Robustness of Multistep Forecasts and Predictive Regressions at Intermediate and Long Horizons.

Guillaume Chevillon*
ESSEC Business School & CREST

July 23, 2017

Abstract

This paper studies the properties of multi-step projections, and forecasts that are obtained using either iterated or direct methods. The models considered are local asymptotic: they allow for a near unit root and a local to zero drift. We treat short, intermediate and long term forecasting by considering the horizon in relation to the observable sample size. We show the implication of our results for models of predictive regressions used in the financial literature. We show here that direct projection methods at intermediate and long horizons are robust to the potential misspecification of the serial correlation of the regression errors. We therefore recommend, for better global power in predictive regressions, a combination of test statistics with and without autocorrelation correction.

Keywords: Multi-step Forecasting, Predictive Regressions, Local Asymptotics, Dynamic Misspecification, Finite Samples, Long Horizons.
JEL *Classification*: C22, C52, C53.

*E-mail: chevillon@essec.edu

1 Introduction and overview

Two parallel literatures have developed or accelerated recently that aim to estimate relationships over a so-called *multi-step* horizon. On the one hand, there has been a renewed interest in assessing the relative merits of two forecasting methods: those of *iterated* and *direct* multi-step forecasts (denoted IMS and DMS). The former technique constitutes the standard in econometrics and consists in estimating a one-step ahead model relating, say, y_t to y_{t-1} in a sample of T observations and using it to forecast y_{T+1} using y_T and extrapolating the relation to generate a forecast for y_{T+2} using the forecast for y_{T+1} that has previously been obtained. Direct multi-step forecasting by contrast will aim to develop a distinct model for each forecast horizon $h \geq 1$: relating, in-sample, y_t to y_{t-h} so that a forecast for y_{T+h} can be obtained 'directly' using y_T. The relative performances of these forecasts was first derived in general settings by Weiss (1991), but it has been been a continuous interest, since, as in e.g., Clements and Hendry (1996), Ing (2003, 2004), Marcellino, Stock, and Watson (2006), and Schorfheide (2005), most recently, Carriero, Clark, and Marcellino (2015) Chevillon (2016), McElroy and McCracken (2017) and Hendry and Martinez (2017).

On the other hand, the seminal work by Fama and French (1988), Campbell and Shiller (1988) and Stambaugh (1999) has spurred a whole literature within finance of authors who aim to assess the predictive power a variable x_t has on another, say z_t over some horizon. The prototypical "predictive" or "long horizon" regression will take the form of a regression of z_{t+h} on x_t, but $\sum_{i=1}^{h} z_{t+i}$ or $\sum_{i=1}^{h} x_{t+i}$ also appear as regressand and regressor in the literature (see e.g. Lanne, 2002, Torous *et al.*, 2004, Valkanov (2003), Boudoukh et *al.*, 2008, Hjalmarsson, 2011, Phillips and Lee, 2013, Phillips, 2015, and the references therein). The long horizon regression literature shares with that on direct multi-step forecasting three key features: (i) the model which is estimated is not *a priori* that which would most efficiently chose (i.e. the one-step ahead model) but one that induces the errors in the regression to be serially correlated; the chosen *multi-step* technique works for the estimated model because (ii) this model is potentially misspecified as the errors are serially correlated (see Ferson *et al.*, 2003, and Pástor and Stambaugh, 2009) and (iii) the variables that are being used are non-stationary or nearly so (in addition to the papers above, see inter alia Stambaugh, 1999, Lettau and Ludvigson, 2001).

In this paper, we propose a local-asymptotic model that builds on the work of Kemp (1999), Valkanov (2003), Torous *et al.* (2004), Chevillon and Hendry (2005) and Hjalmarsson (2011). We prove a new key property of direct multi-step estimators, namely their robustness to misspecification of the serial correlation of the error process. We then show how this property also applies in the case of long-horizon regressions and that it provides a new justification for why they have proved so successful empirically. We show that the bias that was found by Hjalmarsson relies on his assumption that the horizon h is small compared to the observed sample T, $h = o(T)$, but that it vanishes when considering $h = O(T)$, as suggested by Cochrane (2006) and sheds light on his results. Our analytic results lead us to recommending at long horizon the combination of a standard test and that with Heteroskedasticity and Autocorrelation correction (HAC). We show by simulations that the combination achieves better global power.

This paper is organized as follows. Section 2 presents the forecasting and predictive regression models that we consider and the way they are related. We then derive the distributions of iterated and directed multi-step estimators and forecasts in Section 3. The same section applies these results to predictive regressions. A Monte Carlo assessment follows in Section 4. In the paper, row vectors are denoted as $(x_1 : x_2)$ and column vectors as (x_1, x_2). Throughout, we also use the following notations: $\lambda^{\{h\}} \equiv \sum_{i=0}^{h-1} \lambda^i$ for $h \geq 1$, '$\Rightarrow$' denotes weak convergence of the associated probability measure, $W(r)$ is a standard Brownian motion on $C[0,1]$, and $\lfloor w \rfloor$ denotes the integer part of w for any real scalar w.

2 The models and local-asymptotic assumptions

We introduce here the literatures on multistep forecasting and long-horizon regressions. These literatures present similarities which have not always been stressed.

Throughout the paper, we are considering the simple autoregressive model for the process $\{y_t\}$

$$y_t = \tau + \rho y_{t-1} + \epsilon_t \tag{1}$$

for $t \geq 1$, where y_0 has a finite distribution and the error ϵ_t is assumed to satisfy the following condition.

Condition P. *A sequence $\{\epsilon_t\}$ satisfies Condition* P *if and only if*
(i) $E[\epsilon_t] = 0$ *for all* $t \in Z$;
(ii) $\sup_t E|\epsilon_t|^{\beta+\eta} < \infty$ *for some* $\beta > 2$ *and* $\eta > 0$;
(iii) $\{\epsilon_t\}$ *is weakly stationary with covariance function series* $\{\xi_\epsilon(i)\}_{i=-\infty}^{\infty}$ *such that* $\sum_{i=-\infty}^{\infty} i\xi_\epsilon(i) < \infty$.

Condition P allows to derive general results for general distributions of the errors. Here we restrict our attention to weakly stationary ϵ_t as it allows to derive more explicit results. Yet, our results hold replacing *(iii)* above with the less restrictive assumption that ϵ_t is strongly mixing with mixing coefficients α_m such that $\sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty$ and that $\lim_{T\to\infty} T^{-1}\mathsf{Var}\left[\sum_{j=1}^{T} \epsilon_j\right] = \xi_\epsilon < \infty$ (as in Phillips, 1987). In the following we will be led to restricting the serial dependence of ϵ_t and consider the cases where it is a white noise process or follows a moving average or order one, an MA(1). Also, where use the notation $\xi_u = \sum_{i=-\infty}^{\infty} \xi_u(i)$ for the long run variance of any process u_t satisfying the condition. In the specific case of ϵ_t in (1) we write $\sigma^2 = \xi_\epsilon$.

In the time series forecasting literature, the standard multistep forecasting technique consists in estimating the parameters (τ, ρ) of model (1) and then to use the estimators $(\widehat{\tau}, \widehat{\rho})$ to compute forecasts recursively at all horizons $h \geq 1$:

$$\widehat{y}_{t+h|t} = \widehat{\tau} + \widehat{\rho}\widehat{y}_{t+h-1|t} = \widehat{\rho}^{\{h\}}\widehat{\tau} + \widehat{\rho}^h y_t \tag{2}$$

where we let $\widehat{y}_{t|t} = y_t$ and $\widehat{\rho}^{\{h\}} = \sum_{i=0}^{h-1} \widehat{\rho}^i$. This constitutes the plug-in or iterated multi-step (IMS) technique.

Direct multistep (DMS) has often been proposed as an alternative: it which consists in estimating the parameters (τ_h, ρ_h) of the projection of y_{t+h} on y_t,

$$y_t = \tau_h + \rho_h y_{t-h} + w_{h,t}, \text{ for } h \geq 1, \tag{3}$$

with $(\tau_h, \rho_h) = \left(\rho^{\{h\}}\tau, \rho^h\right)$ and $w_{h,t} = \sum_{i=0}^{h-1} \rho^i \epsilon_{t-i}$. The DMS forecasts are obtained from estimators $(\widetilde{\tau}_h, \widetilde{\rho}_h)$ as

$$\widetilde{y}_{t+h|t} = \widetilde{\tau}_h + \widetilde{\rho}_h y_t. \tag{4}$$

To achieve robustness to misspecification, the literature has often considered $(\widehat{\tau}, \widehat{\rho})$ and $(\widetilde{\tau}_h, \widetilde{\rho}_h)$ to be the ordinary least squares (OLS) estimators and we follow this approach here. The rationale for DMS lies in that when ϵ_t in (1) is serially correlated, IMS forecasts are biased and DMS can prove more accurate in terms of mean-square forecast error (MSFE).

The predictive regression literature (since Fama and Schwert, 1977, and Rozeff, 1984, see Stambaugh, 1999) has considered testing the null of not predictability in a bivariate setting: a standard model (see e.g. Valkanov, 2003) lets, for $t = 0, ..., T$

$$\begin{bmatrix} z_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} \alpha \\ \tau \end{bmatrix} + \begin{bmatrix} 0 & \beta \\ 0 & \rho \end{bmatrix} \begin{bmatrix} z_t \\ y_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ \epsilon_{t+1} \end{bmatrix} \tag{5}$$

where both ε_t and ϵ_t are assumed to satisfy Condition P.[1] For instance, in Pástor and Stambaugh (2009), z_t denotes the return on an asset, y_t an imperfect predictor thereof, and the null hypothesis is $\mathsf{H}_0 : \beta = 0$. Model (5) is often expressed, for $h \geq 1$, as

$$z_t = \alpha_h + \beta_h y_{t-h} + \omega_{h,t} \tag{6}$$

with $(\alpha_h, \beta_h) = \left(\alpha + \beta\tau\rho^{\{h-1\}}, \beta\rho^{h-1}\right)$, $\omega_{h,t+h} = \beta \sum_{i=1}^{h-1} \rho^{h-1-i} \epsilon_{t-h+i} + \varepsilon_t$; or using $Z_{t-h+1}^h = \sum_{i=1}^{h} z_{t-h+i}$ as a regressand (see Valkanov, 2003, and references therein). In Expression (6), the hypothesis of interest is $\mathsf{H}_0^h : \beta_h = 0$. The empirical literature has shown that whereas $\mathsf{H}_0^{h=1}$ often does not reject, this is not the case when considering large h, in which case H_0^h may reject and y_{t-h} appears helpful in predicting z_t. The question of how large h should be is an empirical one: Hjalmarsson (2011) studies the case where h is fixed, whereas Valkanov (2003), Torous *et al.* (2004) and Hjalmarsson (2011) have considered letting the horizon grow with the sample size T as respectively $h = O(T)$ and $h = o(T)$. In their setting, Torous *et al.* and Hjalmarsson allowed in addition for the error $(\varepsilon_{t+1}, \epsilon_{t+1})$ to exhibit autocorrelation. This complicates the derivation of the distributions of the estimators and test statistics but it yields insight regarding the role played by h. Indeed, Hjalmarsson shows that the estimators of the regression coefficient of Z_{t-h+1}^h on y_{t-h} suffers from second-order bias generated by the correlation between ε_t and ϵ_t. This result is similar to that of Banerjee *et al.* (1996) in the context of a comparison between iterated and direct multistep forecasting. Here the main insight we gain about predictive regressions from multistep forecasting occurs under predictability ($\beta_h \neq 0$), hence our results allow to devise tests with increased global power.

[1] This precludes assumptions such as in Deng (2013) where ϵ_t exhibits a moving average local unit root.

In the following, we assume that the parameters of (6) are estimated using OLS. This choice assumes that the errors ε_t are martingale difference sequences (MDS) and is common in empirical work, see e.g. Stambaugh (1999). In reality, this assumption may be wrong and ε_t may be autocorrelated as shown by Pástor and Stambaugh (2009) where they follow an MA(1) . Although the literature has also considered variance estimators which are heteroskedasticity and autocorrelation consistent (HAC), we do not study them specifically here. Indeed, these do not correct for the bias in the autoregression. Also, by taking into account the serial correlation in the estimation of the variances of $w_{h,t}$ and $\omega_{h,t}$ they should benefit multistep or long horizon estimators and only strengthen our argument.

In this paper, we aim to capture three key issues that arise both the in predictive regression and multistep forecasting frameworks: (i) the interaction between the horizon h, the available sample size T, (ii) the persistence in the time series y_t and z_t and (iii) the serial dependence in ϵ_t and ε_t. For this we consider the local-asymptotic framework that is now common in the econometric literature.

First, we follow the now standard assumptions that ρ is close to unity: we follow authors such as Phillips (1987) or Campbell and Yogo (2006) and model them as local to unity

$$\rho_T = \exp(\phi/T) = 1 + \phi/T + O\left(T^{-2}\right), \tag{7}$$

Expression (7) implies that y_t is near integrated and that τ acts as a near drift. This latter issue has generally been avoided in the early literature by imposing $\alpha = \tau = 0$ which corresponds to using demeaned variables, but not in the some seminal articles (e.g. Campbell and Yogo, 2006). Owing to the near non-stationary nature of the variables, demeaning may not provide more accurate estimates. In particular if τ is indeed nonzero but very small so that a near linear trend is mistaken for a non zero mean (see for instance Pástor and Stambaugh, 2009, where τ is low when ρ is close to unity). Chevillon and Hendry (2005) have shown that small nonzero drifts can have a significant influence on the multistep forecasts when dealing with non-stationary variables. Also, the literature on returns forecasting has acknowledged the importance of slowly drifting expected returns (see e.g. Lettau and Van Nieuwerburgh, 2007). For these reasons we allow for the parameter τ in (1) to be nonzero but assume that it is small and model it via local-asymptotics as a Pitman drift:

$$\tau_T = \frac{\psi}{\sqrt{T}}. \tag{8}$$

Such a *local* drift would be of low magnitude, justifying the local-asymptotic assumption. Local-to-zero drifts have been used inter alia in Monte Carlo simulations of unit root tests in Vogelsang (1998), Rossi (2005a) and Busetti and Harvey (2008); they have been studied analytically by Haldrup and Hylleberg (1995) and Stock and Watson (1996). Parameterizing the drift as (8) induces a (nonlinear since $\rho_T < 1$) deterministic trend of order $O\left(\sqrt{T}\right)$. In the paper, we denote by y_t the triangular array that is generated by the non-constant parameters (τ_T, ρ_T).

Second, we consider either hold the horizon h constant, or letting it grow as a constant fraction of the sample size T as in the following definition.

Definition 1 *Let $h \geq 1$ denote the horizon of interest.*
We refer to the horizon begin long *with respect to the sample size T if there exists a constant $c \in (0,1)$ such that $h/T \to c$ as $T \to \infty$;*
the horizon is short *if h is constant irrespective of T; and*
the horizon is said intermediate *in a sequential asymptotic setting where $h/T \to c$ as $T \to \infty$ and then we let $c \to 0$.*

Long run forecasting has been studied by Stock (1996), Phillips (1998), Kemp (1999) and in long-run predictive regression by Valkanov (2003), Torous *et al.* (2004), Turner (2004) and Elliott (2006). Although different, the sequential asymptotic intermediate horizon framework relates the setting of Hjalmarsson (2011) where $h/T \to 0$.

Finally, the problem of misspecification may arise even for $h = 1$ if ε_t or ϵ_t exhibits serial correlation and cross correlation. We define their joint autocovariance function

$$\mathsf{E}\begin{bmatrix} \varepsilon_t \\ \epsilon_t \end{bmatrix}\begin{bmatrix} \varepsilon_{t-k} \\ \epsilon_{t-k} \end{bmatrix}' = \Xi_k = \begin{bmatrix} \xi_\varepsilon(k) & \xi_{\varepsilon,\epsilon}(k) \\ \xi_{\epsilon,\varepsilon}(k) & \xi_\epsilon(k) \end{bmatrix}$$

with $\Xi = \sum_{k=-\infty}^{+\infty} \Xi_k$ and denote $\varsigma^2 = \sum_{k=-\infty}^{+\infty} \xi_\varepsilon(k)$, $\varrho^2 = \sum_{k=-\infty}^{+\infty} \xi_{\varepsilon,\epsilon}(k)$, with as before $\sigma^2 = \sum_{k=-\infty}^{+\infty} \xi_\epsilon(k)$.

3 Estimators and Forecasts

This section provides our main results. First, we consider the asymptotic distribution of the OLS estimators $(\widehat{\tau}, \widehat{\rho})$ and $(\widetilde{\tau}_h, \widetilde{\rho}_h)$ under various assumptions on the horizon. Then we derive the implications of our results for forecasting.

3.1 Distributions of empirical moments

Under Condition P, $T^{-1/2}\sum_{i=1}^{\lfloor Tr \rfloor} \epsilon_i \Rightarrow \sigma W(r)$, as $T \to \infty$, where $W(r)$ denotes a Wiener process. We define the Vasiček process[2] $K^*_{\psi,\phi}(r) = \psi f_\phi(r) + \int_0^r e^{\phi(r-s)} dW(s)$,for $r \in [0,1]$, where the functional[3] $f_{(\cdot)} : \mathbb{R} \to C[0,1]$ satisfies $f_\phi(\cdot) : r \to \left(e^{\phi r} - 1\right)/\phi$ for $\phi \in \mathbb{R}\backslash\{0\}$ and $f_0(r) = r$. By extension, for a given $\sigma > 0$, denote by $K_{\psi,\phi}(r)$ the functional $K_{\psi,\phi}(r) = \sigma J_{\psi/\sigma,\phi}(r)$ solution to the linear stochastic differential equation.

$$dK_{\psi,\phi}(r) = [\psi + \phi K_{\psi,\phi}(r)]\, dr + \sigma dW(r) \tag{9}$$

with initial condition $K_{\psi,\phi}(0) = 0$. $K_{\psi,\phi}(r)$ is a Gaussian process for fixed r with expectation $\psi f_\phi(r)$ and variance $\sigma^2 f_{2\phi}(r)$. For $\psi = 0$, it reduces to an Ornstein-Uhlenbeck (OU) $J_\phi(r) = K_{0,\phi}(r)$.

[2]It is standard in the literature to parametrize instead the process imposing $\psi = -\lambda\phi$ for some $\lambda \geq 0$.

[3]We denote by $D[0,1]$ the space of real-valued functions on the interval $[0,1]$ which are right continuous and have finite left limits (càdlàg). $C[0,1]$ is the subspace of $D[0,1]$ of continuous functions. We will straightforwardly extend this definition below to allow for $r > 1$ in forecasting.

First, holding h constant, the variance of the fixed horizon multi-step disturbance $w_{h,t}$ admits the variance

$$\sigma_{w_h}^2 = \lim_{T\to\infty} T^{-1} \sum_{t=h}^{T} w_{h,t}^2 = h\sigma_\epsilon^2 + 2\sum_{i=1}^{h-1} (h-i)\,\xi_\epsilon(i)$$

and let its long-run variance $\sigma_h^2 = \lim_{T\to\infty} T^{-1}\mathsf{Var}\left[\sum_{i=h}^{[Tr]} w_{h,i}\right] = h^2\sigma^2$. Then, simply, for $r \in [0,1]$, $T^{-1/2}\sum_{i=h}^{\lfloor Tr\rfloor} w_{h,i} \Rightarrow \sigma_h W(r) = h\sigma W(r)$, as $T \to \infty$.

