Journal of Econometrics 178 (2014) 4–14

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Testing predictive regression models with nonstationary regressors[☆]

Zongwu Cai[a,b,∗], Yunfei Wang[c]

[a] Department of Economics, University of Kansas, Lawrence, KS 66045, USA

[b] Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China

[c] Direct Banking Decision Sciences, Capital One 360, 1102 Greystone Ln, Newark, DE 19711, USA

ARTICLE INFO

Article history:
Available online 11 August 2013

Keywords:
Endogeneity
Misspecification test
Nearly integrated process
Predictive regression
Unit root

ABSTRACT

Due to nonstationary (nearly integrated or integrated) regressors and the embedded endogeneity, a linear predictive regression model produces biased coefficient estimates, which consequentially leads to the conventional t-test to over-reject the misspecification test. In this paper, our aim is to find an appropriate and easily implemented method for estimating and testing coefficients in predictive regression models. We apply a projection method to remove the embedded endogeneity and then adopt a two-step estimation procedure to manage both highly persistent and nonstationary predictors. The asymptotic distributions of these estimates are established under α-mixing innovations, and different convergence rates among the coefficients are derived for different persistent degrees. We also consider the model with the regressor having a drift in its autoregressive model and show that the asymptotic properties for the estimated coefficients are totally different from the case without drift. To conduct a misspecification test, we rely on the deduced asymptotic distributions and use the Monte Carlo simulation to find the appropriate critical values. A Monte Carlo experiment is then conducted to illustrate the finite sample performance of our proposed estimator and test statistics. Finally, an empirical example is examined to demonstrate the proposed estimation and testing method.

Published by Elsevier B.V.

1. Introduction

In a standard predictive regression model, a dependent variable is regressed on the lagged value of a regressor which can be formulated by an autoregressive model with error correlated to the disturbance from the predictive regression. The predictive regression model has been widely used in economics and finance. For example, during the recent years, it has been utilized to deal with the predictability problem such as the evaluation of the mutual fund performance, the optimization of the asset allocations, the conditional capital asset pricing, and so on. Thereinto, one famous application is to check the predictability of asset returns by various lagged financial variables, such as the log dividend–price (d–p) ratio, the log earnings–price (e–p) ratio, the log book-to-market ratio, the dividend yield, the term spread, the default premium, and the interest rates as well as other financial variables. The question always being asked is the significance of the estimated predicting coefficients. This is equivalent to a misspecification test to test if the slope coefficients are statistically significant. To answer this question, there has been proposed a vast amount of modeling and testing procedures in the literature; see, to name just a few, Mankiw and Shapiro (1986), Stambaugh (1986, 1999), Elliott and Stock (1994), Cavanagh et al. (1995), Viceira (1997), Amihud and Hurvich (2004), Torous et al. (2004), Lewellen (2004), Paye and Timmermann (2006), Campbell and Yogo (2006), Polk et al. (2006), Dangl and Halling (2012), Rossi (2007), and the references therein.

A standard predictive regression has the following linear structural model

$$y_t = \beta_0 + \beta_2 x_{t-1} + \varepsilon_t, \qquad x_t = \rho x_{t-1} + u_t, \quad 1 \le t \le n, \tag{1}$$

where y_t is the dependent variable, say excess stock return at time t, x_{t-1} is a financial variable such as the log dividend–price ratio at time $t-1$, which is commonly formulated by an autoregressive model with order 1 (denoted by AR(1)) as the second equation in (1), and innovations $\{(\varepsilon_t, u_t)\}$ in (1) are usually assumed to be independently and identically distributed (iid) bivariate normal

☆ Cai's research was supported, in part, by the National Nature Science Foundation of China grants #71131008 and #70971113. We are grateful to two anonymous referees and the conference participants at the "International Symposium on Econometrics of Specification Test in 30 Years" in June 2010 in Xiamen, China. All errors are our responsibility.

∗ Corresponding author at: Department of Economics, University of Kansas, Lawrence, KS 66045, USA.

E-mail addresses: zongwucai@gmail.com (Z. Cai), yunfei.wang@capitalone.com (Y. Wang).

0304-4076/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.jeconom.2013.08.002

$N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon u} \\ \sigma_{\varepsilon u} & \sigma_u^2 \end{pmatrix}$. Model (1) is commonly called a predictive regression.

The main interest of this model is to estimate the predicting coefficient β_2 and to test whether the predictability exists or not, i.e. to check the significance of β_2 or to test the null hypothesis H_0 : $\beta_2 = 0$. Mankiw and Shapiro (1986) and Stambaugh (1986) were the first to analyze the econometric and statistical difficulties inherent in the estimation of the above predictive regression model. At first, note that the correlation coefficient between the two innovations ε_t and u_t in (1) is $\delta = \sigma_{\varepsilon u}/\sigma_\varepsilon \sigma_u$, which is unfortunately non-zero for many empirical studies; see, for example, Table 4 in Campbell and Yogo (2006) and Table 1 in Torous et al. (2004) for some real applications. This creates the so-called embedded endogeneity (x_{t-1} and ε_t may be correlated) problem which leads to biased estimates. Another difficulty comes from the parameter ρ, which is the unknown degree of persistence of the variable x_t. That is, x_t is stationary ($|\rho| < 1$); see Viceira (1997), Amihud and Hurvich (2004), Paye and Timmermann (2006) and Amihud et al. (2009); or it is unit root or integrated ($\rho = 1$), denoted by I(1); or it is local-to-unity or nearly integrated ($\rho = 1 + c/n$, where $c < 0$), denoted by NI(1); see, Elliott and Stock (1994), Cavanagh et al. (1995), Torous et al. (2004), Campbell and Yogo (2006), Polk et al. (2006), and Rossi (2007), and among others. This means if the regressor x_t is highly persistent and even nonstationary, it makes econometric modeling and misspecification testing difficult.

