Corrigendum to "Testing predictive regression models with nonstationary regressors"

(Zongwu Cai and Yunfei Wang, Journal of Econometrics 178, 4-14)

Matei Demetrescu*
University of Bonn

January 14, 2014

Abstract

The note re-examines the limiting distribution of the two-step estimators proposed by Cai and Wang (2014, Journal of Econometrics 178, pp. 4-14) for inference in predictive regressions where the regressors are near-integrated and possibly have a trend. It is shown that the distribution is actually not mixed Gaussian for the case where the regressor does not exhibit a deterministic trend. Under deterministically trending regressors, the distribution is Gaussian but with a different covariance matrix.

1 Background

The interest in inferential methods for predictive regressions with persistent regressors has increased exponentially in the past years. See e.g. Elliott and Stock (1994), Campbell and Yogo (2006), Jansson and Moreira (2006) or Phillips and Lee (2013) and the references therein.

Along these lines, Cai and Wang (2014) discuss the baseline predictive model

$$y_t = \beta_0 + \beta_2 x_{t-1} + \varepsilon_t \tag{1}$$

where

$$x_t = \rho x_{t-1} + u_t \tag{2}$$

for $t = 1, \ldots, n$, $\rho = 1 + c/n$ for fixed $c \leq 0$ such that x_t is integrated or near-integrated with initial value $x_0 = 0$, and the innovations ε_t and u_t are a general zero-mean serially dependent and correlated stationary sequences satisfying certain mixing and moment conditions; see Assumptions

*Corresponding author: Hausdorff Center for Mathematics and Department of Economics, Bonn University, Adenauerallee 24-42, D-53113 Bonn, Germany. E-mail address: matei.demetrescu@uni-bonn.de. The author would like to thank Jörg Breitung, Christoph Hanck and Uwe Hassler for comments and suggestions.

A.1-A.3 of Cai and Wang (2014). They argue that augmenting the predictive regression model (1) by u_t eases inference in their setup – in particular under their Assumption A.2 ensuring weak exogeneity after linearly projecting ε_t on u_t, $\varepsilon_t = \beta_1 u_t + v_t$ with $\beta_1 = \frac{\mathrm{Cov}(\varepsilon_t, u_t)}{\mathrm{Var}(u_t)}$. This leads to the augmented predictive regression model

$$y_t = \beta_0 + \beta_1 u_t + \beta_2 x_{t-1} + v_t. \tag{3}$$

Denote by $\tilde{\beta}$ the infeasible OLS estimates from (3); beyond that, the notation of Cai and Wang (2014) is adopted in the following.

Given that u_t is not observed, Cai and Wang propose a two-step version based on estimation of u_t from (2) via OLS autoregression,

$$\hat{u}_t = x_t - \hat{\rho} x_{t-1} = u_t - (\hat{\rho} - \rho)\, x_{t-1} \quad \text{where} \quad \hat{\rho} = \frac{\sum_{t=1}^n x_{t-1} x_t}{\sum_{t=1}^n x_t^2} = \rho + \frac{\sum_{t=1}^n x_{t-1} u_t}{\sum_{t=1}^n x_t^2}, \tag{4}$$

leading to the feasible augmented predictive regression

$$y_t = \beta_0 + \beta_1 \hat{u}_t + \beta_2 x_{t-1} + v_t. \tag{5}$$

Thus, the parameters of (3) – and in particular β_2 – could be estimated and inference could be conducted based on the limiting distribution of the estimators or of the corresponding t statistics. The OLS estimators in (5) are denoted by $\hat{\beta}_{0,1,2}$ as in Cai and Wang (2014).

They also discuss the case where the intercept in the autoregression is nonzero and a deterministic trend dominates the stochastic one in x_t, $x_t = \theta + \rho x_{t-1} + u_t$. For this case, the OLS autoregression includes an intercept,

$$x_t = \hat{\theta} + \hat{\rho} x_{t-1} + \hat{u}_t, \tag{6}$$

and one has $\hat{u}_t = x_t - \hat{\theta} - \hat{\rho} x_{t-1}$ based on the OLS estimates $\hat{\theta}$ and $\hat{\rho}$ from (6).

In the proof of their Theorem 1 (p. 12) Cai and Wang (2014) point out that, given that $\hat{\rho}$ is super-consistent and $\sup_{1 \le t \le n} |x_{t-1}| = O_p(\sqrt{n})$, one obtains a uniform bound for the approximation error,

$$\sup_{1 \le t \le n} |\hat{u}_t - u_t| \le |\hat{\rho} - \rho| \sup_{1 \le t \le T} |x_{t-1}| = O_p\left(\frac{1}{\sqrt{n}}\right). \tag{7}$$

Based on this tight bound (which is only given in models with stationary regressors when the latter are almost surely bounded), they argue that the estimation effect, $\hat{u}_t - u_t$, has a negligible effect on the limiting distribution of $\hat{\beta}$ and derive the limiting distribution of the OLS estimator $\hat{\beta}$ in (5) as if OLS estimation were conducted in (3).

