The role of information in nonstationary regression

Patrick Marsh[1]
School of Economics
University of Nottingham
University Park
NG7 2RD
email: patrick.marsh@nottingham.ac.uk

March 4^{th} 2019

[1]Thanks are due to Fabrizio Iacone, Peter Phillips and participants at seminars given at the Universities of Amsterdam, Nottingham and York.

Abstract

The role of standard likelihood based measures of information and efficiency is unclear when regressions involve nonstationary data. Typically the standardized score is not asymptotically Gaussian and the standardized Hessian has a stochastic, rather than deterministic limit. Here we consider a time series regression involving a deterministic covariate which can be evaporating, slowly evolving or nonstationary. It is shown that conditional information, or equivalently, profile Kullback-Leibler and Fisher Information remain informative about both the accuracy, i.e. asymptotic variance, of profile maximum likelihood estimators, as well as the power of point optimal invariant tests for a unit root. Specifically these information measures indicate fractional, rather than linear trends may minimize inferential accuracy. Such is confirmed in numerical experiment.

1 Introduction

Inference in models involving nonstationary variables is challenging in two important regards. First the standard Cramér-Rao efficiency theory does not apply. Estimators are, generally, not asymptotically normal nor do their covariances converge to Fisher information. Secondly, the asymptotic analysis of such models invariably provides stochastic representations for estimators and tests, rather than their distributional properties. Fisher information, as a probability metric, is not applicable in such models. Some of the asymptotic implications of these issues are explored in Magdalinos [1], while Marsh [2] considers the finite sample properties of Kullback-Leibler divergence.

This paper considers two standard time series specifications, either

$$\text{A) } y_t = d_t + \rho y_{t-1} + \varepsilon_t \quad or \quad \text{B) } y_t = d_t + u_t \ ; \ u_t = \rho u_{t-1} + \varepsilon_t, \tag{1}$$

for $t = 1, .., T$, $\varepsilon_t \sim iidN(0, \sigma^2)$. In these models d_t represents a deterministic component that will be employed to capture the effect of both stationary or ergodic as well as nonstationary covariates. Typically, interest is in inference on ρ, i.e. testing for a unit root, while if $d_t = \alpha' x_t$ for some choice of x_t, then α will be nuisance. In such circumstances conditional information, Bhapkar and Srinivasan [3] and Zhu and Reid [4], ought be employed as a probability metric (see also Gibbs and Su [5] for different choices of such metrics) for inference about the interest parameter. Conditional information is defined for a log-likelihood $l(\theta_1, \theta_2)$ depending on an interest parameter θ_1 and nuisance parameter θ_2 by

$$CI_{\theta_1|\theta_2} = I_{\theta_1\theta_1} - I'_{\theta_1\theta_2} I^{-1}_{\theta_2\theta_2} I_{\theta_1\theta_2}, \tag{2}$$

where $I_{\theta_1\theta_2} = E\left[-\partial^2 l(\theta_1, \theta_2) / \partial\theta_1 \partial\theta_2\right]$.

Since standard information theory does not apply in nonstationary models, here an analogue is defined via expectation of the stochastic limit of the scaled log-likelihood Hessian. This limit is found by first imposing the unit root, giving a preferred point (see Critchley, Marriott and Salmon [6]) probability metric analogue. It is shown that conditional information about ρ in specification A corresponds to profile Kullback-Leibler and profile Fisher information in specification B. Although this metric neither bounds nor equals the asymptotic variance of an unbiased estimator for ρ, it remains informative about inferential accuracy. Specifically, it is found that these can be convex functions: when $d_t = \alpha t^\beta$ they attain a unique minimum at a value of $\beta^* = \left(\sqrt{6} - 1\right)/2$ and at $\beta^+ = \left(\sqrt{10} - 1\right)/2$, when $d_t = \alpha_0 + \alpha_1 t^\beta$. The prediction that inferential accuracy is therefore minimized at these points is supported by numerical experiment.

The analysis of unit root tests began in the context of specification A. More recently, the set-up of specification B has dominated the literature, as it permits straightforward construction of invariant tests, having distributions free of nuisance parameters. In the context of the impact of covariates in unit root testing, Elliott, Rothenberg and Stock [7] characterize the asymptotic power envelope for both a general $d_t = o\left(T^{1/2}\right)$, as well as the linear trend case. Marsh [8] shows that Fisher Information in the maximal invariant (to a linear trend) vanishes under a unit root, while Phillips [9] considers the impact of nonlinear and slowly evolving trends. On the other hand, Hansen [10] (see also Elliott and Jansson [11] and Chrystalleni, Harvey and Leybourne [12]) explores the impact of stationary stochastic regressors in specification A. The results of this paper help shed some light on some of these findings.

The plan for the paper is as follows. Motivation for the results is provided in Section 2 via consideration of the original Dickey-Fuller [13] formulation (i.e. specification A) and the effect of stationary covariates as in Hansen [10]. The main results of the paper are provided for specification B in Section 3 while Section 4 discusses these results and Section 5 concludes. An appendix provides the proofs of the main results as well as tables and graphs for the numerical analysis.

2 Motivation via specification A

The original Dickey-Fuller [13] unit root testing framework considered a model as in specification A. And it is within this context that the power enhancement of stationary covariates, see Hansen [10], is explored. In the simplest possible set-up, suppose that

$$\begin{pmatrix} y_t - \rho y_{t-1} \\ w_t \end{pmatrix} = \begin{pmatrix} u_t \\ v_t \end{pmatrix}, \; u_t \sim iid\left(0, \sigma^2\right), \quad t = 1, .., T$$

and let $R^2 = corr^2\left[u_t, v_t\right]$. In Hansen [10], and also Chrystalleni, Harvey and Leybourne [11], Dickey-Fuller tests of $H_0 : \rho = 1$ in

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

are demonstrated to have powers increasing in R^2. Since in the limit of $R^2 \to 1$ we could, in fact, observe the errors $(y_t - \rho y_{t-1})_1^T$, this result is to be expected, as well as having empirical importance.

Here we explore the effect of the degree of covariate trending in the context of testing $H_0 : \rho = 1$ in the context of the fitted model,

$$y_t = \rho y_{t-1} + \alpha t^\beta + u_t, \qquad t = 1, .., T, \tag{3}$$

with $y_0 = 0$ and where we will assume $\beta \geq -0.5$ and that data is generated via the pure random walk, $\Delta y_t = u_t$. In (3) we attempt to capture the effect of the covariate via the proxy variable $\left\{t^\beta\right\}_{t=1}^T$, i.e. we put $d_t = \alpha t^\beta$. The aim is to capture the influence of different asymptotic covariate behaviour, i.e. whether the sequence $\{d_t\}_{t=1}^T$ diverges or converges and at what rate, on measures of inferential accuracy for the interest parameter ρ.

Specifically, when $-0.5 \leq \beta < 0$ then $\left\{t^\beta\right\}$ is an 'evaporating' trend, and captures the effect of an ergodic regressor, in that when H_0 is true $E\left[\Delta y_t\right]$ converges to a constant (zero, in the simplest case). Instead, when $\beta > 0$, $E\left[\Delta y_t\right]$ diverges. For $0 < \beta < 0.5$ Elliott, Rothenberg and Stock [7] term the trend as being 'slowly evolving', although non-stationary. Since a pure random walk has stochastic order $O\left(T^{1/2}\right)$ we might view the covariate trend being dominant if $\beta > 0.5$, and the stochastic trend being dominant if $\beta < 0.5$. The purpose of the following analysis is to detail the effect of the rate of divergence/convergence of the covariate on inference about ρ.