As shown in Nelson and Kim (1993) and Stambaugh (1999), due to the embedded endogeneity, the ordinary least squares (OLS) estimate of the slope coefficient β_2 in (1) and its standard error are substantially biased in finite samples if x_t is highly persistent, not really exogenous, and even nonstationary. Conventional misspecification tests based on standard t-statistics from OLS estimates tend to over-reject the null hypothesis in Monte Carlo simulations; see Campbell and Yogo (2006). With regard to the estimation procedure, the researchers usually address this problem by correcting the bias of the least squares estimators. There are three methods established in the literature. The first one is the first order bias-correction estimator by Stambaugh (1999), in which the relation between the biases of the OLS estimates of β_2 and ρ are developed, and subsequently the estimate of β_2 is corrected by its first order bias. The second method is the two-stage least squares (linear projection method) estimator proposed in Amihud and Hurvich (2004) and Amihud et al. (2009), where the endogeneity can be removed from the model by using a linear projection of ε_t onto u_t. In such a way, the OLS method can be applied and the second order biases of the estimates are then corrected. However, one basic assumption in their model is to assume that the regressor is stationary, i.e. $|\rho| < 1$, which might not be true in many empirical applications. The third method is called the conservative bias-adjusted estimator by Lewellen (2004), based on the assumption that the true autoregressive coefficient ρ is almost 1, setting it to be, for instance, 0.9999.

The above estimation methods improve the original OLS estimate to some extent. However, none of them overcome the aforementioned limitations, such as the endogeneity and the high persistency of regressor simultaneously. On the other hand, the misspecification (hypothesis) testing for the predictability is also unreliable due to these difficulties. In the literature, it is common to seek more accurate sampling distributions of test statistics. For example, some researches apply the exact finite-sample theory under the assumption of normality, such as Stambaugh (1999) and Lewellen (2004), which might not be applicable in most cases in the real world due to some unrealistic assumptions. Some others employ NI(1) asymptotics to approximate the finite sample distributions; see Cavanagh et al. (1995), Campbell and Yogo (2006), among others. Even though some misspecification test statistics proposed by these papers, say the Q-test based on the Bonferroni confidence interval proposed by Campbell and Yogo (2006), show good performance in finite samples, the implementation of these tests might not be easy and they might be conservative.

To propose an appropriate test statistics for misspecification test on β_2 in the model given in (1), one has to first address the aforementioned estimation issue. Therefore, in this paper, we first consider the estimation of coefficients β_0 and β_2 in (1) by assuming that the regressor x_t is nonstationary, i.e. $\rho = 1 + c/n$, where $c \le 0$. Indeed, many time series in applications, in particular economic and financial time series, exhibit this type of property. To remove the embedded endogeneity, we apply the linear projection method as in Amihud and Hurvich (2004) and Amihud et al. (2009) to estimate the coefficients in the model, and then establish the asymptotic properties for the estimators to show that the limiting distribution is a mixed normal with conditional variance being a function of integrations of an Ornstein–Uhlenbeck (mean-reverting) process. It also demonstrates that the convergence rates for the intercept (the regular rate at $n^{1/2}$) and the slope coefficient for x_{t-1} (a faster rate at n) are totally different due to the I(1) or NI(1) property of regressor. Finally, based on the derived asymptotic distribution of the estimated coefficients, we use the Monte Carlo simulation method to simulate the limiting distribution and then conduct the misspecification test based on the constructed confidence interval for each coefficient.

Another contribution of this paper is to discuss the predictive regression model in (1) when the regressor has a nonzero drift in its AR(1) model. It is common to assume in the predictive model literature that the AR(1) model for the regressor has a zero drift. But this assumption may not hold for all applications. Therefore, we consider the asymptotic properties for two cases: with and without drift. Indeed, in contrast to the mixed normal limiting distribution for the former, the asymptotic distribution for the latter is normal and the convergence rate is much faster at a rate $n^{3/2}$. Similarly, the statistical inferences such as misspecification tests for this case are considered sequentially.

The rest of this paper is organized as follows. Section 2 introduces the predictive regression model and its two-step estimation procedure. In Section 3, we list the regularity conditions and develop the asymptotic properties of the proposed estimators. Also the extended model with a nonzero drift for the AR(1) model of regressor is discussed in the same section. More importantly, the misspecification test is elaborated in this section too. A Monte Carlo study is conducted in Section 4 to evaluate the finite sample performance of the proposed estimation and testing methods. An application of predictive regression with nonstationary regressors to real examples is reported in Section 5 to highlight the practical usefulness of the proposed methods. Section 6 concludes the paper. Proof of the main results is relegated to the Appendix.

2. Estimation procedure

For the predictive regression model in (1), many empirical studies show that the correlation between the two innovations is non-zero, which implies that the "embedded endogeneity" is involved in the regression model. To deal with this problem, by assuming, commonly done in the literature (see Amihud and Hurvich (2004) and Amihud et al. (2009)) that the joint distribution of ε_t and u_t in model (1) is normal, the linear projection of ε_t onto u_t is given as follows:

$$\varepsilon_t = \beta_1 u_t + v_t \quad 1 \le t \le n.$$

Plugging this into the regression model (1), one obtains the following model:

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

where v_t is uncorrelated with u_t and x_{t-1} so that the endogeneity disappears. Different from the stationary assumption on the regressor x_{t-1} in Amihud and Hurvich (2004) and Amihud et al. (2009), in this paper we consider the case that the regressor x_t is nonstationary. That is, x_t satisfies an AR(1) model as $x_t = \rho x_{t-1} + u_t$ with $\rho = 1 + c/n$ with $c \le 0$. Indeed, this assumption is used in various real applications; see Campbell and Yogo (2006). When $c < 0$, the regressor is a nearly integrated process, while when $c = 0$, it is an I(1) process. Finally, note that $\{(u_t, v_t)\}$ is commonly assumed to be iid in the literature. In this paper, the iid assumption is relaxed to a stationary process, particularly a stationary α-mixing process, which covers many known linear and nonlinear time series models in economics and finance as special cases; see Cai (2002) and Cai et al. (2000) for details. Then the above model can be rewritten as follows:

$$y_t = \beta_0 + \beta_1 u_t + \beta_2 x_{t-1} + v_t \equiv \beta^T X_t + v_t,$$
$$x_t = \rho x_{t-1} + u_t, \qquad \rho = 1 + \frac{c}{n}, \quad c \le 0,\ 1 \le t \le n, \tag{2}$$

where $X_t = (1, u_t, x_{t-1})^T$, $\beta = (\beta_0, \beta_1, \beta_2)^T$, and $\{(u_t, v_t)\}$ is a stationary α-mixing process. The error term v_t in expression (2) is uncorrelated with either the regressor x_{t-1} or the innovation u_t and x_{t-1} is an I(1) or NI(1) process. Without loss of generality, we just consider the univariate regressor for simplicity, though the model could be easily extended to multiple regressors as in Amihud et al. (2009).