This note derives the asymptotic distribution of the suitably normalized $\hat{\beta}$ estimated from (5). Throughout, the assumptions of Cai and Wang (2014) are maintained. Still, the distribution is different from that of the infeasible $\tilde{\beta}$ estimated from (3), and the effect of plugging in residuals $\hat{u}_t$ for the unobserved u_t is not negligible. Concretely, mixed Gaussianity is lost for the case without intercept, and an analogous finding holds for the case with intercept in the autoregression for x_t.

2 The effect of plugging in residuals

Note that plugging in $\hat{u}_t$ for u_t is indeed often delivering the desired results, e.g. when estimating the variance of u_t. But even a $n^{-0.5}$ uniform approximation rate as in (7) is not sufficient in general. Consider e.g. the normalized sample average of u_t vs. that of $\hat{u}_t$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\hat{u}_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t - \frac{\hat{\rho}-\rho}{\sqrt{n}}\sum_{t=1}^{n}x_{t-1};$$

then,

$$\frac{\hat{\rho}-\rho}{\sqrt{n}}\sum_{t=1}^{n}x_{t-1} \Rightarrow \frac{\int_0^1 K_c(r)\,\mathrm{d}W_u(r)+\Omega_1}{\int_0^1 K_c^2(r)\,\mathrm{d}r}\int_0^1 K_c(r)\,\mathrm{d}r$$

with $K_c(r)$ an Ornstein-Uhlenbeck [OU] process driven by the Wiener process $W_u(r)$ having variance $\sigma_u^2 = \sum_{h=-\infty}^{\infty}\mathrm{E}(u_t u_{t-h})$ and $\Omega_1 = \sum_{h=1}^{\infty}\mathrm{E}(u_t u_{t-h})$, such that the difference between $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\hat{u}_t$ and $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t$ does not vanish. Moreover,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\hat{u}_t \Rightarrow W_u(1) - \frac{\int_0^1 K_c(r)\,\mathrm{d}W_u(r)+\Omega_1}{\int_0^1 K_c^2(r)\,\mathrm{d}r}\int_0^1 K_c(r)\,\mathrm{d}r$$

is not normally distributed in the limit while

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t \Rightarrow W_u(1)$$

is.

Such an effect occurs in (5) as well: while the errors-in variables bias induced by replacing u_t with $\hat{u}_t$ may not be large enough to affect the first-order limit of $\hat{\beta}_2$, it still causes second-order bias as described in the following.

To analyze the asymptotic behavior of $\hat{\beta}_2$ as opposed to that of $\tilde{\beta}_2$, note that the true model (3) is equivalently written as

$$y_t = \beta_0 + \beta_1\hat{u}_t + \beta_2 x_{t-1} + (v_t + \beta_1(u_t - \hat{u}_t))$$

so the actual error terms in (5) are $v_t + \beta_1(u_t - \hat{u}_t)$ rather than v_t, with the exception of the case where $\beta_1 = 0$. This implies the OLS estimator of (5) to be given as

$$\begin{pmatrix}\hat{\beta}_0-\beta_0\\ \hat{\beta}_1-\beta_1\\ \hat{\beta}_2-\beta_2\end{pmatrix} = \begin{pmatrix} n & \sum_{t=1}^{n}\hat{u}_t & \sum_{t=1}^{n}x_{t-1}\\ \sum_{t=1}^{n}\hat{u}_t & \sum_{t=1}^{n}\hat{u}_t^2 & 0\\ \sum_{t=1}^{n}x_{t-1} & 0 & \sum_{t=1}^{n}x_{t-1}^2\end{pmatrix}^{-1}\begin{pmatrix}\sum_{t=1}^{n}(v_t+\beta_1(u_t-\hat{u}_t))\\ \sum_{t=1}^{n}\hat{u}_t(v_t+\beta_1(u_t-\hat{u}_t))\\ \sum_{t=1}^{n}x_{t-1}(v_t+\beta_1(u_t-\hat{u}_t))\end{pmatrix}$$

whether the autoregression for x_t is estimated with intercept or not.