Consider the Score and Hessian for model (3), initially assuming $\sigma^2 = 1$ for simplicity:

$$S(\rho, \alpha) = \begin{pmatrix} \sum_{t=1}^T y_{t-1} u_t \\ \sum_{t=1}^T t^\beta u_t \end{pmatrix} \quad \& \quad H(\rho, \alpha) = -\begin{pmatrix} \sum_{t=1}^T y_{t-1}^2 & \sum_{t=1}^T t^\beta y_{t-1} \\ \sum_{t=1}^T t^\beta y_{t-1} & \sum_{t=1}^T t^{2\beta} \end{pmatrix}. \tag{4}$$

Imposing $\Delta y_t = u_t$ and $y_0 = 0$ then $E\left[y_{t-1}\right] = 0$ and Fisher information is

$$I(\rho, \alpha) = E\left[-H(\rho, \alpha)\right] = \begin{pmatrix} \frac{T(T-1)}{2} & 0 \\ 0 & \sum_{t=1}^T t^{2\beta} \end{pmatrix}.$$

Using this as an inferential metric would be misleading since it would imply no impact of the covariate on inference on ρ.

Instead, note the standard results,

$$\begin{aligned} T^{-1} \sum_{t=1}^T y_{t-1} u_t &\Rightarrow \int_0^1 W(r) dW(r) =_d \left(\chi_1^2 - 1\right)/2, \quad \text{and} \\ T^{-\beta-1/2} \sum_{t=1}^T t^\beta u_t &\Rightarrow \int_0^1 r^\beta dW(r) =_d N\left(0, \int_0^1 r^{2\beta} dr\right), \end{aligned}$$

where $W(r)$ is standard Brownian motion, χ_1^2 denotes a chi-square random variable with one degree of freedom, $\Rightarrow$ denotes weak convergence and $=_d$ denotes equality in distribution. The Score then obeys the following limit,

$$D_T^{-1} S(\rho, \alpha) \Rightarrow \int_0^1 \begin{pmatrix} W(r) \\ r^\beta \end{pmatrix} dW(r),$$

where $D_T = diag\left\{T, T^{\beta+1/2}\right\}$. Expansion of the Score in the Gaussian case yields,

$$S(\rho, \alpha) = S(\hat{\rho}_{MLE}, \hat{\alpha}_{MLE}) + H(\rho, \alpha) \begin{pmatrix} \hat{\rho}_{MLE} - 1 \\ \hat{\alpha}_{MLE} \end{pmatrix},$$

so that,

$$D_T \begin{pmatrix} \hat{\rho}_{MLE} - 1 \\ \hat{\alpha}_{MLE} \end{pmatrix} = -\left(D_T^{-1} H(\rho, \alpha) D_T^{-1}\right)^{-1} D_T^{-1} S(\rho, \alpha).$$

Now

$$\begin{aligned} -\left(D_T^{-1} H(\rho, \alpha) D_T^{-1}\right)^{-1} &= \sigma^2 \begin{pmatrix} T^{-2} \sum_{t=1}^T y_{t-1}^2 & T^{-\beta-3/2} \sum_{t=1}^T t^\beta y_{t-1} \\ T^{-\beta-3/2} \sum_{t=1}^T t^\beta y_{t-1} & T^{(2\beta+1)} \sum_{t=1}^T t^{2\beta} \end{pmatrix}^{-1} \\ &\Rightarrow \sigma^2 \begin{pmatrix} \sigma^2 \int_0^1 W(r)^2 dr & \sigma \int_0^1 r^\beta W(r) dr \\ \sigma \int_0^1 r^\beta W(r) dr & \int_0^1 r^{2\beta} dr \end{pmatrix}^{-1}, \end{aligned}$$

and hence,

$$T(\hat{\rho}_{MLE} - 1) \Rightarrow \left(\int_0^1 W(r)^2 dr - \frac{\left(\int_0^1 r^\beta W(r) dr\right)^2}{\int_0^1 r^{2\beta} dr} \right)^{-1} \left(\int_0^1 W(r) dW(r) \right) \quad (5)$$

and

$$T(\hat{\alpha}_{MLE}) \Rightarrow \left(\int_0^1 r^{2\beta} dr - \frac{\left(\int_0^1 r^\beta W(r) dr\right)^2}{\int_0^1 W(r)^2 dr} \right)^{-1} \left(\int_0^1 r^\beta dW(r) \right). \quad (6)$$

Note that if we define the limit of the scaled Hessian by,

$$D_T^{-1} H(\rho, \alpha) D_T^{-1} \Rightarrow \bar{H}(\rho, \alpha) = \begin{pmatrix} \bar{H}_{\rho\rho} & \bar{H}_{\rho\alpha} \\ \bar{H}_{\rho\alpha} & \bar{H}_{\alpha\alpha} \end{pmatrix}$$

then the quantities scaling the limit distribution of the components of the Score in (5) and (6) are:

$$\bar{H}_{\rho|\alpha} = \left(\bar{H}_{\rho\rho} - \frac{\left(\bar{H}_{\rho\alpha}\right)^2}{\bar{H}_{\alpha\alpha}} \right) \quad \& \quad \bar{H}_{\alpha|\rho} = \left(\bar{H}_{\alpha\alpha} - \frac{\left(\bar{H}_{\rho\alpha}\right)^2}{\bar{H}_{\rho\rho}} \right),$$

so that $\bar{H}_{\rho|\alpha}$ and $\bar{H}_{\alpha|\rho}$ are the stochastic analogues of Conditional Information[1].

Bhapkar and Srinivasan [3] and Zhu and Reid [4] argue that conditional information (2) should form the basis of any efficiency theory, e.g. application of the Cramér-Rao lower bound to any estimator of ρ. In the current context this would fail since $I_{\rho\alpha} = 0$ would wrongly imply that the value of β does not affect the limit distribution of $\hat{\rho}_{MLE}$. On the other hand the stochastic quantity $\bar{H}_{\rho|\alpha}$ depends explicitly on β and should therefore prove informative about inference on ρ, as a function of β.

Indeed, here the limit in (5) can be interpreted as;

$$T\left(\hat{\rho}_{MLE} - 1\right) \Rightarrow \bar{H}_{\rho|\alpha}^{-1} Z,$$

where $Z \sim \left(\chi_1^2 - 1\right)$. Only $\bar{H}_{\rho|\alpha}$ contains any information on the impact of the covariate on the asymptotic distribution of $\hat{\rho}_{MLE}$. It does not however measure its variance directly, since it is correlated with Z.

Specification A is extremely useful in two regards. First, as in Hansen [10], it exposes the effects of even stationary covariates on tests for nonstationarity. Second, here, a sensible stochastic analogue of conditional information arises naturally and its role in the limit distribution is clear. However, the latter applies only by imposing $\alpha = 0$, while generally the distribution of $\hat{\rho}_{MLE}$ *will* depend explicitly upon α, and any other value will produce different, as well as quickly intractable, limit theory. Specification B, on the other hand, allows construction of invariant statistics and in the next Section it will be shown that $\bar{H}_{\rho|\alpha}$ has far wider applicability, in that context.

3 Profile likelihood and information measures

In the context of specification B, suppose that a process $(u_t)_1^T$ is generated according to

$$u_t = \rho u_{t-1} + \varepsilon_t \quad ; \quad \varepsilon_t \sim iid(0, \sigma^2), \tag{7}$$

and we are interested in testing the null hypothesis $H_0 : \rho = 1$, against $H_1 : \rho = 1 - c/T$, for $c > 0$. In the simplest case we assume that the observed time series data $(y_t)_1^T$ is given by $y_t = u_t$, however we explicitly 'de-trend' the observations according to two non-linear trend models;

$$M_1 : y_t = \alpha t^\beta + u_t \quad \& \quad M_2 : y_t = \alpha_0 + \alpha_1 t^\beta + u_t, \tag{8}$$

with $\beta \geq -0.5$.

[1] The author is grateful to an anonymous referee for steps leading to this interpretation.