Our main interest is testing the predictability of the coefficient β in (2). To this end, the first step is to find an appropriate estimate for β. An intuitive way of doing so is to apply the ordinary least squares (OLS) method to obtain the OLSE of β given by

$$\hat{\beta} = (X^T X)^{-1} X^T y = \left[\sum_{t=1}^{n} X_t X_t^T\right]^{-1} \sum_{t=1}^{n} X_t y_t, \tag{3}$$

where $X = (X_1, X_2, \ldots, X_n)^T$ and $y = (y_1, y_2, \ldots, y_n)^T$. In (3), the matrix X contains the unknown vector $u = (u_1, u_2, \ldots, u_n)^T$. To deal with this, one can replace u_t with its estimated value. Hence, a two step procedure becomes necessary and is described as follows:

Step 1: Run the first order autoregression of x_t to find the estimate of ρ and the estimated residual $\hat{u}_t$.

Step 2: Regress y_t on $\hat{u}_t$ and x_{t-1} with intercept and adopt OLS method again to find the estimate $\hat{\beta}$.

As mentioned before, the linear projection between two innovations can remove the correlation between two innovations from the model so that it can reduce the estimation bias of β. But the distribution of $\hat{\beta}$ might not be normal any more due to the nonstationarity of the regressor x_{t-1}. In the next section, the asymptotic distribution of the two-step estimator $\hat{\beta}$ is developed and then misspecification test is considered.

3. Statistical inferences

3.1. Asymptotic theory

Before we derive the asymptotic distribution of the estimate given in (3), we list some notations and assumptions needed for the theoretical proof. By ignoring the higher order terms, $\rho = 1 + c/n = \exp(c/n) + o(1/n)$ and in what follows, it is assumed without loss of generality that $\rho = \exp(c/n)$ with $c \le 0$. Then, the NI(1) process x_t can be expressed as

$$x_t = \sum_{j=1}^{t} \exp((t-j)c/n)\, u_j + \exp(tc/n)\, x_0,$$

where $x_0 = u_0$ and $\{u_j\}$ is a strictly stationary α-mixing process defined in (2) with its mixing coefficients satisfying the condition listed below (see (4)). Here, it is assumed without loss of generality that $E(x_0) = 0$ or $x_0 = 0$. It follows by Lemma 3.1 in Phillips (1988) that under some regularity conditions such as

(i) $E(u_0) = 0$,

(ii) $E|u_j|^{k_1+k_2} < \infty$ for some $k_1 > 2$ and $k_2 > 0$, (4)

(iii) $\{u_j\}_0^\infty$ is α-mixing with mixing coefficients $\alpha(m)$ satisfying

$$\sum_{m=1}^{\infty} \alpha^{1-2/k_1}(m) < \infty,$$

x_t has the following properties:

$$\text{(a) } n^{-1/2} x_{[nr]} \Rightarrow K_c(r), \qquad \text{(b) } n^{-3/2} \sum_{t=1}^{n} x_t \xrightarrow{d} \int_0^1 K_c(r)dr,$$

$$\text{(c) } n^{-2} \sum_{t=1}^{n} x_t^2 \xrightarrow{d} \int_0^1 K_c^2(r)dr,$$

$$\text{and} \quad \text{(d) } n^{-1} \sum_{t=1}^{n} x_{t-1} u_t \xrightarrow{d} \int_0^1 K_c(r)dW_u(r) + \Omega_1, \tag{5}$$

where $K_c(r) = \int_0^r e^{(r-s)c} dW_u(s)$ is a diffusion process, $W_u(s)$ is a one-dimensional Brownian motion with variance $\sigma_u^2 = \mathrm{Var}(u_t) + 2\,\Omega_1$, and $\Omega_1 = \sum_{k=2}^{\infty} E(u_1 u_k)$. Here and in what it follows, "$\Rightarrow$" represents weak convergence, and "$\xrightarrow{d}$" denotes convergence in distribution. Note that $K_c(\cdot)$ is a special case of the Ornstein–Uhlenbeck process and satisfies the stochastic differential equation system (Black–Scholes model) as $dK_c(r) = c\, K_c(r)dr + dW(r)$. Then, it can be shown easily that $K_c(r) \sim N(0, \sigma_c^2(r))$, where $\sigma_c^2(r) = \sigma_u^2[\exp(2cr) - 1]/2c$ and $\int_0^1 K_c(r)dr \sim N(0, \varsigma(c)^2)$ with $\varsigma(c)^2 = \sigma_u^2/c^2 + \sigma_u^2(e^{2c} - 4e^c + 3)/2c^3$. Clearly, it is easy to obtain that $\lim_{c\to 0} \varsigma(c)^2 = \sigma_u^2/3$. Also, when $c = 0$, $K_c(r)$ becomes $W(r)$. We make the following assumptions.

Assumptions:

A1. X has full column rank. $\{u_t\}$ is a mean zero and strictly stationary strong α-mixing sequence satisfying the regular conditions in (4). Also assume that $x_0 = 0$.

A2. The error term v_t has a finite fourth moment and $E(v_t|X_t) = 0$.

A3. The error term $\{(u_t, v_t)\}$ in (2) is a strictly stationary α-mixing process with δ_1-th moment for some $\delta_1 > 2$, where $\{u_t\}$ and $\{v_t\}$ are uncorrelated. Further, $E(|u_t v_t|^{\delta_2}) \le C < \infty$ for $\delta_2 > \delta_1$ and there exists some $j^* < \infty$ such that for all $j > j^*$, $E|u_0 u_j v_0 v_j| < \infty$. $\alpha(t) = O(t^{-\delta_0})$ for some $\delta_0 > \min\{\delta_2\delta_1/(\delta_2 - \delta_1), 2\delta_3/(2-\delta_3), 2\delta_4/(2-\delta_4)\}$. Also, $\|u_t\|_q = (E|u_t|^q)^{1/q} < \infty$ with $q = \delta_4\delta_3/(\delta_4 - \delta_3)$ for some $1 < \delta_3 < \delta_4 < 2$. Finally, $\sup_k E(u_1^2 v_{t+k}^2) \le C < \infty$.

Among the above conditions, A1 is a basic assumption to guarantee the existence of the asymptotic distribution of the nearly integrated process. Assumption A1 implies that unlike model (1), two innovations $\{(\varepsilon_t, u_t)\}$ are not necessary to be iid or normally distributed. Assumption A2 ensures that regressors u_t and x_{t-1} are exogenous variable. Finally, the α-mixing assumption in A3 is one of the weakest mixing conditions for weakly dependent stochastic processes.