Compared to the OLS estimation of (3), additional terms appear in the expressions for $\hat{\beta}_{0,1,2}$. For instance, the term corresponding to $\sum_{t=1}^{n}x_{t-1}(v_t+\beta_1(u_t-\hat{u}_t))$ is given in the infeasible OLS estimation of (3) by $\sum_{t=1}^{n} x_{t-1}v_t$; noting that $\sum_{t=1}^{n} \hat{u}_t x_{t-1} = 0$ since it is a first-order condition for OLS estimation of a first-order autoregression for x_t (with or without intercept), the term becomes $\sum_{t=1}^{n} x_{t-1}\left(v_t + \beta_1 u_t\right) = \sum_{t=1}^{n} x_{t-1}\varepsilon_t$ rather than the desired $\sum_{t=1}^{n} x_{t-1}v_t$.

The following theorem examines the effect of such "local disturbances" on the limiting distributions when the autoregression for x_t is estimated without intercept.

Theorem 1 *Let $D_n = \operatorname{diag}(1, 1, \sqrt{n})$; under the above assumptions and $\theta = 0$, it holds as $n \to \infty$ that*

$$\sqrt{n}D_n \begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & \int_0^1 K_c(r)\,\mathrm{d}r \\ 0 & \operatorname{Var}(u_t) & 0 \\ \int_0^1 K_c(r)\,\mathrm{d}r & 0 & \int_0^1 K_c^2(r)\,\mathrm{d}r \end{pmatrix}^{-1} \times$$
$$\times \left(\begin{pmatrix} \int_0^1 \mathrm{d}W_v(r) \\ \int_0^1 \mathrm{d}W_{uv}(r) \\ \int_0^1 K_c(r)\,\mathrm{d}W_v(r) \end{pmatrix} + \beta_1 \begin{pmatrix} \frac{\int_0^1 K_c(r)\mathrm{d}r}{\int_0^1 K_c^2(r)\mathrm{d}r} \\ 0 \\ 1 \end{pmatrix} \left(\int_0^1 K_c(r)\,\mathrm{d}W_u(r) + \Omega_1 \right) \right).$$

Proof: *See Section 4.*

In a nutshell, there is an extra additive term due to replacing u_t with $\hat{u}_t$ for the two-step $\sqrt{n}\left(\hat{\beta}_0 - \beta_0\right)$ and $n\left(\hat{\beta}_2 - \beta_2\right)$. The limiting behavior of this difference is given by

$$\beta_1 \begin{pmatrix} 1 & \int_0^1 K_c(r)\,\mathrm{d}r \\ \int_0^1 K_c(r)\,\mathrm{d}r & \int_0^1 K_c^2(r)\,\mathrm{d}r \end{pmatrix}^{-1} \begin{pmatrix} \frac{\int_0^1 K_c(r)\mathrm{d}r}{\int_0^1 K_c^2(r)\mathrm{d}r} \\ 1 \end{pmatrix} \left(\int_0^1 K_c(r)\,\mathrm{d}W_u(r) + \Omega_1 \right)$$

which is not mixed Gaussian since the OU process K_c is driven by W_u. Limiting mixed Gaussianity of $\tilde{\beta}_2$ is thus lost, and it is not clear how inference on β_2 can be conducted, since the nuisance parameter c cannot be estimated consistently (see e.g. Campbell and Yogo, 2006). The estimator $\hat{\beta}_1$ on the other hand has the same limiting distribution as the infeasible $\tilde{\beta}_1$.

Turning our attention to the case where the residuals $\hat{u}_t$ are estimated in (6), the following result holds true.

Theorem 2 *Under the assumptions of Theorem 1, the following weak limits hold as $n \to \infty$.*

1. *If $\theta = 0$, let $D_n = \operatorname{diag}(1, 1, \sqrt{n})$; then,*

$$\sqrt{n}D_n \begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & \int_0^1 K_c(r)\,\mathrm{d}r \\ 0 & \operatorname{Var}(u_t) & 0 \\ \int_0^1 K_c(r)\,\mathrm{d}r & 0 & \int_0^1 K_c^2(r)\,\mathrm{d}r \end{pmatrix}^{-1} \times$$
$$\times \begin{pmatrix} \int_0^1 \mathrm{d}W_v(r) + \beta_1 \int_0^1 \mathrm{d}W_u(r) \\ \int_0^1 \mathrm{d}W_{uv}(r) \\ \int_0^1 K_c(r)\,\mathrm{d}W_v(r) + \beta_1 \left(\int_0^1 K_c(r)\,\mathrm{d}W_u(r) + \Omega_1 \right) \end{pmatrix}.$$