The purpose is to measure the influence of β on our ability to determine whether or not $(u_t)_1^T$ has a unit root. Let $\hat{\alpha}$, $\hat{\alpha}_0$ and $\hat{\alpha}_1$ denote the OLS estimators for α, α_0 and α_1 in (8), respectively. Unit root tests are constructed from detrended data, $(u_t^*)_1^T$ for M_1 and $\left(u_t^+\right)_1^T$ for M_2, where

$$M_1 : u_t^* = y_t - \hat{\alpha} t^\beta \quad ; \quad M_2 : u_t^+ = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t^\beta.$$

The hypotheses H_0 and H_1 are invariant with respect to the groups of transformations defined, respectively, by

$$G_1 : y \to y + \alpha t^\beta \quad ; \quad G_2 : y \to y + \alpha_0 + \alpha_1 t^\beta. \tag{9}$$

Similar to King [14] and Nielsen [15], the maximal invariants under G_1 and G_2 are $v_1 = C_1 y$ and $v_2 = C_2 y$, where C_j satisfies $C_j' C_j = I_{T-j}$, $C_j C_j' = M_j = I_T - X_j \left(X_j' X_j\right)^{-1} X_j'$ and $X_1 = \left(t^\beta\right)_{t=1}^T$ and $X_2 = \left(1, t^\beta\right)_{t=1}^T$. Defining the vectors $U^* = (u_t^*)_{t=1}^T = M_1 v_1$ and $U^+ = \left(u_t^+\right)_{t=1}^T = M_2 v_2$, then all statistics constructed only from u_t^* $\left(u_t^+\right)$ are invariant, having distributions not depending on α or α_0 and α_1, respectively. In particular, any quantity derived via the imposition of $\alpha = \alpha_0 = \alpha_1 = 0$ will, in the context of specification B, still apply more generally, unlike with specification A.

To measure the effect of the trend parameter β on asymptotic inference we will focus upon likelihood based measures constructed from the Gaussian Profile Likelihood:

$$\tilde{L}(\rho, \sigma^2) = \frac{\exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T \left(\tilde{u}_t - \rho \tilde{u}_{t-1}\right)^2\right\}}{(2\pi\sigma^2)^{T/2}}, \tag{10}$$

where $\tilde{u}_t = u_t^*$ for M_1 and $\tilde{u}_t = u_t^+$ for M_2, with likelihood profiled with respect to the nuisance parameters α or (α_0, α_1), respectively, via OLS. Accordingly, define the following profile measures:

Kullback-Leibler divergence

Define the log-likelihood ratio by

$$LR(\rho) = \ln\left[\frac{\tilde{L}(1, \sigma^2)}{\tilde{L}(\rho, \sigma^2)}\right] = \frac{1}{2\sigma^2}\left[\left(\rho^2 - 1\right) \sum_{t=1}^T \left(\tilde{u}_t\right)^2 - 2\left(\rho - 1\right) \sum_{t=1}^T \tilde{u}_t \tilde{u}_{t-1}\right],$$

then the asymptotic profile Kullback-Leibler divergence is given by

$$KL(\rho) = \lim_{T \to \infty} E_{H_0}\left[LR(\rho)\right].$$

Fisher and conditional information

For specification B the profile Score and Hessian are,

$$S\left(\rho,\sigma^2\right) = \left(\begin{array}{c} \frac{1}{\sigma^2}\sum_{t=1}^{T}\tilde{u}_{t-1}\left(\tilde{u}_t-\rho\tilde{u}_{t-1}\right) \\ \frac{1}{2\sigma^4}\sum_{t=1}^{T}\left(\tilde{u}_t-\rho\tilde{u}_{t-1}\right)^2-\frac{T}{2\sigma^2} \end{array}\right)$$

$$-H\left(\rho,\sigma^2\right) = \left(\begin{array}{cc} \frac{1}{\sigma^2}\sum_{t=1}^{T}\tilde{u}_{t-1}^2 & \frac{1}{\sigma^4}\sum_{t=1}^{T}\tilde{u}_{t-1}\left(\tilde{u}_t-\rho\tilde{u}_{t-1}\right) \\ \frac{1}{\sigma^4}\sum_{t=1}^{T}\tilde{u}_{t-1}\left(\tilde{u}_t-\rho\tilde{u}_{t-1}\right) & \frac{1}{\sigma^6}\sum_{t=1}^{T}\tilde{\varepsilon}_t^2-\frac{T}{2\sigma^4} \end{array}\right).$$

The Gaussian profile MLEs satisfy,

$$\tilde{D}_T\left(\begin{array}{c} \hat{\rho}_{PMLE}-\rho \\ \hat{\sigma}^2_{PMLE}-\sigma^2 \end{array}\right) = \left(-\tilde{D}_T^{-1}H\left(\rho,\sigma^2\right)\tilde{D}_T^{-1}\right)^{-1}\tilde{D}_T^{-1}S\left(\rho,\sigma^2\right). \tag{11}$$

Imposing $\rho=1$, noting $\tilde{u}_T=O_p\left(T^{1/2}\right)$ and $\tilde{\varepsilon}_T=\Delta\tilde{u}_T=O_p(1)$, then the limit of the scaled Hessian, $\bar{H}\left(\rho,\sigma^2\right)=\lim_{T\to\infty}\tilde{D}_T^{-1}H\left(\rho,\sigma^2\right)\tilde{D}_T^{-1}$, where $\tilde{D}_T=diag\left\{T,T^{1/2}\right\}$, is diagonal, since $T^{-3/2}\sum_{t=1}^{T}\tilde{u}_{t-1}\tilde{\varepsilon}_t=o_p(1)$, as is its expectation. Asymptotic Fisher information in $(\tilde{u}_t)_{t=1}^T$ about ρ when $\Delta y_t=\varepsilon_t$, is

$$\tilde{I}_1(\beta)=\lim_{T\to\infty}E\left[\frac{T^{-2}}{\sigma^2}\sum_{t=1}^{T}\tilde{u}_{t-1}^2\right],$$

and conditional information in ρ given σ^2 is equal to Fisher information, in this case, i.e. $CI_{\rho|\sigma^2}=\tilde{I}_1(\beta)$.

Before proceeding we will require limiting forms for the OLS estimators of the nuisance parameters $\hat{\alpha}$ and $(\hat{\alpha}_0,\hat{\alpha}_1)$ when $\Delta u_t=\varepsilon_t$. These generalize results found in Durlauf and Phillips [16] and are given in the following Lemma, proved in Appendix I.

Lemma 1: *Let* $y_t:=u_t=u_{t-1}+\varepsilon_t$, $\varepsilon_t\sim iid(0,\sigma^2)$,

$$\begin{aligned} T^{\beta-1/2}\hat{\alpha} &\Rightarrow (2\beta+1)\int_0^1 r^\beta W_\sigma(r)dr \\ T^{\beta-1/2}\hat{\alpha}_1 &\Rightarrow Q_1(\beta)=\frac{(2\beta+1)(\beta+1)^2}{\beta^2}\left[\int_0^1\left(r^\beta-\frac{1}{\beta+1}\right)W_\sigma(r)dr\right] \\ T^{-1/2}\hat{\alpha}_0 &\Rightarrow Q_0(\beta)=\frac{-(2\beta+1)(\beta+1)}{\beta^2}\times \\ &\quad\left[\int_0^1\left(r^\beta-\frac{\left(\beta^4-(2\beta+1)^2(\beta+1)\right)}{\beta^2(2\beta+1)(\beta+1)}\right)W_\sigma(r)dr\right], \end{aligned}$$

where $W_\sigma(r)=_d\sigma W(r)$. ■

Note that, as is well known, $\hat{\alpha}_0$ is never consistent, while neither of $\hat{\alpha}$ or $\hat{\alpha}_1$ are if $\beta < 0.5$. This, for $\hat{\alpha}$, contrasts with the limit for $\hat{\alpha}_{MLE}$ implied by (6) and which could be generalized for $\alpha \neq 0$, if α were the interest parameter, for instance.