To establish the asymptotic results, define $D_n = \mathrm{diag}\{1, 1, \sqrt{n}\}$, $E(u_t^2 v_t^2|X_t) = \sigma_{uv}^2$, and (see Box I)

Then, we have the following theorem with its proof presented in the Appendix.

Theorem 1. *Under Assumptions* A1–A3, *we have*

$$\sqrt{n} D_n \left(\hat{\beta} - \beta\right) \xrightarrow{d} MN(0, \Sigma_\beta),$$

where $MN(0, \Sigma_\beta)$ *is a mixed normal with mean zero and conditional covariance matrix* $\Sigma_\beta = \sigma_v^2 \Omega / \left[\int K_c^2 - (\int K_c)^2\right]^2$ *with* $\sigma_v^2 = Var(v_1) + 2\sum_{k=2}^{\infty} Cov(v_1, v_k)$.

$$\Omega = \begin{pmatrix} \left(\int K_c^2 dr\right)^2 + \int K_c^2 dr \left(\int K_c dr\right)^2 & 0 & -2\int K_c dr \int K_c^2 dr \\ 0 & \sigma_{uv}^2 \left[\int K_c^2 dr - \left(\int K_c dr\right)^2\right]^2 / \sigma_u^4 \sigma_v^2 & 0 \\ -2\int K_c dr \int K_c^2 dr & 0 & \int K_c^2 dr + \left(\int K_c dr\right)^2 \end{pmatrix}$$

Box I.

Remark 1. Based on the theorem above, the asymptotically conditional covariances between $\hat{\beta}_1$ and $\hat{\beta}_2$, and between $\hat{\beta}_1$ and $\hat{\beta}_0$ are both zero, which implies the asymptotic independence between $\hat{\beta}_1$ and $\hat{\beta}_2$, and between $\hat{\beta}_1$ and $\hat{\beta}_0$. However, ignoring the endogeneity could lead to the finite sample problem (higher order bias) and consequently the traditional inferences may not be valid.

In Theorem 1, the focus is only on the case if the regressor x_t is univariate for simplicity, and the multivariate case can be easily extended by the same manner. From Theorem 1, one can see clearly that the convergence rate for the intercept β_0 and slop β_1 is $O(n^{-1/2})$, totally different from that for β_2, which is $O(n)$ and is faster by a factor $n^{1/2}$. This makes sense due to the nearly integrated property of the regressor ($\sum_{t=1}^n x_t^2 = O_p(n^2)$) rather than $O_p(n)$.

3.2. Models with drift in the regressor

In the previous section, we develop the asymptotic distribution of the estimate given in (3) when there is no drift in the AR(1) for x_t. This result could be easily extended to the case if the AR model for regressor x_t has a nonzero drift. That is,

$$y_t = \beta_0 + \beta_1 u_t + \beta_2 x_{t-1} + v_t, \qquad x_t = \theta + \rho x_{t-1} + u_t,$$
$$\rho = 1 + \frac{c}{n}, \quad c \le 0, \ 1 \le t \le n, \tag{6}$$

where $\theta \neq 0$. To estimate β in (6), one can still adopt the two-step procedure proposed in Section 2. The formula for $\hat{\beta}$ should have the same form as in (3), but with different OLS estimates for ρ and u_t. Even though the estimation procedure is the same as if there would be no drift term in the AR(1) model of x_t, the asymptotic distribution of the estimate could be totally different from the previous case. In the following, we examine the asymptotic behavior of x_t and the estimate $\hat{\beta}$ in model (6).

When the persistent parameter c is nonzero, the regressor x_t could be expressed as

$$x_t = \theta \sum_{s=0}^{t-1} \rho^s + \sum_{r=0}^{t-1} \rho^r u_{t-r} = \theta \frac{1-\rho^t}{1-\rho} + o_p(t).$$

Applying the expression $\rho = 1 + c/n = \exp(n^{-1}c) + o(n^{-1})$ to the above equation, it is easy to obtain $n^{-1}x_{[nr]} = \theta[\exp(cr) - 1]/c + o_p(1)$, for any $0 < r \le 1$. Therefore, the convergence rate for x_{t-1} is $O_p(n)$ and is faster by a factor $n^{1/2}$ than the case without drift (see (5)). Similarly, the asymptotic properties of x_t given in (5) are revised as follows.

Proposition 1. *A variable x_t, which satisfies model (6) with a nonzero drift, has the following asymptotic properties:*

(a) $n^{-1}x_{[nr]} \xrightarrow{p} \theta(e^{cr} - 1)/c$,

(b) $n^{-2}\sum_{t=1}^n x_{t-1} \xrightarrow{p} \theta(e^c - 1 - c)/c^2$,

(c) $n^{-3}\sum_{t=1}^n x_{t-1}^2 \xrightarrow{p} \theta^2[(e^c-2)^2 + 2c - 1]/2c^3$, *and*

(d) $n^{-3/2}\sum_{t=1}^n x_{t-1}v_t \to \mathrm{N}(0, \sigma_v^2\theta^2[(e^c-2)^2 + 2c - 1]/2c^3)$, (7)

where σ_v^2 is given in Theorem 1.

The brief proof of this proposition is given in the Appendix. Note that the asymptotic properties given here are different from those listed in Section 6 of Phillips (1988) for the case that $\rho = \exp(n^{-3/2}c)$. When $c = 0$, the limiting processes in above proposition could be replaced by their corresponding limits.

Next, we discuss the asymptotic property of OLS estimate $\hat{\beta}$ in (3) when the drift in the AR(1) model for x_t is non-zero. By recalling that the convergence rate of $x_{[nr]}$ is faster by a factor $n^{-1/2}$, it is necessary to adjust the D_n by $D_n^* = \mathrm{diag}\{1, 1, n\}$. In addition, define

$$\Omega^* = \begin{pmatrix} \Omega_{11}^* & 0 & \Omega_{13}^* \\ 0 & \Omega_{22}^* & 0 \\ \Omega_{13}^* & 0 & \Omega_{33}^* \end{pmatrix}$$

with $\Omega_{11}^* = \frac{\theta^2}{2c^3}[(e^c-2)^2 + 2c - 1]\{\frac{1}{2}[(e^c-2)^2 + 2c - 1] + \frac{1}{c}(e^c - c - 1)^2\}$, $\Omega_{13}^* = -\frac{\theta}{c^2}(e^c - c - 1)[(e^c-2)^2 + 2c - 1]$, $\Omega_{22}^* = \frac{\theta^2}{4c^5}[(c-2)e^{2c} + 4e^c - c - 2]^2\sigma_{uv}^2/\sigma_u^4\sigma_v^2$, and $\Omega_{33}^* = \frac{1}{2}[(e^c-2)^2 + 2c - 1] + \frac{1}{c}(e^c - c - 1)^2$. Then, we have the following theorem and its proof is relegated to the Appendix.