When letting h be a fraction of the sample size $T : h = O(T)$, $w_{h,t}$ becomes a non-stationary series exhibiting a stochastic trend and the usual scaling factors for integrated processes no longer hold. We define the operator δ_ϕ^c which, for any diffusion process $Z(r)$ defined on $C[0,\eta]$, $\eta \geq c$ lets $\delta_\phi^c Z$ on $C[0,\eta]$ be such that: $\delta_\phi^c Z(r) = Z(r) - e^{\phi c} Z(r-c)$, for $c \leq r \leq \eta$ and $\delta_\phi^c Z(r) = Z(r)$, for $0 \leq r < c$. A proposition follows that provides all the asymptotic convergence properties that we require in the paper.

Proposition 1 *Let y_t be generated as (1) under Condition P with local asymptotic parameters (7) and (8). Then, the following holds as $T \to \infty$,*
under ***short horizon,*** $h \in [1,T)$ *is constant,*
(a_h) $T^{-1/2}y_{\lfloor Tr\rfloor} \Rightarrow K_{\psi,\phi}$, $\quad T^{-3/2}\sum_{t=h}^{T} y_t \Rightarrow \int_0^1 K_{\psi,\phi}$, $\quad T^{-2}\sum_{t=h}^{T} y_t^2 \Rightarrow \int_0^1 K_{\psi,\phi}^2$;
(b_h) $T^{-1}\sum_{t=h}^{T} y_{t-h}w_{h,t} \Rightarrow h\sigma\int_0^1 K_{\psi,\phi}dW + \frac{1}{2}\left[h\sigma^2 - \sigma_{w_h}^2\right]$;
(c_h) $T^{-1/2}\sum_{i=h}^{\lfloor Tr\rfloor} w_{h,i} \Rightarrow h\sigma W(r)$, $\quad T^{-1}\sum_{t=h}^{T} w_{h,t}^2 \to \sigma_{w_h}^2$.
under ***long horizon*** $h/T \to c \in (0,1)$, *as* $T \to \infty$,
(a_c) $T^{-1/2}y_{\lfloor Tr\rfloor} \Rightarrow K_{\psi,\phi}(r)$, $\quad T^{-3/2}\sum_{t=h}^{T} y_t \Rightarrow \int_c^1 K_{\psi,\phi}$, $\quad T^{-2}\sum_{t=h}^{T} y_t^2 \Rightarrow \int_c^1 K_{\psi,\phi}^2$;
(b_c) $T^{-2}\sum y_{t-h}w_{h,t} \Rightarrow \int_c^1 K_{\psi,\phi}(r-c)\,\delta_\phi^c J_\phi(r)\,dr - \frac{1}{2}c\psi^2 e^{-\phi c}\left[f_\phi(c)\right]^2$
(c_c) $T^{-1/2}w_{h,\lfloor Tr\rfloor} \Rightarrow \delta_\phi^c J_\phi(r)$, $\quad T^{-3/2}\sum_{t=h}^{T} w_{h,t} \Rightarrow \int_c^1 \delta_\phi^c J_\phi(r)\,dr$, $\quad T^{-2}\sum_{t=h}^{T} w_{h,t}^2 \Rightarrow \int_c^1 \left[\delta_\phi^c J_\phi(r)\right]^2 dr$.

To allow for a comparison between short and long horizons in Proposition 1, the following corollary considers the intermediate horizon setting. The notation $A \underset{c\to 0}{\overset{L}{=}} B$ means that as $c \to 0^+$, for all $x \in \mathbb{R}$, $\Pr(A < x)/\Pr(B < x) \to 1$.

Corollary 2 (Intermediate Horizon) *Under the assumption of 1, the asymptotic distributions under long horizon settings satisfy as* $c \to 0$:
(a_c) $\int_c^1 K_{\psi,\phi} \underset{c\to 0}{\overset{L}{=}} \int_0^1 K_{\psi,\phi}$, $\quad \int_c^1 K_{\psi,\phi}^2 \underset{c\to 0}{\overset{L}{=}} \int_0^1 K_{\psi,\phi}^2$;
(b_c) $\int_c^1 K_{\psi,\phi}(r-c)\,\delta_\phi^c J_\phi(r)\,dr - \frac{1}{2}c\psi^2\lambda^{-c}\left[f_\phi(c)\right]^2 \underset{c\to 0}{\overset{L}{=}} \sqrt{c}\sigma\int_0^1 K_{\psi,\phi}(r)\,dW(r)$
(c_c) $\delta_\phi^c J_\phi(r+c) \underset{c\to 0}{\overset{L}{=}} \sigma\left[W(r+c) - W(r)\right]$, $\quad \int_c^1 \delta_\phi^c J_\phi(r+c)\,dr \underset{c\to 0}{\overset{L}{=}} \sqrt{c}\sigma W(1)$, *and*
$\int_c^1 \delta_{2\phi}^c J_\phi^2(r)\,dr \underset{c\to 0}{\overset{L}{=}} \sigma^2 c$.

One of the key implications of the different behaviors under short and long horizons relates to the sample covariance between the regressors and the disturbances in expressions (1) and (3). Under short horizons:

$$h^{-1}T^{-1}\left(\sum_{t=h}^{T} y_{t-h}w_{h,t} - h\sum_{t=1}^{T} y_{t-1}\epsilon_t\right) \Rightarrow -\sum_{i=1}^{h-1}(h-i)\,\xi_\epsilon(i). \qquad (10)$$

so differences between multi-step and scaled one-step moments only arise asymptotically when the error ϵ_t is autocorrelated.

Expressions under long horizons are more involved, but intermediate horizons allow for an easy comparison:

$$\begin{array}{ccc} \lim_{T\to\infty} T^{-2}\sum y_{t-\lfloor cT\rfloor} w_{\lfloor cT\rfloor,t} & \underset{c\to 0}{\overset{L}{=}} & c^{1/2}\sigma \int_0^1 K_{\psi,\phi}(r)\, dW(r), \\ T^{-2}\left(\lfloor cT\rfloor \sum_{t=1}^{T} y_{t-1}\epsilon_t\right) & \Rightarrow & c\left(\sigma\int_0^1 K_{\psi,\phi} dW + \frac{1}{2}\left[\sigma^2 - \sigma_\epsilon^2\right]\right). \end{array} \tag{11}$$

This latter expression show that in OLS estimation, there will be a trade-off of bias and efficiency between one-step ahead and multistep projections. Indeed the expectation of $\lim_{T\to\infty} T^{-2}\sum y_{t-\lfloor cT\rfloor} w_{\lfloor cT\rfloor,t}$ is nonzero but $o(c)$ whether or not the error is autocorrelated.[4] By contrast the corresponding expectation of $T^{-2}\left(\lfloor cT\rfloor \sum_{t=1}^{T} y_{t-1}\epsilon_t\right)$ is zero in the absence of misspecification but $O(c)$ otherwise. In terms of variance, the order is reversed: the multistep moments have asymptotic variance in the $O(c)$ and the scaled one-step in $O\left(c^2\right)$ for $c \to 0$.

The previous analysis show that whether the horizon is short or long will have a significant impact on the estimators. Short horizons multistep estimation will be affected by misspecification, and this may be beneficial or detrimental. By contrast, long horizon multi-step estimation will be mostly unaffected by the misspecification. This is due to the fact that as $h \to \infty$, the multi-step error $w_{h,t}$ becomes an integrated process whose autocovariance function is constant; in other terms, the correction $h\sigma^2 - \sigma_{w_h}^2 \to 0$ in Proposition 1-(b_h).

3.2 OLS Estimators

To emphasize the different behaviors, define the scaled deviations of OLS slope and intercept estimators from the parameters as:[5]

$$\gamma_T = T\left(\widehat{\rho}_T - 1\right) \Rightarrow \gamma_0, \quad \text{and} \quad \pi_T = T^{1/2}\left(\widehat{\tau}_T - \tau_T\right) \Rightarrow \pi_0. \tag{12}$$

We define, for notational ease, the following stochastic matrix

$$D_c = \begin{bmatrix} 1-c & \int_c^1 K_{\psi,\phi} \\ \int_c^1 K_{\psi,\phi} & \int_c^1 K_{\psi,\phi}^2 \end{bmatrix}.$$

The one-step OLS estimator is then characterized by

$$\begin{bmatrix} \pi_0 \\ \gamma_0 - \phi \end{bmatrix} = D_0^{-1} \begin{bmatrix} \sigma W(1) \\ \sigma \int_0^1 K_{\psi,\phi} dW + \frac{\sigma^2 - \sigma_\epsilon^2}{2} \end{bmatrix} \tag{13}$$

The presence of a local-to-zero drift implies that the stochastic and the deterministic trends have identical asymptotic orders of magnitude, both $O_p\left(T^{1/2}\right)$. The unit-root estimator is super-consistent but the corresponding error is of order $O_p\left(T^{-1}\right)$ and not the $O_p\left(T^{-3/2}\right)$ observed

[4] The term $\frac{1}{2}c\psi^2 e^{-\phi c}\left[f_\phi(c)\right]^2$ that arises in proposition 1-(b_c) is $O\left(c^3\right)$. It is zero in the absence of a drift, i.e. when $\psi = 0$.

[5] We define γ_T as deviation for $\widehat{\rho}_T$ for unity rather than from ρ_T for ease of notation in the long horizon setting.

in the presence of a true linear trend. When ϵ_t is white noise, $\sigma = \sigma_\epsilon$ and we denote the estimators of a true AR(1) as $\left(\pi_0^\times, \gamma_0^\times\right) = \sigma D_0^{-1}\left(W(1), \int_0^1 K_{\psi,\phi} dW\right)$. In the presence of dependent errors, $\left(\sigma^2 - \sigma_\epsilon^2\right)/2 = \sum_{i=1}^{\infty} \xi_\epsilon(i)$ and this is the channel through which misspecification of the innovations affects the estimators. The correlation between π_0 and γ_0 is then a positive function of $-\int_0^1 K_{\psi,\phi} dr$ and the latter's expectation has a sign opposite that of ψ.

Now, the previous results may be used for computing IMS and DMS estimators of the multistep parameters $\left(\tau_{h,T}, \rho_{h,T}\right)' = \left(\rho_T^{\{h\}} \tau_T, \rho_T^h\right)'$. The IMS estimators are naturally defined as $\left(\widehat{\tau}_{\{h\},T}, \widehat{\rho}_{\{h\},T}\right)' = \left(\widehat{\rho}_T^{\{h\}} \widehat{\tau}_T, \widehat{\rho}_T^h\right)'$. The DMS estimators $\left(\widetilde{\tau}_{h,T}, \widetilde{\rho}_{h,T}\right)'$ are computed via OLS of (1) over a sample of size T. We denote the asymptotic limits as follows. Under short horizon (fixed h),

$$\text{IMS}: \left(\sqrt{T}\left(\widehat{\tau}_{\{h\},T} - \tau_{h,T}\right), T\left(\widehat{\rho}_T^h - 1\right)\right) \Rightarrow \left(\pi_{\{h\}}, \gamma_{\{h\}}\right),$$

$$\text{DMS}: \left(\sqrt{T}\left(\widetilde{\tau}_{h,T} - \tau_{h,T}\right), T\left(\widetilde{\rho}_{h,T} - 1\right)\right) \Rightarrow \left(\pi_h, \gamma_h\right),$$

and under long horizon, as $h/T \to c$,

$$\text{IMS}: \left(T^{-1/2}\left(\widehat{\tau}_{\{h\},T} - \tau_{h,T}\right), \widehat{\rho}_T^h - 1\right) \Rightarrow \left(\pi_{\{c\}}, \gamma_{\{c\}}\right),$$

$$\text{DMS}: \left(T^{-1/2}\left(\widetilde{\tau}_{h,T} - \tau_{h,T}\right), \widetilde{\rho}_{h,T} - 1\right) \Rightarrow \left(\pi_c, \gamma_c\right).$$

Using the results above, the following Proposition relates the distribution of the multi-step estimators to those of the one-step.

Proposition 3 *Let y_t be generated as (1) under Condition* P *with local asymptotic parameters (7) and (8). Then the following holds as $T \to \infty$,*
under ***short horizon,*** *$h \in [1, T)$ is constant, and the limits are,*

$$\text{IMS}: \begin{bmatrix} \pi_{\{h\}} \\ \gamma_{\{h\}} \end{bmatrix} = h \begin{bmatrix} \pi_0 \\ \gamma_0 \end{bmatrix},$$

$$\text{DMS}: \begin{bmatrix} \pi_h \\ \gamma_h \end{bmatrix} = h \begin{bmatrix} \pi_0 \\ \gamma_0 \end{bmatrix} + \frac{1}{2}\left(h\sigma_\epsilon^2 - \sigma_{w_h}^2\right) D_0^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix};$$

under ***long horizon,*** *$h/T \to c \in (0,1)$, the limits are*

$$\text{IMS}: \begin{bmatrix} \pi_{\{c\}} \\ \gamma_{\{c\}} \end{bmatrix} = \begin{bmatrix} \left[f_{\gamma_0}(c) - f_\phi(c)\right]\psi + f_{\gamma_0}(c)\pi_0 \\ f_{\gamma_0}(c)\gamma_0 \end{bmatrix},$$

$$\text{DMS}: \begin{bmatrix} \pi_c \\ \gamma_c \end{bmatrix} = \begin{bmatrix} 0 \\ \phi f_\phi(c) \end{bmatrix} + D_c^{-1} \begin{bmatrix} \int_c^1 \delta_\phi^c J_\phi(r)\, dr \\ \int_c^1 K_{\psi,\phi}(r-c)\,\delta_\phi^c J_\phi(r)\, dr - \frac{1}{2} c\psi^2 e^{-\phi c}\left[f_\phi(c)\right]^2 \end{bmatrix}.$$

Proposition 3 allows for a comparison of IMS and DMS estimation accuracy. Both estimators are consistent for the multistep parameters at short but not at long horizons. Indeed for the latter, the estimators must be scaled by an additional T (or h) to ensure they weakly converge. At short horizons, IMS and DMS yield identical asymptotic distributions when ϵ_t follows a white noise and these are simply h times the one-step. By contrast, serial correlation in ϵ_t implies that

DMS distributions which are not h times that of the one-step model. To see the impact of the autocorrelation of ϵ_t, consider the differences between $\left(\pi_{\{h\}}, \gamma_{\{h\}}\right)$ and the corresponding random variables $\left(\pi^{\times}_{\{h\}}, \gamma^{\times}_{\{h\}}\right)$ when ϵ_t is white noise (we define $\left(\pi^{\times}_h, \gamma^{\times}_h\right)$ similarly): then

$$\left(\pi_{\{h\}} - \pi^{\times}_{\{h\}}, \gamma_{\{h\}} - \gamma^{\times}_{\{h\}}\right) = h\left(\pi_0 - \pi^{\times}_0, \gamma_0 - \gamma^{\times}_0\right)$$
$$\left(\pi_h - \pi^{\times}_h, \gamma_h - \gamma^{\times}_h\right) = \frac{h\sigma^2 - \sigma^2_{w_h}}{\sigma^2 - \sigma^2_{\epsilon}}\left(\pi_0 - \pi^{\times}_0, \gamma_0 - \gamma^{\times}_0\right)$$

where $\frac{h\sigma^2 - \sigma^2_{w_h}}{\sigma^2 - \sigma^2_{\epsilon}} = h - \left(\sum_{i=1}^{\infty} \xi_{\epsilon}(i)\right)^{-1} \sum_{i=1}^{h-1} (h-i)\, \xi_{\epsilon}(i)$. In particular, if ϵ_t follows an MA(q), then for $h > q$, $h\sigma^2 - \sigma^2_{w_h} = q\sigma^2 - \sigma^2_{w_q}$. This shows that if ϵ_t follows a moving average (as in e.g. Pástor and Stambaugh, 2009, where it is an MA(1)) then the impact of the serial correlation in ϵ_t is increasing linearly in the horizon for IMS but bounded by that at horizon q for DMS. Banerjee, Hendry, and Mizon (1996) find a similar result. Now the actual distribution of $\left(\pi_0 - \pi^{\times}_0, \gamma_0 - \gamma^{\times}_0\right)$ depends on the parameters of the DGP but its expectation has the sign of $(-\psi, 1)\left(\sigma^2 - \sigma^2_{\epsilon}\right) = (-\psi, 1) \sum_{i=1}^{h-1} (h-i)\, \xi_{\epsilon}(i)$. Since, in general, the bias in autoregressive parameter estimators is negative an AR(1) with a near unit root, this implies that $\mathsf{E}\left(\gamma^{\times}_0\right) < 0$. Hence if ϵ_t is negatively autocorrelated then the probability $\mathrm{E}\left[\gamma_0 - \gamma^{\times}_0\right] < 0$ so the distribution of γ_0 is shifted to the left, i.e $\widehat{\rho}$ further away from unity, with a larger absolute bias than when ϵ_t is white noise. As the horizon grows, then IMS compounds the bias but that of DMS remains bounded (if ϵ_t follows a moving average). Given the negative expected correlation between the intercept and slope estimators, positive ψ will have the same effect on the bias of the multistep intercept estimator. This is what Chevillon and Hendry (2005) found in their simulations.

Proposition 3 also allows for a comparison of the estimators at long horizons, but nonlinearities render the analysis of analytical results difficult. For this reason, the following corollary considers intermediate horizons.

Corollary 4 (Intermediate Horizon) *Under the assumptions of Proposition 3,the asymptotic distributions under long horizon settings satisfy as* $c \to 0$ *:*

$$\begin{bmatrix} \pi_{\{c\}} \\ \gamma_{\{c\}} - c\phi \end{bmatrix} \underset{c\to 0}{\overset{L}{=}} c \begin{bmatrix} \pi_0 \\ \gamma_0 - \phi \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \pi_c \\ \gamma_c - c\phi \end{bmatrix} \underset{c\to 0}{\overset{L}{=}} \sqrt{c} \begin{bmatrix} \pi^{\times}_0 \\ \gamma^{\times}_0 - \phi \end{bmatrix}$$

Corollary 4 confirms the analysis that was made previously that intermediate horizon DMS estimators are robust to serial correlation of ϵ_t since their distribution is a proportional to the unbiased $\left(\pi^{\times}_0, \gamma^{\times}_0\right)$. This is not the case for IMS which are biased. Yet, as $c \to 0$, $\sqrt{c}$ is of higher magnitude than c, so DMS suffers from higher variance than IMS.

3.3 Forecasting

We now derive the distributions of the forecast errors. Parameter estimates are used to forecast the series h steps ahead from an end-of-sample forecast origin y_T using the expressions of Section 2. Define the IMS forecast errors under short horizon as $\widehat{e}_{h|T} = y_{T+h} - \widehat{y}_{T+h|T}$ and under long horizons as $\widehat{e}^{*}_{c,T} = h^{-1/2}\widehat{e}_{h|T}$. Denote the corresponding DMS forecast errors as $\widetilde{e}_{h|T}$ and $\widetilde{e}^{*}_{c,T}$.