Applying the results of Lemma 1 to the appropriate profile likelihood yields explicit expressions for the profile Kullback-Leibler, Fisher and conditional information as given below. For each model we find that these are all asymptotically equivalent and depend upon the degree of trending, β, in exactly the same way. The findings are summarized in the following theorem, which is also proved in Appendix I.

Theorem 1: Part I) *Let $y_t := u_t = u_{t-1} + \varepsilon_t$ and suppose that we de-trend y_t according to M_1, with $u_t^* = y_t - \hat{\alpha} t^\beta$, then:*

(a)

$$\begin{aligned} T^{-2}\sum_1^T \left(u_{t-1}^*\right)^2 &\Rightarrow \int_0^1 W_\sigma^2(r)dr - (2\beta+1)\left(\int_0^1 r^\beta W_\sigma(r)dr\right)^2 \\ T^{-2}\sum_1^T u_t^* u_{t-1}^* &\Rightarrow \int_0^1 W_\sigma^2(r)dr - (2\beta+1)\left(\int_0^1 r^\beta W_\sigma(r)dr\right)^2. \end{aligned}$$

(b) Letting $\rho = 1 - c/T$, for $c > 0$, we have

$$\begin{aligned} I_1^*(\beta) &= CI_{1|\sigma^2}^* = \frac{1}{2}\left[\frac{2\beta^2 - \beta + 2}{(2\beta+3)(\beta+2)}\right] \\ KL^*(\beta) &= \frac{c^2}{4}\left[\frac{2\beta^2 - \beta + 2}{(2\beta+3)(\beta+2)}\right]. \end{aligned}$$

(c) In model M_1 Kullback-Leibler divergence and, therefore both information measures, are minimized for trends of the form t^{β^*}, where $\beta^* = \left(\sqrt{6} - 1\right)/2$.

Part II) *Now let $y_t := u_t = u_{t-1} + \varepsilon_t$ and suppose that we de-trend y_t according to M_2, with $u_t^+ = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t^\beta$, then:*

(a) Both $T^{-2}\sum_1^T \left(u_{t-1}^+\right)^2$ and $T^{-2}\sum_1^T u_t^+ u_{t-1}^+$ have the same asymptotic stochastic representation, with

$$\begin{aligned} T^{-2}\sum_1^T \left(u_{t-1}^+\right)^2 \Rightarrow & \int_0^1 W_\sigma(r)^2 dr - \left(\int_0^1 W_\sigma(r)dr\right)^2 \\ & - \frac{(2\beta+1)(\beta+1)^2}{\beta^2}\left[\int_0^1 r^\beta W_\sigma(r)dr - \frac{1}{\beta+1}\int_0^1 W_\sigma(r)dr\right]^2. \end{aligned}$$

(b) Letting $\rho = 1 - c/T$, for $c > 0$, we have

$$\begin{aligned} I_1^+(\beta) &= CI^+_{1|\sigma^2} = \frac{1}{6}\left[\frac{\left(2\beta^2+\beta+5\right)}{(2\beta+3)(\beta+3)}\right] \\ KL^+(\beta) &= \frac{c^2}{12}\left[\frac{\left(2\beta^2+\beta+5\right)}{(2\beta+3)(\beta+3)}\right]. \end{aligned}$$

(c) In model M_2 Kullback-Leibler divergence and, therefore both information measures, are minimized for trends of the form t^{β^+}, where $\beta^+ = \left(\sqrt{10}-1\right)/2$. ■

4 Discussion and Analysis

1) Returning to the original Dickey-Fuller [13] Model (i.e. specification A in (1)), then we find that the expectation of the limit of the Conditional Hessian is

$$\begin{aligned} E\left[\lim_{T\to\infty} H_{\rho|\alpha}\right] &= E\left[\int_0^1 W^2(r)\,dr - (2\beta+1)\left(\int_0^1 r^\beta W(r)\,dr\right)\right]^2 \\ &= \frac{1}{2}\left[\frac{2\beta^2-\beta+2}{(2\beta+3)(\beta+2)}\right] = I_1^*(\beta). \end{aligned}$$

That is the measure of conditional information derived for specification A is identical to profile Fisher information in specification B. This finding can generalized, at some considerable algebraic cost, to the case of $d_t = \alpha_0 + \alpha t^\beta$.

2) In all cases it is clear that the covariate is relevant for inference on ρ, whether it is evaporating or nonstationary, whether slowly evolving or explosive. For instance, in M_1 with $\beta = 0$, we have $I_1^*(0) = CI^*_{1|\sigma^2} = 1/6$, and $KL^*(0) = c^2/12$. The outcomes can also be compared with the benchmark of a pure random walk (i.e. the likelihood does not need profiling), in which case we find $I_1 = 1/2$ and $KL = c^2/4$. In the case of M_1, $I_1^*(\beta) < 1/2$ for all $-0.5 < \beta < \infty$, although $I_1^*(-0.5) = 1/2$ and $\lim_{\beta\to\infty} I_1^*(\beta) = 1/2$. That is, profiling with respect to the limiting evaporating or explosive covariate has, effectively, no effect on information. For M_2 the benchmark case can be taken as M_1 with $\beta = 0$. Once again we find $I_1^+(\beta) < 1/6$ for all $-0.5 < \beta < \infty$, but $I_1^+(-0.5) = 1/6$ and $\lim_{\beta\to\infty} I_1^+(\beta) = 1/6$.

3) In order to demonstrate that these findings are genuinely informative about the effect of regressing out t^β on unit root inference we examine the power envelope. Adding scale invariance to the groups of transformations G_1 and G_2 defined above, then from King [14] the maximal invariant (under (9)) for testing $H_0 : \rho = 1$ in (7) is $v_j = C_j' y/\sqrt{y' M_j y}$, where C_j and M_j are defined above. The statistic v_j has density

(up to normalized Haar measure on the surface of the unit $T-j$ sphere), as

$$f(v_j;\rho) = |A_{j,\rho}|^{-1/2}\left(v' A_{j,\rho}^{-1} v\right)^{-\frac{(T-j)}{2}} \quad ; \quad A_{j,\rho} = C_j'\left(\Delta_\rho^{-1}\right)\left(\Delta_\rho^{-1}\right)' C_j,$$

where $\Delta_\rho = I - \rho L$, and L is the lag-operator matrix. The Neyman-Pearson tests for H_0 against the alternative $H_1 : \rho \neq 1$ are to reject H_0 if

$$NP_j = \frac{v_j' A_{j,\rho}^{-1} v_j}{v_j' A_{j,1}^{-1} v_j} < k_\delta, \tag{12}$$

where k_δ is chosen so that the size is δ.

In Table 1 (in Appendix II) the resulting power envelope was simulated for $T = 250$, for $\rho = 1 - c/T$ with $c = 1, 2, .., 10$ and for different values of β. The simulations were carried out with 2 million replications. Note that $\beta = T$ is used to approximate the limiting case of $\beta \to \infty$. In Table 1 a clear prediction is supported; in M_1 power is not maximized when $\beta = 0$, detrending with respect to an evaporating trend can imply as much or even more power. It is not quite possible, in this context, to confirm the prediction that $\beta^*$and β^+ minimize power. This is for two reasons. Firstly the powers are clearly very close and insignificantly different even with two million replications. Second the properties of the power envelope are determined by behavior of tests under both the null and alternative, whereas Theorem 1 applies only under the null.