Theorem 2. *Under Assumptions* A1–A3 *and model* (6) *with $c < 0$, we have*

$$\sqrt{n}D_n^*\left(\hat{\beta} - \beta\right) \xrightarrow{d} N(0, \Sigma_\beta^*),$$

where $N(0, \Sigma_\beta^*)$ is a normal distribution with mean zero and covariance matrix

$$\Sigma_\beta^* = 4c^5\sigma_v^2\Omega^*/\theta^2[(c-2)e^{2c} + 4e^c - c - 2]^2.$$

In particular, when $c = 0$, Σ_β^* becomes

$$\Sigma_\beta^* = \sigma_v^2 \begin{pmatrix} 28 & 0 & -48/\theta \\ 0 & \sigma_{uv}^2/\sigma_u^4\sigma_v^2 & 0 \\ -48/\theta & 0 & 84/\theta^2 \end{pmatrix},$$

where σ_v^2 is given in Theorem 1.

An immediate consequence of Theorem 2 is that adding an intercept into the AR(1) model for the regressor does not affect the rate of convergence for $\hat{\beta}_0$ or $\hat{\beta}_1$, while the rate for $\hat{\beta}_2$ is faster by a factor $\sqrt{n}$. In addition, the asymptotic variances of $\hat{\beta}_0$ and $\hat{\beta}_2$ do not depend on the diffusion process $K_c(r)$. Instead, they depend only on the drift θ. Finally, the asymptotic distribution is normal with a deterministic asymptotic covariance matrix.

3.3. Misspecification tests

In real applications, it is important and of interest to do the misspecification tests about the coefficients in model (2). Especially, one may want to construct the confidence interval for coefficients, or to test whether a regressor has predicting ability or not. For example, to check if the regressor is significant or not, it is of interest to consider the hypothesis $H_0 : \beta_2 = 0$ versus $H_a : \beta_2 \neq 0$.

From Theorem 1, it is easy to see that the alternative expression for the limiting distribution of $\hat{\beta}$ is given by

$$\sqrt{n}(\hat{\beta}_0 - \beta_0) \xrightarrow{d} \frac{W_v(1)\int K_c^2 dr - \left(\int K_c dr\right)\left(\int K_c dW_v(r)\right)}{\int K_c^2 dr - \left(\int K_c dr\right)^2} \quad (8)$$

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, \sigma_{uv}^2/\sigma_u^2) \quad (9)$$

$$\text{and} \quad n(\hat{\beta}_2 - \beta_2) \xrightarrow{d} \frac{\int K_c dW_v(r) - W_v(1)\int K_c dr}{\int K_c^2 dr - \left(\int K_c dr\right)^2}. \quad (10)$$

From (9), the asymptotic variance of $\hat{\beta}_1$ can be estimated by its corresponding sample variance, say, $\hat{\sigma}_{\beta_1}^2 = \sum \hat{u}_t^2 \hat{v}_t^2 / \sum \hat{u}_t^2$, which can be easily shown to be a consistent estimate of $\sigma_{\beta_1}^2$. As for $\hat{\beta}_0$ and $\hat{\beta}_2$, their limiting distributions can be estimated by using a Monte Carlo simulation method as mentioned in Chapter 17 in Hamilton (1994), as long as the parameters involved are known or can be estimated. Therefore, the critical values can be found by a Monte Carlo simulation method respectively. In particular, the asymptotic distribution in (10) can be used to test the predictability such as $H_0 : \beta_2 = 0$.

When there is an intercept in the AR(1) for x_t, our interest is to consider a two-sided test with null hypothesis $\beta_2 = 0$ against the alternative $\beta_2 \neq 0$. Similar to the previous discussion, the test statistics for the model with a nonzero drift could be considered based on the asymptotic distribution in Theorem 2, say,

$$\sqrt{n}(\hat{\beta}_0 - \beta_0) \xrightarrow{d} N\left(0, 4c^5\sigma_v^2\Omega_{11}^*/\theta^2[(c-2)e^{2c} + 4e^c - c - 2]^2\right)$$

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, \sigma_{uv}^2/\sigma_u^2)$$

$$\text{and} \quad \sqrt{n^3}(\hat{\beta}_2 - \beta_2) \xrightarrow{d} N\left(0, 4c^5\sigma_v^2\Omega_{33}^*/\theta^2[(c-2)e^{2c} + 4e^c - c - 2]^2\right). \quad (11)$$

In practice, σ_u^2, σ_v^2 and σ_{uv}^2 could be estimated by $\hat{\sigma}_u^2 = n^{-1}\sum \hat{u}_t^2$, $\hat{\sigma}_v^2 = n^{-1}\sum \hat{v}_t^2$ and $\hat{\sigma}_{uv}^2 = n^{-1}\sum \hat{u}_t^2 \hat{v}_t^2$ respectively, and the values of θ and c can be replaced by their OLS estimates.

4. Monte Carlo simulations

In this section, we conduct a Monte Carlo simulation study to illustrate the finite sample performance of the proposed two-step estimation procedure and the misspecification test method.

Example 1. Consider a model with the following data generating process:

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

where $\hat{u}_t$ is the OLS estimated residual of the AR(1) model: $x_t = \rho x_{t-1} + u_t$. We choose $\beta_0 = 0$, $\beta_1 = \delta = -0.75$ or -0.95 (note that β_1 captures the correlation between two innovations), $\beta_2 = 0$, and $\rho = 1 + c/n$ with $c = 0$, $c = -2$ or $c = -20$. The selection of β_1 is based on the relevant range of the correlation between two innovations in the real example of predicting the stock return by the log dividend–price or the log earning–price ratio and the three levels of persistence correspond to the cases when the regressor x_t is a unit root process, a nearly unit root process and a stationary process, respectively. The two innovations u_t and v_t are generated from the following AR(1) models:

$$u_t = 0.3u_{t-1} + e_{1t} \quad \text{and} \quad v_t = 0.3v_{t-1} + e_{2t},$$

where e_{1t} and e_{2t} are independently generated from normal distributions with mean zero and variance $\sigma_{e_1}^2 = 0.91$ and $\sigma_{e_2}^2 = 1 - \delta^2 = 0.4375$ when $\delta = -0.75$ and $\sigma_{e_2}^2 = 0.0975$ when $\delta = -0.95$, respectively, to guarantee that u_t and ε_t are standard normally distributed.