In short-horizon forecasting, consistency of the estimators imply that the asymptotic limit of the forecast error is simply $\widehat{e}_{h|T} - \sum_{j=0}^{h-1} \epsilon_{T+h-j} \underset{T\to\infty}{\overset{p}{\to}} 0$ and similarly for $\widetilde{e}_{h|T}$. Hence, for a comparison we derive the short horizon distributions as deviations from $\sum_{j=0}^{h-1} \epsilon_{T+h-j}$. For the long horizon case, we need to extend the definition of $K_{\psi,\phi}(r)$ to cover $r \in [0, 1+\ell]$ for some $\ell \in (c, 1)$. The following proposition provides asymptotic distributions of the forecast errors.

Proposition 5 *Let y_t be generated as (1) under Condition P with local asymptotic parameters (7) and (8). Then the following holds as $T \to \infty$,*
under ***short horizons*** $h \in [1, T)$,

$$\begin{aligned} T^{1/2}\left(\widehat{e}_{h|T} - \textstyle\sum_{j=0}^{h-1} \rho_T^j \epsilon_{T+h-j}\right) &\Rightarrow -\pi_{\{h\}} - \left(\gamma_{\{h\}} - h\phi\right) K_{\psi,\phi}(1) \\ T^{1/2}\left(\widetilde{e}_{h|T} - \textstyle\sum_{j=0}^{h-1} \rho_T^j \epsilon_{T+h-j}\right) &\Rightarrow -\pi_h - \left(\gamma_h - h\phi\right) K_{\psi,\phi}(1) \end{aligned}$$

and under ***long horizons*** $h/T \to c \in (0, 1)$,

$$\begin{aligned} \sqrt{c}\widehat{e}^{*}_{c,T} &\Rightarrow -\left[\pi_{\{c\}} + \left(\gamma_{\{c\}} - \phi f_\phi(c)\right) K_{\psi,\phi}(1)\right] + \delta^c_\phi J_\phi(1+c), \\ \sqrt{c}\widetilde{e}^{*}_{c,T} &\Rightarrow -\left[\pi_c + \left(\gamma_c - \phi f_\phi(c)\right) K_{\psi,\phi}(1)\right] + \delta^c_\phi J_\phi(1+c). \end{aligned}$$

The key to forecast accuracy is here the correlation between the slope estimator and the demeaned forecast origin. Indeed, whereas for stationary processes it has been customary to assume that the correlation between the forecast origin and the estimators has little impact, this assumption does not hold in the presence of trending behavior (see Ing, 2004). In short horizon forecasting, the proposition implies that

$$T^{1/2}\left(\widehat{e}_{h|T} - \widetilde{e}_{h|T}\right) \Rightarrow -h \sum_{i=1}^{h-1} (1 - i/h)\, \xi_\epsilon(i)\, \frac{K^\mu_{\psi,\phi}(1)}{\int \left(K^\mu_{\psi,\phi}\right)^2}. \tag{14}$$

where $K^\mu_{\psi,\phi} = K_{\psi,\phi}(r) - \int_0^1 K_{\psi,\phi}(u)\, du$. This expression shows that for $\epsilon_t \sim MA(q)$, whichever method is more precise at horizon $q+1$ will tend also to be so for $h \geq q+1$, and the difference in forecast errors is close to being linear in h. When ϵ_t is white noise and the horizon short, both methods are asymptotically equivalent. Expression (14) also shows that if $\mathsf{E}\left[K^\mu_{\psi,\phi}(1) / \int \left(K^\mu_{\psi,\phi}\right)^2\right] > 0$, such as when $\psi > 0$, then negatively autocorrelated ϵ_t imply that $\mathsf{E}\left[\widehat{e}_{h|T} - \widetilde{e}_{h|T}\right] > 0$. In particular, if ϵ_t follows an MA(1) , then

$$sign\left(\mathsf{E}\left[\widehat{e}_{h|T} - \widetilde{e}_{h|T}\right]\right) = -sign\left(\xi_\epsilon(1)\, \psi\right). \tag{15}$$

Heuristically, if $\mathrm{E}\left[\widehat{e}_{h|T}\right]$ and $\mathrm{E}\left[\widetilde{e}_{h|T}\right]$ have the sign of ψ, then $\xi_\epsilon(1) < 0$ implies that forecast biases favor DMS: $\mathsf{E}\left[\left|\widehat{e}_{h|T}\right|\right] > \mathsf{E}\left[\left|\widetilde{e}_{h|T}\right|\right]$.

Next, we consider intermediate and long horizon settings. For low c, the forecast errors from either method do not behave comparably with respect to the horizon:

Corollary 6 (Intermediate Horizon) *Under the assumptions of Proposition 5, the limiting distributions as $c \to 0$ satisfy:*

$$\widehat{e}^{*}_{c,T} \underset{c\to 0}{\overset{L}{=}} -\sqrt{c}\left(\gamma_0 - \phi\right) K_{\psi,\phi}\left(1\right) - \sqrt{c}\pi_0 + \sigma\left[\frac{W\left(1+c\right) - W\left(1\right)}{\sqrt{c}}\right] \tag{16a}$$

$$\widetilde{e}^{*}_{c,T} \underset{c\to 0}{\overset{L}{=}} -\left(\gamma_0^{\times} - \phi\right) K_{\psi,\phi}\left(1\right) - \pi_0^{\times} + \sigma\left[\frac{W\left(1+c\right) - W\left(1\right)}{\sqrt{c}}\right] \tag{16b}$$

The corollary shows the insight we drew from the estimators carry over to the forecasts: (i) since DMS estimator biases are not affected by serial correlation of ϵ_t at intermediate horizons nor are the forecast; yet (ii) DMS forecasts have higher variance. There therefore exists a trade-off between DMS robustness to dynamic misspecification and the compounded variance due to the horizon effect. The corollary shows though that at intermediate horizons in the presence of serial correlation of ϵ_t, biases differ by an order of magnitude: $\mathsf{E}\left[\widehat{e}^{*}_{c,T}\right]/\mathsf{E}\left[\widetilde{e}^{*}_{c,T}\right] = O\left(\sqrt{c}\right)$, but variances are comparable $\mathsf{Var}\left[\widehat{e}^{*}_{c,T}\right] = O\left(1\right)$, $\mathsf{Var}\left[\widetilde{e}^{*}_{c,T}\right] = O\left(1\right)$. If $\left(\gamma_0^{\times} - \phi\right) K_{\psi,\phi}\left(1\right)$ has zero expectation, then DMS is unbiased but not IMS so the contribution of the IMS bias to the MSFE is of order c.

3.4 Predictive Regressions

The results that were derived in the multi-step autoregression can be used to obtain the distributions of the estimators in the predictive regression of z_t on y_{t-h}. Define the bivariate Brownian motion $\mathbf{H}\left(r\right)$ such that $T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\left(\varepsilon_t, \epsilon_t\right)' \Rightarrow \mathbf{H}\left(r\right) = \left(H\left(r\right), \sigma W\left(r\right)\right)'$ where we write $H = \varsigma U + \sigma\delta W$. In Expression (6), $\alpha_1 = \left(1-\rho\right)\alpha + \beta\tau$ hence since $\left(1-\rho_T\right) = O\left(T^{-1}\right)$, only τ needs to be considered local asymptotic. To match the results from Proposition 1, we let $\mathbf{G}_{\phi}\left(r\right) = \int_0^r e^{\phi(r-s)}d\mathbf{H}\left(s\right) = \left(G_{\phi}, J_{\phi}\right)$. A proposition follows.

Proposition 7 *Let $\{z_t, y_t\}$ generated by (5), where ϵ_t and ε_t satisfy Condition* P *and with local asymptotic parameters (7) and (8). Then the following holds, as $T \to \infty$, if $\beta \neq 0$*
*under **short horizon**, for $h \in [1, T)$ constant, there exist*[6] $\varpi_h \in \mathbb{R}$ *such that*
(a_h) $T^{-1/2}\sum_{t=h}^{T}\omega_{h,t} \Rightarrow H\left(1\right) + \sigma\left(h-1\right)\beta W\left(r\right)$
(b_h) $T^{-1}\sum_{t=h}^{T} y_{t-h}\omega_{h,t} \Rightarrow \int_0^1 K_{\psi,\phi} d\left(H + \sigma\beta\left(h-1\right)W\right) + \varpi_h$
*under **long horizon**, for $h/T \to c \in (0,1)$,*

[6]The definition of ϖ_h is

$$\varpi_h = \frac{1}{2\beta}\Big(\left[\left(h-1\right)\xi_{\varepsilon} + \xi_{\varepsilon}\left(0\right)\right] + \rho_T^{2(h-1)}\left[\left(h-1\right)\xi_{\varepsilon} - \xi_{\varepsilon}\left(0\right)\right]\Big) - \frac{1}{\beta}\rho_T^{h-1}\left[\left(h-1\right)\xi_{\varepsilon} - \xi_{\varepsilon}\left(h\right)\right] - \frac{1}{\beta}\xi_{\varepsilon}\left(h-1\right) \tag{17}$$
$$+ \sum_{i=1}^{\infty}\xi_{\varepsilon,\epsilon}\left(-i\right) - \sum_{i=1}^{h-1}\xi_{\varepsilon,\epsilon}\left(1-i\right) + \left[\sum_{j=1}^{h-1}\rho_T^{j-1}\Big[\left(h-1\right)\xi_{\varepsilon,\epsilon} - \xi_{\varepsilon,\epsilon}\left(j\right) - \rho_T^{h-1}\left[\left(h-1\right)\xi_{\varepsilon,\epsilon} - \xi_{\varepsilon,\epsilon}\left(j-\left(h-1\right)\right)\right]\Big]\right]$$
$$+ \frac{\beta}{2}\sum_{j=1}^{h-1}\sum_{k=1}^{h-1}\rho_T^{j+k-2}\left[\left(h-1\right)\xi_{\epsilon} - \xi_{\epsilon}\left(k-j\right)\right]$$

(a_c) $T^{-3/2}\sum_{t=h}^{T}\omega_{h,t}\Rightarrow\beta\int_c^1\delta_\phi^c J_\phi ds$
(b_c) $T^{-2}\sum_{t=h}^{T}y_{t-h}\omega_{h,t}\Rightarrow\beta\int_c^1 K_{\psi,\phi}(r-c)\,\delta_\phi^c J_\phi(r)\,dr-\frac{1}{2}c\beta\psi^2e^{-\phi c}\left[f_\phi(c)\right]^2$

Corollary 8 *Under the assumptions of Proposition 7, if $\beta=0$ the results simplify to*
(a_h^*) $T^{-1/2}\sum_{t=h}^{T}\omega_{h,t}\Rightarrow H(1)$, (b_h^*) $T^{-1}\sum_{t=h}^{T}y_{t-h}\omega_{h,t}\Rightarrow\int_0^1 K_{\psi,\phi}dH+\sum_{i=h}^{\infty}\xi_{\varepsilon,\epsilon}(i)$
(a_c^*) $T^{-1/2}\sum_{t=h}^{T}\omega_{h,t}\Rightarrow H(1)-H_1(c)$, (b_c^*) $T^{-1}\sum_{t=h}^{T}y_{t-h}\omega_{h,t}\Rightarrow\int_c^1 K_{\psi,\phi}(s-c)\,dH(s)$

Elements (b_h^*) and (b_c^*) show that only long horizons are robust to the cross-correlation $\sum_{i=h}^{\infty}\xi_{\varepsilon,\epsilon}(i)$. Proposition 7 shows that the results that were derived for multi-step forecasting can be used for the analysis of the predictive regression. In particular, the scaled empirical moments converge to distributions that are very close to those of DMS. They share the similar properties that when $h=\lfloor cT\rfloor$ misspecification of the regression errors has a negligible impact. By contrast, if a modeler had attempted to forecast using a one-step predictive regression, she would have been subject to errors comparable to those found in IMS forecasting.

Indeed consider $\widetilde{\beta}_h$ the OLS estimator of β_h, in the regression $z_t=\alpha_h+\beta_h y_{t-h}+\omega_{h,t}$. Let $\xi_{\varepsilon,\epsilon}^{(h)}=\sum_{i=h-1}^{\infty}\xi_{\varepsilon,\epsilon}(i)$, then a straightforward application of Proposition 7 yields the following proposition.

Proposition 9 *Under the assumptions of Proposition 7, the following holds:*
first if $\beta\neq 0$,

$$T\left(\widehat{\beta}-\beta\right)\Rightarrow\left(\int_0^1\left(K_{\varphi,\phi}^{\mu}\right)^2dr\right)^{-1}\left(\int_0^1 K_{\varphi,\phi}^{\mu}d\left(\varsigma U+\sigma\delta W\right)+\varpi_1\right)\overset{def}{=}\lambda_\beta$$

$$T\left(\widetilde{\beta}_h-\beta\right)\Rightarrow\left(\int_0^1\left(K_{\varphi,\phi}^{\mu}\right)^2dr\right)^{-1}\left[\int_0^1 K_{\varphi,\phi}^{\mu}d\left(\varsigma U+\sigma\left[\delta+\beta\left(h-1\right)\right]W\right)+\varpi_h\right]\overset{def}{=}\lambda_{\beta,\{h\}}$$

$$\widetilde{\beta}_{\lfloor cT\rfloor}-\beta e^{\phi c}\Rightarrow\beta\left(\int_0^{1-c}\left(K_{\varphi,\phi}^{\mu}\right)^2dr\right)^{-1}\left[\int_c^1 K_{\varphi,\phi}^{\mu}(r-c)\,\delta_\phi^c J_\phi(r)\,dr-\frac{1}{2}c\psi^2e^{-\phi c}\left[f_\phi(c)\right]^2\right]\overset{def}{=}\lambda_{\beta,c}$$

and if $\beta=0$

$$\left(T\widehat{\beta},T\widetilde{\beta}_h,T\widetilde{\beta}_{\lfloor cT\rfloor}\right)\Rightarrow\left(\frac{\int_0^1 K_{\varphi,\phi}^{\mu}dH+\xi_{\varepsilon,\epsilon}^{(i)}}{\int_0^1\left(K_{\varphi,\phi}^{\mu}\right)^2dr},\frac{\int_0^1 K_{\varphi,\phi}^{\mu}dH+\xi_{\varepsilon,\epsilon}^{(h)}}{\int_0^1\left(K_{\varphi,\phi}^{\mu}\right)^2dr},\frac{\int_c^1 K_{\psi,\phi}^{\mu}(s-c)dH(s)}{\int_0^{1-c}\left(K_{\varphi,\phi}^{\mu}\right)^2dr}\right)$$

Corollary 10 *Under the assumptions of Proposition 9, at* ***intermediate horizons****,*
$\lambda_{\beta,c}\underset{c\to 0}{\overset{L}{=}}\sqrt{c}\frac{\sigma\beta\int_0^1 K_{\varphi,\phi}^{\mu}dW}{\int_0^1\left(K_{\varphi,\phi}^{\mu}\right)^2dr}$ *if $\beta\neq 0$, and* $\lambda_{0,c}\underset{c\to 0}{\overset{L}{=}}\frac{\int_0^1 K_{\varphi,\phi}^{\mu}dH}{\int_0^1\left(K_{\varphi,\phi}^{\mu}\right)^2dr}$ *otherwise.*

Corollary 11 *Consider the regression of $h^{-1}\sum_{k=1}^{h}z_{t+k}$ on a constant and x_t, the estimator of the coefficient of x_t admits the following distribution:*
if h is fixed, $T\left(\widehat{\beta}_{\{h\}}-\beta\right)\Rightarrow h^{-1}\sum_{k=1}^{h}\lambda_{\beta,\{k\}}$;
if $h/T\to c$, $\widetilde{\beta}_c-\beta f_\phi(c)\Rightarrow\int_0^c\lambda_{\beta,s}ds$.

Proposition 9 shows that intermediate and long horizon predictive regressions are robust to dynamic misspecification (yet not to contemporary correlation of the errors). As $c\to 0$, the behavior of $c^{-1/2}\lambda_{\beta,c}$ is close to that of λ_β provided that all Ξ_k, are diagonal. The main difference is that the former involves the stochastic integral of $K^{\mu}_{\varphi,\phi}$ with respect to increments in W whereas that of λ_β involves the increments of H. Hence, when $\beta \neq 0$, $\lambda_{\beta,c}$ is immune at intermediate horizons to the long run endogeneity and serial correlation of the errors in the predictive regression.

4 Monte Carlo

In order to illustrate the theoretical results presented above, we perform some simple simulations. We first compare the distributions of IMS and DMS forecast error under dynamic specification to those under correct specification. For this, we simulate an ARMA$(1,1)$ data generating process (DGP) $y_t = \tau + \rho y_{t-1} + \epsilon_t + \theta\epsilon_{t-1}$ where $\epsilon_t \overset{i.i.d}{\sim} N(0,1)$ as well as an AR(1) with the same long run variance $y_t = \tau + \rho y_{t-1} + (1+\theta)\,\epsilon_t$. Parameters vary as follows: and $\rho \in \{\pm.99, \pm.95, \pm.6, 0\}$, $\theta \in \{\pm.9, \pm.4\}$ and $T \in \{100, 250\}$ (with an initialization of 200 observations) with h ranging from 1 to $\lfloor T/3 \rfloor$. For each DGP, we compute 5,000 replications of the IMS and DMS forecast errors based on an AR(1) model. We report the p-values of a Kolmogorov-Smirnov test for the null of equal distributions of the forecast errors under the ARMA$(1,1)$ and AR(1) DGP. Non-rejection of the null is interpreted as evidence that for the DGP and horizon considered, the forecasting method is robust to the dynamic misspecification considered.

Figures 1 and 2 report the p-values of the Kolmogorov-Smirov test as a function of the horizon and for, respectively, $T = 100$ and 250 observations. The simulations all confirm that the p-values reject equal distributions of the forecast errors (and hence robustness to dynamic misspecification) at very low horizons in the presence of severe misspecification (large $|\theta|$). Yet the p-values increase rapidly with h when ρ is positive. This is especially true of DMS; this is less so for IMS: for instance, when $\rho = .99$ and $\theta = -.4$, the test rejects at the 10% level for $h \leq 0.15 \times T$.

The figures also report cases where $\rho < 0$ and we see that the forecasts then tend to be less robust, in particular when $\theta > 0$ and ρ is close to -1.

To assess how the results on multistep forecasting carry over to predictive regressions, we simulate Model (5) where $\alpha = \tau = 0$ and $\beta = 1$. Under dynamic misspecification ε_t follows an MA(1) process with parameter θ and standard Gaussian white noise innovations, whereas under correct specification $\varepsilon_t \overset{i.i.d}{\sim} N\left(0, (1+\theta)^2\right)$. We consider both the case of $\mathsf{Corr}(\varepsilon_t, \epsilon_t) = 0$ (no endogeneity) and of $\mathsf{Corr}(\varepsilon_t, \epsilon_t) = 1/\sqrt{2} \approx .7$ (endogenous case). We let $\epsilon_t \overset{i.i.d}{\sim} N(0,1)$ and consider various values of ρ and h as before. We only record the case of $T = 250$. Results are reported in Figures 3 and 4, respectively for the exogenous and endogenous situations: the graphs present the p-values of the Kolmogorov-Smirnov test for the null that the standardized $\widehat{\beta}$ (i.e. divided by their estimated standard error, without autocorrelation correction) have identical distributions for $\varepsilon_t \sim MA(1)$ or $\varepsilon_t \sim iid$. The figures report patterns similar to those observed under multi-step forecasting.