4) Instead, consider the profile maximum likelihood estimators for ρ in M_1 and M_2,

$$\hat{\rho}_1 = \frac{\sum_{t=2}^T u_t^* u_{t-1}^*}{\sum_{t=2}^T \left(u_{t-1}^*\right)^2} \quad \& \quad \hat{\rho}_2 = \frac{\sum_{t=2}^T u_t^+ u_{t-1}^+}{\sum_{t=2}^T \left(u_{t-1}^+\right)^2},$$

where u_t^* and u_t^+ are defined above. Figures 1 and 2, in Appendix II, plot the simulated (with $T = 250$ and two million replications) variances of $T\left(\hat{\rho}_1 - 1\right)$and $T\left(\hat{\rho}_2 - 1\right)$, respectively, for different values of the trend parameter β. Plotted also are vertical lines at β^* and β^+. These figures help confirm, finally, the third prediction that there is a value which minimizes the inferential accuracy and, crucially, this value is not equal to 1.

5 Conclusions

This paper argues that likelihood based measures of information and efficiency remain informative about inferential accuracy even in regressions involving nonstationary data. This, even though such models obey none of the required assumptions for consistent and efficient, asymptotically normal estimation.

The equivalency of conditional information in a lagged dependent variable justifies use of the simpler Kullback-Leibler, or Fisher information applied to profile likelihood in the case of unit root inference in the presence of a general covariate. These are informative, in that clear predictions including maximum inferential efficiency for 'evaporating' trends and minimum efficiency for fractional, not linear, trends are clearly supported through numerical experiment.

References

[1] Magdalinos T. On the inconsistency of the unrestricted estimator of the information matrix near a unit root. Econometrics Journal. 2007;10:245–262.

[2] Marsh P. The properties of Kullback-Leibler divergence for the unit root hypothesis. Econometric Theory. 2009; 25:1662-1681.

[3] Bhapkar VP, Srinivasan C. On Fisher information inequalities in the presence of nuisance parameters. Ann. Inst. Statist. Math. 1994;46:593-604.

[4] Zhu Y, Reid N. Information, ancillarity and sufficiency in the presence of nuisance parameters. Canadian Journal of Statistics. 1994;22:111-123.

[5] Gibbs, AL, Su, FE (2002): Choosing and bounding probability metrics. International Statistical Review. 2002;70:419-435.

[6] Critchley F, Marriott P, Salmon M. Preferred point geometry and statistical manifolds. Annals of Statistics. 1993;21:1197-1224.

[7] Elliott G, Rothenberg TJ, Stock JH. Efficient tests for an autoregressive unit root. Econometrica. 1996;64:813-836.

[8] Marsh P. The available information for invariant tests of a unit root. Econometric Theory. 2007;23:686-710.

[9] Phillips PCB. Regression with slowly varying regressors and nonlinear trends. Econometric Theory. 2007;23:557-614.

[10] Hansen BE. Rethinking the univariate approach to unit root testing: using covariates to increase power. Econometric Theory. 1995;11:1148-1171.

[11] Elliott G, Jansson M. Testing for unit roots with stationary covariates. Journal of Econometrics. 2003;115:75-89.

[12] Chrystalleni, A, Harvey DI, and Leybourne SJ. The Impact of the initial condition on covariate augmented unit root tests. Journal of Time Series Econometrics. 2017;9.

[13] Dickey DA and Fuller WA. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. J.A.S.A. 1979;366: 427-431.

[14] King, ML. Robust tests for spherical symmetry and their application to least squares regression. Annals of Statistics. 1980;8:1265–1271.

[15] Nielsen, B. Power of tests for unit roots in the presence of a linear trend. Oxford Bulletin of Economics and Statistics. 2008;70:619-644.
[16] Durlauf, SN, Phillips PCB. Trends versus random walks in time series analysis. Econometrica, 1988;56:1333-1354.

Appendix I

Proof of Lemma 1:

Since $u_t = \sum_1^t \varepsilon_j$, then first note first the following standard results $T^{-3/2}\sum_1^T u_t \Rightarrow \int_0^1 W_\sigma(r)dr$, $T^{-(\beta+3/2)}\sum_1^T t^\beta u_t \Rightarrow \int_0^1 r^\beta W_\sigma(r)dr$ and $T^{-(\beta+1)}\sum_1^T t^\beta \to \int_0^1 r^\beta dr = \frac{1}{\beta+1}$, which can then be immediately applied to the OLS estimators. Immediately, therefore we have,

$$T^{\beta-1/2}\hat{\alpha} = \frac{T^{-(\beta+3/2)}\sum_1^T y_t t^\beta}{T^{-(2\beta+1)}\sum_1^T t^{2\beta}} \Rightarrow \frac{\int_0^1 r^\beta W_\sigma(r)dr}{\int_0^1 r^{2\beta+1}dr} = (2\beta+1)\int_0^1 r^\beta W_\sigma(r)dr.$$

Then,

$$\begin{aligned}
T^{\beta-1/2}\hat{\alpha}_1 &= \frac{T^{-(\beta+3/2)}\sum_1^T y_t t^\beta - T^{-3/2}\sum_1^T y_t T^{-(\beta+1)}\sum_1^T t^\beta}{T^{-(2\beta+1)}\sum_1^T t^{2\beta} - T^{-2(\beta+1)}\left(\sum_1^T t^\beta\right)^2} \\
&\Rightarrow \frac{\int_0^1 r^\beta W_\sigma(r)dr - \int_0^1 W_\sigma(r)dr\int_0^1 r^\beta dr}{\int_0^1 r^{2\beta+1}dr - \left(\int_0^1 r^\beta dr\right)^2} \\
&\equiv \frac{(2\beta+1)(\beta+1)^2}{\beta^2}\left[\int_0^1 r^\beta W_\sigma(r)dr - \frac{1}{\beta+1}\int_0^1 W_\sigma(r)dr\right] \\
&\equiv \frac{(2\beta+1)(\beta+1)^2}{\beta^2}\int_0^1 \left(r^\beta - \frac{1}{\beta+1}\right) W_\sigma(r)dr \equiv Q_1(\beta)
\end{aligned}$$

as required and finally,

$$\begin{aligned}
T^{-1/2}\hat{\alpha}_0 &= T^{-3/2}\sum_1^T u_t - T^{\beta-1/2}\hat{\alpha}_1 T^{-(\beta+1)}\sum_1^T t^\beta \\
&\Rightarrow \int_0^1 W_\sigma(r)dr - Q_1(\beta)\int_0^1 r^\beta dr \\
&\equiv \int_0^1 W_\sigma(r)dr - \frac{(2\beta+1)(\beta+1)}{\beta^2}\int_0^1 \left(r^\beta - \frac{1}{\beta+1}\right) W_\sigma(r)dr \\
&\equiv \frac{-(2\beta+1)(\beta+1)}{\beta^2}\left[\int_0^1 \left(r^\beta - \frac{\left(\beta^4 - (2\beta+1)^2(\beta+1)\right)}{\beta^2(2\beta+1)(\beta+1)}\right) W_\sigma(r)dr\right]. \quad \blacksquare
\end{aligned}$$

Proof of Theorem 1:

Part I): a) The OLS detrended data is

$$u_t^* = y_t - \hat{\alpha} t^\beta = y_t - \frac{\sum_1^T y_t t^\beta}{\sum_1^T t^{2\beta}} t^\beta,$$

so that when $y_t := u_t = u_{t-1} + \varepsilon_t$

$$\begin{aligned} \sum_1^T \left(u_{t-1}^*\right)^2 &= \sum_1^T \left(u_{t-1} - \frac{\sum_1^T u_{t-1}(t-1)^\beta}{\sum_1^T (t-1)^{2\beta}} (t-1)^\beta\right)^2 \\ &= \sum_1^T u_{t-1}^2 - \frac{\left(\sum_1^T u_{t-1}(t-1)^\beta\right)^2}{\sum_1^T (t-1)^{2\beta}}. \end{aligned}$$