Table 1
Medians (Standard deviations) of 5000 ADE values by the two-step estimation procedure in Example 1.

δ	c	n	ADE_{β_0}	ADE_{β_1}	ADE_{β_2}
		50	0.1301 (0.1649)	0.0783 (0.0695)	0.0288 (0.0295)
	0	100	0.0961 (0.0999)	0.0493 (0.0503)	0.0130 (0.0162)
		250	0.0582 (0.0582)	0.0328 (0.0291)	0.0051 (0.0058)
		50	0.1088 (0.1261)	0.0762 (0.0669)	0.0324 (0.0361)
−0.75	−2	100	0.0867 (0.0870)	0.0467 (0.0423)	0.0173 (0.0169)
		250	0.0483 (0.0524)	0.0298 (0.0276)	0.0061 (0.0063)
		50	0.0943 (0.0829)	0.0711 (0.0622)	0.0578 (0.0532)
	−20	100	0.0648 (0.0585)	0.0480 (0.0438)	0.0290 (0.0275)
		250	0.0390 (0.0361)	0.0310 (0.0279)	0.0119 (0.0111)
		50	0.0614 (0.0778)	0.0370 (0.0328)	0.0136 (0.0139)
	0	100	0.0453 (0.0472)	0.0233 (0.0237)	0.0061 (0.0077)
		250	0.0275 (0.0312)	0.0155 (0.0137)	0.0024 (0.0027)
		50	0.0513 (0.0595)	0.0360 (0.0316)	0.0153 (0.0170)
−0.95	−2	100	0.0409 (0.0411)	0.0220 (0.0199)	0.0082 (0.0080)
		250	0.0228 (0.0247)	0.0141 (0.0130)	0.0029 (0.0030)
		50	0.0445 (0.0392)	0.0336 (0.0294)	0.0273 (0.0251)
	−20	100	0.0306 (0.0276)	0.0226 (0.0207)	0.0137 (0.0130)
		250	0.0184 (0.0170)	0.0147 (0.0132)	0.0056 (0.0053)

We consider different sample sizes as $n = 50$, 100 and 250 and repeat the simulation 5000 times for each sample size. The sample sizes used in the simulation are relatively small since the method could be expected to perform well for samples with size larger than 500, as mentioned in Campbell and Yogo (2006). We use the two-step estimation method proposed in Section 2 to estimate the three coefficients. The estimation procedure is evaluated by the absolute deviation error (ADE) as $ADE_{\beta_i} = |\hat{\beta}_i - \beta_i|$ for $0 \leq i \leq 2$. The medians and standard deviations (in parentheses) of the 5000 ADE values obtained from the proposed estimation procedure are reported in Table 1.

The results reported in Table 1 correspond to the case when the correlation between the two innovations is relative weak ($\delta = -0.75$). Fixed the value of c, one can see clearly that both median and standard deviation of 5000 ADE values for each coefficient decline with the increase of sample size. It is also obvious that the decreasing rate for ADE_{β_2} is faster than that for ADE_{β_0} and ADE_{β_1}. This is consistent with the asymptotic theory that the convergence rate for $\hat{\beta}_2$ is faster than that for the other two coefficients by a factor $n^{1/2}$. Furthermore, comparing the medians and the standard deviations for ADE_{β_2} values at different values of c, one can see that the bias of the estimate for β_2 is smaller when the predictor is closer to a unit root process. This observation also makes sense since the convergence rate for $\hat{\beta}_2$ is faster when the regressor is higher persistent. The bottom panel of Table 1 is for the case $\delta = -0.95$. The results show a similar pattern to the case when $\delta = -0.75$ but with smaller ADE medians and standard deviations. When the two innovations are more correlated, the subtraction of more innovation to the regressor from the innovation to the dependent variable reduces the estimation biases of the coefficients.

Next, we use the same data generating process to examine the finite sample performance of the misspecification test proposed in Section 3.3. Of interest is to test the significance of each coefficient in model (2) respectively, say, $H_0 : \beta_i = 0$ against $H_a : \beta_i \neq 0$ for $0 \leq i \leq 2$. The test statistics and their asymptotic distributions are given in (8)–(10), based on which the critical values for all tests

Table 2
Test sizes for testing predictability in Example 1 at nominal size 5%.

		$\delta = -0.75$			$\delta = -0.95$		
		β_0	β_1	β_2	β_0	β_1	β_2
$c = 0$	$n = 50$	0.0600	0.0673	0.0633	0.0593	0.0673	0.0633
	$n = 100$	0.0547	0.0587	0.0563	0.0547	0.0587	0.0567
	$n = 250$	0.0527	0.0537	0.0527	0.0520	0.0533	0.0530
$c = -2$	$n = 50$	0.0587	0.0640	0.0580	0.0587	0.0640	0.0580
	$n = 100$	0.0547	0.0580	0.0540	0.0547	0.0580	0.0540
	$n = 250$	0.0533	0.0533	0.0527	0.0533	0.0533	0.0527
$c = -20$	$n = 50$	0.0540	0.0573	0.0367	0.0540	0.0573	0.0367
	$n = 100$	0.0527	0.0533	0.0440	0.0520	0.0533	0.0440
	$n = 250$	0.0520	0.0513	0.0480	0.0520	0.0513	0.0480

Table 3
Test powers in Example 1 at nominal size 5%.