Finally, we assess the implications of the results above for the test of the null $\mathsf{H}_0 : \beta = 0$ at the 10% significance level in the predictive regression model with $\rho = 0.99$. For this, we consider the simple situation where critical values of the test statistic is obtained by parametric bootstrap over

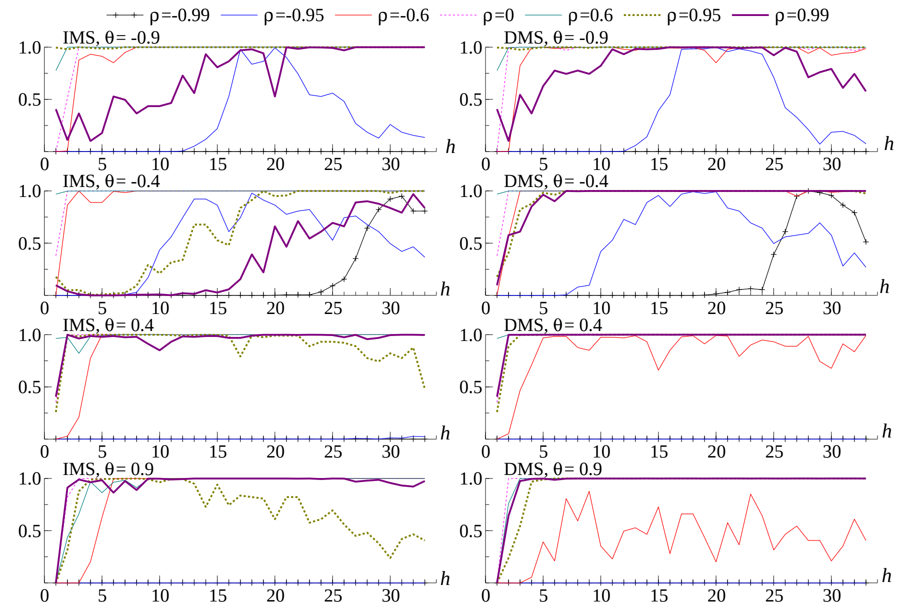

Figure 1: The figure reports p-values of the Kolmogorov-Smirnov test that the distributions of forecast errors (IMS, left and DMS, right) are the same in the models with misspeficied and correctly specified error dynamics. The horizontal axis is the horizon h. The sample size is $T = 100$ observations.

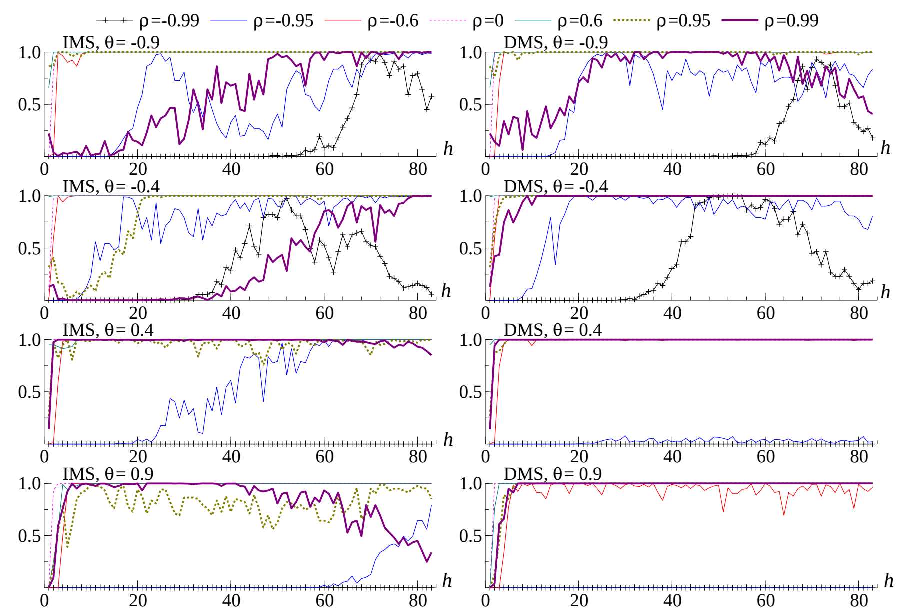

Figure 2: The figure reports p-values of the Kolmogorov-Smirnov test that the distributions of forecast errors (IMS, left and DMS, right) are the same in the models with misspeficied and correctly specified error dynamics. The horizontal axis is the horizon h. The sample size is $T = 250$ observations.

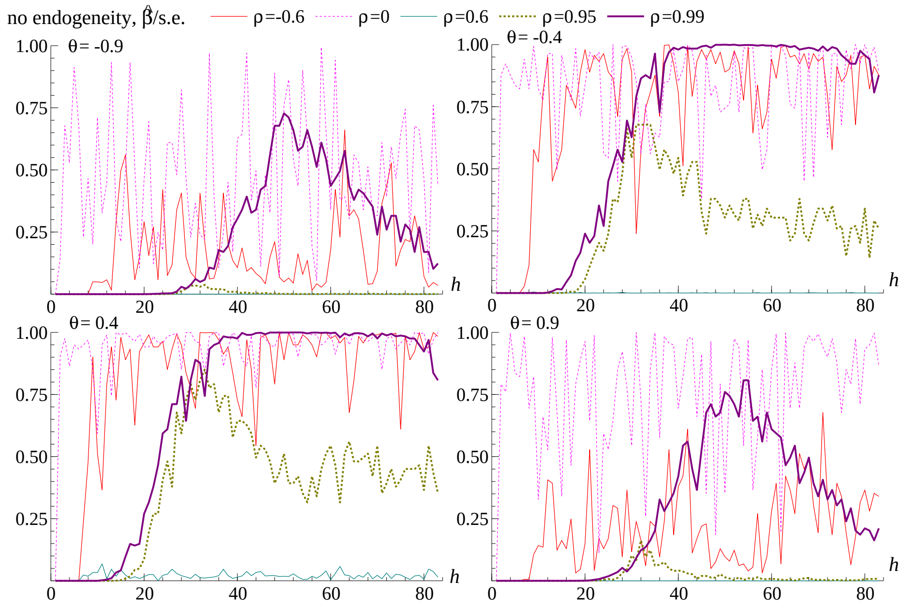

Figure 3: The figure reports p-values of the Kolmogorov-Smirnov test that the distributions of standardized $\widehat{\beta}$ are the same in the predictive regression models with misspecified and correctly specified error dynamics (without long run endogeneity). The horizontal axis is the horizon h. The sample size is $T = 250$ observations.

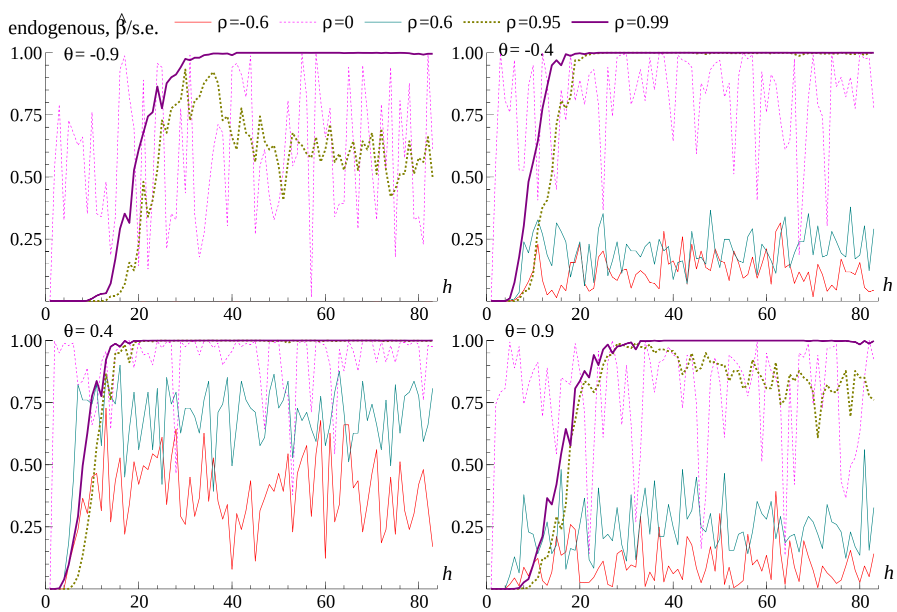

Figure 4: The figure reports p-values of the Kolmogorov-Smirnov test that the distributions of standardized $\widehat{\beta}$ are the same in the predictive regression models with mispeficied and correctly specified error dynamics (with long run endogeneity). The horizontal axis is the horizon h. The sample size is $T = 250$ observations.

a sample of $T = 250$ observations under the assumption that ε_t is iid and normal.[7]

Figures 5-8 report the rejection probabilities of four statistics: t_0 is obtained as as simple t-test where the DGP shows no serial correlation in the errors, $t_{0,\ HAC}$ is computed with a New-West HAC correction (in a DGP with no serial correlation), t is the statistic where ε_t follows an MA(1) with parameter θ (where θ =-0.4 in Figures 5 and 6, θ =-0.9 in Figures 7 and 8) and t_{HAC} is the statistic with Newey-West HAC correction where $\varepsilon_t \sim MA(1)$. In all DGPs considered the long run variance of ε_t is $(1+\theta)^2$, and we consider both the exogenous case ($\mathsf{Corr}(\varepsilon_t, \epsilon_t) = 0$ in Figures 5 and 7) and the presence of endogeneity ($\mathsf{Corr}(\varepsilon_t, \epsilon_t) = 1/\sqrt{2}$ in Figures 6 and 8).

The figures show that misspecifying the dynamics of ε_t yields very low local power for the standard t statistic close to the null $\beta = 0$ at all horizons when $\text{Corr}(\varepsilon_t, \epsilon_t) = 0$. In the exogenous case, t_{HAC} is slightly undersized but shows better local power than t. As the horizon h grows though, HAC corrections lower the power, whether or not ε_t is serially correlated. By contrast, standard t test do not suffer from this upper limit and the power tends to unity as $|\beta|$ gets larger. Hence a combined test that rejects if either t or t_{HAC} rejects will yield better local and global power at all horizons. When θ =-0.9 so the degree of misspecification is large, the local power remains low though.

Similar results hold for the endogenous case where $\mathsf{Corr}(\varepsilon_t, \epsilon_t) = 1/\sqrt{2}$. The main difference is that both t and t_{HAC} are locally biased and skewed at low h. Both are unreliable here when $h = 1$ (t_{HAC} become very liberal).

Overall, our simulations show that the robustness of long horizon projections to dynamic misspecification advocates the use of the non HAC corrected statistic. To ensure better power, this statistic should be combined with its HAC version which the empirical literature has usually considered: the combined test rejects occurs if either statistic does.

5 Conclusions

In this paper, we have studied the properties of iterated and direct multi-step forecasts in the presence of model misspecification and non-stationarity (both stochastic and deterministic trends). We have shown that in this framework, most general random walk estimation results apply when standard Brownian motions are replaced with trending Ornstein-Uhlenbeck processes. This allowed us to characterize the non-linear patterns exhibited by both estimators and forecasts. In particular, by letting the forecast horizon h grow with the sample size, we were able to show how much IMS and DMS differ in terms of long range forecasting. A Monte Carlo simulation illustrated the analytical results that were derived from the weak trend framework. Namely, that DMS is exhibits robustness to dynamic misspecification at intermediate horizons, and, these can be possibly very short in finite samples.

[7] This critical value is unobtainable in practice since we compute it under a known ρ so this constitutes an unfeasible bound where we do not need to resort to the corrections considered in the literature, e.g., Bonferroni as in Rossi (2005) or the IVX of Phillips and Magdalinos (2009) and Kostakis et al. (2015)

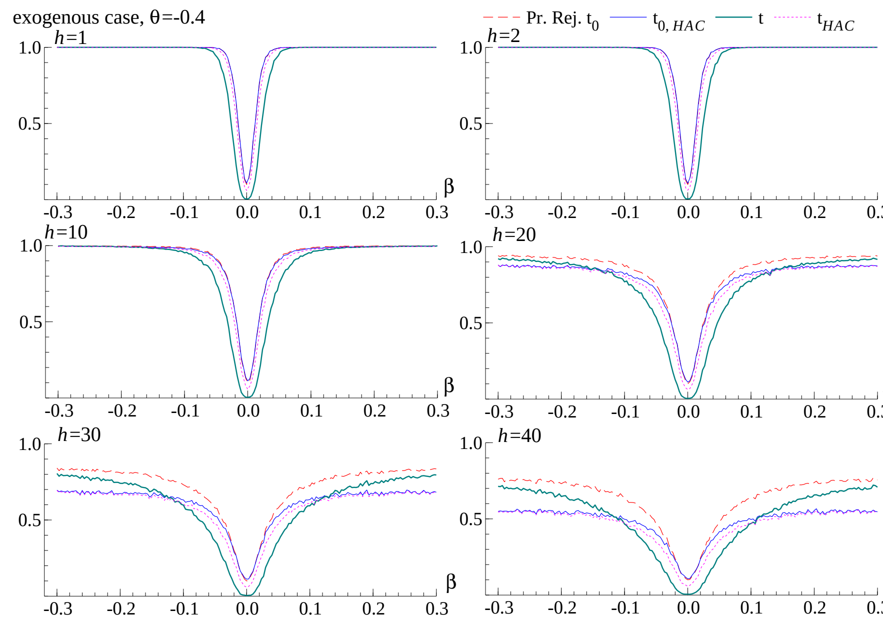

Figure 5: The figure reports the rejection probabilities of four test statistics ($t_0, t_{0,HAC}, t$ and t_{HAC}) for the null that $\beta = 0$. The sample size is $T = 250$ observations, $\theta = -0.4$ and there is no endogeneity.

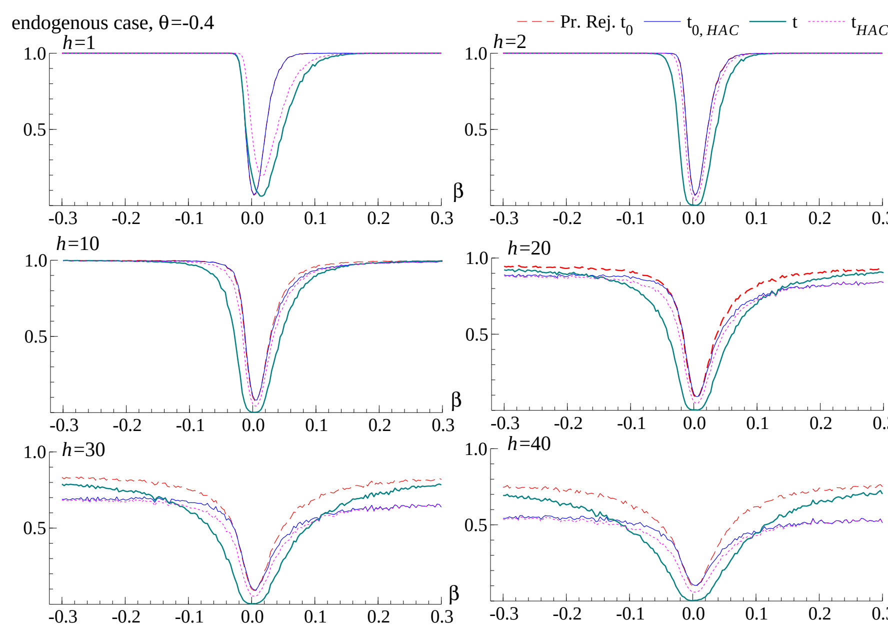

Figure 6: The figure reports the rejection probabilities of four test statistics ($t_0, t_{0,HAC}, t$ and t_{HAC}) for the null that $\beta = 0$. The sample size is $T = 250$ observations, $\theta = -0.4$ and there is contemporaneous endogeneity: $\mathsf{Corr}(\varepsilon_t, \epsilon_t) = 1/\sqrt{2}$.

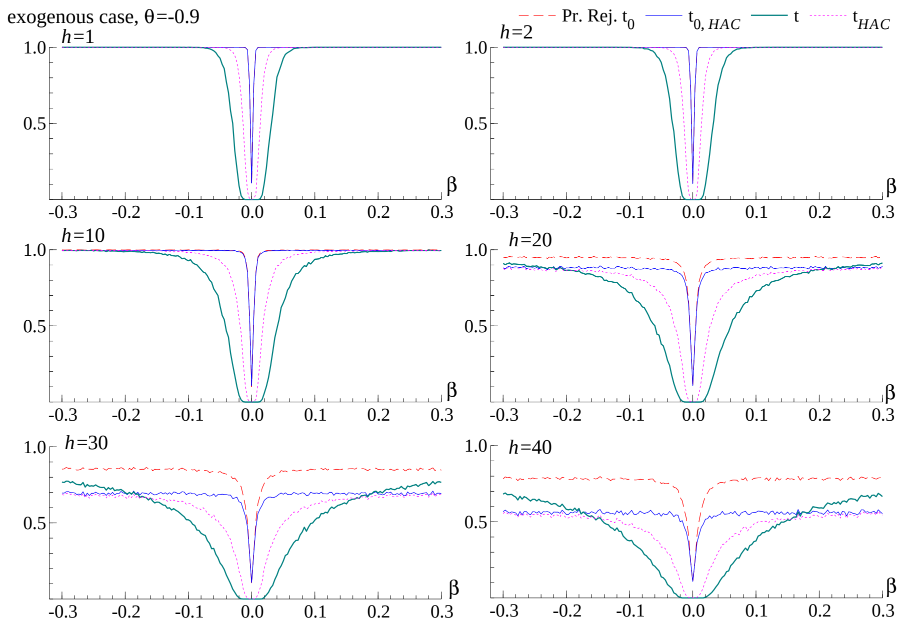

Figure 7: The figure reports the rejection probabilities of four test statistics ($t_0, t_{0,HAC}, t$ and t_{HAC}) for the null that $\beta = 0$. The sample size is $T = 250$ observations, $\theta = -0.9$ and there is no endogeneity.