Since, $T^{-2} \sum_1^T u_{t-1}^2 \Rightarrow \int_0^1 W_\sigma^2(r)dr$, $\lim_{T\to\infty} T^{-(2\beta+1)} \sum_1^T (t-1)^{2\beta} \to (2\beta+1)^{-1}$ and $T^{-(\beta+3/2)} \sum_1^T u_{t-1}(t-1)^\beta \Rightarrow \int_0^1 r^\beta W_\sigma(r)dr$ then

$$T^{-2} \sum_1^T \left(u_{t-1}^*\right)^2 \Rightarrow \int_0^1 W_\sigma^2(r)dr - (2\beta+1)\left(\int_0^1 r^\beta W_\sigma(r)dr\right)^2.$$

Similarly, we have

$$\begin{aligned} \sum_1^T u_t^* u_{t-1}^* &= \sum_1^T \left(u_t - \frac{\sum_1^T u_t t^\beta}{\sum_1^T t^{2\beta}} t^\beta\right)\left(u_{t-1} - \frac{\sum_1^T u_{t-1}(t-1)^\beta}{\sum_1^T (t-1)^{2\beta}} (t-1)^\beta\right) \\ &= \sum_1^T u_t u_{t-1} + \frac{\left(\sum_1^T u_t t^\beta\right)\left(\sum_1^T u_{t-1}(t-1)^\beta\right)\sum_1^T t^\beta (t-1)^\beta}{\sum_1^T t^{2\beta} \sum_1^T (t-1)^{2\beta}} \\ &\quad - \frac{\left(\sum_1^T u_t (t-1)^\beta\right)\left(\sum_1^T u_{t-1}(t-1)^\beta\right)}{\sum_1^T (t-1)^{2\beta}} - \frac{\left(\sum_1^T u_{t-1} t^\beta\right)\left(\sum_1^T u_t t^\beta\right)}{\sum_1^T t^{2\beta}}. \end{aligned}$$

Consequently, using

$$\begin{aligned} T^{-2} \sum_1^T u_t u_{t-1} &= T^{-2}\left(\sum_1^T u_{t-1}^2 + \sum_1^T \varepsilon_t u_{t-1}\right) = T^{-2} \sum_1^T u_{t-1}^2 + o_p(1) \\ &\Rightarrow \int_0^1 W_\sigma^2(r)dr, \end{aligned} \tag{13}$$

as well as

$$\begin{aligned} T^{-(\beta+3/2)} \sum_1^T u_t (t-1)^\beta &= T^{-(\beta+3/2)} \sum_1^T u_{t-1}(t-1)^\beta + o_p(1) \\ &\Rightarrow \int_0^1 r^\beta W_\sigma(r)dr, \end{aligned} \tag{14}$$

and similar for $T^{-(\beta+3/2)}\sum_1^T u_{t-1}t^\beta$, and $\lim_{T\to\infty} T^{-(2\beta+1)}\sum_1^T t^\beta(t-1)^\beta = (2\beta+1)^{-1}$, then we have,

$$T^{-2}\sum_1^T u_t^* u_{t-1}^* \Rightarrow \int_0^1 W_\sigma^2(r)dr - (2\beta+1)\left(\int_0^1 r^\beta W_\sigma(r)dr\right)^2.$$

(b) Since Kullback-Leibler divergence is defined as $KL = E\left[LR\right]$, where

$$LR = \frac{1}{2\sigma^2}\left[\left(\rho^2-1\right)\sum_1^T \left(u_t^*\right)^2 - 2\left(\rho-1\right)\sum_1^T u_t^* u_{t-1}^*\right],$$

and expectations are taken under the unit root null. Consequently, since $T^{-2}\sum_1^T \left(u_t^*\right)^2$ and $T^{-2}\sum_1^T u_t^* u_{t-1}^*$ have the same asymptotic representation, then we have

$$T^{-2}LR \Rightarrow \frac{1}{2}\left[(1-\rho)^2\left(\int_0^1 W^2(r)dr - (2\beta+1)\left(\int_0^1 r^\beta W(r)dr\right)^2\right)\right],$$

or letting $\rho = 1 - c/T$,

$$LR \Rightarrow \frac{c^2}{2}\left[\int_0^1 W^2(r)dr - (2\beta+1)\left(\int_0^1 r^\beta W(r)dr\right)^2\right].$$

Since,

$$E\left[\int_0^1 W^2(r)dr\right] = \frac{1}{2},$$

and

$$\begin{aligned}
E\left[\left(\int_0^1 r^\beta W(r)dr\right)^2\right] &= E\left[\int_0^1 r^\beta W(r)dr\int_0^1 r^s W(s)ds\right] \\
&= E\left[\int_0^1\int_0^1 r^\beta s^\beta W\left(r\right)W\left(s\right)dsdr\right] \\
&= \int_0^1 r^\beta\int_0^r s^{\beta+1}dsdr + \int_0^1 r^{\beta+1}\int_0^r s^\beta dsdr \\
&= \frac{2}{\left(\beta+2\right)\left(2\beta+3\right)},
\end{aligned}$$

since $E\left[W\left(r\right)W\left(s\right)\right] = \min[r,s]$, and so,

$$KL = \frac{c^2}{2}\left[\frac{1}{2} - \frac{2\left(2\beta+1\right)}{\left(2\beta+3\right)\left(\beta+2\right)}\right],$$

which when rearranged gives the expression as in the statement of the Theorem.

For the information measures, we have, immediately that Fisher information is

$$\begin{aligned} I_1 &= CI_{1|\sigma^2} = \lim_{T\to\infty} E\left[\frac{1}{2}T^{-2}\sum_1^T \left(u_{t-1}^*\right)^2\right] \\ &= E\left[\int_0^1 W^2(r)dr - (2\beta+1)\left(\int_0^1 r^\beta W(r)dr\right)^2\right] = \left[\frac{1}{2} - \frac{2(2\beta+1)}{(2\beta+3)(\beta+2)}\right]. \end{aligned}$$

(c) Immediate from the definition of KL^*.

Part II)

a) For M_2 and with $y_t = u_t$ we have

$$u_t^+ = u_t - \hat{\alpha}_0 - \hat{\alpha}_1 t^\beta,$$

where $\hat{\alpha}_0$ and $\hat{\alpha}_0$ are defined above. Using well known results for the simplest OLS regression, we define

$$\tilde{u}_t = u_t - T^{-1}\sum_1^T u_t \quad \text{and} \quad \tilde{t} = t^\beta - T^{-1}\sum_1^T t^\beta,$$

so that we can write

$$\hat{\alpha}_1 = \frac{\sum_1^T \tilde{u}_t\tilde{t}}{\sum_1^T \tilde{t}^2} \quad \text{and} \quad u_t^+ = \tilde{u}_t - \frac{\sum_1^T \tilde{u}_t\tilde{t}}{\sum_1^T \tilde{t}^2}\tilde{t},$$

so that

$$\sum_1^T \left(u_t^+\right)^2 = \sum_1^T\left(\tilde{u}_t - \frac{\sum_1^T \tilde{u}_t\tilde{t}}{\sum_1^T \tilde{t}^2}\tilde{t}\right)^2 = \sum_1^T \tilde{u}_t^2 - \frac{\left(\sum_1^T \tilde{u}_t\tilde{t}\right)^2}{\sum_1^T \tilde{t}^2}.$$

Using results found in the proof of (a), we first find

$$\begin{aligned} T^{-2}\sum_1^T \tilde{u}_t^2 &= T^{-2}\sum_1^T u_t^2 - T^{-3}\left(\sum_1^T u_t\right)^2 \\ &\Rightarrow \int_0^1 W_\sigma(r)^2 dr - \left(\int_0^1 W_\sigma(r)dr\right)^2. \end{aligned} \tag{15}$$