2. If $\theta \neq 0$, let $D_n = \mathrm{diag}(1,1,n)$ and $\tau_{\theta,c}(r) = \theta\frac{e^{cr}-1}{c}$; then,

$$\sqrt{n}D_n \begin{pmatrix} \hat\beta_0-\beta_0 \\ \hat\beta_1-\beta_1 \\ \hat\beta_2-\beta_2 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & \int_0^1 \tau_{\theta,c}(r)\,\mathrm{d}r \\ 0 & \mathrm{Var}(u_t) & 0 \\ \int_0^1 \tau_{\theta,c}(r)\,\mathrm{d}r & 0 & \int_0^1 \tau^2_{\theta,c}(r)\,\mathrm{d}r \end{pmatrix}^{-1} \times$$
$$\times \left(\begin{pmatrix} \int_0^1 \mathrm{d}W_v(r) \\ \int_0^1 \mathrm{d}W_{uv}(r) \\ \int_0^1 \tau_{\theta,c}(r)\,\mathrm{d}W_v(r) \end{pmatrix} + \beta_1 \begin{pmatrix} \int_0^1 \mathrm{d}W_u(r) \\ 0 \\ \int_0^1 \tau_{\theta,c}(r)\,\mathrm{d}W_u(r) \end{pmatrix} \right).$$

Proof: *See Section 4.*

In fact, the proof of the theorem employs the fact that

$$\begin{pmatrix} \hat\beta_0-\beta_0 \\ \hat\beta_1-\beta_1 \\ \hat\beta_2-\beta_2 \end{pmatrix} = \begin{pmatrix} n & 0 & \sum_{t=1}^n x_{t-1} \\ 0 & \sum_{t=1}^n \hat u_t^2 & 0 \\ \sum_{t=1}^n x_{t-1} & 0 & \sum_{t=1}^n x_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^n \varepsilon_t \\ \sum_{t=1}^n u_t v_t + O_p(1) \\ \sum_{t=1}^n x_{t-1}\varepsilon_t \end{pmatrix},$$

such that the two-step estimators for β_0 and β_2 are numerically identical with the OLS estimators in (1), which are given by

$$\begin{pmatrix} \bar\beta_0 \\ \bar\beta_2 \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} n & \sum_{t=1}^n x_{t-1} \\ \sum_{t=1}^n x_{t-1} & \sum_{t=1}^n x_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^n \varepsilon_t \\ \sum_{t=1}^n x_{t-1}\varepsilon_t \end{pmatrix}$$

irrespective of whether $\theta = 0$ or not.

This explains why one obtains for $\hat\beta_0$ and $\hat\beta_2$ the same limiting distribution as from OLS estimation of the initial predictive regression (1) without any augmentation, which, for a zero intercept $\theta = 0$, is not mixed Gaussian unless $\beta_1 = 0$. Like in the case without intercept in the OLS autoregression for x_t, the limiting distribution of $\hat\beta_1$ is the same as that of $\tilde\beta_1$, but $\hat\beta_0$ and in particular $\hat\beta_2$ have different limiting distributions than the unfeasible $\tilde\beta_0$ and $\tilde\beta_2$ unless $\beta_1 = 0$.

For $\theta \neq 0$, the limiting distribution is Gaussian indeed since the weak limit $\tau_{\theta,c}(r)$ of the suitably normalized x_t is deterministic, yet with a different covariance matrix than given in Theorem 2 of Cai and Wang (2014) who only consider the term $\left(\int_0^1 \mathrm{d}W_v(r), \int_0^1 \mathrm{d}W_{uv}(r), \int_0^1 \tau_{\theta,c}(r)\,\mathrm{d}W_v(r)\right)'$.

Since the cases $\theta = 0$ and $\theta \neq 0$ lead to different limiting distributions for $\hat\beta_2$ or its t statistic, where the distributions are nonstandard and depend on the nuisance parameter c for $\theta = 0$, inference on β_2 is again affected by c which cannot be consistently estimated.[1]

3 Finite sample illustration

To illustrate Theorem 1 above, we simulate samples of length $T = 100$, $T = 200$ and $T = 1000$ from (1) with $\beta_2 = 0$ and x_t a random walk generated without intercept, i.e. $\rho = 1$ in (2).

[1] Kurozumi and Aono (2011) conjecture that using a *different* estimator for ρ in (4)—they consider an IV estimator using $\mathrm{sgn}(x_{t-1})$ as instrument for x_{t-1}—may recover mixed Gaussianity of $\hat\beta_2$.

The shocks u_t and v_t are independent iid standard normal sequences, and we consider the cases $\beta_1 = 0$ and $\beta_1 = 0.5$.