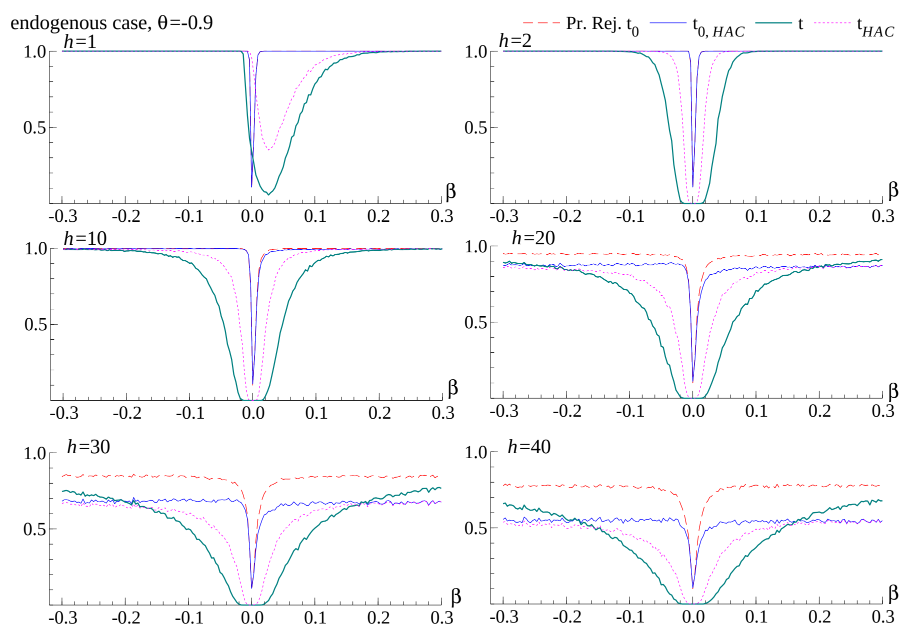

Figure 8: The figure reports the rejection probabilities of four test statistics ($t_0, t_{0,HAC}, t$ and t_{HAC}) for the null that $\beta = 0$. The sample size is $T = 250$ observations, $\theta = -0.9$ and there is contemporaneous endogeneity: $\mathsf{Corr}(\varepsilon_t, \epsilon_t) = 1/\sqrt{2}$.

The recommendations that we were able to derive are as follow. A forecaster who is confident that her model is well-specified ought to use iterated multi-step forecasts when the horizon is small compared to the sample size. If she must obtain long horizon forecasts using the available data, she should then resort to DMS. By contrast, should she suspect that her model might be misspecified, then DMS ought to be used at all horizons.

The Direct Multi-Step Forecasting framework has also been show to be useful for the analysis of predictive regressions as found in the literature. It follows that long-horizon regressions can be understood to work well when the model is misspecified for the serial correlation of the regression errors. Using simple simulations, we were able to show that, at intermediate or long horizons, a combination of the HAC test often considered in the empirical literature with the non-HAC version of the statistic achieves better global power than either separately. The literature has also considered alternative test, (optimal under Gaussianity in the case of Jansson and Moreira, 2006) or, e.g., Campbell and Yogo, 2005) or finite sample distributional adjustments (McCloskey, 2012). Although we do not explicitly study them here, our theoretical analysis seem to indicate that similar results are likely to hold.

References

Andrews, D. W. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric theory 4*(3), 458–467.

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.

Boudoukh, J., M. Richardson, and R. F. Whitelaw (2008). The myth of long-horizon predictability. *The Review of Financial Studies 21*(4), 1577–1605.

Busetti, F. and A. Harvey (2008). Testing for trend. *Econometric Theory 24*(1), 72–87.

Campbell, J. and R. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies 1*, 195–228.

Campbell, J. Y. and M. Yogo (2006). Efficient tests of stock return predictability. *Journal of Financial Economics 81*(1), 27–60.

Carriero, A., T. E. Clark, and M. Marcellino (2015). Bayesian vars: specification choices and forecast accuracy. *Journal of Applied Econometrics 30*(1), 46–73.

Chevillon, G. (2016). Multistep forecasting in the presence of location shifts. *International Journal of Forecasting 32*, 121–37.

Chevillon, G. and D. F. Hendry (2005). Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting 21*, 201–18.

Clements, M. P. and D. F. Hendry (1996). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics 58*, 657–683.

Deng, A. (2013). Understanding spurious regression in financial economics. *Journal of Financial Econometrics 12*(1), 122–150.

Elliott, G. (2006). Forecasting with trending data. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Chapter 11, pp. 555–604. North-Holland.

Fama, E. and K. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics 22*, 3–25.

Fama, E. F. and G. W. Schwert (1977). Asset returns and inflation. *Journal of financial economics 5*(2), 115–146.

Ferson, W. E., S. Sarkissian, and T. T. Simin (2003). Spurious regressions in financial economics? *Journal of Finance 58*(4), 1393–1413.

Haldrup, N. and S. Hylleberg (1995). A note on the distribution of the least squares estimator of a random walk with drift: Some analytical evidence. *Economics Letters 48*, 221–8.

Hendry, D. F. and A. B. Martinez (2017). Evaluating multi-step system forecasts with relatively few forecast-error observations. *International Journal of Forecasting 33*(2), 359–72.

Hjalmarsson, E. (2011). New methods for inference in long-horizon regressions. *Journal of Financial and Quantitative Analysis 46*(3), 815–839.

Ing, C.-K. (2003). Multistep prediction in autoregressive processes. *Econometric Theory 19*, 254–279.

Ing, C.-K. (2004). Selecting optimal multistep predictors for autoregressive process of unknown order. *The Annals of Statistics 32*, 693–722.

Jansson, M. and M. J. Moreira (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica 74*(3), 681–714.

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.

Lanne, M. (2002). Testing the predictability of stock returns. *Review of Economics and Statistics 84*, 407–15.

Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance 56*(3), 815–849.

Lettau, M. and S. Van Nieuwerburgh (2007). Reconciling the return predictability evidence: The review of financial studies: Reconciling the return predictability evidence. *The Review of Financial Studies 21*(4), 1607–1652.

Marcellino, M., J. Stock, and M. Watson (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of Econometrics 135*, 499–526.

McCloskey, A. (2012). Bonferroni-based size-correction for nonstandard testing problems. Technical report, Brown University.

McElroy, T. and M. W. McCracken (2017). Multistep ahead forecasting of vector time series. *Econometric Reviews 36*(5), 495–513.

Pástor, L. and R. F. Stambaugh (2009). Predictive systems: Living with imperfect predictors. *The Journal of Finance 64*(4), 1583–1628.

Phillips, P. C. (2015). Halbert white jr. memorial jfec lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics 13*(3), 521–555.

Phillips, P. C. and J. H. Lee (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics 177*(2), 250–264.

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

Phillips, P. C. B. (1998). Impulse response and forecast error variance asymptotics in nonstationary VARs. *Journal of Econometrics 83*, 21–56.

Rossi, B. (2005a, October). Optimal Tests For Nested Model Selection With Underlying Parameter Instability. *Econometric Theory 21*(05), 962–990.

Rossi, B. (2005b). Testing long-horizon predictive ability with high persistence, and the meese-rogoff puzzle. *International Economic Review 46*(1), 61–92.

Rozeff, M. S. (1984). Dividend yields are equity risk premiums. *The Journal of Portfolio Management 11*(1), 68–75.

Schorfheide, F. (2005). VAR forecasting under local misspecification. *Journal of Econometrics 128*(1), 99–136.

Stambaugh, R. (1999). Predictive regressions. *Journal of Financial Economics 54*, 375–421.

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

Stock, J. H. and M. W. Watson (1996). Confidence sets in regressions with highly serially correlated regressors. mimeo, Princeton University.

Torous, W., R. Valkanov, and S. Yan (2004). On predicting stock returns with nearly integrated explanatory variables. *Journal of Business 77*, 937–66.

Turner, J. L. (2004). Local to unity, long-horizon forecasting thresholds for model selection in the AR(1). *Journal of Forecasting 23*, 513–539.

Valkanov, R. (2003). Long-horizon regressions: theoretical results and aplications. *Journal of Financial Economics 68*, 201–32.

Vogelsang, T. (1998). Trend function hypothesis testing in the presence of serial correlation. *Econometrica 66*, 123–148.

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

Appendices

A Proof of Proposition 1

A.1 Short horizon

For $r \in [0,1]$, we write the series as the sum of a moving average and a deterministic component:

$$T^{-1/2} y_{\lfloor Tr \rfloor} = T^{-1/2} e^{\phi \lfloor Tr \rfloor / T} y_0 + \left[\sum\nolimits_{i=0}^{\lfloor Tr \rfloor - 1} e^{i\phi/T} \right] \psi T^{-1} + T^{-1/2} \sum\nolimits_{i=0}^{\lfloor Tr \rfloor - 1} e^{i\phi/T} \epsilon_{\lfloor Tr \rfloor - i}.$$

Hence,

$$T^{-1/2} y_{\lfloor Tr \rfloor} = e^{\phi \lfloor Tr \rfloor / T} \frac{y_0}{\sqrt{T}} + \psi f_\phi(r) \left(1 + O\left(T^{-1}\right)\right) + T^{-1/2} \sum\nolimits_{i=1}^{\lfloor Tr \rfloor} e^{(\lfloor Tr \rfloor - i)\phi/T} \epsilon_i,$$

where $T^{-1/2} \sum_{i=1}^{\lfloor Tr \rfloor} e^{(\lfloor Tr \rfloor - i)\phi/T} \epsilon_i \Rightarrow J_\phi(r)$ Phillips (1987). Proof of (a_h) follows. Now, we write the statistic (b_h) as a functional on $D[0,1]$. We first square $T^{-1/2} y_t$:

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

Hence

$$\begin{aligned} 2T^{-1} e^{h\phi/T} y_{t-h,T} w_{h,t}. = {} & T^{-1} \left[y_{t,T}^2 - e^{2h\phi/T} y_{t-h,T}^2 \right] - T^{-2} \left(\frac{1 - e^{h\phi/T}}{1 - e^{\phi/T}} \right)^2 \psi^2 \\ & - T^{-1} w_{h,t}^2 - 2T^{-3/2} \left(e^{h\phi/T} \frac{1 - e^{h\phi/T}}{1 - e^{\phi/T}} \right) \psi y_{t-h,T} - 2T^{-3/2} \psi w_{h,t}. \end{aligned}$$

We notice that, summing over t,

$$T^{-1} \sum_{t=h}^{T} \left(y_{t,T}^2 - e^{2h\phi/T} y_{t-h,T}^2 \right) = T^{-1} \sum_{t=T-h+1}^{T} y_{t,T}^2 - T^{-1} \sum_{t=0}^{h-1} y_{t,T}^2 - \left(e^{2h\phi/T} - 1 \right) T^{-1} \sum_{t=h}^{T} y_{t-h,T}^2$$

$$\begin{aligned} T^{-1} \sum_{t=h}^{T} \left(y_{t,T}^2 - e^{2h\phi/T} y_{t-h,T}^2 \right) = {} & \left(T^{-1} \sum_{t=T-h+1}^{T} y_{t,T}^2 - 2\phi h T^{-2} \sum_{t=h}^{T} y_{t-h,T}^2 - T^{-1} \sum_{t=h}^{2h-1} y_{t-h,T}^2 \right) \\ & + \frac{1}{T^2} \left(\sum_{i=0}^{\infty} \frac{(2h\phi)^{i+2}}{(i+2)! T^i} \right) T^{-1} \sum_{t=h}^{T} y_{t-h,T}^2, \end{aligned}$$

hence, as $T \to \infty$

$$T^{-1} \sum_{t=h}^{T} \left(y_{t,T}^2 - e^{2h\phi/T} y_{t-h,T}^2 \right) \Rightarrow h \left\{ K_{\psi,\phi}(1) \right\}^2 - 2\phi \int_0^1 \left[K_{\psi,\phi}(r) \right]^2 dr$$

Collecting the elements we find:

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

Now, using Itô's lemma:

$$dK_{\psi,\phi}^{2}(r) = \left(2K_{\psi,\phi}(r)\left[\psi + \phi K_{\psi,\phi}(r)\right] + \sigma^{2}\right) dr + 2\sigma K_{\psi,\phi}(r)\, dW(r)\,, \tag{A.2}$$

$$\left\{K_{\psi,\phi}(1)\right\}^{2} = \sigma^{2} + 2\int_{0}^{1}\left[\psi K_{\psi,\phi}(r) + \phi K_{\psi,\phi}^{2}(r)\right] dr + 2\sigma\int_{0}^{1} K_{\psi,\phi}(r)\, dW(r)\,. \tag{A.3}$$

whence the result, using (A.1) and the definition of $\sigma_{w_h}^2$.

A.2 Long Horizon

Preliminary results Item (a_c) is clear using the Functional Central Limit theorem (FCLT) and the Continuous Mapping theorem (CMT) respectively. As regard (b_c), we first derive the asymptotic distribution of sample moments of the multi-step residuals $w_{h,t}$ (this constitutes the proof of (c_c)). They follow an MA$(h-1)$

$$w_{h,t} = \sum_{i=0}^{h-1} \rho_T^i \epsilon_{t-i} = \sum_{j=t-h+1}^{t} e^{(t-j)\phi/T}\epsilon_j,$$

which, using $U_T(r) = T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\epsilon_t$, can be rewritten so that we let appear a stochastic integral:

$$\begin{aligned} T^{-1/2} w_{\lfloor cT\rfloor,\lfloor rT\rfloor} &= \sum_{j=\lfloor (r-c)T\rfloor+1}^{\lfloor rT\rfloor} e^{\phi(\lfloor rT\rfloor - j)/T}\int_{(j-1)/T}^{j/T} dU_T(s) \\ &= \sum_{j=\lfloor T(r-c)\rfloor+1}^{\lfloor Tr\rfloor}\int_{(j-1)/T}^{j/T} e^{\phi(\lfloor rT\rfloor/T - s)} dU_T(s) = \int_{r-c}^{r} e^{\phi(\lfloor rT\rfloor/T-s)} dU_T(s) \\ &\Rightarrow \sigma\int_{r-c}^{r} e^{\phi(r-s)} dW(s) = \sigma J_\phi(r) - \sigma e^{\phi c} J_\phi(r-c)\,. \end{aligned}$$

We recognize the quasi-difference of an Ornstein-Uhlenbeck process:

$$T^{-1/2} w_{\lfloor cT\rfloor,\lfloor rT\rfloor} \Rightarrow \delta_\phi^c J_\phi(r) = \delta_0^c J_\phi(r) - \phi f_\phi(c)\, J_\phi(r-c)\,. \tag{A.4}$$

Using the continuous mapping theorem, we obtain the limit distributions of empirical moments of $T^{-1/2}w_{h,\lfloor Tr\rfloor}$, first the sample mean: $T^{-3/2}\sum_{t=h}^{T} w_{h,t} = T^{-3/2}\sum_{j=\lfloor cT\rfloor}^{T} w_{\lfloor cT\rfloor,\lfloor Tr\rfloor}$

$$T^{-3/2}\sum_{j=\lfloor cT\rfloor}^{T} w_{\lfloor cT\rfloor,j} \Rightarrow \sigma\int_c^1 \delta_\phi^c J_\phi(r)\, dr, \tag{A.5}$$

and the sum of squares:

$$T^{-2}\sum_{j=\lfloor cT\rfloor}^{T} w_{\lfloor cT\rfloor,j}^2 \Rightarrow \sigma\int_c^1 \left[\delta_\phi^c J_\phi(r)\right]^2 dr. \tag{A.6}$$

A useful lemma

Lemma 12 *Using the definition of* δ_θ^c *and* $K_{\psi,\phi}(r)$ *from the main text, then for nonzero* ϕ

$$\begin{aligned}\int_c^1 \left[\delta_\theta^c K_{\psi,\phi}(r)\right]^2 dr &= \int_c^1 K_{\psi,\phi}^2(r)\,dr - \theta^{2c}\int_c^1 K_{\psi,\phi}^2(r-c)\,dr \\ &- \frac{\theta^c}{\phi}K_{\psi,\phi}(1-c)\,\delta_\theta^c K_{\psi,\phi}(1) - \theta^{2c}(1-c)\frac{\sigma^2}{\phi} \\ &+ \theta^c\frac{\psi}{\phi}\int_c^1 \left\{\delta_\theta^c K_{\psi,\phi}(r) + (1-\theta^c)\,K_{\psi,\phi}(r-c)\right\} dr \\ &+ \frac{\sigma\theta^c}{\phi}\left(\int_c^1 \delta_\theta^c K_{\psi,\phi}(r)\,dW(r-c) + \int_c^1 K_{\psi,\phi}(r-c)\,d\left[\delta_\theta^c W(r)\right]\right)\end{aligned}$$

Proof. Develop

$$\begin{aligned}\left[\delta_\theta^c K_{\psi,\phi}(r)\right]^2 &= K_{\psi,\phi}^2(r) + \theta^{2c}K_{\psi,\phi}^2(r-c) - 2\theta^c K_{\psi,\phi}(r)\,K_{\psi,\phi}(r-c) \\ &= K_{\psi,\phi}^2(r) - \theta^{2c}K_{\psi,\phi}^2(r-c) - 2\theta^c\left[\delta_\theta^c K_{\psi,\phi}(r)\right]K_{\psi,\phi}(r-c) \qquad \text{(A.7)}\end{aligned}$$

When taking the integral over $(c,1)$ with respect to r, we recognize the sum of $\int_c^1 \delta_\theta^{2c}K_{\psi,\phi}^2(r)\,dr$ and of $-2\theta\int_c^1\left[\delta_\theta^c K(r)\right]K(r-c)\,dr$. We analyze them in turn. First, Expression (A.2) implies that

$$\theta^{2c}dK_{\psi,\phi}^2(r-c) = \left(2\theta^{2c}K_{\psi,\phi}(r-c)\left[\psi+\phi K_{\psi,\phi}(r-c)\right] + \theta^{2c}\sigma^2\right)dr + 2\sigma\theta^{2c}K_{\psi,\phi}(r-c)\,dW(r-c) \qquad \text{(A.8)}$$

hence

$$\begin{aligned}&d\left[K_{\psi,\phi}^2(r) - \theta^{2c}K_{\psi,\phi}^2(r-c)\right] \\ &= \left(2\left[\psi\left\{K_{\psi,\phi}(r) - \theta^{2c}K_{\psi,\phi}(r-c)\right\} + \phi\left\{K_{\psi,\phi}^2(r) - \theta^{2c}K_{\psi,\phi}^2(r-c)\right\}\right]^2\right)dr \\ &+ \sigma\left(1-\theta^{2c}\right)dr + 2\sigma\left[K_{\psi,\phi}(r)\,dW(r) - \theta^{2c}K_{\psi,\phi}(r-c)\,dW(r-c)\right].\end{aligned}$$

Integrating over $(c,1)$

$$\begin{aligned}\int_c^1 d\left[K_{\psi,\phi}^2(r) - \theta^{2c}K_{\psi,\phi}^2(r-c)\right] &= \left(1-\theta^{2c}\right)(1-c)\,\sigma^2 + 2\psi\int_c^1\left\{K_{\psi,\phi}(r) - \theta^{2c}K_{\psi,\phi}(r-c)\right\}dr \\ &+ 2\phi\int_c^1\left\{K_{\psi,\phi}^2(r) - \theta^{2c}K_{\psi,\phi}^2(r-c)\right\}dr \\ &+ 2\sigma\left[\int_c^1 K_{\psi,\phi}(r)\,dW(r) - \theta^{2c}\int_c^1 K_{\psi,\phi}(r-c)\,dW(r-c)\right].\end{aligned}$$

Therefore

$$\begin{aligned}2\phi\int_c^1\left\{K_{\psi,\phi}^2(r) - \theta^{2c}K_{\psi,\phi}^2(r-c)\right\}dr &= 2\phi\int_c^1 \delta_\theta^{2c}K_{\psi,\phi}^2(r)\,dr \\ &= K_{\psi,\phi}^2(1) - \theta^{2c}K_{\psi,\phi}^2(1-c) - K_{\psi,\phi}^2(c) - \left(1-\theta^{2c}\right)(1-c)\,\sigma^2 \\ &- 2\psi\int_c^1\left\{K_{\psi,\phi}(r) - \theta^{2c}K_{\psi,\phi}(r-c)\right\}dr \\ &- 2\sigma\left[\int_c^1 K_{\psi,\phi}(r)\,dW(r) - \theta^{2c}\int_c^1 K_{\psi,\phi}(r-c)\,dW(r-c)\right].\end{aligned}$$

