} K_{\psi,\phi}(r) &= \psi f_\phi(r) + \sigma\int_0^1 e^{\phi(r-s)}dW(s) \\ &= \psi f_\phi(r) + (1+\theta)J_\phi(r), \end{aligned}$$

where we define $J_\phi(r) = \sigma^{-1}K_{0,\phi}(r)$, to make explicit how the parameters influence $K_{\psi,\phi}(r)$. Then:

$$K_{\psi,\phi}^2(r) = \psi^2 f_\phi^2(r) + 2\psi\sigma f_\phi(r)J_\phi(r) + \sigma^2 J_\phi^2(r).$$

We define the operators:

$$\begin{aligned} \mathcal{D}(f,g) &= \int_0^1 f(r)g(r)\,dr - \left(\int_0^1 f(r)\,dr\right)\left(\int_0^1 g(r)\,dr\right), \\ \mathcal{Q}(f,g) &= W(1)\int_0^1 f(r)g(r)\,dr - \left(\int_0^1 f(r)\,dr\right)\left(\int_0^1 g(r)\,dW(r)\right), \\ \mathcal{L}(f) &= \int_0^1 f(r)\,dW(r) - W(1)\int_0^1 f(r)\,dr, \\ \mathcal{I}(f) &= \int_0^1 f(r)\,dr, \end{aligned}$$

where for fixed r,

$$\int_0^1 K_{\psi,\phi}(r)\,dr \sim \mathsf{N}\left(\psi\int_0^1 f_\phi(r)\,dr, \frac{1}{2}\sigma^2\left(e^\phi-1\right)^2\left(1-e^{-2\phi}\right)/\phi^3\right),$$

and write $\mathcal{L}(f,f) = \mathcal{L}(f)$, and similarly for the other operators. Recall that $\mathcal{D}(g) \overset{a.s.}{>} 0$ for all g. $\mathcal{D}(\cdot)$ and $\mathcal{Q}(\cdot)$ are quadratic in (ψ,σ) as

$$\begin{aligned} \mathcal{D}(K_{\psi,\phi}) &= \psi^2\mathcal{D}(f_\phi) + 2\psi\sigma\mathcal{D}(f_\phi,J_\phi) + \sigma^2\mathcal{D}(J_\phi), \\ \mathcal{Q}(K_{\psi,\phi}) &= \psi^2\mathcal{Q}(f_\phi) + \psi\sigma\left[\mathcal{Q}(f_\phi,J_\phi) + \mathcal{Q}(J_\phi,f_\phi)\right] + \sigma^2\mathcal{Q}(J_\phi), \end{aligned}$$

and $\mathcal{L}(\cdot)$ and $\mathcal{I}(\cdot)$ are linear:

$$\begin{aligned} \mathcal{L}(K_{\psi,\phi}) &= \psi\mathcal{L}(f_\phi) + \sigma\mathcal{L}(J_\phi), \\ \mathcal{I}(K_{\psi,\phi}) &= \psi\mathcal{I}(f_\phi) + \sigma\mathcal{I}(J_\phi), \end{aligned}$$

where $\mathsf{E}[\mathcal{I}(f_\phi)] \geq 0$,

$$\mathcal{I}(K_{\psi,\phi}) \sim \mathsf{N}\left(\psi\left(e^{\phi}-\phi-1\right), \frac{1}{2}\sigma^2\left(e^{\phi}-1\right)^2\left(1-e^{-2\phi}\right)/\phi^3\right), \text{ if } \phi \neq 0,$$

and $\mathcal{L}(f_\phi) \sim \mathsf{N}(0, \mathcal{D}(f_\phi))$. Notice that $\mathcal{L}(K_{\psi,\phi})$ is the limit of the scaled sample covariance of y_t and ϵ_t, whose true population value is σ_ϵ^2. We can re-write (26) as

$$\begin{bmatrix} \sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right) \\ T\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right) \end{bmatrix} \Rightarrow \left[\mathcal{D}(K_{\psi,\phi})\right]^{-1}\left\{h\sigma\begin{bmatrix}\mathcal{Q}(K_{\psi,\phi})\\ \mathcal{L}(K_{\psi,\phi})\end{bmatrix}+\theta\begin{bmatrix}\mathcal{I}(K_{\psi,\phi})\\ -1\end{bmatrix}\right\}.$$

The h-step DMS biases are, thus, non-linear in (ψ,σ). For θ close to -1, i.e. σ near zero, the dependence in h is not strong and we can predict that the biases will be close at all horizons h. We know, besides, that for $\theta \simeq -1$, $T\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right)$ will tend to a distribution close to that of:

$$-\theta\left[\psi^2\mathcal{D}(f_\phi)+2\psi\sigma\mathcal{D}(f_\phi,J_\phi)+\sigma^2\mathcal{D}(J_\phi)\right]^{-1},$$

which is the inverse of a quadratic function with positive—stochastic—coefficients $\mathcal{D}(f_\phi)$ and $\mathcal{D}(J_\phi)$, and given one realization, achieves a global maximum in the (ψ,σ) space at $(0,0)$ and monotonically decreases as $\|\psi,\sigma\|^2 = \psi^2+\sigma^2$ increases. As regards the intercept bias when $\theta \simeq -1$, and assuming that $\mathcal{I}(K_{\psi,\phi})$ and $\mathcal{D}(K_{\psi,\phi})$ are mildly correlated, its expectation is increasing in ψ for low σ. For larger σ, it is no longer possible to neglect $\mathcal{Q}(K_{\psi,\phi})$, but since:

$$\mathcal{Q}(K_{\psi,\phi}) = W(1)\mathcal{D}(K_{\psi,\phi}) - \mathcal{I}(K_{\psi,\phi})\mathcal{L}(K_{\psi,\phi}),$$

we can conclude that:

$$\sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right)+\mathcal{I}(K_{\psi,\phi})T\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right) \Rightarrow h\sigma W(1),$$

and the behavior patterns of both biases are likely to be opposite as $\|\psi,\sigma\|$ increases. Given that a larger $\|\psi,\sigma\|$ means that the trend coefficient is larger or that there is less residual autocorrelation, the unit-root bias is then likely to be lower, so that both intercept and slope biases are then decreasing in $\|\psi,\sigma\|$. The conclusion form this heuristic analysis is that *the unit-root bias is decreasing overall in* $\|\psi,\sigma\|$ *and the intercept bias is first increasing then decreasing in* $\|\psi,\sigma\|$ *and there is a set of coefficients* (ψ,σ) *for which the bias achieves a local (or even global) maximum.*

If we compare both methods, (27) becomes:

$$\begin{bmatrix} \sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right) \\ T\left(\widetilde{\rho}_{h,T}-1\right) \end{bmatrix} - \begin{bmatrix} \sqrt{T}\left(\widehat{\tau}_T^{\{h\}}-\tau_{h,T}\right) \\ T\left(\widehat{\rho}_T^{h}-1\right) \end{bmatrix} \Rightarrow (h-1)\theta\begin{bmatrix}\left[\mathcal{D}(K_{\psi,\phi})\right]^{-1}\mathcal{I}(K_{\psi,\phi})\\ -\left[\mathcal{D}(K_{\psi,\phi})\right]^{-1}\end{bmatrix}, \quad (28)$$

and under our heuristic assumption, the difference between the two is decreasing in $\|\psi,\sigma\|$ for the slope and increasing for the intercept. The DMS unit-root estimator is, moreover, larger than the IMS and the sign of $\widetilde{\tau}_{h,T}-\widehat{\tau}_T^{\{h\}}$ is, asymptotically, opposite that of the intercept, and given the plausibility that the slope is under-estimated—which implies that the absolute value of the intercept is over-estimated—the conclusion of this analysis, so far, is that *the multi-step estimators are more accurate than the powered-up one-step when* $||\psi,\sigma||$ *is low.*

6 Inference

6.1 Unit-root and drift t-tests

Assume that the model is estimated by least-squares and provides the estimates $\left(\widetilde{\tau}_{h,T}, \widetilde{\rho}_{h,T}\right)$. It is now wished to test some hypotheses about their true values, namely, to determine whether the intercept is significantly different from zero and whether the series exhibit a unit-root, and the combination of both. A modeler who would posit the potential presence of a nonzero drift would normally resort to a joint test together with that of a unit root. Here, using an F-test would render the results less easy to observe and we resort to single hypothesis testing. Under the null hypothesis of no intercept and a unit-root, t-statistics are constructed where

$$t_{\tau_h} = \frac{\widetilde{\tau}_{h,T}}{\widetilde{\sigma}_{\tau_h}} = \frac{T^{1/2}\left(\tau_{h,T} + \widetilde{\tau}_{h,T} - \tau_{h,T}\right)}{\left\{T\widetilde{\sigma}^2_{\tau_h}\right\}^{1/2}}, \quad \text{and} \quad t_{\rho_h} = \frac{\widetilde{\rho}_{h,T} - 1}{\widetilde{\sigma}_{\rho_h}} = \frac{T\left(\widetilde{\rho}_{h,T} - 1\right)}{\left\{T^2\widetilde{\sigma}^2_{\rho_h}\right\}^{1/2}},$$

with

$$\Sigma_{h,T} = \begin{bmatrix} T-h+1 & \sum_{t=h}^{T} y_{t-h,T} \\ \sum_{t=h}^{T} y_{t-h,T} & \sum_{t=h}^{T} y^2_{t-h,T} \end{bmatrix},$$

and $T\widetilde{\sigma}_{\tau_h} = s^2_{h,T}\begin{bmatrix} T^{1/2} & 0 \end{bmatrix}\Sigma^{-1}_{h,T}\begin{bmatrix} T^{1/2} \\ 0 \end{bmatrix}$, $T^2\widetilde{\sigma}_{\rho_h} = s^2_{h,T}\begin{bmatrix} 0 & T \end{bmatrix}\Sigma^{-1}_{h,T}\begin{bmatrix} 0 \\ T \end{bmatrix}$. The usual formula is used for:

$$s^2_{h,T} = \frac{1}{T-h-1}\sum_{t=h}^{T}\left(y_{t,T} - \widetilde{\tau}_{h,T} - \widetilde{\rho}_{h,T}y_{t-h,T}\right)^2,$$