We estimate models (3) and (5) and plot for the case $\beta_1 = 0.5$ the estimated densities of the difference between the normalized feasible and infeasible estimators, $n\hat{\beta}_2$ and $n\tilde{\beta}_2$. Since the difference between $n\hat{\beta}_2$ and $n\tilde{\beta}_2$ should vanish when $\beta_1 = 0$, we plot for this no-endogeneity case the estimated densities of the difference between the feasible and the infeasible $n^{3/2}\hat{\beta}_2$ and $n^{3/2}\tilde{\beta}_2$ using normalization with a higher power of the sample size.

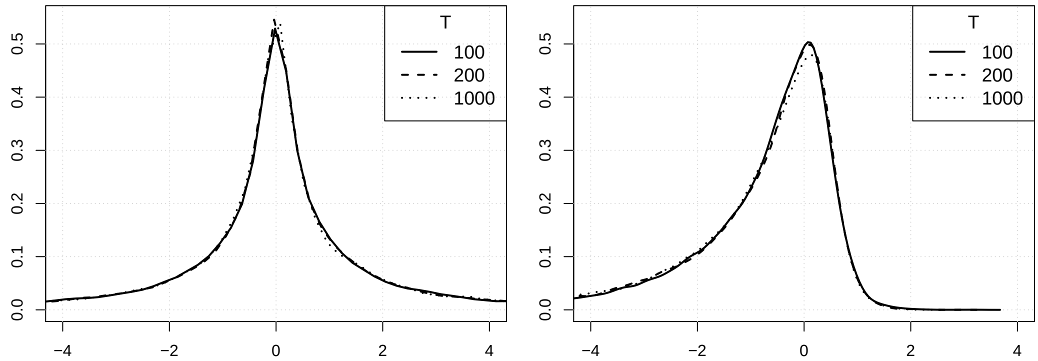

Figure 1: Distribution of the difference between normalized feasible and infeasible slope parameter estimators: $n^{3/2}\left(\hat{\beta}_2 - \tilde{\beta}_2\right)$ when $\beta_1 = 0$ (left) and $n\left(\hat{\beta}_2 - \tilde{\beta}_2\right)$ when $\beta_1 = 0.5$ (right), no intercept in the autoregression for x_t

We observe in Figure 1 that for $\beta_1 = 0$ the distribution of $n^{3/2}\hat{\beta}_2 - n^{3/2}\tilde{\beta}_2$ appears to converge, which illustrates the vanishing difference between $n\hat{\beta}_2$ and $n\tilde{\beta}_2$. This is not the case for $\beta_1 = 0.5$, where the distribution of $n\hat{\beta}_2 - n\tilde{\beta}_2$ converges to a proper limit as predicted by Theorem 1.

For the case with intercept in the autoregression for x_t, we examine the more interesting case where $\theta \neq 0$ in (6): here, with $\theta = 1$. We plot the densities of $n^{3/2}\hat{\beta}_2 - n^{3/2}\tilde{\beta}_2$ for $\beta_1 \neq 0$ since the convergence rate of $\hat{\beta}_2$ is higher due to the dominating deterministic trend, and correspondingly $n^2\hat{\beta}_2 - n^2\tilde{\beta}_2$ for $\beta_1 = 0$.

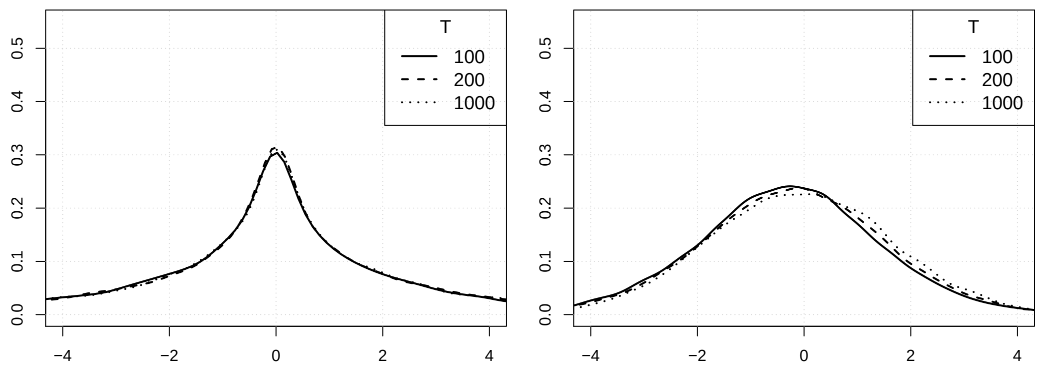

Figure 2: Distribution of the difference between feasible and infeasible slope parameter estimators: $n^2\left(\hat{\beta}_2 - \tilde{\beta}_2\right)$ when $\beta_1 = 0$ (left) and $n^{3/2}\left(\hat{\beta}_2 - \tilde{\beta}_2\right)$ when $\beta_1 = 0.5$ (right), with intercept $\theta = 1$ in the autoregression for x_t