Now, for $\int_c^1 [\delta_\theta^c K_{\psi,\phi}(r)] K_{\psi,\phi}(r-c)\,dr$, using the formula for stochastic integration by parts, if $c \neq 0$:

$$d[K_{\psi,\phi}(r) K_{\psi,\phi}(r-c)] = [\psi\{K_{\psi,\phi}(r) + K_{\psi,\phi}(r-c)\} + 2\phi K_{\psi,\phi}(r) K_{\psi,\phi}(r-c)]\,dr \\ + \sigma K_{\psi,\phi}(r)\,dW(r-c) + \sigma K_{\psi,\phi}(r-c)\,dW(r),$$

since $dW(r)$ and $dW(r-c)$ are independent.
Combining (A.2) and the previous expression, the difference

$$d[K_{\psi,\phi}(r-c)\,\delta_\theta^c K_{\psi,\phi}(r)] = d[K_{\psi,\phi}(r) K_{\psi,\phi}(r-c)] - \theta^c dK_{\psi,\phi}^2(r-c)$$

is then equal to

$$\left\{\psi\{\delta_\theta^c K_{\psi,\phi}(r) + (1-\theta^c) K_{\psi,\phi}(r-c)\} + 2\phi K_{\psi,\phi}(r-c)[\delta_\theta^c K_{\psi,\phi}(r)] - \sigma^2\theta^c\right\}dr \\ + \sigma\{[\delta_\theta^c K_{\psi,\phi}(r)]\,dW(r-c) + K_{\psi,\phi}(r-c)\,d[\delta_\theta^c W(r)]\}.$$

We then re-express $2\phi K_{\psi,\phi}(r-c)[\delta_\theta^c K_{\psi,\phi}(r)]\,dr$ as

$$d[K_{\psi,\phi}(r) K_{\psi,\phi}(r-c)] - \theta^c dK_{\psi,\phi}^2(r-c) - \left\{\psi\{\delta_\theta^c K_{\psi,\phi}(r) + (1-\theta^c) K_{\psi,\phi}(r-c)\} - \sigma^2\theta^c\right\}dr \\ - \sigma\{[\delta_\theta^c K_{\psi,\phi}(r)]\,dW(r-c) + K_{\psi,\phi}(r-c)\,d\delta_\theta^c W(r)\}.$$

The expression for $2\phi\int_c^1 [\delta_\theta^c K_{\psi,\phi}(r)] K_{\psi,\phi}(r-c)\,dr$ is therefore

$$[K_{\psi,\phi}(1) K_{\psi,\phi}(1-c)] - \theta^c K_{\psi,\phi}^2(1-c) - \int_c^1 \psi\{\delta_\theta^c K_{\psi,\phi}(r) + (1-\theta^c) K_{\psi,\phi}(r-c)\}\,dr - \sigma^2\theta^c(1-c) \\ - \sigma\int_c^1 \{[\delta_\theta^c K_{\psi,\phi}(r)]\,dW(r-c) + K_{\psi,\phi}(r-c)\,d\delta_\theta^c W(r)\}$$

and the result follows using (A.7). ■

Proof of (b_c)**.** We can now move to finding the expression for (b_c). For nonzero ϕ, we square $y_{t,T}$ and express it as the sum $(\tau_{h,T} + \rho_{h,T} y_{t-h,T} + w_{h,t})^2$, or:

$$y_{t,T}^2 = \psi^2\left(\frac{1-e^{\phi h/T}}{1-e^{\phi/T}}\right)^2 T^{-1} + e^{2\phi h/T} y_{t-h,T}^2 + w_{h,t}^2 \\ + 2\psi\left(\frac{1-e^{\phi h/T}}{1-e^{\phi/T}}\right) T^{-1/2} e^{\phi h/T} y_{t-h,T} + 2e^{\phi h/T} y_{t-h,T} w_{h,t} + 2\psi\left(\frac{1-e^{\phi h/T}}{1-e^{\phi/T}}\right) T^{-1/2} w_{h,t}.$$

Summing over t ranging from h to T and rearranging yields

$$2e^{\phi h/T} T^{-2}\sum y_{t-h,T} w_{h,t} = T^{-2}\sum y_{t,T}^2 - e^{2\phi h/T} T^{-2}\sum y_{t-h,T}^2 \\ - \psi^2\left(T^{-1}\frac{1-e^{\phi h/T}}{1-e^{\phi/T}}\right)^2 - T^{-2}\sum w_{h,t}^2 \\ - 2\psi e^{\phi h/T}\left(T^{-1}\frac{1-e^{\phi h/T}}{1-e^{\phi/T}}\right) T^{-3/2}\sum y_{t-h,T} \\ - 2\psi\left(T^{-1}\frac{1-e^{\phi h/T}}{1-e^{\phi/T}}\right) T^{-3/2}\sum w_{h,t}. \tag{A.9}$$

Deterministic components admit the following limits as $T \to \infty$

$$T\left(1 - e^{\phi/T}\right) \to -\phi$$
$$T^{-1}\frac{1 - e^{\phi h/T}}{1 - e^{\phi/T}} \to f_{\phi}(c).$$

Using the continuous mapping theorem, and Slutsky's formula for weak convergence (since $e^{2\phi h/T} \to e^{2\phi c}$),

$$T^{-2}\sum y_{t,T}^{2} - e^{2\phi h/T}T^{-2}\sum y_{t-h,T}^{2} \Rightarrow \int_{c}^{1}\left[K_{\psi,\phi}^{2}(r)\, e^{2\phi c}K_{\psi,\phi}^{2}(r-c)\right]dr. \tag{A.10}$$

Combing (A.5), (A.6) and (A.10) in Expression (A.9), we obtain

$$\begin{aligned}2e^{c\phi}T^{-2}\sum y_{t-h,T}w_{h,t} \Rightarrow & \int_{c}^{1}\delta_{e^{\phi}}^{2c}K_{\psi,\phi}^{2}(r)\,dr \\ & -\psi^{2}f_{\phi}(c)^{2} - \int_{c}^{1}\left[\delta_{\phi}^{c}J_{\phi}(r)\right]^{2}dr \\ & -2\psi e^{\phi c}f_{\phi}(c)\int_{0}^{1-c}K_{\psi,\phi}(r)\,dr \\ & -2\psi f_{\phi}(c)\int_{c}^{1}\delta_{\phi}^{c}J_{\phi}(r)\,dr.\end{aligned} \tag{A.11}$$

Now, from lemma 12, we let for ease of notation

$$\begin{aligned}F &= \int_{c}^{1}\delta_{e^{\phi}}^{2c}K_{\psi,\phi}^{2}(r)\,dr - \int_{c}^{1}\left[\delta_{\phi}^{c}K_{\psi,\phi}(r)\right]^{2}dr \\ &= \frac{e^{\phi c}}{\phi}K_{\psi,\phi}(1-c)\,\delta_{\phi}^{c}K_{\psi,\phi}(1) + e^{2\phi c}(1-c)\frac{\sigma^{2}}{\phi} \\ &- e^{\phi c}\frac{\psi}{\phi}\int_{c}^{1}\left\{\delta_{\phi}^{c}K_{\psi,\phi}(r) + \left(1 - e^{\phi c}\right)K_{\psi,\phi}(r-c)\right\}dr \\ &- \frac{\sigma e^{\phi c}}{\phi}\left(\int_{c}^{1}\delta_{\phi}^{c}K_{\psi,\phi}(r)\,dW(r-c) + \int_{c}^{1}K_{\psi,\phi}(r-c)\,d\left[\delta_{\phi}^{c}W(r)\right]\right)\end{aligned}$$

so the right-hand side of Expression (A.11) is equal to

$$\begin{aligned}&\int_{c}^{1}\left[\delta_{\phi}^{c}K_{\psi,\phi}(r)\right]^{2}dr + F \\ &-\psi^{2}f_{\phi}(c)^{2} - \int_{c}^{1}\left[\delta_{\phi}^{c}J_{\phi}(r)\right]^{2}dr - 2\psi e^{\phi c}f_{\phi}(c)\int_{0}^{1-c}K_{\psi,\phi}(r)\,dr - 2\psi f_{\phi}(c)\int_{c}^{1}\delta_{\phi}^{c}J_{\phi}(r)\,dr.\end{aligned}$$

Recall that $K_{\psi,\phi}(r) = \psi f_{\phi}(r) + J_{\phi}(r)$ so that the previous expression is equal to

$$\begin{aligned}&\int_{c}^{1}\left[\psi\delta_{\phi}^{c}f_{\phi}(r)\right]^{2}dr + 2\psi\int_{c}^{1}\left[\delta_{\phi}^{c}f_{\phi}(r)\right]\left[\delta_{\phi}^{c}J_{\phi}(r)\right]dr + F \\ &-\psi^{2}f_{\phi}(c)^{2} - 2\psi e^{\phi c}f_{\phi}(c)\int_{0}^{1-c}K_{\psi,\phi}(r)\,dr - 2\psi f_{\phi}(c)\int_{c}^{1}\delta_{\phi}^{c}J_{\phi}(r)\,dr\end{aligned}$$

Notice that $\delta_{\phi}^{c} f_{\phi}(r) = f_{\phi}(r) - e^{\phi c} f_{\phi}(r-c) = f_{\phi}(c)$ for all r, hence the previous expression becomes

$$\begin{aligned}
&\psi^2 f_{\phi}(c)^2 (1-c) + F \\
&- \psi^2 f_{\phi}(c)^2 - 2\psi e^{\phi c} f_{\phi}(c) \int_0^{1-c} K_{\psi,\phi}(r)\, dr \\
&= -c\psi^2 f_{\phi}(c)^2 - 2\psi e^{\phi c} f_{\phi}(c) \int_0^{1-c} K_{\psi,\phi}(r)\, dr + F
\end{aligned}$$

Now we replace F with its expression and get:

$$\begin{aligned}
&- c\psi^2 f_{\phi}(c)^2 - 2\psi e^{\phi c} f_{\phi}(c) \int_0^{1-c} K_{\psi,\phi}(r)\, dr \\
&+ \frac{e^{\phi c}}{\phi} K_{\psi,\phi}(1-c)\, \delta_{\phi}^{c} K_{\psi,\phi}(1) + e^{\phi 2c}(1-c)\frac{\sigma^2}{\phi} \\
&- e^{\phi c}\frac{\psi}{\phi} \int_c^1 \left\{ \delta_{\phi}^{c} K_{\psi,\phi}(r) + \left(1 - e^{\phi c}\right) K_{\psi,\phi}(r-c) \right\} dr \\
&- \frac{\sigma e^{\phi c}}{\phi} \left(\int_c^1 \delta_{\phi}^{c} K_{\psi,\phi}(r)\, dW(r-c) + \int_c^1 K_{\psi,\phi}(r-c)\, d\left[\delta_{\phi}^{c} W(r)\right] \right)
\end{aligned}$$

We rewrite $\delta_{\phi}^{c} K_{\psi,\phi}(r) + \left(1 - e^{\phi c}\right) K_{\psi,\phi}(r-c)$ as $K_{\psi,\phi}(r) + \left(1 - 2e^{\phi c}\right) K_{\psi,\phi}(r-c)$ and the previous expression rewrites as

$$\begin{aligned}
&- c\psi^2 f_{\phi}(c)^2 - \left[2\psi e^{\phi c} f_{\phi}(c) - e^{\phi c}\frac{\psi}{\phi}\left(1 - 2e^{\phi c}\right) \right] \int_0^{1-c} K_{\psi,\phi}(r)\, dr \\
&+ \frac{e^{\phi c}}{\phi} K_{\psi,\phi}(1-c)\, \delta_{\phi}^{c} K_{\psi,\phi}(1) + e^{2\phi c}(1-c)\frac{\sigma^2}{\phi} \\
&- e^{\phi c}\frac{\psi}{\phi} \int_c^1 K_{\psi,\phi}(r)\, dr \\
&- \frac{\sigma e^{\phi c}}{\phi} \left(\int_c^1 \delta_{\phi}^{c} K_{\psi,\phi}(r)\, dW(r-c) + \int_c^1 K_{\psi,\phi}(r-c)\, d\left[\delta_{\phi}^{c} W(r)\right] \right)
\end{aligned}$$

and $2e^{\phi c} f_{\phi}(c) + \left(1 - 2e^{\phi c}\right)\lambda^c/\phi = -e^{\phi c}/\phi$ hence the result:

$$\begin{aligned}
T^{-2} \sum y_{t-h,T} w_{h,t} \Rightarrow & -\frac{\sigma^2}{2\phi}(1-c)\, e^{\phi c} - \frac{c}{2} e^{-\phi c} \psi^2 f_{\phi}(c)^2 \\
& + \frac{1}{2\phi} K_{\psi,\phi}(1-c)\, \delta_{\phi}^{c} K_{\psi,\phi}(1) \\
& + \frac{\psi}{2\phi} \left[\int_0^c K_{\psi,\phi}(r)\, dr - \int_{1-c}^1 K_{\psi,\phi}(r)\, dr \right] \\
& - \frac{\sigma}{2\phi} \left(\int_c^1 \delta_{\phi}^{c} K_{\psi,\phi}(r)\, dW(r-c) + \int_c^1 K_{\psi,\phi}(r-c)\, d\left[\delta_{\phi}^{c} W(r)\right] \right).
\end{aligned} \tag{A.12}$$

We notice that $e^{-\phi c} f_{\phi}^2(c) = \phi^{-1}\left(f_{\phi}(c) + f_{\phi}(-c)\right).$

We can simplify the result further using Expression (A.3:

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

$$\{K_{\psi,\phi}(1)\}^2 - \{K_{\psi,\phi}(c)\}^2 = (1-c)\,\sigma^2 + 2\int_c^1 \left[\psi K_{\psi,\phi}(r) + \phi K_{\psi,\phi}^2(r)\right] dr + 2\sigma \int_c^1 K_{\psi,\phi}(r)\, dW(r)$$

$$\{K_{\psi,\phi}(1-c)\}^2 = (1-c)\,\sigma^2 + 2\int_0^{1-c} \left[\psi K_{\psi,\phi}(r) + \phi K_{\psi,\phi}^2(r)\right] dr + 2\sigma \int_0^{1-c} K_{\psi,\phi}(r)\, dW(r)$$

and

$$\begin{aligned} d\left[K_{\psi,\phi}(r)\, K_{\psi,\phi}(r-c)\right] &= \left[\psi\left\{K_{\psi,\phi}(r) + K_{\psi,\phi}(r-c)\right\} + 2\phi K_{\psi,\phi}(r)\, K_{\psi,\phi}(r-c)\right] dr \\ &\quad + \sigma K_{\psi,\phi}(r)\, dW(r-c) + \sigma K_{\psi,\phi}(r-c)\, dW(r), \end{aligned}$$

which, when integrating over $[c, 1]$ yields

$$\begin{aligned} K_{\psi,\phi}(1)\, K_{\psi,\phi}(1-c) &= \int_c^1 \psi\left\{K_{\psi,\phi}(r) + K_{\psi,\phi}(r-c)\right\} dr + 2\phi \int_c^1 K_{\psi,\phi}(r)\, K_{\psi,\phi}(r-c)\, dr \\ &\quad + \sigma \int_c^1 K_{\psi,\phi}(r)\, dW(r-c) + \sigma \int_c^1 K_{\psi,\phi}(r-c)\, dW(r). \end{aligned}$$

This implies that

$$\begin{aligned} K_{\psi,\phi}(1-c)\, \delta_\phi^c K_{\psi,\phi}(1) &= K_{\psi,\phi}(1)\, K_{\psi,\phi}(1-c) - e^{\phi c} K_{\psi,\phi}(1-c)^2 \\ &= \int_c^1 \psi\left\{K_{\psi,\phi}(r) + K_{\psi,\phi}(r-c)\right\} dr + 2\phi \int_c^1 K_{\psi,\phi}(r)\, K_{\psi,\phi}(r-c)\, dr \\ &- 2e^{\phi c} \int_0^{1-c} \left[\psi K_{\psi,\phi}(r) + \phi K_{\psi,\phi}^2(r)\right] dr - e^{\phi c}(1-c)\,\sigma^2 \\ &+ \sigma \int_c^1 K_{\psi,\phi}(r)\, dW(r-c) + \sigma \int_c^1 K_{\psi,\phi}(r-c)\, dW(r) \\ &- 2e^{\phi c}\sigma \int_0^{1-c} K_{\psi,\phi}(r)\, dW(r) \end{aligned}$$

We rearrange the previous expression as

$$\begin{aligned} &K_{\psi,\phi}(1-c)\, \delta_\phi^c K_{\psi,\phi}(1) \\ &- \sigma\left(\int_c^1 \delta_\phi^c K_{\psi,\phi}(r)\, dW(r-c) + \int_c^1 K_{\psi,\phi}(r-c)\, d\left[\delta_\phi^c W(r)\right]\right) \\ &= \int_{1-c}^1 \psi K_{\psi,\phi}(r)\, dr - \int_0^c \psi K_{\psi,\phi}(r)\, dr + 2\left(1 - e^{\phi c}\right) \int_0^{1-c} \psi K_{\psi,\phi}(r)\, dr \\ &+ 2\phi \int_c^1 K_{\psi,\phi}(r)\, K_{\psi,\phi}(r-c)\, dr - 2e^{\phi c} \int_0^{1-c} \phi K_{\psi,\phi}^2(r)\, dr - e^{\phi c}(1-c)\,\sigma^2. \end{aligned}$$

So that the right-hand side of (A.12) rewrites as

$$
\begin{aligned}
&-\frac{c}{2}\left[e^{-\phi c}\psi^2 f_\phi(c)^2+\frac{\sigma^2}{\phi}e^{\phi c}\right]+\frac{1}{2}\frac{\sigma^2}{\phi}e^{\phi c}\\
&+\frac{1}{2\phi}\left[\int_{1-c}^{1}\psi K_{\psi,\phi}(r)\,dr-\int_0^c\psi K_{\psi,\phi}(r)\,dr+2\left(1-e^{\phi c}\right)\int_0^{1-c}\psi K_{\psi,\phi}(r)\,dr+\right.\\
&\qquad\left.+2\phi\int_c^1 K_{\psi,\phi}(r)\,K_{\psi,\phi}(r-c)\,dr-2e^{\phi c}\int_0^{1-c}\phi K^2_{\psi,\phi}(r)\,dr\right]\\
&-\frac{1}{2\phi}e^{\phi c}(1-c)\,\sigma^2\\
&+\frac{\psi}{2\phi}\left[\int_0^c K_{\psi,\phi}(r)\,dr-\int_{1-c}^1 K_{\psi,\phi}(r)\,dr\right]\\
&=-\frac{1}{2}c\psi^2e^{-\phi c}f_\phi(c)^2-f_\phi(c)\int_0^{1-c}\psi K_{\psi,\phi}(r)\,dr+\int_0^{1-c}\left[K_{\psi,\phi}(r+c)-e^{\phi c}K_{\psi,\phi}(r)\right]K_{\psi,\phi}(r)\,dr,\\
&=-\frac{1}{2}c\psi^2e^{-\phi c}f_\phi(c)^2+\int_c^1 K_{\psi,\phi}(r-c)\,\delta_\phi^c J_\phi(r)\,dr \qquad \text{(A.13)}
\end{aligned}
$$

When $\phi=0$, (A.12) rewrites as $\sigma\int_c^1\left[\delta_1^c W(r)\right]K_{\psi,\phi}(r-c)\,dr-\frac{1}{2}\psi^2c^3$ which implies that Expression (A.13) also holds for $\phi=0$.