and since $s^2_{h,T} = \frac{1}{T-h-1}\sum_{t=h}^{T}\left(w_{h,t} + O_p\left(T^{-1/2}\right)\right)^2$, obviously $s^2_{h,T} \xrightarrow{p} \sigma^2_{w_h}$. Letting

$$\varpi_h = \frac{1}{2}\left(1 - \sigma^2_\epsilon/\sigma^2\right) - \left(h\sigma^2\right)^{-1}\left[\sum_{i=1}^{h-1}(h-i)\,\gamma^{(\epsilon)}_i\right],$$

which is zero if ε_t is white noise, the continuous mapping theorem implies that:

$$t_{\tau_h} \Rightarrow \sqrt{h}\frac{\psi + \sigma^2\left\{\mathcal{D}\left(K_{\psi,\phi}\right)\right\}^{-1}\left[\sigma^{-1}\mathcal{Q}\left(K_{\psi,\phi}\right) - \varpi_h\int_0^1 K_{\psi,\phi}\left(r\right)dr\right]}{\left\{\mathcal{D}\left(K_{\psi,\phi}\right)\right\}^{-1/2}\left\{\sigma^2_\epsilon + 2h^{-1}\sum_{i=1}^{h-1}(h-i)\,\gamma^{(\epsilon)}_i\right\}^{1/2}\left\{\int_0^1 K^2_{\psi,\phi}\left(r\right)dr\right\}^{1/2}},$$

$$t_{\rho_h} \Rightarrow \sqrt{h}\frac{\phi + \sigma^2\left\{\mathcal{D}\left(K_{\psi,\phi}\right)\right\}^{-1}\left[\sigma^{-1}\mathcal{L}\left(K_{\psi,\phi}\right) + \varpi_h\right]}{\left\{\mathcal{D}\left(K_{\psi,\phi}\right)\right\}^{-1/2}\left\{\sigma^2_\epsilon + 2h^{-1}\sum_{i=1}^{h-1}(h-i)\,\gamma^{(\epsilon)}_i\right\}^{1/2}}.$$

We, thus, notice that, although the process can exhibit a deterministic trend in small samples, the t_{τ_h} statistic under the wrong null hypothesis has a finite distribution. This is to be compared with the case when the trend is strong. In this latter case, $t^*_{\tau_h} = O_p\left(\sqrt{T}\right)$ although, under the null of a unit root and zero intercept, the distribution used for testing is a Dickey-Fuller. The immediate

consequence of this result is that, in finite samples, using the t-statistics will lead to over-rejection of the presence of a deterministic trend. Indeed in the case of a non–weak trend, where $k=0$ in (1), the biases in (21) imply that, under the null of a unit root and of, now, $\tau_{h,T}=\alpha$:

$$t^*_{\tau_h} = \frac{\widetilde{\tau}_{h,T}-\alpha}{\widetilde{\sigma}_{\tau_h}} = \frac{T^{1/2}\left(\tau_{h,T}-\alpha+\widetilde{\tau}_{h,T}-\tau_{h,T}\right)}{\left\{T\widetilde{\sigma}^2_{\tau_h}\right\}^{1/2}}$$

$$t_{\tau_h} - \sqrt{h}T^{1/2}\frac{(\tau-\alpha/h)}{2\sqrt{\sigma^2_\epsilon+2\sum_{i=1}^{h-1}(1-i/h)\,\gamma^{(\epsilon)}_i}} \xrightarrow{L} \mathsf{N}\left[0, h\frac{\sigma^2_\epsilon+2\sum_{i=1}^{\infty}\gamma^{(\epsilon)}_i}{\sigma^2_\epsilon+2\sum_{i=1}^{h-1}(1-i/h)\,\gamma^{(\epsilon)}_i}\right]$$

$$\text{and } t^*_{\rho_h} = \frac{\widetilde{\rho}^*_h-1}{\widetilde{\sigma}^*_{\rho_h}} = \frac{T^{3/2}\left(\widetilde{\rho}^*_h-1\right)}{\left\{T^3\widetilde{\sigma}^{*2}_{\rho_h}\right\}^{1/2}} \xrightarrow{L} \mathsf{N}\left(0, h\frac{\sigma^2_\epsilon+2\sum_{i=1}^{\infty}\gamma^{(\epsilon)}_i}{\sigma^2_\epsilon+2\sum_{i=1}^{h-1}(1-i/h)\,\gamma^{(\epsilon)}_i}\right)$$

This results in both $t^*_\tau/t^*_{\tau_h}$ and $t^*_\rho/t^*_{\rho_h}$ converging in probability to

$$\frac{1}{\sqrt{h}}\sqrt{1+2\sum_{i=1}^{h-1}(1-i/h)\,\gamma^{(\varepsilon)}_i/\sigma^2_\varepsilon}.$$

Proof. Notice that, again, $T\widetilde{\sigma}^*_{\tau_h} = s^2_{h,T}\begin{bmatrix}T^{1/2} & 0\end{bmatrix}\Sigma^{-1}_{h,T}\begin{bmatrix}T^{1/2}\\0\end{bmatrix}$, i.e.

$$T\widetilde{\sigma}^*_{\tau_h} \to 4\sigma^2_{w_h},$$

and $T^3\widetilde{\sigma}_{\rho_h} = s^2_{h,T}\begin{bmatrix}0 & T^{3/2}\end{bmatrix}\Sigma^{-1}_{h,T}\begin{bmatrix}0\\T^{3/2}\end{bmatrix}$, so that:

$$T^3\widetilde{\sigma}^{*2}_{\rho_h} \to \frac{12}{(\tau^*)^2}\sigma^2_{w_h},$$

and, since $T^{3/2}\left(\widetilde{\rho}^*_h-1\right) \xrightarrow{L} \mathsf{N}\left(0, \frac{12}{(\tau^*)^2}h^2\sigma^2\right)$, hence the limiting distributions of $t^*_{\tau_h}$ and $t^*_{\rho_h}$. Now

$$\left(\widetilde{\sigma}^*_\tau/\widetilde{\sigma}^*_{\tau_h}\right)^2 \xrightarrow{p} \sigma^2_\varepsilon/\sigma^2_{w_h} = \frac{1}{h}$$

and noticing that $\sum_{t=h}^{T} w_{h,t} = \sum_{i=0}^{h-1}\left(\sum_{t=1}^{T}\varepsilon_t - \sum_{t=T-i+1}^{T}\varepsilon_t - \sum_{t=1}^{h-i-1}\varepsilon_t\right)$,then

$$\frac{\sum_{t=1}^{T}\varepsilon_t}{\sum_{t=h}^{T}w_{h,t}} \xrightarrow{p} \frac{1}{h}$$

and hence the results. ■

The behaviors of the statistics are consistent with the general results of local asymptotics, whereby, in our context, under the null of a 'weak' trend, the statistics are centered on some parameter whose limit is finite as the sample size increases, whereas in the case of a strong trend, this parameter itself tends to infinity. The main issue when dealing with local asymptotics is that the weak trend parameter cannot be consistently estimated since it itself tends to zero and, thus, the modeler cannot test which of strong or weak drift frameworks is the more appropriate. However, if the drift is weak but the 'strong' framework is wrongly used for testing, under the null of zero drift—which is the assumption that a modeler would make here, since she would otherwise

include a deterministic trend in the estimated model—then *the conventional distribution of the test statistics leads to under-rejection of the null of no intercept, when the true value is non-zero*. In fact, the modeler would use a joint F-test of zero drift and unit-root, but our analysis of the t-statistics will translate to the same for the F-statistics.

The main issue is therefore for the modeler to determine which of the weak or strong trend framework is more appropriate. This is where multi-step estimation can help. Indeed, the DMS test statistics exhibit different patterns under the two assumptions: under the weak trend hypothesis,

$$t_\tau/t_{\tau_h} \quad \Rightarrow \quad h^{-1/2}\left\{1+2\left(h\sigma_\epsilon^2\right)^{-1}\sum_{i=1}^{h-1}(h-i)\,\gamma_i^{(\epsilon)}\right\}^{1/2} \tag{29}$$

$$\times\left(1-\frac{\left(h\sigma^2\right)^{-1}\left[\sum_{i=1}^{h-1}(h-i)\,\gamma_i^{(\epsilon)}\right]\int_0^1 K_{\psi,\phi}(r)\,dr}{\psi+\sigma^2\left\{\mathcal{D}\left(K_{\psi,\phi}\right)\right\}^{-1}\left[\sigma^{-1}\mathcal{Q}\left(K_{\psi,\phi}\right)-\varpi_h\int_0^1 K_{\psi,\phi}(r)\,dr\right]}\right), \tag{30}$$

and

$$\begin{array}{l} t_\rho/t_{\rho_h}\Rightarrow \\ h^{-1/2}\left\{1+2\left(h\sigma_\epsilon^2\right)^{-1}\sum_{i=1}^{h-1}(h-i)\,\gamma_i^{(\epsilon)}\right\}^{1/2} \\ \times\left(1+\dfrac{\left(h\sigma^2\right)^{-1}\sum_{i=1}^{h-1}(h-i)\,\gamma_i^{(\epsilon)}}{\phi+\sigma^2\left\{\mathcal{D}\left(K_{\psi,\phi}\right)\right\}^{-1}\left[\sigma^{-1}\mathcal{L}\left(K_{\psi,\phi}\right)+\varpi_h\right]}\right). \end{array} \tag{31}$$

Both ratios are of order $O_p\left(h^{-1/2}\right)$, since the multi-step method implies that the errors follow a MA($h-1$). We notice that in both (29) and (31), there appears a coefficient which is the same as in the strong drift case, but that t_τ/t_{τ_h} is shifted downwards and t_ρ/t_{ρ_h} upwards. Thus; in the presence of error autocorelation:

$$\frac{t_\rho/t_{\rho_h}}{t_\tau/t_{\tau_h}}\Rightarrow\frac{\left(1+\dfrac{\left(h\sigma^2\right)^{-1}\sum_{i=1}^{h-1}(h-i)\,\gamma_i^{(\epsilon)}}{\phi+\sigma^2\left\{\mathcal{D}\left(K_{\psi,\phi}\right)\right\}^{-1}\left[\sigma^{-1}\mathcal{L}\left(K_{\psi,\phi}\right)+\varpi_h\right]}\right)}{\left(1-\dfrac{\left(h\sigma^2\right)^{-1}\left[\sum_{i=1}^{h-1}(h-i)\,\gamma_i^{(\epsilon)}\right]\int_0^1 K_{\psi,\phi}(r)\,dr}{\psi+\sigma^2\left\{\mathcal{D}\left(K_{\psi,\phi}\right)\right\}^{-1}\left[\sigma^{-1}\mathcal{Q}\left(K_{\psi,\phi}\right)-\varpi_h\int_0^1 K_{\psi,\phi}(r)\,dr\right]}\right)},$$

and in this case, we see that we would be able to contruct a test for the presence of a weak trend vs a strong.