Also we have

$$\begin{aligned} T^{-(\beta+3/2)}\sum_1^T \tilde{u}_t\tilde{t} &= T^{-(\beta+3/2)}\sum_1^T\left(u_t - T^{-1}\sum_1^T u_t\right)\left(t^\beta - T^{-1}\sum_1^T t^\beta\right) \\ &= T^{-(\beta+3/2)}\sum_1^T t^\beta u_t - T^{-(\beta+1)}\sum_1^T t^\beta T^{-3/2}\sum_1^T u_t \\ &\Rightarrow \int_0^1 r^\beta W_\sigma(r)dr - \frac{1}{\beta+1}\int_0^1 W_\sigma(r)dr, \end{aligned} \tag{16}$$

while

$$T^{-(2\beta+1)}\sum_{1}^{T}\tilde{t}^{2}=T^{-(2\beta+1)}\sum_{1}^{T}\left(t^{\beta}-T^{-1}\sum_{1}^{T}t^{\beta}\right)^{2}\rightarrow\frac{(2\beta+1)\,(\beta+1)^{2}}{\beta^{2}}. \tag{17}$$

Consequently, combining (15), (16) and (17), we have

$$\begin{aligned} T^{-2}\sum_{1}^{T}\left(u_{t}^{+}\right)^{2} &\Rightarrow \int_{0}^{1}W_{\sigma}(r)^{2}dr-\left(\int_{0}^{1}W_{\sigma}(r)dr\right)^{2} \\ &\quad -\frac{(2\beta+1)\,(\beta+1)^{2}}{\beta^{2}}\left[\int_{0}^{1}r^{\beta}W_{\sigma}(r)dr-\frac{1}{\beta+1}\int_{0}^{1}W_{\sigma}(r)dr\right]^{2} \\ &= \left\{\int_{0}^{1}W_{\sigma}(r)^{2}dr-\left(1+\frac{2\beta+1}{\beta^{2}}\right)\left(\int_{0}^{1}W_{\sigma}(r)dr\right)^{2}\right. \\ &\quad -\frac{(2\beta+1)\,(\beta+1)^{2}}{\beta^{2}}\left(\int_{0}^{1}r^{\beta}W_{\sigma}(r)dr\right)^{2} \\ &\quad \left.+\frac{2\,(2\beta+1)\,(\beta+1)}{\beta^{2}}\int_{0}^{1}r^{\beta}W_{\sigma}(r)dr\int_{0}^{1}W_{\sigma}(r)dr\right\}. \end{aligned} \tag{18}$$

Once again it is straight forward to show that $T^{-2}\sum_{1}^{T}u_{t}^{+}u_{t-1}^{+}$ has the same asymptotic limit, via

$$\sum_{1}^{T}u_{t}^{+}u_{t-1}^{+}=\sum_{1}^{T}\left(\tilde{u}_{t}-\frac{\sum_{1}^{T}\tilde{u}_{t}\tilde{t}^{\beta}}{\sum_{1}^{T}\tilde{t}^{2\beta}}\tilde{t}^{\beta}\right)\left(\tilde{u}_{t-1}-\frac{\sum_{1}^{T}\tilde{u}_{t-1}(\tilde{t}_{-1})^{\beta}}{\sum_{1}^{T}(\tilde{t}_{-1})^{2\beta}}(\tilde{t}_{-1})^{\beta}\right),$$

where $\tilde{t}_{-1}=(t-1)^{\beta}-T^{-1}\sum_{1}^{T}t^{\beta}$, and using the results in equations (13) and (14), above.

Again, to calculate Kullback-Leibler divergence, we require $\lim_{T\rightarrow\infty}E\left[\frac{T^{-2}}{\sigma^{2}}\sum_{1}^{T}\left(u_{t}^{+}\right)^{2}\right]$. As above, we have

$$E\left[\int_{0}^{1}W(r)^{2}dr\right] = \frac{1}{2},\quad E\left[\left(\int_{0}^{1}r^{\beta}W(r)dr\right)^{2}\right]=\frac{2}{(2\beta+3)(\beta+2)}\quad\text{and also}$$

$$E\left[\left(\int_{0}^{1}W(r)dr\right)^{2}\right] = \frac{1}{3}.$$

For the remaining expectation, consider

$$\lim_{T\rightarrow\infty}E\left[\left(T^{-(\beta+3/2)}\sum_{1}^{T}t^{\beta}u_{t}\right)\left(T^{-(\gamma+3/2)}\sum_{1}^{T}t^{\gamma}u_{t}\right)\right]=E\left[\int_{0}^{1}r^{\beta}W(r)dr\int_{0}^{1}r^{\gamma}W(r)dr\right]. \tag{19}$$

We can write $\sum_1^T t^\beta u_t \sum_1^T t^\gamma u_t = \sum_{t,s=1}^T t^\beta s^\gamma u_t u_s$, so that noting that $E[u_t u_s] = \min[s,t]$, we have

$$\begin{aligned} T^{-(\beta+\gamma+3)} E\left[\sum_1^T t^\beta u_t \sum_1^T t^\gamma u_t\right] &= T^{-(\beta+\gamma+3)}\left(\sum_{t=1}^T \sum_{s=1}^t t^\beta s^{\gamma+1} + \sum_{t=1}^T \sum_{s=t+1}^T t^{\beta+1} s^\gamma\right) \\ &= T^{-(\beta+\gamma+3)}\left(\sum_{t=1}^T \frac{t^{\beta+\gamma+2}}{\gamma+2} + \sum_{t=1}^T \frac{t^{\beta+1}}{\gamma+1}\left[T^{\gamma+1} - t^{\gamma+1}\right]\right), \end{aligned}$$

and then

$$\lim_{T\to\infty} T^{-(\beta+\gamma+3)} E\left[\sum_1^T t^\beta u_t \sum_1^T t^\gamma u_t\right] = \frac{\gamma+\beta+4}{(\beta+2)(\gamma+2)(\beta+\gamma+3)}. \tag{20}$$

Consequently, we have

$$KL^+ = \lim_{T\to\infty} E\left[\frac{1}{2\sigma^2}\left[\left(\rho^2-1\right)\sum_{t=1}^T \left(u_t^+\right)^2 - 2(\rho-1)\sum_{t=1}^T u_t^+ u_{t-1}^+\right]\right],$$

so that with $\rho = (1 - c/T)$, and using both (19) and (20) in (18) we have

$$\begin{aligned} KL^+ &= \frac{c^2}{2} E\left[\int_0^1 W(r)^2 dr - \left(1 + \frac{2\beta+1}{\beta^2}\right)\left(\int_0^1 W(r)dr\right)^2 \left(\int_0^1 W(r)dr\right)^2\right. \\ &\quad - \frac{(2\beta+1)(\beta+1)^2}{\beta^2}\left(\int_0^1 r^\beta W_\sigma(r)dr\right)^2 \\ &\quad \left.- \frac{2(2\beta+1)(\beta+1)}{\beta^2}\int_0^1 r^\beta W_\sigma(r)dr \int_0^1 W_\sigma(r)dr\right] \\ &= \frac{c^2}{2}\left(\frac{1}{2} - \frac{1}{3}\left(1 + \frac{2\beta+1}{\beta^2}\right) - \frac{2(2\beta+1)(\beta+1)^2}{\beta^2(2\beta+3)(\beta+2)}\right) \\ &\quad + \frac{c^2}{2}\left(\frac{2(2\beta+1)(\beta+1)}{\beta^2}\frac{\beta+4}{2(\beta+2)(\beta+3)}\right) \\ &= \frac{c^2\left(2\beta^2+\beta+5\right)}{12(\beta+3)(2\beta+3)}. \end{aligned}$$

Moreover, by arguments almost identical to those given above in the proof of Part I; for M_2 we have

$$I_1^+ = CI_{1|\sigma^2} = \lim_{T\to\infty} E\left[T^{-2}\sum_{t=1}^T \left(u_{t-1}^+\right)^2\right] = \frac{\left(2\beta^2+\beta+5\right)}{6(\beta+3)(2\beta+3)}.$$