	b	$\delta = -0.75$			$\delta = -0.95$		
		$n = 50$	$n = 100$	$n = 250$	$n = 50$	$n = 100$	$n = 250$
$c = 0$	0	0.0633	0.0563	0.0527	0.0633	0.0567	0.0530
	1	0.0773	0.0707	0.0670	0.1367	0.1313	0.1273
	2	0.1273	0.1220	0.1180	0.4607	0.4593	0.4613
	3	0.2327	0.2273	0.2247	0.8353	0.8400	0.8420
	4	0.4087	0.4073	0.4087	0.9600	0.9633	0.9647
	5	0.6240	0.6253	0.6280	0.9900	0.9920	0.9927
	6	0.7933	0.7973	0.7993	0.9973	0.9980	0.9987
	7	0.8920	0.8973	0.8993	0.9993	1.0000	1.0000
	8	0.9447	0.9487	0.9500	1.0000	1.0000	1.0000
	9	0.9720	0.9747	0.9760	1.0000	1.0000	1.0000
	10	0.9853	0.9873	0.9887	1.0000	1.0000	1.0000
$c = -2$	0	0.0580	0.0540	0.0527	0.0580	0.0540	0.0527
	1	0.0700	0.0660	0.0640	0.1147	0.1107	0.1093
	2	0.1087	0.1047	0.1027	0.3387	0.3353	0.3367
	3	0.1827	0.1807	0.1790	0.6873	0.6880	0.6893
	4	0.3040	0.3013	0.3027	0.9020	0.9047	0.9070
	5	0.4647	0.4620	0.4647	0.9740	0.9760	0.9767
	6	0.6333	0.6340	0.6353	0.9933	0.9940	0.9947
	7	0.7740	0.7760	0.7780	0.9980	0.9987	0.9987
	8	0.8713	0.8733	0.8757	0.9993	1.0000	1.0000
	9	0.9300	0.9320	0.9333	1.0000	1.0000	1.0000
	10	0.9620	0.9647	0.9653	1.0000	1.0000	1.0000
$c = -20$	0	0.0367	0.0440	0.0480	0.0367	0.0440	0.0480
	1	0.0407	0.0487	0.0533	0.0567	0.0647	0.0707
	2	0.0540	0.0627	0.0680	0.1240	0.1347	0.1407
	3	0.0773	0.0877	0.0933	0.2520	0.2620	0.2660
	4	0.1133	0.1247	0.1307	0.4347	0.4377	0.4397
	5	0.1633	0.1733	0.1797	0.6347	0.6287	0.6260
	6	0.2260	0.2367	0.2420	0.8007	0.7900	0.7873
	7	0.3027	0.3120	0.3153	0.9087	0.8993	0.8953
	8	0.3907	0.3963	0.3993	0.9640	0.9580	0.9553
	9	0.4853	0.4860	0.4863	0.9873	0.9847	0.9833
	10	0.5807	0.5767	0.5767	0.9960	0.9950	0.9940

can be estimated by using a Monte Carlo simulation method at different levels of persistence (c) and correlation coefficients (δ). The number of replications is $m = 5000$, and the rejection rates are reported at nominal significance levels 1%, 5% and 10%. The results corresponding to nominal size 5% are summarized in Table 2, and those with nominal sizes 1% and 10% are quite similar, and thus omitted.

Table 2 shows that the proposed test has good finite sample sizes at the nominal significance level 5%. The observed sizes converge to the nominal size with the increase of sample size and the rejection rates are almost equal to the nominal size when $n = 250$. Compared to Table 3 in Campbell and Yogo (2006), which reports the simulated finite-sample rejection rates by using different test methods, it is obvious that our proposed test produces the smallest distortions under the same setting. Furthermore, the proposed test procedure here is much easier to implement than other suitable tests such as the Bonferroni Q-test or the Sup-bound Q-test proposed in Campbell and Yogo (2006).

Table 4
Comparison of the results by the two-step estimation method with OLS estimation in Example 1.

δ	c		True	Two-step estimate (95% CI)	OLS estimate
−0.75	0	β_0	0	0.0473 (−0.1188, 0.2143)	0.0052
		β_1	−0.75	−0.7763 (−0.6870, −0.8656)	NA
		β_2	0	0.0093 (−0.0115, 0.0301)	0.0145
	−2	β_0	0	0.0490 (−0.1171, 0.2160)	0.2322⋆
		β_1	−0.75	−0.7818 (−0.6925, −0.8711)	NA
		β_2	0	−0.0036 (−0.0245, 0.0171)	0.0027
	−20	β_0	0	0.0381 (−0.0788, 0.1557)	0.0454
		β_1	−0.75	−0.7826 (−0.6933, −0.8720)	NA
		β_2	0	0.0139 (−0.0228, 0.0507)	0.0026
−0.95	0	β_0	0	0.0268 (−0.0678, 0.1215)	0.2392⋆
		β_1	−0.95	−0.9664 (−0.9242, −1.0086)	NA
		β_2	0	−0.0017 (−0.0067, 0.0100)	0.0158⋆
	−2	β_0	0	0.0229 (−0.0555, 0.1019)	0.0482
		β_1	−0.95	−0.9634 (−0.9213, −1.0056)	NA
		β_2	0	0.0038 (−0.0061, 0.0136)	0.0055
	−20	β_0	0	0.0191 (−0.0361, 0.0746)	−0.0150
		β_1	−0.95	−0.9344 (−0.8923, −0.9766)	NA
		β_2	0	0.0058 (−0.0115, 0.0231)	−0.0383⋆

We also evaluate the power of the proposed misspecification test in finite samples. A natural way of doing so is to consider the ability of the test to reject local alternatives. We consider a sequence of alternatives with form $H_a : \beta_2 = b/n$, where $b \in [1, 10]$. The sample size is added to the alternative hypothesis because $\hat{\beta}_2$ has a convergence rate n; thus, the effect of sample size on the power could be canceled out by using above local alternatives. Note that in the above alternative form, if $b = 0$, the alternative collapses into the null hypothesis and the power becomes the test size. The simulated local asymptotic powers at the nominal significance level 5% are reported in Table 3. The results shown in the table imply that the powers rise and approach to one when the value of b increases, while the difference among various sample sizes is not clear due to the elimination of the sample size effect by the local alternatives. In addition, the power goes to one much quicker when $\delta = -0.95$ and the test is more powerful when the regressor is more persistent. These results are consistent with both the above analysis and the asymptotic theorem in Section 3.

Now we choose a typical sample to show the estimate $\hat{\beta}$ when the sample size $n = 250$ with different values of c and δ. This typical sample is selected in such a way that its ADE values (ADE_{β_i}, $0 \le i \le 2$) are equal to their corresponding medians of the 5000 replications. Then we use the proposed two-step estimation procedure to estimate the coefficients. We also consider the 95% confidence interval associated with each coefficient, which is useful to check whether the estimate is different from the true value of the parameter and in some further inferences. Furthermore, we compare the estimates $\hat{\beta}_0$ and $\hat{\beta}_2$ by using the two-step estimation method proposed in this paper to those obtained by applying the OLS estimation method directly to the regression of y_t onto x_{t-1}, say,

$$y_t = \beta_0 + \beta_2 x_{t-1} + \varepsilon_t. \quad (12)$$

We use $\hat{\beta}_{0,OLS}$ and $\hat{\beta}_{2,OLS}$ to denote the OLS estimates for (12). The estimation results by using above two methods are listed in Table 4.