6.2 Limit distributions as $\psi\to\pm\infty$ and $\phi\to\pm\theta$

It seems interesting to study the limiting behavior of the asymptotic theory in §5, as the non-centrality parameter of the weak trend approaches the boundaries of its domain of definition. In this case, we can expect that the deterministic trend will dominate in estimation and that the resulting distributions should be close to those of the case of a strong drift.

The central results are contained in the following lemma:

Lemma 3 *As $\psi\to\pm\infty$, the behavior of the estimation biases resemble those of a strong drift case*

and

$$\begin{bmatrix} \sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right) \\ T^2\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right) \end{bmatrix} \Rightarrow \begin{bmatrix} \mathsf{N}\left(0, h^2\sigma^2 \dfrac{\phi\left[e^{2\phi}-4e^{\phi}+2\phi+3\right]\left[[\phi-2]\,e^{2\phi}+4e^{\phi}-[\phi+2]\right]}{\left[2\left(e^{\phi}+1\right)\left[\frac{\phi}{2}\left(e^{\phi}-1\right)-\left(e^{\phi}+1\right)\right]\right]^2}\right) \\ \dfrac{h^2}{\phi^{-4}\left(e^{\phi}+1\right)\left[\phi\left(e^{\phi}-1\right)-2\left(e^{\phi}+1\right)\right]} \end{bmatrix},$$

with the special case of an integrated process:

$$\frac{1}{4h\sigma}\left[\sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right)\right] \underset{\psi\to\pm\infty}{\overset{L}{\to}} \mathsf{N}\left(0,1\right) \quad \text{if } \phi=0,$$

which is exactly that of a strong deterministic trend. If we then let $\phi\to\pm\infty$:

$$\begin{aligned} \tfrac{1}{h\sigma}\left[\sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right)\right] &\underset{\psi\to\pm\infty,\ then\ \phi\to+\infty}{\overset{L}{\to}} \mathsf{N}\left(0,1\right), \\ \tfrac{1}{h\sigma\sqrt{-2\phi}}\left[\sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right)\right] &\underset{\psi\to\pm\infty,\ then\ \phi\to-\infty}{\overset{L}{\to}} \mathsf{N}\left(0,1\right). \end{aligned}$$

and

$$\begin{aligned} T^2\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right) &\underset{\psi\to\pm\infty,\ then\ \phi\to+\infty}{\overset{L}{\to}} 0, \\ \phi^{-3}T^2\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right) &\underset{\psi\to\pm\infty,\ then\ \phi\to-\infty}{\overset{L}{\to}} h^2. \end{aligned}$$

(Proof in appendix).

The results above are obtained by studying asymptotic behavior in successive limits: first $T\to\infty$, then the parameters ψ and eventually ϕ, if applicable. Notice that in this case, the slope estimator achieves very strong consistency as it converges in $O_p\left(T^{-2}\right)$ whereas the bias is of order $O_p\left(T^{-3/2}\right)$ in the presence of a 'strong' drift. The asymptotic behavior is different from that described in Phillips (1987b) where he let $\phi\to\pm\infty$ when $\psi=0$, and in which he shows some similarities with stationary or explosive processes, but where definite comparisons cannot be made easily.

7 Forecasting a weakly trending process

The aim of this section is to derive the distribution of the forecast errors when the data generating process is given as in (1) and lemma 1 and the AR(1) model with an intercept is used as above for estimation either by one-step or by multi-step OLS. The parameter estimates are then used to forecast the series h steps ahead from an end-of-sample forecast origin y_T. We use the following notation: $c_T=h/T$, $Y_T\left(r\right)=T^{-1/2}y_{[Tr]}$ for $r\in\mathbb{R}_+$, and $\lambda=e^{\phi}$. Define $\widehat{e}^{*}_{c,T}=h_T^{-1/2}\widehat{e}_{h_T|T}$, with the forecast error $\widehat{e}_{h|T}=y_{T+h}-\widehat{y}_{T+h|T}$.

There are two ways to express the forecast error, either for fixed horizon as in the forecast error taxonomy:

$$\begin{aligned} \widehat{e}_{h|T}= \quad & -\left(\widehat{\rho}_T^{\{h\}}-\rho_T^{\{h\}}\right)\tau_T-\left(\widehat{\rho}_T^{h}-\rho_T^{h}\right)y_T && (i)\ \textit{slope estimation} \\ & -\rho_T^{\{h\}}\left(\widehat{\tau}_T-\tau_T\right) && (ii)\ \textit{intercept estimation} \\ & +\textstyle\sum_{j=0}^{h-1}\rho_T^{j}\epsilon_{T+h-j} && (iii)\ \textit{error accumulation} \\ & -\left(\widehat{\rho}_T^{\{h\}}-\rho_T^{\{h\}}\right)\left(\widehat{\tau}_T-\tau_T\right) && (iv)\ \textit{second-order error} \end{aligned}$$

or letting h increase with the sample size, and let c be constant so that $h_T = [cT]$ as in Kemp (1999). In the context of a weak trend, this latter forecasting technique seems the more interesting since our aim is to derive the asymptotic distributions and to verify whether these can be used as good approximations for small samples. We therefore define:

$$\gamma_T = T\left(\widehat{\rho}_T - 1\right) \Rightarrow \gamma_0, \quad \text{and} \quad \pi_T = T^{1/2}\left(\widehat{\tau}_T - \tau_T\right) \Rightarrow \pi_0,$$

and similarly:

$$\pi_{h,T} = h^{-1/2} c_T^{-1/2}\left(\widetilde{\tau}_{h,T} - \tau_{h,T}\right) \Rightarrow \pi_c, \quad \text{and} \quad \gamma_{h,T} = c_T^{-1}\left(\widetilde{\rho}_{h,T} - \rho_{h,T}\right) \Rightarrow \gamma_c.$$

If we extend the definition of $K_{\psi,\phi}(r)$ to cover $r \in [0, 1+\delta]$ for some $\delta \in (0,1)$—so that the forecast horizon h is always less than the sample size T, we can establish the following results regarding the asymptotic distribution of the forecast errors.

Theorem 2 *With the notations and assumptions above, and $c_T \to c > 0$ as $T \to \infty$:*

$$\begin{aligned}\widehat{e}^*_{c,T} \quad \Rightarrow \quad & c^{-1/2}\left[K_{\psi,\phi}(1+c) - \lambda^c K_{\psi,\phi}(1) - f_\phi(c)\,\psi\right] \\ & -c^{-1/2}\left[\left\{\frac{e^{c\gamma_0}-1}{\gamma_0} - f_\phi(c)\right\}\psi + \left(e^{c\gamma_0}-1\right)\frac{\pi_0}{\gamma_0} + \left\{e^{c\gamma_0} - \lambda^c\right\} K_{\psi,\phi}(1)\right],\end{aligned}$$

where $c^{-1/2} f_\phi(c)\,\psi$ is non-stochastic and the vectors $(\pi_0, \gamma_0, K_{\psi,\phi}(1))$ and $[K_{\psi,\phi}(1+c) - \lambda^c K_{\psi,\phi}(1)]$ are independent if $\forall i \geq 1, \quad \gamma_i^{(\epsilon)} = 0$, which implies that $\sigma = \sigma_\epsilon$.

The direct multi-step forecast errors are given by:

$$\begin{aligned}\widetilde{e}^*_{c,T} \Rightarrow c^{-1/2}\left[K_{\psi,\phi}(1+c) - \lambda^c K_{\psi,\phi}(1) - f_\phi(c)\,\psi\right] \\ -c^{1/2}\left[\pi_c + \gamma_c K_{\psi,\phi}(1)\right],\end{aligned}$$

with the notation as above. The vectors $(\pi_c, \gamma_c, K_{\psi,\phi}(1))$ and $[K_{\psi,\phi}(1+c) - \lambda^c K_{\psi,\phi}(1)]$ are independent if the ϵ_t are innovations with respect to $\mathcal{F}_{t-1}$. the σ-field generated by $\{y_i, \epsilon_i\}_{i<t}$.

The forecast errors are asymptotically biased in both cases, and the biases are given by the second line on the right-hand side of both expressions. (Proof in appendix).

By comparison, if we had not let h tend to infinity, the distributions of the forecast errors would have been:

$$\begin{aligned}\widehat{e}_{h|T} \quad = \quad & h^{1/2}\left\{h^{-1/2}\sum_{j=0}^{h-1}\rho_T^j \epsilon_{T+h-j}\right\} \\ & -T^{-1/2}\left[\left\{\rho_T^{\{h\}} T^{1/2}\left(\widehat{\tau}_T - \tau_T\right)\right\} + h\left\{h^{-1}T\left(\widehat{\rho}_T^h - \rho_T^h\right) Y_T(1)\right\}\right] \\ & -hT^{-3/2}\left[\left\{h^{-1}T\left(\widehat{\rho}_T^{\{h\}} - \rho_T^{\{h\}}\right)\right\}\psi + \left\{h^{-1}T\left(\widehat{\rho}_T^{\{h\}} - \rho_T^{\{h\}}\right)\right\}\left\{T^{1/2}\left(\widehat{\tau}_T - \tau_T\right)\right\}\right],\end{aligned}$$

and it would be possible to approximate each component individually for small samples. Similarly, the multi-step forecast error can be decomposed into its components as in:

$$\begin{aligned}\widetilde{e}_{h|T} \quad = \quad & h^{1/2}\left\{h^{-1/2}\sum_{j=0}^{h-1}\rho_T^j \epsilon_{T+h-j}\right\} \\ & -hT^{-1/2}\left\{h^{-1}T^{1/2}\left(\widetilde{\tau}_{h,T} - \tau_{h,T}\right) + h^{-1}T\left(\widetilde{\rho}_{h,T} - \rho_T^h\right) Y_T(1)\right\}.\end{aligned}$$

But, in both cases, the consistency of the estimators imply that the forecast errors are asymptotically normally distributed with mean zero and that they are the same for both methods.