(c) Immediate from the definition of KL^+. ■

Appendix II

Table 1: Power Envelopes for Models $M_1$and M_2 with different trends

	M_1					M_2					Pure $AR(1)$
c \ β	0	-0.5	T	1	β^*	0	-0.5	T	1	β^+	
1	.080	.080	.080	.052	.053	.079	.077	.076	.054	.054	.080
2	.121	.122	.121	.062	.062	.119	.115	.116	.061	.061	.122
3	.172	.178	.174	.074	.075	.169	.154	.167	.073	.073	.178
4	.234	.244	.236	.095	.094	.225	.198	.226	.093	.093	.246
5	.300	.321	.317	.114	.115	.298	.258	.289	.109	.109	.323
6	.365	.416	.406	.143	.145	.365	.311	.353	.141	.141	.417
7	.457	.520	.508	.178	.178	.448	.373	.434	.169	.169	.520
8	.528	.611	.596	.219	.217	.521	.450	.515	.214	.214	.610
9	.591	.687	.673	.262	.258	.579	.512	.593	.260	.259	.687
10	.668	.770	.754	.326	.317	.658	.569	.654	.296	.297	.771

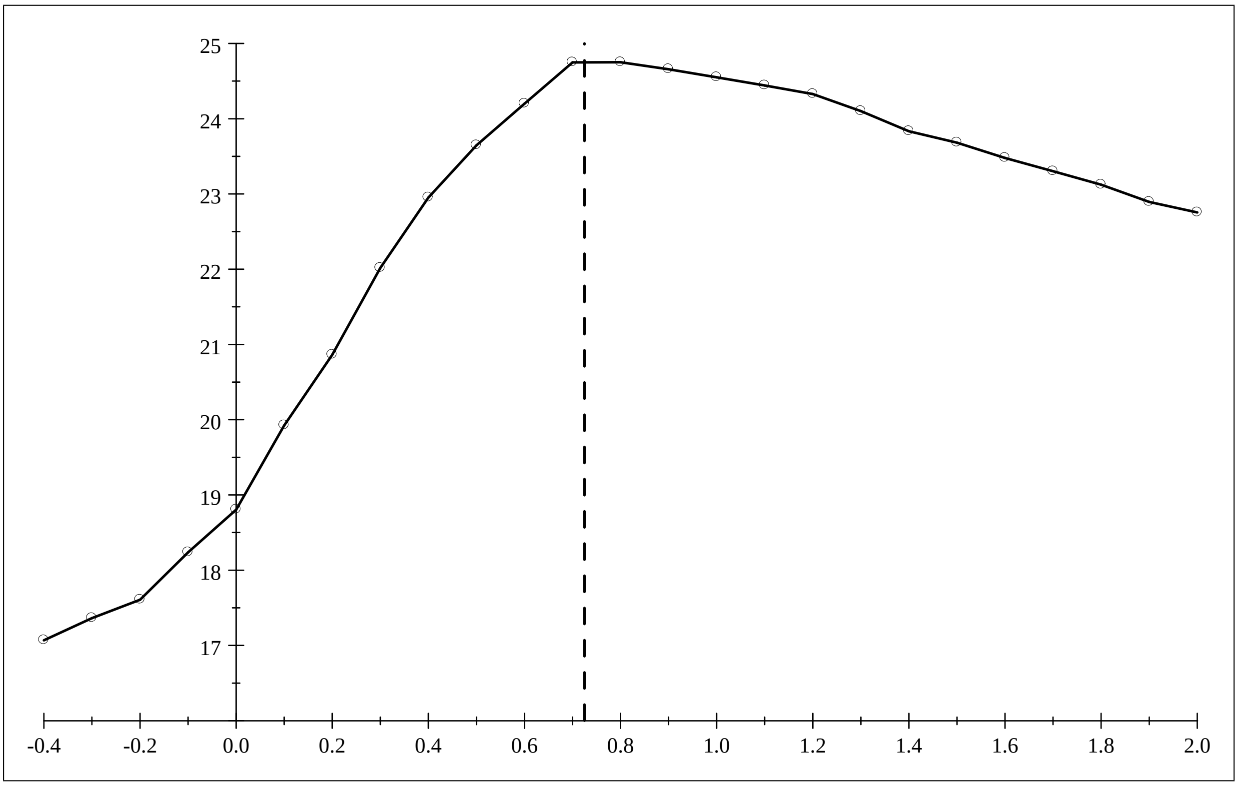

Figure 1: Plot of $V[\hat{\rho}_1] \times 10000$ vs. β. Vertical line at $\beta^* = \left(\sqrt{6} - 1\right)/2$.

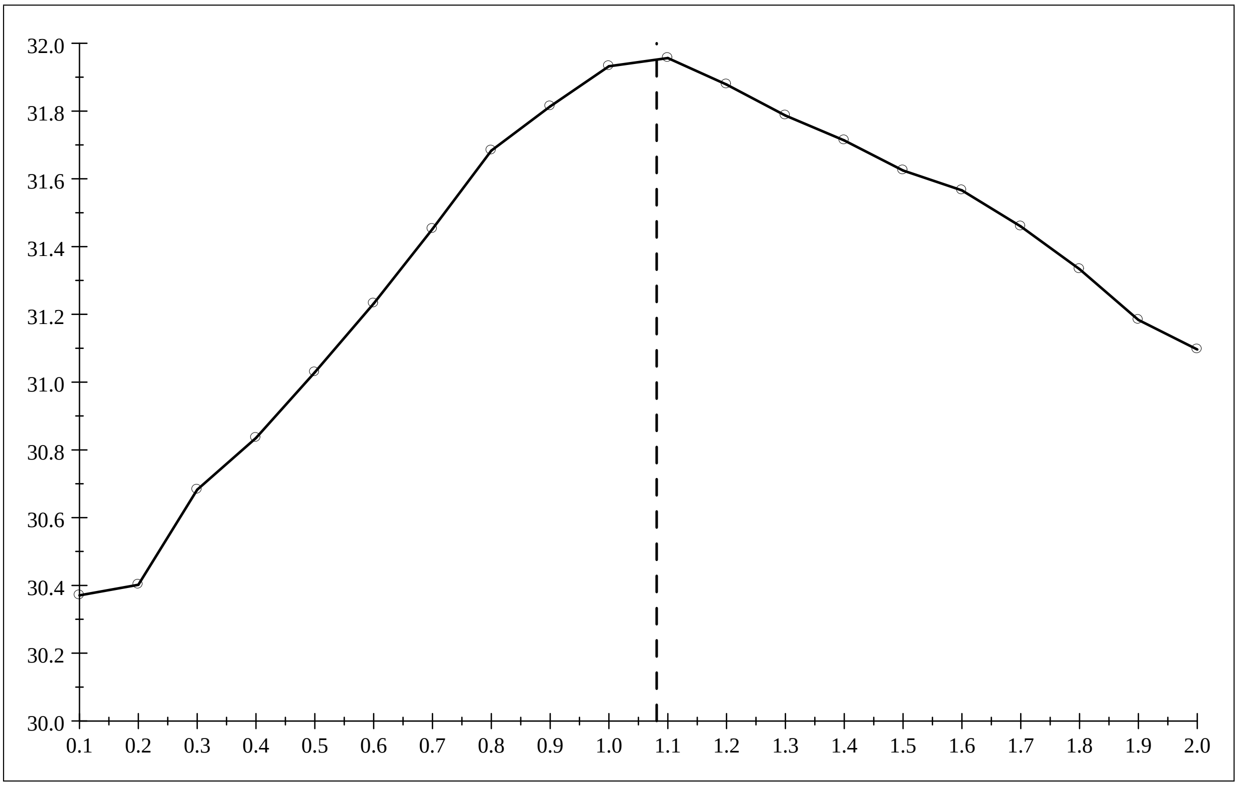

Figure 2: Plot of $V[\hat{\rho}_2] \times 10000$ vs. β. Vertical line at $\beta^+ = \left(\sqrt{10} - 1\right)/2$