Figure 2 indicates that $n^2\hat{\beta}_2 - n^2\tilde{\beta}_2$ appears to converge in distribution as n grows, such that $n^{3/2}\hat{\beta}_2 - n^{3/2}\tilde{\beta}_2$ vanishes indeed when $\beta_1 = 0$, but $n^{3/2}\hat{\beta}_2 - n^{3/2}\tilde{\beta}_2$ does not vanish for $\beta_1 = 0.5$, as predicted by Theorem 2.

4 Proofs

Let us begin with the case without intercept, $\hat{u}_t = x_t - \hat{\rho}x_{t-1}$. Write the feasible estimators as the sum of two components,

$$\begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} = A_1 + A_2$$

where

$$A_1 = \begin{pmatrix} n & \sum_{t=1}^n \hat{u}_t & \sum_{t=1}^n x_{t-1} \\ \sum_{t=1}^n \hat{u}_t & \sum_{t=1}^n \hat{u}_t^2 & \sum_{t=1}^n \hat{u}_t x_{t-1} \\ \sum_{t=1}^n x_{t-1} & \sum_{t=1}^n \hat{u}_t x_{t-1} & \sum_{t=1}^n x_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^n v_t \\ \sum_{t=1}^n u_t v_t \\ \sum_{t=1}^n x_{t-1} v_t \end{pmatrix}$$

and

$$\begin{aligned} A_2 &= \beta_1 \begin{pmatrix} n & \sum_{t=1}^n \hat{u}_t & \sum_{t=1}^n x_{t-1} \\ \sum_{t=1}^n \hat{u}_t & \sum_{t=1}^n \hat{u}_t^2 & \sum_{t=1}^n \hat{u}_t x_{t-1} \\ \sum_{t=1}^n x_{t-1} & \sum_{t=1}^n \hat{u}_t x_{t-1} & \sum_{t=1}^n x_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^n (u_t - \hat{u}_t) \\ \sum_{t=1}^n \hat{u}_t (u_t - \hat{u}_t) \\ \sum_{t=1}^n x_{t-1} (u_t - \hat{u}_t) \end{pmatrix} \\ &\quad + \beta_1 \begin{pmatrix} n & \sum_{t=1}^n \hat{u}_t & \sum_{t=1}^n x_{t-1} \\ \sum_{t=1}^n \hat{u}_t & \sum_{t=1}^n \hat{u}_t^2 & \sum_{t=1}^n \hat{u}_t x_{t-1} \\ \sum_{t=1}^n x_{t-1} & \sum_{t=1}^n \hat{u}_t x_{t-1} & \sum_{t=1}^n x_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \sum_{t=1}^n v_t (\hat{u}_t - u_t) \\ 0 \end{pmatrix}. \end{aligned}$$

Recall that $\sum_{t=1}^n \hat{u}_t = O_p(\sqrt{n})$ and $\sum_{t=1}^n \hat{u}_t x_{t-1} = 0$; with $D_n = \operatorname{diag}(1, 1, \sqrt{n})$ we then have

$$\begin{aligned} & nD_n \begin{pmatrix} n & \sum_{t=1}^n \hat{u}_t & \sum_{t=1}^n x_{t-1} \\ \sum_{t=1}^n \hat{u}_t & \sum_{t=1}^n \hat{u}_t^2 & 0 \\ \sum_{t=1}^n x_{t-1} & 0 & \sum_{t=1}^n x_{t-1}^2 \end{pmatrix}^{-1} D_n \\ &= \begin{pmatrix} 1 & \frac{1}{n}\sum_{t=1}^n \hat{u}_t & \frac{1}{n^{1.5}}\sum_{t=1}^n x_{t-1} \\ \frac{1}{n}\sum_{t=1}^n \hat{u}_t & \frac{1}{n}\sum_{t=1}^n \hat{u}_t^2 & 0 \\ \frac{1}{n^{1.5}}\sum_{t=1}^n x_{t-1} & 0 & \frac{1}{n^2}\sum_{t=1}^n x_{t-1}^2 \end{pmatrix}^{-1} \\ &\Rightarrow \begin{pmatrix} 1 & 0 & \int_0^1 K_c(r)\,\mathrm{d}r \\ 0 & \operatorname{Var}(u_t) & 0 \\ \int_0^1 K_c(r)\,\mathrm{d}r & 0 & \int_0^1 K_c^2(r)\,\mathrm{d}r \end{pmatrix}^{-1}, \end{aligned}$$

where weak convergence of $n^{-1/2}x_{[rn]}$ has been used together with the continuous mapping theorem.