Corollary 3 *It is possible to separate the various impacts of the components defined as above*

$$\begin{aligned}
\widehat{e}^*_{c,T} \Rightarrow \quad & -c^{-1/2}\left[\tfrac{e^{c\gamma_0}-1}{\gamma_0} - f_\phi(c)\right]\psi - c^{-1/2}\lambda^c\left\{e^{c(\gamma_0-\phi)}-1\right\}K_{\psi,\phi}(1) && (i)\ \text{slope estimation}\\
& -c^{-1/2} f_\phi(c)\,\pi_0 && (ii)\ \text{intercept estimation}\\
& +c^{-1/2}\left\{K_{0,\phi}(1+c) - \lambda^c K_{0,\phi}(1)\right\} && (iii)\ \text{error accumulation}\\
& -c^{-1/2}\left[\tfrac{e^{c\gamma_0}-1}{\gamma_0} - f_\phi(c)\right]\pi_0 && (iv)\ \text{second-order error}
\end{aligned}$$

and

$$\begin{aligned}
\widehat{e}^*_{c,T} \Rightarrow \quad & -c^{1/2}\gamma_c K_{\psi,\phi}(1) && (i)\ \text{slope estimation}\\
& -c^{1/2}\pi_c && (ii)\ \text{intercept estimation}\\
& +c^{-1/2}\left\{K_{0,\phi}(1+c) - \lambda^c K_{0,\phi}(1)\right\} && (iii)\ \text{error accumulation}
\end{aligned}$$

which show how different the behaviors of the slope estimation components are with respect to c, the forecast horizon, especially the slope and intercept estimations. As c increases, the DMS estimation components grow slowly (in $c^{1/2}$), whereas it is more difficult to tell what happens to IMS: it depends on the signs of γ_0 and ϕ; the former is likely to be nonpositive if the latter so is, and in this case the influence of the components is decreasing in c (the converse is also true for positive ϕ). (Proof in appendix).

Notice that the various powers of the coefficient c in the decompositions in corollary **??** provide the asymptotic rates of convergence. Yet, because of the nonlinearities, it is quite difficult to determine it precisely. Notice, though, that all the IMS components exhibit the coefficient $c^{-1/2}$, whereas the DMS slope and intercept estimation impacts are products of $c^{1/2}$. This implies that the forecast error is a complex function of the horizon. Using a Taylor expansion of the functions of c, the IMS forecast error becomes:

$$\begin{aligned}
\widehat{e}^*_{c,T} \Rightarrow \quad & -c^{1/2}(\gamma_0-\phi)K_{\psi,\phi}(1) && (i)\\
& -c^{1/2}\pi_0 && (ii)\\
& +c^{-1/2}\left\{K_{0,\phi}(1+c) - \lambda^c K_{0,\phi}(1)\right\} && (iii)\\
& -\tfrac{1}{2}c^{3/2}(\gamma_0-\phi)\left\{1-(\gamma_0-3\phi)K_{\psi,\phi}(1)\right\} && (i)\\
& -\tfrac{1}{2}c^{3/2}\phi\gamma_0 && (ii)+(iv)\\
& +O_p\left(c^{5/2}\right). &&
\end{aligned}$$

We notice here that for short horizons (i.e. c close to zero), both forecasting techniques behave in a similar way (terms in order $O_p\left(c^{1/2}\right)$). Additional components enter the iterated forecast error as the horizon increases. This is why we turn to a Monte Carlo simulation in the next section to clarify matters.

Remark 4 *Notice that Theorem 2 and Corollary 3 together imply that*

$$K_{\psi,\phi}(1+c) - \lambda^c K_{\psi,\phi}(1) \equiv K_{0,\phi}(1+c) - \lambda^c K_{0,\phi}(1) + f_\phi(c)\,\psi.$$

This can otherwise be derived directly.

8 Monte Carlo

In order to observe the validity of the weak trend approach, we now present the results from a Monte Carlo simulation which compares the results derived analytically to those obtained from estimation and forecasting over small samples of observations, here $T = 25$ and the forecast horizon varies between $h = 1$ and $h = 4$. Simulation proves the only way to compare the actual distributions of the statistics owing to their rather intricate expressions, involving many OU-d, or K, processes Computations were carried through using OxEdit and the Ox programming language. Given the non-stationary feature of the simulated data, the origin of the sample cannot be randomly drawn from a common distribution, we therefore resort to setting it to zero for all replications.

Owing to the large number of parameters, it seems difficult to present a thorough assessment of the distributional equivalence between small sample estimation, inference and forecasting properties and their weak trend asymptotic approximation, and we refer the reader to Chevillon (2004) for a more complete comparison. Here, we resort to 3D graphs representing the ratios of the small sample Monte Carlo estimates (appropriately scaled biases, t-test statistics and forecast errors over 10,000 replications) over their asymptotic weak trend counterparts (2,500 replications for $K(1)$ processes and integrals); the horizontal axes refer to the drift parameter ψ, varying between 0 and 2, and to the distribution quartile. For each figure, we present six graphs for which the moving average parameter θ takes values 0, -0.3 and -0.6 and the horizon $h = 1$ or 4. Altogether, these figure produce a concise view of the approximation properties.

8.1 Estimation

The weak trend framework provides a good approximation to the small sample intercept bias as figure 4 quadrant *a* shows: quartiles—between $x = 0.1$ and 0.9—are the same for the two when the model is not misspecified for the error autocorrelation. The non-linearity that appears in this graph and becomes more significant in the others is essentially caused by quartile values too close to zero and which alter the ratios (they are removed when too large). From this we observe that if the distribution is centered on zero in the absence of a drift, a positive value of the latter implies that the bias distribution is shifted towards positive values. When estimating the multi-step parameter, using DMS, the weak trend framework is as accurate (fig. 4 *b*), except for the lower tail and ψ close to zero, in which case the 'weak' distribution is thinner tailed. If we progress downwards in the figure and observe the graphs in the presence of a negative moving average component, then the approximation is less accurate, but we notice that it is still reasonable for $h = 4$ as the ratio is then still close to unity.

Results for the slope bias are similar as we notice again on figure 5 that estimate ratios in a well-specified model are close to unity. The unit root estimator is negatively biased, ratios for the upper quartiles were truncated as denominators were almost zero. When $\theta < 0$, the approximation is again more valid at horizon 4 than for $h = 1$. The negative values of the ratios indicate that the weak framework does not represent the strongly skewed distribution correctly as too much of the interquartile range leads to a positive slope bias. By contrast lower tail approximations are very accurate.

8.2 Inference

How do these fairly accurate bias representations translate into test statistics? We report in figures 6 and 7 the same ratios as before but now for the t test statistics under the null of no intercept and under the true null. With respect to the previous figures, we have now included $x = 0.01$, 0.05, 0.95 and 0.99. This results in kinks in the surfaces for the uppermost quartiles, with the same spike as before for biases close to zero. Ratios are now closer to 2 on average implying that the asymptotic weak distribution does not exhibit fat enough tails. The main difference between the two figures lies in the accuracy of the one-step and DMS estimators: the approximation is better for the null of no intercept when $h = 1$ and for the true null when $h = 4$.

8.3 Forecasting

Forecast errors are the most complex of the Monte Carlo estimates as they combine results from the two previous subsections. As expected, their approximation turns out the least accurate. Figure 8 presents the IMS forecast error ratios. These are only close to unity for low ψ and in the lower tail. A magnified version of the $h = 1$ case is shown in the left-hand side column of fig. 9 and we see that the most regular approximation actually corresponds to the presence of a negative moving average of low absolute value (second row). Contrary to the IMS case, increasing the horizon does not worsen the DMS ratios.

The reason for the approximation to be less valid for forecasting than in the previous subsections could lie in the vary large number of replications needed for an appropriate Monte Carlo estimation. Indeed the asymptotic biases enter in the forecast error, yielding a higher degree of sample variability.

9 Conclusions

In this paper, we have presented a method of approximation which aims at representing the highly non-linear patterns of estimation biases and forecast errors which had been found in Chevillon and Hendry (2004). For this purpose, we have introduced the concept of weak trend, allowing the asymptotic behaviors of sample statistics to mimic the small sample interaction between a stochastic and a deterministic trend, whereas in the traditional framework, the latter dominates. Here by contrast, the effect of the trend falls in a continuum and can be tuned to reflect the finite

sample behavior of the estimators, by letting the parameter k in $\tau = \psi/T^k$ vary. We have focused in the specific case where $k = 1/2$ so that the series is of order $O_p\left(\sqrt{T}\right)$ as a pure random walk.

We have shown that in this framework, most general random walk estimation results apply when standard Brownian motions are replaced with drifting Ornstein-Uhlenbeck processes. This allowed us to characterize the non-linear patterns exhibited by both estimators and forecasts. Unfortunately, as in most cases of local asymptotics, it proves difficult to test for which framework—strong or weak drift—is most relevant in an empirical example, but the use of multi-step t statistics allows for a specification analysis and could, in the presence of serial correlation of the error process, lead to testing which of the strong or weak trend framework is more appropriate. Yet, one of the most interesting aspect of our framework lies in the interaction between the stochastic and the deterministic trends and when it shows how differently the direct and iterated multi-step methods differ as far as forecasting is concerned. We noticed that the components of the forecast errors as derived in the taxonomy present opposite patterns with respect to the forecast horizon but that it is not the case when the latter is small compared to the sample size. A Monte Carlo simulation showed that the weak trend framework appropriately represents the behaviors of estimation biases and specification test statistics in small samples. Yet the approximation is less valid for forecast errors, but this may be due to a Monte Carlo variability and ought to be checked with a more powerful processor.

The weak trend framework can be extended for $\tau = \sum_{k=\underline{k}}^{\overline{k}} \frac{\psi_k}{T^k}$ which we could combine with the results from Phillips (1998) who show that a stochastically trending process can be represented as a expansion of deterministic functions of time with random coefficients. This would provide an expansion of the regression estimators.

References

Banerjee, A., D. F. Hendry, and G. E. Mizon (1996). The econometric analysis of economic policy. *Oxford Bulletin of Economics and Statistics* **58**, 573–600.

Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis, Forecasting an Control* (2nd ed.). San Francisco, CA: Holden–Day. First published, 1970.