B Proof of Corollary 2

First note that

$$
\begin{aligned}
J_\phi(r+c)-e^{\phi c}J_\phi(r)&=J_\phi(r+c)-J_\phi(r)-\phi f_\phi(c)\,J_\phi(r)\\
&=\int_0^c dJ_\phi(r+u)-\phi f_\phi(c)\,J_\phi(r)
\end{aligned}
$$

hence for low horizon, such that replacing c with cdr

$$
\begin{aligned}
\delta_\phi^{cdr}J_\phi(r+cdr)&=J_\phi(r+cdr)-J_\phi(r)-\phi cJ_\phi(r)\,dr+o_p(cdr)\\
&=c\phi J_\phi(r)\,dr+\sigma\left[W(r+cdr)-W(r)\right]-\phi cJ_\phi(r)\,dr+o_p(cdr)\\
&=\sigma\left[W(r+cdr)-W(r)\right]+o_p(cdr)
\end{aligned}
$$

and $W(r+cdr)-W(r)\sim\mathsf{N}(0,cdr)$, with the definition $dW(r)=W(r+dr)-W(r)$. Then (e_c) rewrites for low horizons as

$$
\begin{aligned}
&\delta_\phi^{cdr}J_\phi(r+cdr)\underset{c\to0}{\overset{L}{=}}\sqrt{c}\sigma dW(r)+o_p(cdr)\\
&\delta_\phi^{c}J_\phi(r+c)\underset{c\to0}{\overset{L}{=}}\sigma\left[W(r+c)-W(r)\right]+o_p(c) \qquad \text{(B.14)}
\end{aligned}
$$

Hence (f_c) becomes

$$
\int_0^{1-c}\delta_\phi^c J_\phi(r+c)\,dr\underset{c\to0}{\overset{L}{=}}\sqrt{c}\sigma W(1)+o_p(c)\,. \qquad \text{(B.15)}
$$

Now, as regards (d_c),

$$\int_0^{1-c} K_{\psi,\phi}(r)\, \delta_\phi^c J_\phi(r+c)\, dr \underset{c\to 0}{\overset{L}{=}} \sqrt{c}\sigma \int_0^1 K_{\psi,\phi}(r)\, dW(r) + o_p(c) \tag{B.16}$$

Finally, for (g_c) we see that, $\left[\delta_\phi^{cdr} J_\phi(r+cdr)\right]^2 = \sigma^2 cdr + o_p(cdr)$ so (g_c) rewrites as.

$$\int_c^1 \left[\delta_\phi^c J_\phi(r)\right]^2 dr \underset{c\to 0}{\overset{L}{=}} \sigma^2 c + o_p(c)$$

C Proof of Proposition 3

C.1 Short horizon

Results for DMS follow from Proposition 1 since estimators are computed by OLS. Now for IMS, We first consider fixed horizons $\widehat{\rho}_T^h = \left(1+T^{-1}\gamma_T\right)^h = 1 + hT^{-1}\gamma_T + O_p\left(T^{-2}\right)$, hence the result for the slope estimator:

$$T\left(\widehat{\rho}_T^h - 1\right) = h\gamma_T + O_p\left(T^{-1}\right). \tag{C.17}$$

As regards the intercept:

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

since $\left(\tau_T + T^{-1/2}\pi_T\right) = O_p\left(T^{-1/2}\right)$. Now

$$\tau_{\{h\},T} = \sum_{i=0}^{h-1} \rho_T^i \tau_T = \frac{f_\phi(h/T)}{f_\phi(1/T)} \tau_T = h\frac{\psi}{\sqrt{T}}\left(1 + \frac{h-1}{2T} + O\left(\frac{1}{T^2}\right)\right)$$

and

$$T^{1/2}\left(\widehat{\tau}_{\{h\},T} - \tau_{\{h\},T}\right) = h\pi_0 + O_p\left(T^{-1}\right). \tag{C.18}$$

Now, letting $h = \lfloor cT \rfloor$, we see that $\rho_T^h = e^{h\phi/T} = e^{c_T\phi}$ and $h^{-1/2}\rho_T^{\{h\}}\tau_T = T^{-1}\dfrac{f_\phi(c_T)}{f_\phi(1/T)} c_T^{-1/2}\psi$. Also, by definition $\gamma_T = T\left(\widehat{\rho}_T - 1\right) \Rightarrow \gamma_0$. The estimated slope therefore converges to

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

hence

$$\widehat{\rho}_T^h - 1 \Rightarrow e^{c\gamma_0} - 1 = f_{\gamma_0}(c)\,\gamma_0.$$

Similarly, we obtain the asymptotic distribution:

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

and hence the convergence:

$$\begin{aligned}
T^{-1/2}\left(\widehat{\rho}_T^{\{h\}}\widehat{\tau}_T - \rho_T^{\{h\}}\tau_T\right) &\Rightarrow \left[c\frac{f_{c\gamma_0}}{f_0\left(1\right)} - \frac{f_\phi\left(c\right)}{f_0\left(1\right)}\right]\psi + c\frac{f_{c\gamma_0}}{f_0\left(1\right)}\pi_0 \\
&= \left[f_{\gamma_0}\left(c\right) - f_\phi\left(c\right)\right]\psi + f_{\gamma_0}\left(c\right)\pi_0. \qquad \text{(C.19)}
\end{aligned}$$

C.2 Long Horizon

We first focus on the IMS forecast error and apply the short horizon results derive previously to $\widehat{e}_{h|T} = y_{T+h} - \left(\widehat{\tau}_{\{h\},T} + \widehat{\rho}_T^h y_T\right)$. Then,

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

Given that $\rho_T^h = e^{h\phi/T} = e^{c_T\phi}$ and $h^{-1/2}\rho_T^{\{h\}}\tau_T = T^{-1}\frac{f_\phi\left(c_T\right)}{f_\phi\left(1/T\right)}c_T^{-1/2}\psi$, the scaled forecast error $\widehat{e}_{c,T}^* = h^{-1/2}\widehat{e}_{h|T}$ can be decomposed into

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

where the estimated slope converges as $\widehat{\rho}_T^h \Rightarrow e^{c\gamma_0}$. Recall the definition of $\left(\pi_{\{c\}}, \gamma_{\{c\}}\right)$, then

$$\begin{aligned}
\sqrt{c_T}\widehat{e}_{c,T}^* &\Rightarrow K_{\psi,\phi}\left(1+c\right) - e^{c\phi}K_{\psi,\phi}\left(1\right) \\
&- f_\phi\left(c\right)\psi - \pi_{\{c\}} \\
&- \left(\gamma_{\{c\}} - \phi f_\phi\left(c\right)\right)K_{\psi,\phi}\left(1\right) \qquad \text{(C.20)}
\end{aligned}$$

and the IMS part of the theorem follows, using Expression (C.19).

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

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

but now the rates of convergence of estimators differ since $T^{-1/2} (\widetilde{\tau}_{h,T} - \tau_{h,T}) \Rightarrow \pi_c$ and $\left(\widetilde{\rho}_{h,T} - \rho_T^h\right) \Rightarrow \gamma_c + 1 - \lambda^c$. This leads to

$$T^{-1/2} \widetilde{e}_{h|T} \Rightarrow -\pi_c - (\gamma_c + 1 - \lambda^c) K_{\psi,\phi}(1) + K_{0,\phi}(1+c) - \lambda^c K_{0,\phi}(1)$$

or, rewriting with the scaled forecast error, since $\lfloor cT \rfloor^{-1/2} \widetilde{e}_{h|T} = h^{-1/2} \widetilde{e}_{h|T}$

$$\widetilde{e}^*_{c,T} \Rightarrow c^{-1/2} \left\{ -\pi_c - (\gamma_c + 1 - \lambda^c) K_{\psi,\phi}(1) + K_{0,\phi}(1+c) - \lambda^c K_{0,\phi}(1) \right\} \tag{C.21}$$

and hence the results.

Independence between $K_{\psi,\phi}(1+c) - e^{c\phi} K_{\psi,\phi}(1)$ and the remainders of the RHS of expressions (C.20) and (C.20) follows from uncorrelatedness and Gaussianity.

D Proof of Proposition 7

We use the definition:

$$z_t = \alpha_h + \beta_h y_{t-h} + \omega_{h,t} \tag{D.22}$$

with

$$(\alpha_h, \beta_h) = \left(\alpha + \beta \tau_T \rho_T^{\{h-1\}}, \beta \rho_T^{h-1}\right), \quad \omega_{h,t} = \beta \sum_{i=1}^{h-1} \rho_T^{h-1-i} \epsilon_{t-h+i} + \varepsilon_t;$$

and let $\nu_t = (\varepsilon_t, \epsilon_t)$ so $\omega_{h,t} = \left(1, \beta \sum_{i=1}^{h-1} \rho_T^{i-1} L^i\right)' \nu_t$. Let first $h \in [1, T)$ be fixed, then $T^{-1/2} \sum_{t=h}^{\lfloor Tr \rfloor} \nu_t \Rightarrow \mathbf{H}(r)$ implies

$$T^{-1/2} \sum_{t=h}^{\lfloor Tr \rfloor} \omega_{h,t} \Rightarrow (1, (h-1)\beta)' \mathbf{H}(r) \tag{D.23}$$

whereas if $h = \lfloor cT \rfloor$, then

$$\left\{ \begin{array}{c} T^{-1/2} \omega_{\lfloor cT \rfloor, \lfloor rT \rfloor} \Rightarrow (0:\beta) \delta_\phi^c \mathbf{G}_\phi(r) \\ T^{-3/2} \sum_{t=h}^{\lfloor Tr \rfloor} \omega_{h,t} \Rightarrow (0:\beta) \int_c^r \delta_\phi^c \mathbf{G}_\phi(s) \, ds. \end{array} \right. \tag{D.24}$$

This proves (a_h) and (a_c) using the definition of $\mathbf{J}_\phi$. In the following, we use $\mathbf{H} = (H, \sigma W)$ and for all i,

$$\mathbf{\Xi}_i = \left[\begin{array}{cc} \xi_\varepsilon(i) & \xi_{\varepsilon,\epsilon}(i) \\ \xi_{\epsilon,\varepsilon}(i) & \xi_\epsilon(i) \end{array} \right].$$

Now for (b_h) and (b_c), i.e. the distribution of $\sum_{t=h}^{\lfloor Tr \rfloor} x_{t-h}\omega_{h,t}$. We assume throughout that $\beta \neq 0$. Using $y_{t-h} = \beta^{-1}(z_{t-h+1} - \alpha - \varepsilon_{t-h+1})$, replacing it in Expression (D.22) we get an expression similar to the multi-step forecasting model

$$z_t = \left(\alpha\left(1-\rho_T^h\right) + \beta\tau_T\rho_T^{\{h\}}\right) + \rho_T^h z_{t-h} + w_{h,t} \tag{D.25}$$

where $w_{h,t} = \omega_{h+1,t} - \rho_T^h \varepsilon_{t-h}$ and $w_{0,t} = 0$. Letting $h = 1$ in Expression (D.25) yields $z_t = (\alpha(1-\rho_T) + \beta\tau_T) + \rho_T z_{t-1} + w_{1,t}$, where $w_{1,t} = \varepsilon_t - \rho\varepsilon_{t-1} + \beta\epsilon_{t-1}$. The intercept above is the sum of two terms, $\beta\tau_T = O\left(T^{-1/2}\right)$ and $\alpha(1-\rho_T) = O\left(T^{-1}\right)$ so we may disregard the impact of α when using the results we derived in univariate forecasting.

To apply the results from the forecasting section, we need to compute the covariance function of $w_{h,t}$,

$$\begin{aligned}
\xi_{w_h}(i) &= \mathsf{Cov}\left(\varepsilon_t - \rho_T^h\varepsilon_{t-h} + \beta\sum_{j=1}^{h}\rho_T^{j-1}\epsilon_{t-j}, \varepsilon_{t-i} - \rho_T^h\varepsilon_{t-h-i} + \beta\sum_{j=1}^{h}\rho_T^{j-1}\epsilon_{t-j-i}\right) \\
&= \xi_\varepsilon(i) - \rho_T^h\xi_\varepsilon(h+i) + \beta\sum_{j=1}^{h}\rho_T^{j-1}\xi_{\varepsilon,\epsilon}(j+i) \\
&- \rho_T^h\xi_\varepsilon(h-i) + \rho_T^{2h}\xi_\varepsilon(i) - \rho_T^h\beta\sum_{j=1}^{h}\rho_T^{j-1}\xi_{\varepsilon,\epsilon}(j-h+i) \\
&+ \beta\sum_{j=1}^{h}\rho_T^{j-1}\xi_{\varepsilon,\epsilon}(j-i) - \rho_T^h\beta\sum_{j=1}^{h}\rho_T^{j-1}\xi_{\varepsilon,\epsilon}(j-h-i) + \beta^2\sum_{j=1}^{h}\sum_{k=1}^{h}\rho_T^{j+k-2}\xi_\epsilon(k-j+i) \\
&= \left(1+\rho_T^{2h}\right)\xi_\varepsilon(i) - \rho_T^h\left[\xi_\varepsilon(h+i) + \xi_\varepsilon(h-i)\right] \\
&+ \beta\left[\sum_{j=1}^{h}\rho_T^{j-1}\left[\xi_{\varepsilon,\epsilon}(j+i) + \xi_{\varepsilon,\epsilon}(j-i) - \rho_T^h\left[\xi_{\varepsilon,\epsilon}(j-h+i) + \xi_{\varepsilon,\epsilon}(j-h-i)\right]\right]\right] \\
&+ \beta^2\sum_{j=1}^{h}\sum_{k=1}^{h}\rho_T^{j+k-2}\xi_\epsilon(k-j+i)
\end{aligned}$$

with variance

$$\begin{aligned}
\xi_{w_h}(0) &= \left(1+\rho_T^{2h}\right)\xi_\varepsilon(0) - 2\rho_T^h\xi_\varepsilon(h) + 2\beta\left[\sum_{j=1}^{h}\rho_T^{j-1}\left[\xi_{\varepsilon,\epsilon}(j) - \rho_T^h\xi_{\varepsilon,\epsilon}(j-h)\right]\right] \\
&+ \beta^2\sum_{j=1}^{h}\sum_{k=1}^{h}\rho_T^{j+k-2}\xi_\epsilon(k-j),
\end{aligned}$$

and long run variance $\xi_{w_h} = \left(1-\rho_T^h\right)^2\xi_\varepsilon + 2\beta\frac{\left(1-\rho_T^h\right)^2}{1-\rho_T}\xi_{\varepsilon,\epsilon} + \beta^2\left(\frac{1-\rho_T^h}{1-\rho_T}\right)^2\xi_\epsilon$. Specifically for $h = 1$

the expressions become

$$\xi_{w_1}(0) = \left(1+\rho_T^2\right)\sigma_\varepsilon^2 + \beta^2\sigma_\epsilon^2 - 2\beta\left(\rho_T\xi_{\varepsilon,\epsilon}(0) - \xi_{\varepsilon,\epsilon}(1)\right)$$

$$\begin{aligned}\xi_{w_1}(i) = {} & \left(1+\rho_T^2\right)\xi_\varepsilon(i) - \rho_T\left[\xi_\varepsilon(i+1) + \xi_\varepsilon(i-1)\right] \\ & - \beta\left[\rho_T\left(\xi_{\varepsilon,\epsilon}(i) + \xi_{\varepsilon,\epsilon}(-i)\right) - \left(\xi_{\varepsilon,\epsilon}(i+1) + \xi_{\varepsilon,\epsilon}(1-i)\right)\right] \\ & + \beta^2\xi_\epsilon(i)\end{aligned}$$

with long run variance $\xi_{w_1} = \sum_{i=-\infty}^{\infty}\xi_{w_1}(i) = (1-\rho_T)^2\xi_\varepsilon + 2(1-\rho_T)\beta\xi_{\varepsilon,\epsilon} + \beta^2\xi_\epsilon$. From the assumption $w_{1,t} = \left((1-\rho_T L), \beta_T L\right)'\nu_t$, it follows that

$$\begin{aligned}T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor} w_{1,t} &= T^{-1/2}\left(\varepsilon_{\lfloor Tr\rfloor} + (1-\rho_T,\beta)'\sum_{t=1}^{\lfloor Tr\rfloor -1}\nu_t + (-\rho_T,\beta)'\varepsilon_0\right) \\ &\Rightarrow (0,\beta)'\mathbf{H}(r) = \beta\sigma W(r)\end{aligned}$$

hence $T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor} z_t \Rightarrow \beta K_{\psi,\phi}(r)$. Also for fixed h, $T^{-1/2}\sum_{t=h}^{\lfloor rT\rfloor} w_{h,t} \Rightarrow h\beta\sigma W(r)$, whereas for $h/T \to c > 0$

$$T^{-1/2}w_{h,\lfloor Tr\rfloor} \Rightarrow \beta\delta_\phi^c J_\phi(r), \quad T^{-3/2}\sum_{t=h}^{T} w_{h,t} \Rightarrow \beta\int_c^1 \delta_\phi^c J_\phi(r)\,dr, \quad T^{-2}\sum_{t=h}^{T} w_{h,t}^2 \Rightarrow \beta\int_c^1 \left[\delta_\phi^c J_\phi(r)\right]^2 dr.$$

To derive the required results using those on univariate forecasting, we rewrite:

$$\begin{aligned}\sum_{t=h}^{\lfloor Tr\rfloor} y_{t-h}\omega_{h,t} &= \beta^{-1}\sum_{t=h}^{\lfloor Tr\rfloor}\left(z_{t-(h-1)} - \alpha - \varepsilon_{t-h+1}\right)\left(w_{h-1,t} + \rho_T^{h-1}\varepsilon_{t-h+1}\right) \\ &= \beta^{-1}\left(\sum_{t=h}^{\lfloor Tr\rfloor} z_{t-(h-1)}w_{h-1,t} + \rho_T^{h-1}\sum_{t=h}^{\lfloor Tr\rfloor} z_{t-(h-1)}\varepsilon_{t-(h-1)} - \alpha\sum_{t=h}^{\lfloor Tr\rfloor}\omega_{h,t} - \sum_{t=h}^{\lfloor Tr\rfloor}\varepsilon_{t-(h-1)}\omega_{h,t}\right)\end{aligned} \tag{D.26}$$