Moreover,

$$\sqrt{n}D_nA_1 = nD_n\begin{pmatrix} n & \sum_{t=1}^n \hat{u}_t & \sum_{t=1}^n x_{t-1} \\ \sum_{t=1}^n \hat{u}_t & \sum_{t=1}^n \hat{u}_t^2 & 0 \\ \sum_{t=1}^n x_{t-1} & 0 & \sum_{t=1}^n x_{t-1}^2 \end{pmatrix}^{-1} D_n\frac{1}{\sqrt{n}}D_n^{-1}\begin{pmatrix} \sum_{t=1}^n v_t \\ \sum_{t=1}^n u_tv_t \\ \sum_{t=1}^n x_{t-1}v_t \end{pmatrix}$$
$$\Rightarrow \begin{pmatrix} 1 & 0 & \int_0^1 K_c(r)\,\mathrm{d}r \\ 0 & \mathrm{Var}(u_t) & 0 \\ \int_0^1 K_c(r)\,\mathrm{d}r & 0 & \int_0^1 K_c^2(r)\,\mathrm{d}r \end{pmatrix}^{-1}\begin{pmatrix} \int_0^1 \mathrm{d}W_v(r) \\ \int_0^1 \mathrm{d}W_{uv}(r) \\ \int_0^1 K_c(r)\,\mathrm{d}W_v(r) \end{pmatrix}$$

like in Cai and Wang (2014). This is the weak limit of the OLS estimators $\tilde{\beta}$ in the infeasible regression (3).[2] Thus, A_1 is the limiting distribution for estimation based on the unobservables u_t, while A_2 can now be interpreted as expressing the effect of plugging in $\hat{u}_t$.

To examine A_2, recall that $u_t - \hat{u}_t = (\hat{\rho} - \rho)\,x_{t-1}$ such that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n v_t(\hat{u}_t - u_t) = -\frac{\hat{\rho}-\rho}{\sqrt{n}}\sum_{t=1}^n v_tx_{t-1} = O_p\left(\frac{1}{\sqrt{n}}\right);$$

with $\hat{u}_t$ OLS residuals from (2),

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n \hat{u}_t(u_t - \hat{u}_t) = -\frac{\hat{\rho}-\rho}{\sqrt{n}}\sum_{t=1}^n \hat{u}_tx_{t-1} = 0$$

such that, using Cai and Wang (2014, properties (a), (b) and (c) on p. 6),

$$\frac{1}{\sqrt{n}}D_n\left(\begin{pmatrix} \sum_{t=1}^n (u_t-\hat{u}_t) \\ \sum_{t=1}^n \hat{u}_t(u_t-\hat{u}_t) \\ \sum_{t=1}^n x_{t-1}(u_t-\hat{u}_t) \end{pmatrix} + \begin{pmatrix} 0 \\ \sum_{t=1}^n v_t(\hat{u}_t-u_t) \\ 0 \end{pmatrix}\right) \Rightarrow \begin{pmatrix} \frac{\int_0^1 K_c(r)\mathrm{d}W_u(r)+\Omega_1}{\int_0^1 K_c^2(r)\mathrm{d}r}\int_0^1 K_c(r)\,\mathrm{d}r \\ 0 \\ \frac{\int_0^1 K_c(r)\mathrm{d}W_u(r)+\Omega_1}{\int_0^1 K_c^2(r)\mathrm{d}r}\int_0^1 K_c^2(r)\,\mathrm{d}r \end{pmatrix}$$

and thus

$$\sqrt{n}D_nA_2 \Rightarrow \beta_1\begin{pmatrix} 1 & 0 & \int_0^1 K_c(r)\,\mathrm{d}r \\ 0 & \mathrm{Var}(u_t) & 0 \\ \int_0^1 K_c(r)\,\mathrm{d}r & 0 & \int_0^1 K_c^2(r)\,\mathrm{d}r \end{pmatrix}^{-1}\begin{pmatrix} \frac{\int_0^1 K_c(r)\mathrm{d}r}{\int_0^1 K_c^2(r)\mathrm{d}r} \\ 0 \\ 1 \end{pmatrix}\left(\int_0^1 K_c(r)\,\mathrm{d}W_u(r) + \Omega_1\right)$$

as required for Theorem 1.