Chevillon, G. (2004). Multi-step estimation for forecasting economic processes. D.Phil. thesis, Economics department, University of Oxford.

Chevillon, G. and D. F. Hendry (2004). Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting* **forthcoming**.

Clements, M. P. and D. F. Hendry (2001). Forecasting with difference-stationary and trend-stationary models. *Econometrics Journal* **4**, S1–S19.

Diebold, F. X. and A. S. Senhadji (1996). Deterministic vs. stochastic trend in U.S. GNP, yet again. *NBER Working Paper Series* **5481**.

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

Kemp, G. C. R. (1999). The behavior of forecast errors from a nearly integrated AR(1) model as both sample size and forecast horizon become large. *Econometric Theory* **15**, 238–256.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica* **55**(2), 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika* **74**(3), 535–547.

Phillips, P. C. B. (1988). Regression theory for near-integrated time series. *Econometrica* **56**(5), 1021–1043.

Phillips, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica* **66**, 1299–1326.

Phillips, P. C. B. (2004). Challenges of trending time series econometrics. Discussion paper no. 1472, Cowles Foundation.

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.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica* **65**, 557–586.

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

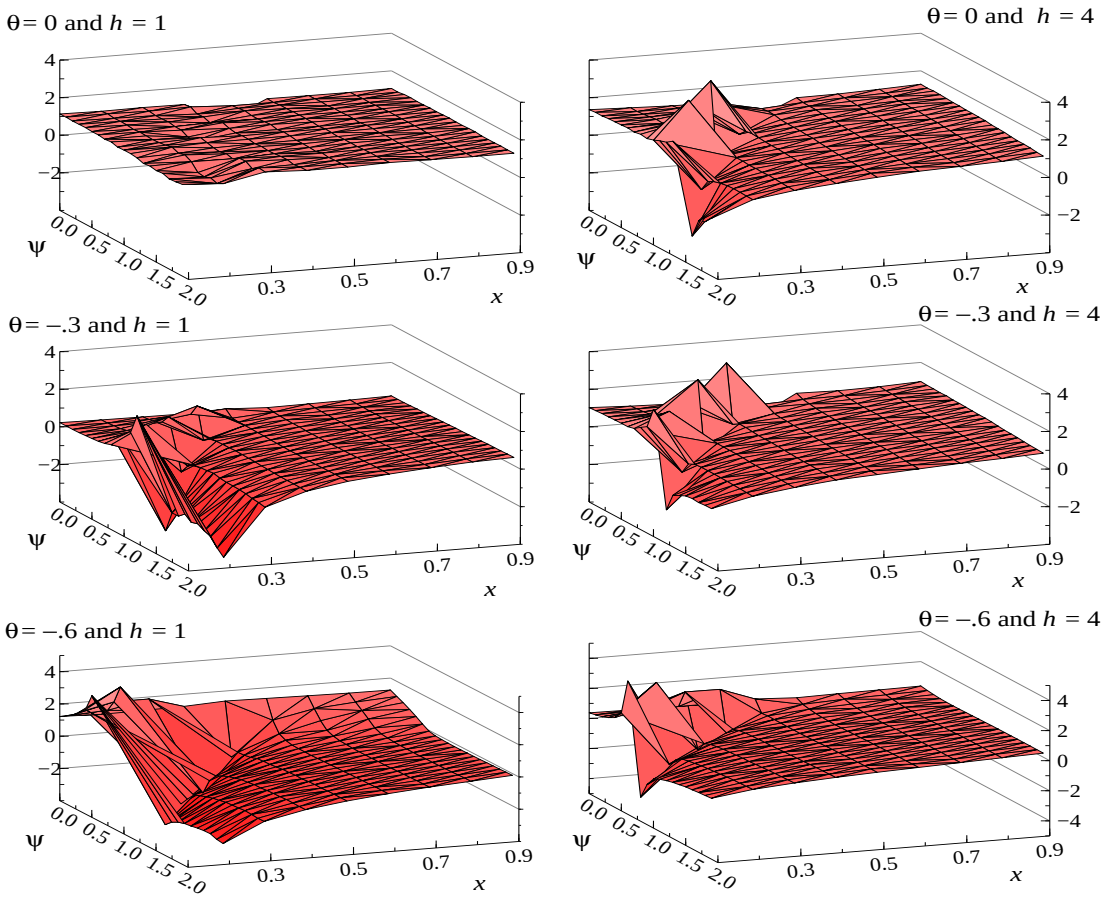

Figure 4: Ratios of the Monte Carlo values of the distribution of the small sample intercept bias over its asymptotically corresponding 'weak trend' counterpart, for a sample of $T = 25$ observations, $10,000$ replications and varying ψ parameter and quantiles x.

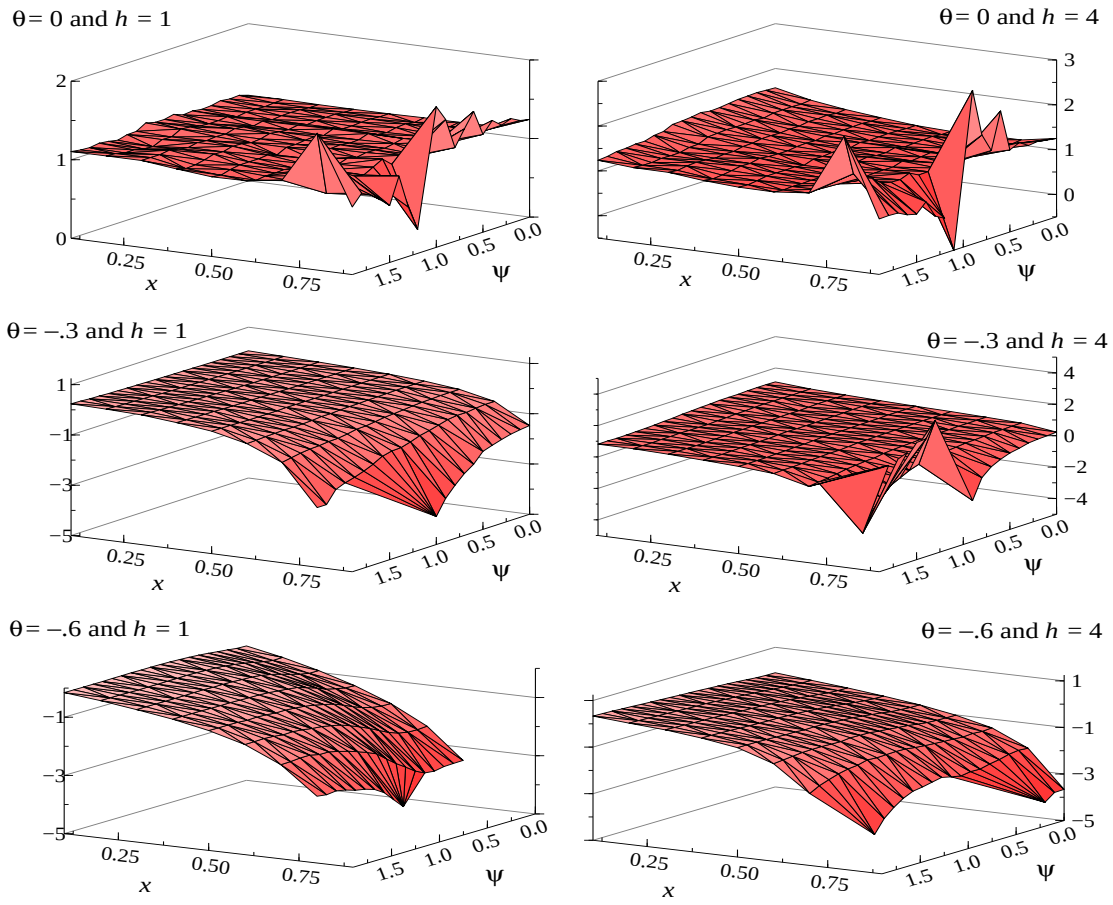

Figure 5: Ratios of the Monte Carlo values of the distribution of the small sample slope bias over its asymptotically corresponding 'weak trend' counterpart, for a sample of $T = 25$ observations, $10,000$ replications and varying ψ paramater and quantiles.

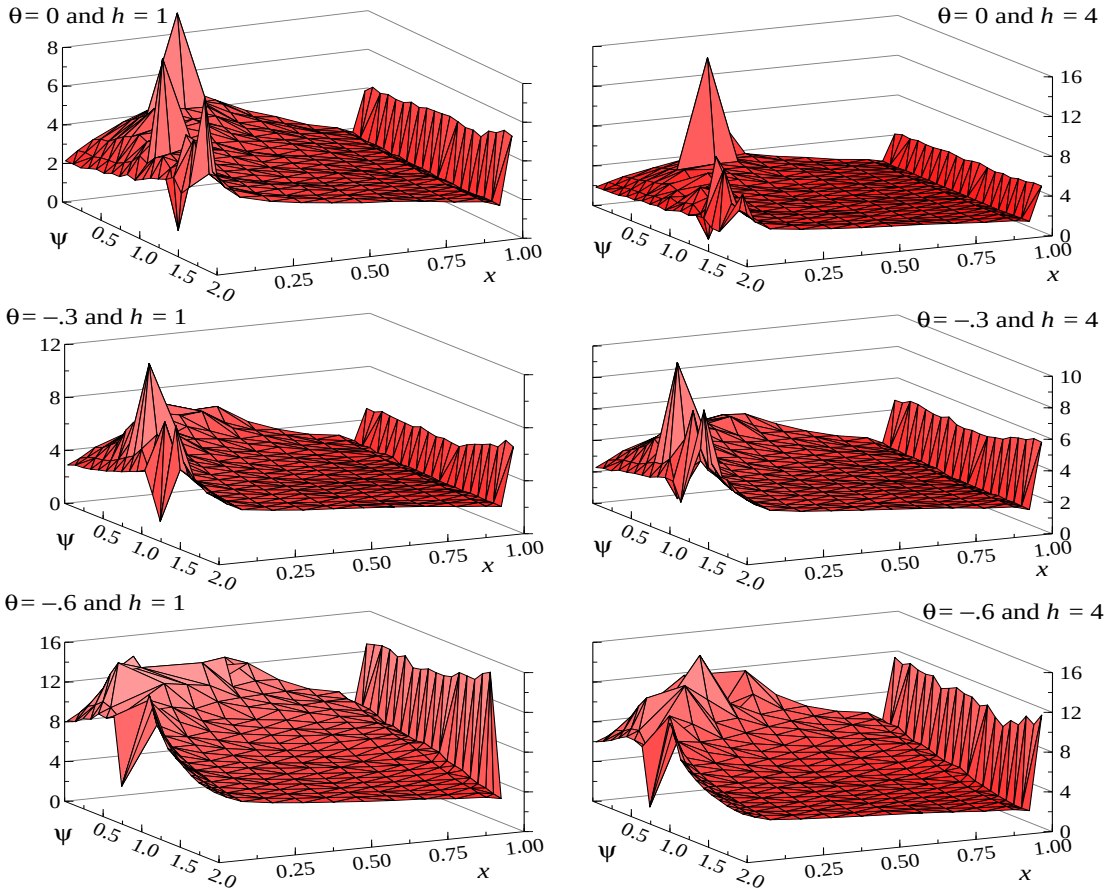

Figure 6: Ratios of the Monte Carlo values of the distribution of the small sample intercept t test statistic (under the Null of no intercept) over its asymptotically corresponding 'weak trend' counterpart, for a sample of $T = 25$ observations, $10,000$ replications and varying ψ parameter and quantiles x.