The asymptotic distribution of the first term on the RHS is derived from the multi-step forecasting model, Proposition 1,

$$T^{-1}\sum_{t=h-1}^{T} z_{t-(h-1)}w_{h-1,t} \Rightarrow (h-1)\beta^2\sigma\int_0^1 K_{\psi,\phi}dW + \frac{1}{2}\left[(h-1)\xi_{w_{h-1}} - \xi_{w_{h-1}}(0)\right]$$

$$T^{-2}\sum_{t=\lfloor cT-1\rfloor}^{T} z_{t-}w_{\lfloor cT-1\rfloor,t} \Rightarrow \beta^2\int_c^1 K_{\psi,\phi}(r-c)\,\delta_\phi^c J_\phi(r)\,dr - \frac{1}{2}c\beta^2\psi^2 e^{-\phi c}\left[f_\phi(c)\right]^2$$

For the other terms, we start with with

$$\sum_{t=h}^{\lfloor Tr\rfloor} z_{t-(h-1)}\varepsilon_{t-(h-1)} = \alpha\sum_{t=h}^{\lfloor Tr\rfloor}\varepsilon_{t-(h-1)} + \beta\sum_{t=h}^{\lfloor Tr\rfloor} y_{t-h}\varepsilon_{t-(h-1)} + \sum_{t=h}^{\lfloor Tr\rfloor}\varepsilon_{t-(h-1)}^2$$

where

$$\sum_{t=h}^{\lfloor Tr \rfloor} y_{t-h}\varepsilon_{t-(h-1)} = \sum_{t=h}^{\lfloor Tr \rfloor}\sum_{i=1}^{t-h} \rho_T^{t-i}\left(\tau_T + \epsilon_i\right)\varepsilon_{t-(h-1)} + o_p\left(T\right)$$

$$= \tau_T \sum_{t=h}^{\lfloor Tr \rfloor}\left(\sum_{i=h}^{t-1} \rho_T^{i}\right)\varepsilon_{t-(h-1)} + \sum_{t=h}^{\lfloor Tr \rfloor}\sum_{i=1}^{t-h} \rho_T^{t-i}\epsilon_i\varepsilon_{t-(h-1)} + o_p\left(T\right)$$

Hence, starting with h fixed, noticing that

$$\mathsf{E}\sum_{t=h}^{T}\sum_{i=1}^{t-h} \rho_T^{t-i}\epsilon_i\varepsilon_{t-(h-1)} = \sum_{t=h}^{T}\sum_{i=1}^{t-h} \rho_T^{t-i}\xi_{\epsilon,\varepsilon}\left(t-(h-1)-i\right)$$

$$= \rho_T^{h-1}\sum_{i=1}^{T-h}\left(T-(h-1)-i\right)\rho_T^{i}\xi_{\epsilon,\varepsilon}\left(i\right)$$

$$= \left[T-(h-1)\right]\rho_T^{h-1}\sum_{i=1}^{T-h}\rho_T^{i}\xi_{\epsilon,\varepsilon}\left(i\right) - \rho_T^{h-1}\sum_{i=1}^{T-h} i\rho_T^{i}\xi_{\epsilon,\varepsilon}\left(i\right)$$

hence for fixed h since $\sum_{i=1}^{\infty} i\xi_{\epsilon,\varepsilon}\left(i\right) < \infty$, and using the weak law or large numbers for non-identically distributed processes (see e.g. Andrews, 1988), $T^{-1}\sum_{t=h}^{T}\sum_{i=1}^{t-h}\rho_T^{t-i}\epsilon_i\varepsilon_{t-(h-1)} \xrightarrow{p} \sum_{i=1}^{\infty}\xi_{\varepsilon,\epsilon}\left(-i\right)$, and

$$T^{-1}\sum_{t=h}^{T} y_{t-h}\varepsilon_{t-(h-1)} \Rightarrow \int_0^1 K_{\psi,\phi}\left(s\right)dH\left(s\right) + \sum_{i=1}^{\infty}\xi_{\varepsilon,\epsilon}\left(-i\right)$$

$$T^{-1}\sum_{t=h}^{\lfloor Tr \rfloor} z_{t-(h-1)}\varepsilon_{t-(h-1)} \Rightarrow \beta\int_0^1 K_{\psi,\phi}\left(s\right)dH_1\left(s\right) + \beta\sum_{i=1}^{\infty}\xi_{\epsilon,\varepsilon}\left(-i\right) + \sigma_\varepsilon^2.$$

Now if $h = \lfloor cT \rfloor$, $\sum_{i=h-1}^{\lfloor Tr \rfloor - h}\rho_T^{t-i}\left(\epsilon_i\varepsilon_{t-(h-1)} - \xi_{\epsilon,\varepsilon}\left(t-(h-1)-i\right)\right) = O_p\left(\sqrt{T}\right)$ and

$$\sum_{i=1}^{T-h}\rho_T^{i}\xi_{\epsilon,\varepsilon}\left(-i\right) = \sum_{i=1}^{T-h}\xi_{\epsilon,\varepsilon}\left(-i\right) + \frac{\phi}{T}\sum_{i=1}^{T-h}\left(1 + \frac{i}{2!T} + ...\right)i\xi_{\varepsilon,\epsilon}\left(-i\right)$$

$$= \sum_{i=1}^{\infty}\xi_{\epsilon,\varepsilon}\left(-i\right) + O\left(T^{-1}\right)$$

hence $T^{-1}\sum_{t=h}^{T} y_{t-h}\varepsilon_{t-(h-1)} \Rightarrow \int_0^{1-c} K_{\psi,\phi}\left(s\right)dH\left(s\right) + \left(1-c\right)e^{\phi c}\sum_{i=1}^{\infty}\xi_{\epsilon,\varepsilon}\left(-i\right)$, and

$$T^{-1}\sum_{t=h}^{T} z_{t-(h-1)}\varepsilon_{t-(h-1)} \Rightarrow \beta\int_0^{1-c} K_{\psi,\phi}\left(s\right)dH_1\left(s\right) + \beta\left(1-c\right)e^{\phi c}\sum_{i=1}^{\infty}\xi_{\epsilon,\varepsilon}\left(-i\right) + \left(1-c\right)\sigma_\varepsilon^2.$$

Now the third term in (D.26) is $\sum_t \omega_{h,t}$ whose asymptotic behavior depends on the rate of h: for fixed h

$$T^{-1/2}\sum_{t=h}^{\lfloor rT \rfloor}\omega_{h,t} \Rightarrow H\left(r\right) + h\beta\sigma W\left(r\right)$$

whereas for $h/T \to c > 0$

$$T^{-1/2}\omega_{\lfloor cT\rfloor,\lfloor Tr\rfloor} \Rightarrow \beta\delta_\phi^c J_\phi(r), \quad T^{-3/2}\sum_{t=\lfloor cT\rfloor}^{T}\omega_{\lfloor cT\rfloor,t} \Rightarrow \beta\int_c^1 \delta_\phi^c J_\phi(r)\,dr,$$

$$T^{-2}\sum_{t=\lfloor cT\rfloor}^{T}\omega_{\lfloor cT\rfloor,t}^2 \Rightarrow \beta\int_c^1 \left[\delta_\phi^c J_\phi(r)\right]^2 dr.$$

The fourth term in (D.26) is

$$\begin{aligned}\sum_t \varepsilon_{t-(h-1)}\omega_{h,t} &= \sum_t\left(\beta\sum_{i=1}^{h-1}\rho_T^{i-1}\varepsilon_{t-(h-1)}\epsilon_{t-i} + \varepsilon_{t-(h-1)}\varepsilon_t\right)\\ &= \beta\sum_t\sum_{i=1}^{h-1}\rho_T^{h-1-i}\epsilon_{t-(h-i)}\varepsilon_{t-(h-1)} + \sum_t\varepsilon_{t-(h-1)}\varepsilon_t\end{aligned}$$

so for h fixed, $T^{-1}\sum_{t=h}^{\lfloor Tr\rfloor}\varepsilon_{t-(h-1)}\omega_{h,t} \Rightarrow \beta\sum_{i=1}^{h-1}\xi_{\varepsilon,\epsilon}(1-i) + \xi_\varepsilon(h-1)$, and if $h = \lfloor cT\rfloor$, $\sum_{t=h}^{\lfloor Tr\rfloor}\sum_{i=1}^{h-1}\rho_T^{h-1-i}\epsilon_{t-h+i}\varepsilon_{t-(h-1)} = O_p(T)$.

We now collect all the terms, starting with $h/T \to c$

$$T^{-2}\sum_{t=h}^{T} y_{t-h}\omega_{h,t} = \beta\int_c^1 K_{\psi,\phi}(r-c)\,\delta_\phi^c J_\phi(r)\,dr - \frac{1}{2}c\beta\psi^2 e^{-\phi c}\left[f_\phi(c)\right]^2$$

and if h fixed,

$$T^{-1}\sum_{t=h}^{T} y_{t-h}\omega_{h,t} = \int_0^1 K_{\psi,\phi}\,d\left(H_1 + (h-1)\beta\sigma W\right) + \varpi_h$$

with

$$\begin{aligned}\varpi_h &= \frac{1}{2\beta}\left[(h-1)\,\xi_{w_{h-1}} - \xi_{w_{h-1}}(0)\right] + \frac{1}{\beta}\left[\sigma_\varepsilon^2 - \xi_\varepsilon(h-1)\right]\\ &\quad + \sum_{i=1}^{\infty}\xi_{\varepsilon,\epsilon}(-i) - \sum_{i=1}^{h-1}\xi_{\varepsilon,\epsilon}(1-i)\end{aligned}$$

where

$$\begin{aligned}\xi_{w_{h-1}} &= \left(1-\rho_T^{h-1}\right)^2\xi_\varepsilon + \beta\frac{\left(1-\rho_T^{h-1}\right)^2}{1-\rho_T}\xi_{\varepsilon,\epsilon} + \beta^2\left(\frac{1-\rho_T^{h-1}}{1-\rho_T}\right)^2\xi_\epsilon.\\ \xi_{w_{h-1}}(0) &= \left(1+\rho_T^{2(h-1)}\right)\xi_\varepsilon(0) - 2\rho_T^{h-1}\xi_\varepsilon(h)\\ &\quad + 2\beta\left[\sum_{j=1}^{h-1}\rho_T^{j-1}\left[\xi_{\varepsilon,\epsilon}(j) - \rho_T^{h-1}\xi_{\varepsilon,\epsilon}(j-(h-1))\right]\right]\\ &\quad + \beta^2\sum_{j=1}^{h-1}\sum_{k=1}^{h-1}\rho_T^{j+k-2}\xi_\epsilon(k-j)\end{aligned}$$

so

$$
\begin{aligned}
&(h-1)\,\xi_{w_{h-1}} - \xi_{w_{h-1}}(0) \\
&= \left(1+\rho_T^{2(h-1)}\right)\left[(h-1)\,\xi_{\varepsilon} - \xi_{\varepsilon}(0)\right] - 2\rho_T^{h-1}\left[(h-1)\,\xi_{\varepsilon} - \xi_{\varepsilon}(h)\right] \\
&+ 2\beta\left[\sum_{j=1}^{h-1}\rho_T^{j-1}\left[(h-1)\,\xi_{\varepsilon,\epsilon} - \xi_{\varepsilon,\epsilon}(j) - \rho_T^{h-1}\left[(h-1)\,\xi_{\varepsilon,\epsilon} - \xi_{\varepsilon,\epsilon}(j-(h-1))\right]\right]\right] \\
&+ \beta^2\sum_{j=1}^{h-1}\sum_{k=1}^{h-1}\rho_T^{j+k-2}\left[(h-1)\,\xi_{\epsilon} - \xi_{\epsilon}(k-j)\right]
\end{aligned}
$$

and

$$
\begin{aligned}
\varpi_h &= \frac{1}{2\beta}\left[(h-1)\,\xi_{w_{h-1}} - \xi_{w_{h-1}}(0)\right] + \frac{1}{\beta}\left[\sigma_{\varepsilon}^2 - \xi_{\varepsilon}(h-1)\right] + \sum_{i=h+1}^{\infty}\xi_{\varepsilon,\epsilon}(-i) - \xi_{\varepsilon,\epsilon}(0) \\
&= \frac{1}{2\beta}\left(\left[(h-1)\,\xi_{\varepsilon} + \xi_{\varepsilon}(0)\right] + \rho_T^{2(h-1)}\left[(h-1)\,\xi_{\varepsilon} - \xi_{\varepsilon}(0)\right]\right) \\
&- \frac{1}{\beta}\rho_T^{h-1}\left[(h-1)\,\xi_{\varepsilon} - \xi_{\varepsilon}(h)\right] - \frac{1}{\beta}\xi_{\varepsilon}(h-1) \\
&+ \sum_{i=1}^{\infty}\xi_{\varepsilon,\epsilon}(-i) - \sum_{i=1}^{h-1}\xi_{\varepsilon,\epsilon}(1-i) \\
&+ \left[\sum_{j=1}^{h-1}\rho_T^{j-1}\left[(h-1)\,\xi_{\varepsilon,\epsilon} - \xi_{\varepsilon,\epsilon}(j) - \rho_T^{h-1}\left[(h-1)\,\xi_{\varepsilon,\epsilon} - \xi_{\varepsilon,\epsilon}(j-(h-1))\right]\right]\right] \\
&+ \frac{\beta}{2}\sum_{j=1}^{h-1}\sum_{k=1}^{h-1}\rho_T^{j+k-2}\left[(h-1)\,\xi_{\epsilon} - \xi_{\epsilon}(k-j)\right]
\end{aligned}
$$

Notice that if ε_t is not autocorrelated then

$$
(h-1)\,\xi_{\varepsilon} + \xi_{\varepsilon}(0) = h\sigma_{\varepsilon}^2 \quad \text{and} \quad (h-1)\,\xi_{\varepsilon} - \xi_{\varepsilon}(0) = (h-2)\,\sigma_{\varepsilon}^2, \tag{D.27}
$$

if $\xi_{\varepsilon,\epsilon} = \xi_{\varepsilon,\epsilon}(0) = \sigma_{\varepsilon,\epsilon}$, i.e. the cross correlation is only contemporaneous:

$$
\begin{aligned}
&\sum_{i=1}^{\infty}\xi_{\varepsilon,\epsilon}(-i) - \sum_{i=1}^{h-1}\xi_{\varepsilon,\epsilon}(1-i) + \sum_{j=1}^{h-1}\rho_T^{j-1}\left[(h-1)\,\xi_{\varepsilon,\epsilon} - \rho_T^{h-1}\left[(h-1)\,\xi_{\varepsilon,\epsilon} - \xi_{\varepsilon,\epsilon}(j-(h-1))\right]\right] \\
&= \begin{cases} \frac{\left(1-\rho_T^{h-1}\right)^2}{1-\rho_T}(h-1)\,\sigma_{\varepsilon,\epsilon} - \rho_T^{2h-3}\sigma_{\varepsilon,\epsilon} - \sigma_{\varepsilon,\epsilon} & h \geq 2 \\ 0 & h = 1 \end{cases}
\end{aligned}
$$

and finally if ϵ_t is not autocorrelated

$$
\frac{\beta}{2}\sum_{j=1}^{h-1}\sum_{k=1}^{h-1}\rho_T^{j+k-2}\left[(h-1)\,\xi_{\epsilon} - \xi_{\epsilon}(k-j)\right] = \frac{\beta}{2}\frac{1-\rho_T^{2(h-1)}}{1-\rho_T^2}h\xi_{\epsilon}
$$

Under these assumptions for $h \geq 2$

$$\begin{aligned}
\varpi_h &= \frac{1}{2\beta}\left(h - 2\rho_T^{h-1}(h-1) + \rho_T^{2(h-1)}(h-2)\right)\sigma_\varepsilon^2 \\
&+ \frac{\left(1-\rho_T^{h-1}\right)^2}{1-\rho_T}(h-1)\,\sigma_{\varepsilon,\epsilon} - \left(1+\rho_T^{2h-3}\right)\sigma_{\varepsilon,\epsilon} + \frac{\beta}{2}\frac{1-\rho_T^{2(h-1)}}{1-\rho_T^2}h\xi_\epsilon \\
&= \frac{\sigma_\varepsilon^2}{2\beta}\left(1-\rho_T^{h-1}\right)\left(h\left(1-\rho_T^{h-1}\right)+2\rho_T^{h-1}\right) \\
&+ \frac{\left(1-\rho_T^{h-1}\right)^2}{1-\rho_T}(h-1)\,\sigma_{\varepsilon,\epsilon} - \left(1+\rho_T^{2h-3}\right)\sigma_{\varepsilon,\epsilon} + \frac{\beta}{2}\frac{1-\rho_T^{2(h-1)}}{1-\rho_T^2}h\sigma_\epsilon^2
\end{aligned}$$

and $\varpi_1 = 0$.

E Proof of Corollary 8

We assume $\beta = 0$. From the assumption $\nu_t = \left((1-\rho_T L), 0\right)'(\varepsilon_t, \epsilon_t)$, it follows that

$$\begin{aligned}
\sum_{t=1}^{\lfloor Tr \rfloor}\nu_t &= \varepsilon_{\lfloor Tr \rfloor} + (1-\rho_T)\sum_{t=1}^{\lfloor Tr \rfloor - 1}\varepsilon_t + (-\rho_T : 0)\,\varepsilon_0 \\
&= \varepsilon_{\lfloor Tr \rfloor} - \varepsilon_0 + o_p(1) = O_p(1)
\end{aligned}$$

From (5), $z_{t+h} = \alpha + \varepsilon_{t+h} = \alpha + \omega_{h,t+h}$ with $\omega_{h,t+h} = \varepsilon_{t+h}$. Therefore, if as $T \to \infty$, $h \in [1, T)$ is fixed

$$T^{-1/2}\sum_{t=h}^{\lfloor Tr \rfloor}\omega_{h,t} \Rightarrow H(r) \tag{E.28}$$

and if $h/T \to c$, then

$$\begin{cases} T^{-1/2}\omega_{\lfloor cT \rfloor, \lfloor rT \rfloor} \Rightarrow 0 \\ T^{-1/2}\sum_{t=h}^{\lfloor Tr \rfloor}\omega_{h,t} \Rightarrow H(r) - H(c). \end{cases} \tag{E.29}$$

This proves (a_h) and (a_c) using the definition of J_ϕ. Now for (b_h) and (b_c) ,i.e. the distribution of

$$\sum_{t=h}^{T} y_{t-h}\omega_{h,t} = \sum_{t=h}^{T}\left(e^{\phi(t-h)/T}y_0 + \psi T^{-1/2}\frac{f_\phi\left((t-h)/T\right)}{f_\phi(1/T)} + \sum_{i=0}^{t-h-1} e^{\phi i/T}\epsilon_{t-h-i}\right)\varepsilon_t$$

Hence for h fixed $T^{-1}\sum_{t=h}^{T} y_{t-h}\omega_{h,t} \Rightarrow \int_0^1 K_{\psi,\phi}(s)\,dH(s) + \sum_{i=h}^{\infty}\xi_{\varepsilon,\epsilon}(i)$, and as $h/T \to c$, $\sum_{i=h}^{\infty}\xi_{\varepsilon,\epsilon}(i) \to 0$ so

$$T^{-1}\sum_{t=h}^{T} y_{t-h}\omega_{h,t} \Rightarrow \int_c^1 K_{\psi,\phi}(s-c)\,dH(s).$$