Moving on to the case with intercept in Theorem 2, the OLS first-order conditions of the autoregression (6) indicate that $\sum_{t=1}^n \hat{u}_t = 0$ in addition to $\sum_{t=1}^n x_{t-1}\hat{u}_t = 0$. Thus,

$$\sum_{t=1}^n (v_t + \beta_1(u_t - \hat{u}_t)) = \sum_{t=1}^n (v_t + \beta_1u_t) = \sum_{t=1}^n \varepsilon_t,$$

[2]Examining the vector $\left(\int_0^1 \mathrm{d}W_v(r), \int_0^1 \mathrm{d}W_{uv}(r), \int_0^1 K_c(r)\,\mathrm{d}W_v(r)\right)'$, note that its covariance matrix is not diagonal, unlike stated in the proof of Theorem 1 of Cai and Wang (2014, p. 13). E.g. the quadratic covariation of the first and the third elements is given by $\sigma_v^2\int_0^1 K_c(r)\,\mathrm{d}r$, which is nonzero w.p. 1.

$$\sum_{t=1}^{n} x_{t-1}\left(v_t+\beta_1\left(u_t-\hat{u}_t\right)\right)=\sum_{t=1}^{n} x_{t-1}\left(v_t+\beta_1 u_t\right)=\sum_{t=1}^{n} x_{t-1}\varepsilon_t$$

and, since $\sum_{t=1}^{n}\hat{u}_t\left(u_t-\hat{u}_t\right)=0$ (see above),

$$\sum_{t=1}^{n}\hat{u}_t\left(v_t+\beta_1\left(u_t-\hat{u}_t\right)\right)=\sum_{t=1}^{n}\hat{u}_t v_t=\sum_{t=1}^{n} u_t v_t+O_p\left(1\right).$$

One thus obtains

$$\begin{pmatrix}\hat{\beta}_0-\beta_0\\ \hat{\beta}_1-\beta_1\\ \hat{\beta}_2-\beta_2\end{pmatrix}=\begin{pmatrix} n & 0 & \sum_{t=1}^{n} x_{t-1}\\ 0 & \sum_{t=1}^{n}\hat{u}_t^2 & 0\\ \sum_{t=1}^{n} x_{t-1} & 0 & \sum_{t=1}^{n} x_{t-1}^2\end{pmatrix}^{-1}\begin{pmatrix}\sum_{t=1}^{n}\varepsilon_t\\ \sum_{t=1}^{n} u_t v_t+O_p\left(1\right)\\ \sum_{t=1}^{n} x_{t-1}\varepsilon_t\end{pmatrix}.$$

With $\varepsilon_t=v_t+\beta_1 u_t$, the first item follows immediately along the lines of the proof of Theorem 1 since $n^{-1/2}x_{[rT]}\Rightarrow K_c\left(r\right)$ when $\theta=0$. To establish the second, recall from Proposition 1 of Cai and Wang (2014) that

$$\frac{1}{n}x_{[rT]}\Rightarrow\tau_{\theta,c}\left(r\right)$$

such that the desired second result follows immediately.[3]

References

Cai, Z. and Y. Wang (2014). Testing predictive regression models with nonstationary regressors. *Journal of Econometrics 178*(1), 4–14.

Campbell, J. Y. and M. Yogo (2006). Efficient Tests of Stock Return Predictability. *Journal of Financial Economics 81*(1), 27–60.

Elliott, G. and J. H. Stock (1994). Inference in Time Series Regression when the Order of Integration of a Regressor Is Unknown. *Econometric Theory 10*(3-4), 672–700.

Jansson, M. and M. J. Moreira (2006). Optimal Inference in Regression Models with Nearly Integrated Regressors. *Econometrica 74*(3), 681–714.

Kurozumi, E. and K. Aono (2011). Estimation and inference in predictive regressions. Technical report, Institute of Economic Research, Hitotsubashi University.

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

[3]Examining the Gaussian vector $\left(\int_0^1 \mathrm{d}W_v\left(r\right),\int_0^1 \mathrm{d}W_{uv}\left(r\right),\int_0^1\tau_{\theta,c}\left(r\right)\mathrm{d}W_v\left(r\right)\right)'$, note that its covariance matrix is not diagonal, unlike stated in the proof of Theorem 2 of Cai and Wang (2014, p. 14). E.g. the covariance of the first and the third elements is given by $\sigma_v^2\int_0^1\tau_{\theta,c}\left(r\right)\mathrm{d}r$, which equals $\sigma_v^2\theta\frac{e^c-1-c}{c^2}$ according to Proposition 1 b) of Cai and Wang.