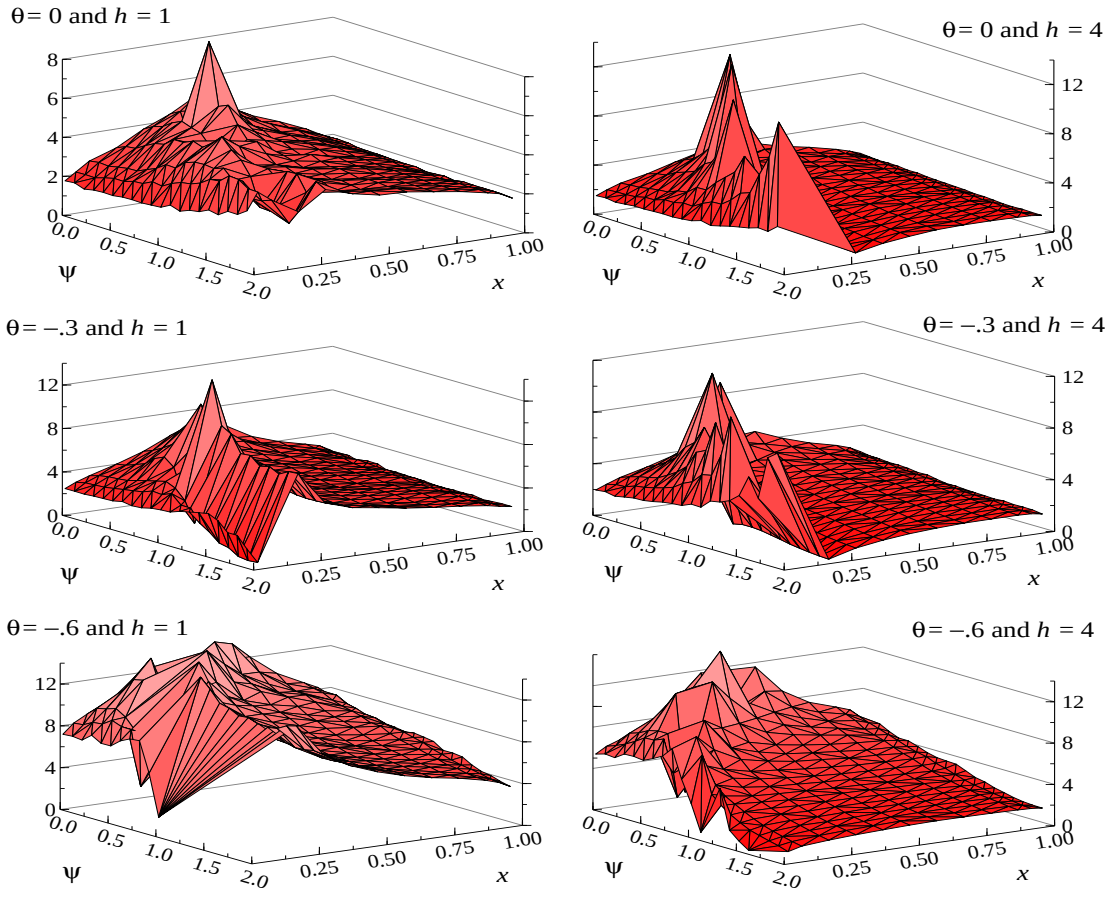

Figure 7: Ratios of the Monte Carlo values of the distribution of the small sample intercept t test statistic (under the true Null) over its asymptotically corresponding 'weak trend' counterpart, for a sample of $T = 25$ observations, $10,000$ replications and varying ψ parameter and quantiles x.

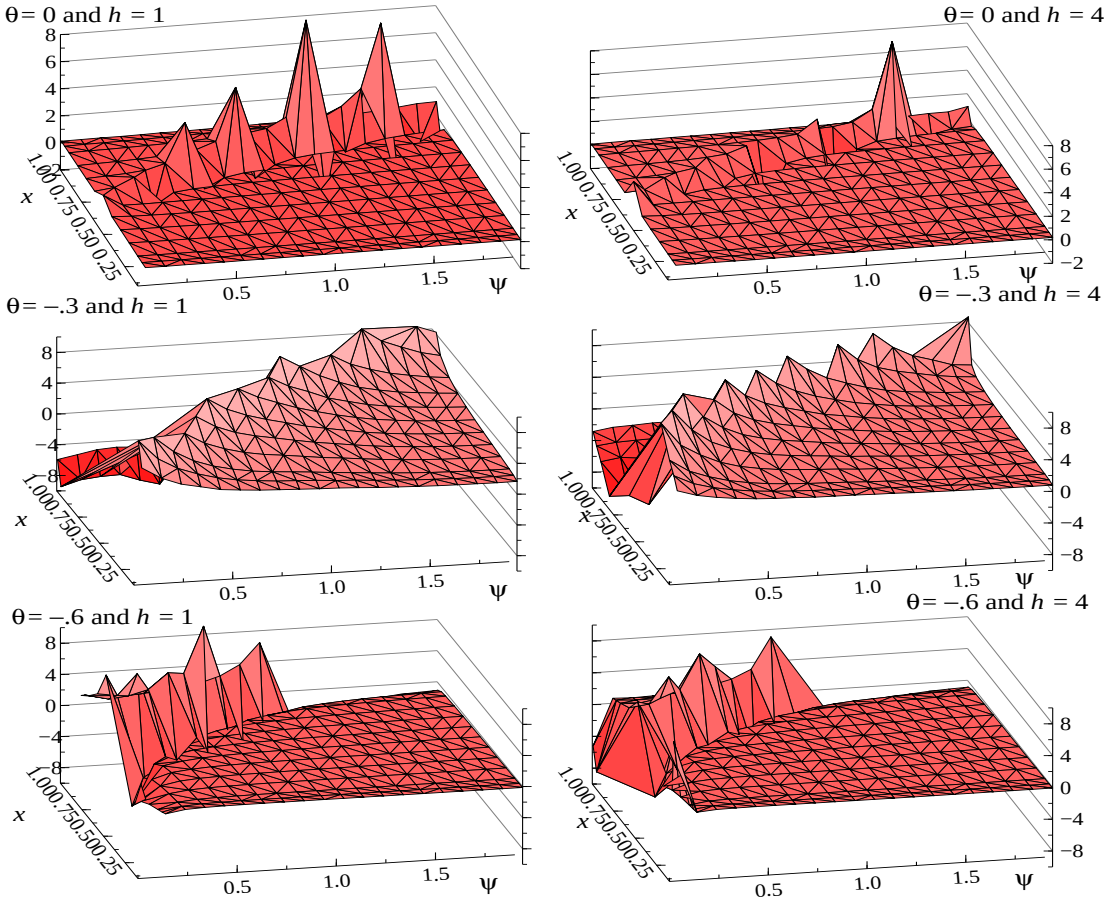

Figure 8: Ratios of the Monte Carlo values of the distribution of the small sample IMS forecast error over its asymptotically corresponding 'weak trend' counterpart, for a sample of $T = 25$ observations, $10,000$ replications and varying ψ parameter and quantiles x.

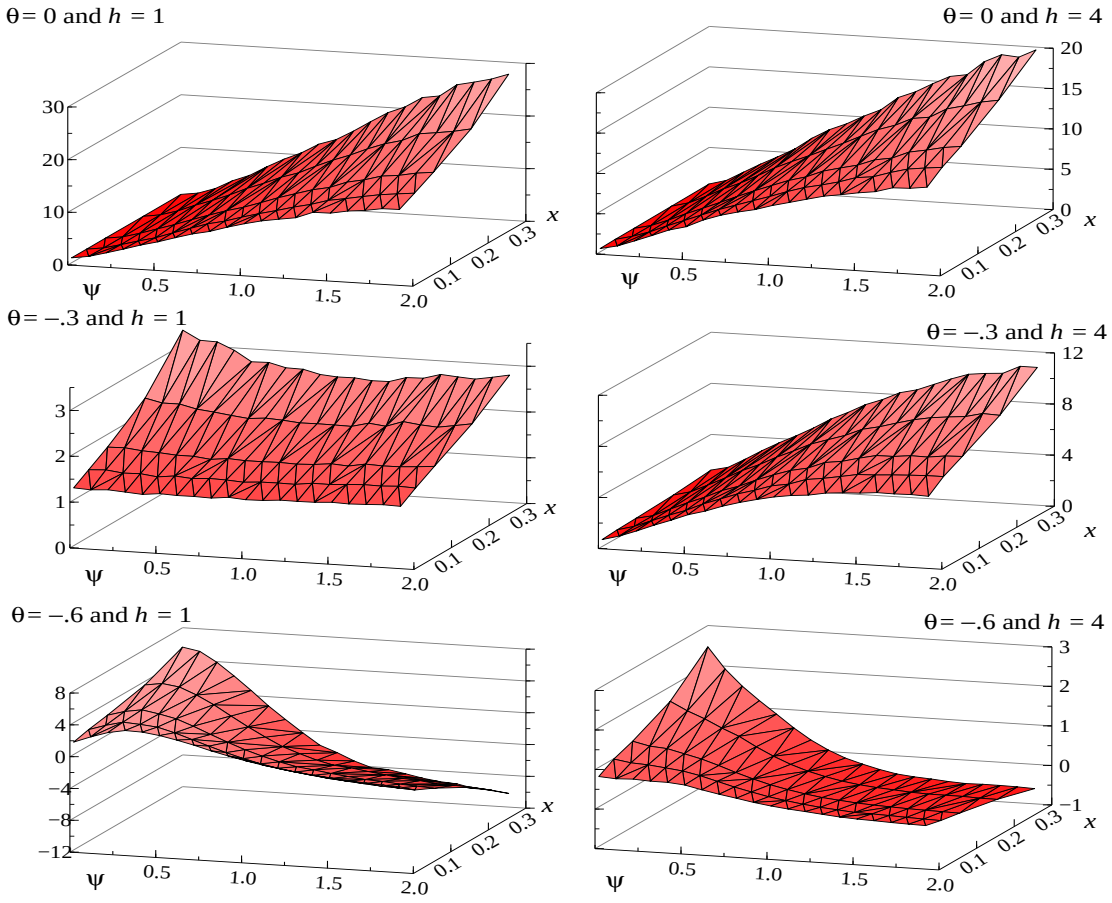

Figure 9: Ratios of the Monte Carlo values of the distribution of the small sample DMS forecast error over its asymptotically corresponding 'weak trend' counterpart, for a sample of $T = 25$ observations, $10,000$ replications and varying ψ parameter and quantiles x.

Appendices

A Proof of Theorem 1

Proofs of (a_h)–(c_h) are standard. We write the statistic (d_h) as a functional of $X_{h,T}$ on $D[0,1]$. We first square $T^{-1/2}y_{t,T}$:

$$\begin{aligned} T^{-1}y_{t,T}^2 &= T^{-1}\left(\tau_{h,T}+\rho_{h,T}y_{t-h,T}+w_{h,t}\right)^2 \\ &= T^{-2}\left(\sum_{i=0}^{h-1}e^{i\phi/T}\right)^2\psi^2+T^{-1}e^{2h\phi/T}y_{t-h,T}^2+T^{-1}w_{h,t}^2 \\ &\quad +2T^{-3/2}\left(\sum_{i=0}^{h-1}e^{(i+h)\phi/T}\right)\psi y_{t-h,T}+2T^{-3/2}\psi w_{h,t}+2T^{-1}e^{h\phi/T}y_{t-h,T}w_{h,t}. \end{aligned}$$

Summing over t, and letting $T\to\infty$ yield:

$$\begin{aligned} T^{-1}\sum y_{t-h,T}w_{h,t} &\Rightarrow \frac{h}{2}\left\{\left\{K_{\psi,\phi}(1)\right\}^2-2\phi\int_0^1\left[K_{\psi,\phi}(r)\right]^2dr-h^{-1}\sigma_{w_h}^2-2\psi\int_0^1K_{\psi,\phi}(r)\,dr\right\} \\ &= \frac{h}{2}\left\{\left\{K_{\psi,\phi}(1)\right\}^2-2\int_0^1\left[\psi K_{\psi,\phi}+\phi K_{\psi,\phi}^2(r)\right]dr-h^{-1}\sigma_{w_h}^2\right\}, \end{aligned}$$

whence the result, using (14) and (24).

B Proof of Lemma 3

As $\psi\to\pm\infty$, we notice that

$$\begin{bmatrix}\sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right)\\ T\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right)\end{bmatrix}\Rightarrow h\sigma\begin{bmatrix}\dfrac{\mathcal{Q}(f_\phi)}{\mathcal{D}(f_\phi)}\\ 0\end{bmatrix},$$

if $\sigma\neq0$, and $\sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right)\Rightarrow0$ otherwise. Notice that the notation used is that of §5.3, but no assumption is made here about a particular form of disturbances. The distribution of $\{\mathcal{D}(f_\phi)\}^{-1}\mathcal{Q}(f_\phi)$ is given by Slutsky's formula since$\int_0^1 f_\phi(r)\,dr=\left(e^\phi-\phi-1\right)/\phi^2$ for $\phi\neq0$ and $\int_0^1 f_0(r)\,dr=1/2$; and

$$\begin{aligned} \mathcal{D}(f_\phi) &= \phi^{-4}\left(e^\phi+1\right)\left[\frac{\phi}{2}\left(e^\phi-1\right)-\left(e^\phi+1\right)\right], \\ \mathcal{D}(f_0) &= \frac{1}{12}, \\ \mathcal{Q}(f_\phi) &= \phi^{-3}\left[\frac{e^{2\phi}}{2}-2e^\phi+\phi+\frac{3}{2}\right]W(1)-\phi^{-3}\left[e^\phi-\phi-1\right]\left(\int_0^1\left(e^{\phi r}-1\right)dW(r)\right), \\ \mathsf{E}\left[\mathcal{Q}(f_\phi)\right] &= 0, \\ \mathsf{Var}\left[\mathcal{Q}(f_\phi)\right] &= \frac{1}{4}\phi^{-7}\left[e^{2\phi}-4e^\phi+2\phi+3\right]\left[\left[\phi-2\right]e^{2\phi}+4e^\phi-\left[\phi+2\right]\right], \\ \mathsf{Var}\left[\mathcal{Q}(f_0)\right] &= \mathsf{Var}\left[\frac{1}{3}W(1)-\frac{1}{2}\int_0^1 rdW(r)\right]=\frac{1}{9}, \end{aligned}$$

since $\mathsf{E}\left[\int_0^1\int_0^1\left(e^{\phi r}-1\right)dW\left(r\right)dW\left(s\right)\right]=\int_0^1\mathsf{E}\left[\left(e^{\phi r}-1\right)dr\right]$ and hence:

$$\begin{array}{rcl}
\left[\sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right)\right] & \underset{\psi\to\infty}{\overset{L}{\to}} & \mathsf{N}\left(0,h^2\sigma^2\frac{\phi\left[e^{2\phi}-4e^{\phi}+2\phi+3\right]\left[\left[\phi-2\right]e^{2\phi}+4e^{\phi}-\left[\phi+2\right]\right]}{\left[2(e^{\phi}+1)\left[\frac{\phi}{2}(e^{\phi}-1)-(e^{\phi}+1)\right]\right]^2}\right),\\
\text{and }\frac{1}{4h\sigma}\left[\sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right)\right] & \underset{\psi\to\pm\infty}{\overset{L}{\to}} & \mathsf{N}\left(0,1\right),\quad\text{if }\phi=0,\\
\frac{1}{h\sigma}\left[\sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right)\right] & \underset{\psi\to\pm\infty,\text{ then }\phi\to\infty}{\overset{L}{\to}} & \mathsf{N}\left(0,1\right),\\
\frac{1}{h\sigma\sqrt{-2\phi}}\left[\sqrt{T}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right)\right] & \underset{\psi\to\pm\infty,\text{ then }\phi\to\infty}{\overset{L}{\to}} & \mathsf{N}\left(0,1\right)
\end{array}$$

As regards the slope estimator, we first notice that

$$\begin{array}{rl}
& T\left\{T^{-1}\sum y_{t-h,T}w_{h,t}\right.\\
& \left.-\frac{1}{2}e^{-h\phi/T}\left[h\left\{K_{\psi,\phi}\left(1\right)\right\}^2-2h\phi\int_0^1\left[K_{\psi,\phi}\left(r\right)\right]^2dr-\sigma_{w_h}^2-2\psi h\int_0^1K_{\psi,\phi}\left(r\right)dr\right]\right\}\\
\Rightarrow\quad & -h\psi\left(\frac{h\psi}{2}+\sigma W\left(1\right)\right)
\end{array}$$

so that, whereas $T\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right)\underset{\psi\to\infty}{\to}0$,

$$T^2\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right)\underset{\psi\to\infty}{\to}\frac{h^2}{\phi^{-4}\left(e^{\phi}+1\right)\left[\phi\left(e^{\phi}-1\right)-2\left(e^{\phi}+1\right)\right]}$$

C Proof of Theorem 2

Re-write the forecast error as

$$\begin{array}{rcl}
h^{-1/2}\widehat{e}_{h|T} & = & c_T^{-1/2}Y_T\left(1+c_T\right)-h^{-1/2}\widehat{\rho}_T^{\{h\}}\widehat{\tau}_T-c_T^{-1/2}\widehat{\rho}_T^{h}Y_T\left(1\right)\\
& = & c_T^{-1/2}Y_T\left(1+c_T\right)-c_T^{-1/2}\rho_T^{h}Y_T\left(1\right)-h^{-1/2}\rho_T^{\{h\}}\tau_T\\
& & -h^{-1/2}\left(\widehat{\rho}_T^{\{h\}}\widehat{\tau}_T-\rho_T^{\{h\}}\tau_T\right)-c_T^{-1/2}\left(\widehat{\rho}_T^{h}-\rho_T^{h}\right)Y_T\left(1\right).
\end{array}$$

Notice that $\rho_T^{h_T}=e^{h_T\phi/T}=e^{c_T\phi}$, and that $h^{-1/2}\rho_T^{\{h\}}\tau_T=T^{-1}\frac{f_\phi(c_T)}{f_\phi(1/T)}c_T^{-1/2}\psi$, hence

$$\begin{array}{rcl}
\widehat{e}_{c,T}^{*} & = & c_T^{-1/2}\left(Y_T\left(1+c_T\right)-e^{c_T\phi}Y_T\left(1\right)\right)\\
& & -\frac{1/T}{f_\phi\left(1/T\right)}f_\phi\left(c_T\right)c_T^{-1/2}\psi-c_T^{-1/2}T^{-1/2}\left(\widehat{\rho}_T^{\{h\}}\widehat{\tau}_T-\rho_T^{\{h\}}\tau_T\right)\\
& & -c_T^{-1/2}\left(\left(1+T^{-1}\gamma_T\right)^hY_T\left(1\right)-e^{c_T\phi}Y_T\left(1\right)\right),
\end{array}$$

where $\gamma_T=T\left(\widehat{\rho}_T-1\right)\Rightarrow\gamma_0$ as above. And then:

$$\widehat{\rho}_T^{h}=\left(1+T^{-1}\gamma_T\right)^h=\left(1+h^{-1}c_T\gamma_T\right)^h=e^{h\log\left(1+h^{-1}c_T\gamma_T\right)}\Rightarrow e^{c\gamma_0}.$$

Similarly:

$$\begin{array}{rcl}
\left(\widehat{\rho}_T^{\{h\}}\widehat{\tau}_T-\rho_T^{\{h\}}\tau_T\right) & = & \left[\left(\sum_{i=0}^{h-1}\left\{e^{\log\left(1+h^{-1}c_T\gamma_T\right)}\right\}^i\right)\widehat{\tau}_T-\left(\sum_{i=0}^{h-1}e^{i\phi/T}\right)\tau_T\right]\\
& = & \left[\frac{f_{\log(1+h^{-1}c_T\gamma_T)}\left(h\right)}{f_{\log(1+h^{-1}c_T\gamma_T)}\left(1\right)}\left(\tau_T+T^{-1/2}\pi_T\right)-\frac{f_\phi\left(h/T\right)}{f_\phi\left(1/T\right)}\tau_T\right]\\
& = & T^{1/2}\left[c_T\frac{f_{h\log(1+h^{-1}c_T\gamma_T)}\left(1\right)}{f_{\log(1+h^{-1}c_T\gamma_T)}\left(1\right)}-\frac{f_\phi\left(h/T\right)}{f_{\phi/T}\left(1\right)}\right]\psi\\
& & +T^{1/2}c_T\frac{f_{h\log(1+h^{-1}c_T\gamma_T)}\left(1\right)}{f_{\log(1+h^{-1}c_T\gamma_T)}\left(1\right)}\pi_T,
\end{array}$$

and hence

$$T^{-1/2}\left(\widehat{\rho}_T^{\{h\}}\widehat{\tau}_T-\rho_T^{\{h\}}\tau_T\right)\Rightarrow\left[\frac{e^{c\gamma_0}-1}{\gamma_0}-f_\phi\left(c\right)\right]\psi+\left(e^{c\gamma_0}-1\right)\frac{\pi_0}{\gamma_0}.$$

Letting $\lambda=e^{\phi}$,

$$\begin{aligned}\widehat{e}_{c,T}^{*} &= c_T^{-1/2}\left(Y_T\left(1+c_T\right)-\lambda^{c_T}Y_T\left(1\right)\right)\\ &\quad -c_T^{-1/2}\left\{\frac{T^{-1}}{f_\phi\left(T^{-1}\right)}f_\phi\left(c_T\right)+\left[c_T\frac{f_{h\log(1+h^{-1}c_T\gamma_T)}\left(1\right)}{f_{\log(1+h^{-1}c_T\gamma_T)}\left(1\right)}-\frac{f_\phi\left(h/T\right)}{f_{\phi/T}\left(1\right)}\right]\right\}\psi\\ &\quad -c_T^{1/2}\frac{f_{h\log(1+h^{-1}c_T\gamma_T)}\left(1\right)}{f_{\log(1+h^{-1}c_T\gamma_T)}\left(1\right)}\pi_T\\ &\quad -c_T^{-1/2}\left(e^{h\log\left(1+h^{-1}c_T\gamma_T\right)}Y_T\left(1\right)-\lambda^{c_T}Y_T\left(1\right)\right).\end{aligned}$$

By contrast the multi-step forecast error is much simpler:

$$\widetilde{e}_{h|T}=y_{T+h}-\widetilde{y}_{T+h|T}=\left(\tau_{h,T}-\widetilde{\tau}_{h,T}\right)+\left(\rho_T^h-\widetilde{\rho}_{h,T}\right)y_T+\sum_{j=0}^{h-1}\rho_T^j\epsilon_{T+h-j},$$

and if we let h grow with T,

$$\begin{aligned}\widetilde{e}_{c,T}^{*} &= h_T^{-1/2}T^{1/2}\left[Y_T\left(1+c_T\right)-\widetilde{\rho}_{h,T}Y_T\left(1\right)\right]-h_T^{-1/2}\widetilde{\tau}_{h,T}\\ &= c_T^{-1/2}\left[Y_T\left(1+c_T\right)-\rho_{h,T}Y_T\left(1\right)\right]\\ &\quad -\frac{T^{-1}}{f_\phi\left(T^{-1}\right)}f_\phi\left(c_T\right)c_T^{-1/2}\psi-c_T^{1/2}h^{-1}T\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right)Y_T\left(1\right)\\ &\quad -c_T^{1/2}h^{-1}T^{1/2}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right),\end{aligned}$$

and hence the results. Independence follows from uncorrelatedness and Gaussianity.

D Proof of Corollary 3

Recall that

$$\begin{aligned}\widehat{e}_{h|T}= &-\left(\widehat{\rho}_T^{\{h\}}-\rho_T^{\{h\}}\right)\tau_T-\left(\widehat{\rho}_T^{h}-\rho_T^{h}\right)y_T\\ &-\rho_T^{\{h\}}\left(\widehat{\tau}_T-\tau_T\right)\\ &+\textstyle\sum_{j=0}^{h-1}\rho_T^j\epsilon_{T+h-j}\\ &-\left(\widehat{\rho}_T^{\{h\}}-\rho_T^{\{h\}}\right)\left(\widehat{\tau}_T-\tau_T\right).\end{aligned}$$

and $h^{-1/2}\left(\widehat{\rho}_T^{\{h\}}-\rho_T^{\{h\}}\right)\tau_T=h^{-1/2}T^{-1/2}\sum_{i=0}^{h-1}\left(\widehat{\rho}_T^{i}-\rho_T^{i}\right)\psi$. Moreover

$$\begin{aligned}T^{-1}\sum_{i=0}^{h-1}\left(\widehat{\rho}_T^{i}-\rho_T^{i}\right) &\Rightarrow \frac{e^{c\gamma_0}-1}{\gamma_0}-f_\phi\left(c\right),\\ h^{-1/2}\left(\widehat{\rho}_T^{\{h\}}-\rho_T^{\{h\}}\right)\tau_T &\Rightarrow c^{-1/2}\left[\frac{e^{c\gamma_0}-1}{\gamma_0}-f_\phi\left(c\right)\right]\psi.\end{aligned}$$

Similarly, $h^{-1/2}\rho_T^{\{h\}}\left(\widehat{\tau}_T-\tau_T\right)=T^{1/2}h^{-1/2}T^{-1}\left(\sum_{i=0}^{h-1}\rho_T^i\right)\left\{T^{1/2}\left(\widehat{\tau}_T-\tau_T\right)\right\}$ and

$$h^{-1/2}\rho_T^{\{h\}}\left(\widehat{\tau}_T-\tau_T\right)\Rightarrow c^{-1/2}f_\phi\left(c\right)\pi_0.$$

Finally $\left(\widehat{\rho}_T^{\{h\}}-\rho_T^{\{h\}}\right)(\widehat{\tau}_T-\tau_T)=\sum_{i=0}^{h-1}\left(\widehat{\rho}_T^{i}-\rho_T^{i}\right)(\widehat{\tau}_T-\tau_T)$ and

$$\begin{aligned}\sum_{i=0}^{h-1}\left(\widehat{\rho}_T^{i}-\rho_T^{i}\right)(\widehat{\tau}_T-\tau_T) &= \sum_{i=0}^{h-1}\left[i\left(\widehat{\rho}_T-\rho_T\right)+o_p\left(T^{-1}\right)\right](\widehat{\tau}_T-\tau_T)\\ &= T^{-3/2}\frac{h\left(h-1\right)}{2}T\left(\widehat{\rho}_T-\rho_T\right)T^{1/2}\left(\widehat{\tau}_T-\tau_T\right)+o_p\left(T^{-3/2}\right),\end{aligned}$$

so that

$$h^{-1/2}\left(\widehat{\rho}_T^{\{h\}}-\rho_T^{\{h\}}\right)(\widehat{\tau}_T-\tau_T)\Rightarrow c^{-1/2}\left[\frac{e^{c\gamma_0}-1}{\gamma_0}-f_\phi\left(c\right)\right]\pi_0.$$

The results follow.

E Principal notations

$$y_{t,T}=\tau_T+\rho_T y_{t-1,T}+\epsilon_t,\quad (t,T=1,2,...)$$

$$\tau_T=\frac{\psi}{T^k}\ \text{ and }\ k=1/2$$

$$\rho_T=e^{\phi/T}\ \text{ and }\ \lambda=e^{\phi}$$

$$\gamma_i^{(\epsilon)}=\mathsf{Cov}\left[\epsilon_t,\epsilon_{t-j}\right]$$

$$u_T=\sum\nolimits_{j=1}^{T}\epsilon_j\ \text{ and }\ \sigma^2=\lim_{T\to\infty}T^{-1}\mathsf{E}\left[u_T^2\right]$$

$$f_\phi\left(\cdot\right):r\to\frac{e^{\phi r}-1}{\phi}\ \text{ and }\ f_0\left(r\right)=r$$

$$K_{\psi,\phi}\left(r\right)=\psi f_\phi\left(r\right)+\sigma\int_0^r e^{\phi(r-s)}dW\left(s\right)$$

$$\rho_T^{\{h\}}=\sum_{i=0}^{h-1}\rho_T^{i}$$

$$w_{h,t}=\sum_{j=0}^{h-1}\rho_T^{i}\epsilon_{t-j}\ \text{ and }\ \sigma_{w_h}^2=\lim_{T\to\infty}T^{-1}\mathsf{E}\left[\sum_{t=h}^{T}w_{h,t}^2\right]$$

$$v_{h,[Tr]}=\sum_{i=h}^{[Tr]}w_{h,i}\ \text{ and }\ \sigma_h^2=\lim_{T\to\infty}T^{-1}\mathsf{E}\left[v_{h,T}^2\right]$$

$$\varpi_h=\frac{1}{2}\left(1-\sigma_\epsilon^2/\sigma^2\right)-\left(h\sigma^2\right)^{-1}\left[\sum_{i=1}^{h-1}\left(h-i\right)\gamma_i^{(\epsilon)}\right]$$

$$c_T=h/T\Rightarrow c$$

$$h^{-1/2}c_T^{-1/2}\left(\widetilde{\tau}_{h,T}-\tau_{h,T}\right)\Rightarrow\pi_c,\quad\text{ and }\quad\gamma_{h,T}=c_T^{-1}\left(\widetilde{\rho}_{h,T}-\rho_{h,T}\right)\Rightarrow\gamma_c$$