From the table, the true value of each parameter is included in the 95% confidence interval created by adopting the proposed two-step estimation method. Therefore, one cannot reject the null hypothesis that each coefficient β_i, $0 \le i \le 2$, is equal to the corresponding true value. As for the OLS estimates $\hat{\beta}_{j,OLS}$, where $j = 0$, 2, the previous discussion implies that the conventional t-test is invalid to check the significance of the OLS estimates. While since $\hat{\beta}_{0,OLS}$, is not included in the 95% confidence interval of $\hat{\beta}_0$ when

Table 5
Medians (Standard deviations) of 5000 ADE values by the two-step estimation procedure in Example 2.

δ	c	n	ADE_{β_0}	ADE_{β_1}	ADE_{β_2}
		50	0.1496 (0.1514)	0.0745 (0.0623)	0.0188 (0.0252)
	0	100	0.1185 (0.1127)	0.0487 (0.0438)	0.0061 (0.0075)
		250	0.0758 (0.0720)	0.0308 (0.0268)	0.0018 (0.0020)
		50	0.1439 (0.1536)	0.0737 (0.0682)	0.0262 (0.0269)
−0.75	−2	100	0.1128 (0.1251)	0.0502 (0.0448)	0.0105 (0.0125)
		250	0.0779 (0.0885)	0.0339 (0.0272)	0.0036 (0.0041)
		50	0.0982 (0.0906)	0.0685 (0.0632)	0.0568 (0.0529)
	−20	100	0.0748 (0.0682)	0.0491 (0.0444)	0.0291 (0.0264)
		250	0.0569 (0.0545)	0.0305 (0.0283)	0.0118 (0.0111)
		50	0.0706 (0.0715)	0.0352 (0.0294)	0.0089 (0.0119)
	0	100	0.0559 (0.0532)	0.0230 (0.0207)	0.0029 (0.0036)
		250	0.0358 (0.0340)	0.0146 (0.0126)	0.0009 (0.0010)
		50	0.0679 (0.0725)	0.0348 (0.0322)	0.0124 (0.0127)
−0.95	−2	100	0.0533 (0.0591)	0.0237 (0.0211)	0.0050 (0.0059)
		250	0.0368 (0.0418)	0.0160 (0.0128)	0.0017 (0.0019)
		50	0.0464 (0.0428)	0.0323 (0.0298)	0.0268 (0.0250)
	−20	100	0.0353 (0.0322)	0.0232 (0.0210)	0.0138 (0.0125)
		250	0.0269 (0.0257)	0.0144 (0.0133)	0.0056 (0.0052)

$c = -2$ and $\delta = -0.75$, at least we could conclude that $\hat{\beta}_{0,OLS}$ and $\hat{\beta}_0$ are different in this case. This is also applicable to other OLS estimates with asterisk characters in Table 4. The above analysis also verifies the previous statement that the OLS estimation method could lead to biased estimate in the finite sample case when the embedded endogeneity exists in the model, and the two-step estimation method proposed in this paper overcomes the endogeneity bias problem.

Example 2. In this example, we consider the case when the intercept in the AR(1) model for the regressor is non-zero. We add an intercept to the previous data generating process:

$$y_t = \beta_0 + \beta_1 \hat{u}_t + \beta_2 x_{t-1} + v_t, \qquad x_t = \theta + \rho x_{t-1} + u_t,$$

where $\theta = -0.3$, and all other settings are exactly the same as those in Example 1. We repeat the analysis in the above example by computing the ADE values in 5000 simulations, and also the empirical size and power of the proposed misspecification test statistics are computed for each simulation.

At first, we apply the proposed two-step estimation method to obtain the estimates of the three coefficients. The medians and standard deviations of 5000 ADE values are reported in Table 5. In this table, a similar conclusion to that from Table 1 can be made. When the sample size increases, both median and the standard deviation decrease, and the rate of convergence for ADE_{β_2} is faster than that for ADE_{β_0} and ADE_{β_1}. Next, the bias of $\hat{\beta}_2$ looks smaller when the regressor is more persistent. Finally, the biases of all three estimates are smaller when the correlation between the two innovations is stronger (when $\delta = -0.95$). One can also observe some differences between Tables 1 and 5. When there is a drift embedded in the AR(1) model for the regressor, keeping all other settings the same, one can see that the biases of $\hat{\beta}_0$ have larger magnitudes than without a drift, while those for $\hat{\beta}_2$ are much smaller. Furthermore, as the sample size increases, ADE_{β_2} declines much faster than it does in Table 1. This is consistent with the results in Theorem 2, which shows that the convergence rate for $\hat{\beta}_2$ should be $n^{3/2}$ with a drift, compared to a rate of n without a drift.

In the following we illustrate the finite sample performance of the test statistics shown in (11). Again we would like to test the hypothesis $H_0 : \beta_i = 0$ versus $H_0 : \beta_i \neq 0, 0 \leq i \leq 2$, respectively. In order to emphasize the effect of the sample size, three sample sizes $n = 100$, 250 and 500 are considered. In addition, we only focus on the cases when $c = 0$ and $c = -2$, which means the regressor is highly persistent or non-stationary. We repeat the testing procedure as in Example 1 with test statistics given in (11), and the simulated sizes corresponding to nominal significance level 5% are displayed in Table 6. From the table, one can conclude that the simulated sizes converge to corresponding nominal size with the increase of the sample size in most cases, which is similar to what observed in Example 1. The distortions of the simulated sizes when the sample size is small ($n = 100$ for example) look obvious in the tests for β_0 and β_2, but these sizes converge to the corresponding nominal sizes quickly with the rise of sample size. This observation makes sense since when there is a nonzero drift in the AR(1) model, one expects larger domains for x_t and y_t, which leads to higher type I errors in the hypothesis testing. While according to Theorem 2, these distortions should disappear quickly when the sample size increases.

Finally, we discuss the local asymptotic powers of the proposed misspecification test statistics. Similar to the above size simulation, here we also consider the sample sizes as $n = 100$, 250 and 500, and again the persistence level of the regressor is chosen as $c = 0$ or -2. We also consider the local alternatives, but the form should be changed to $H_a : \beta_2 = b/n^{3/2}$, since the convergence rate for $\hat{\beta}_2$ is $n^{3/2}$ in this case. Further, in order to show the effectiveness of the test, we choose $b \in [10, 100]$ with an increment by 10. The simulated test powers are shown in Table 7. The results follow similar patterns to what observed from Example 1.

Table 6
Test sizes for testing predictability in Example 2 at nominal size 5%.

		$\delta = -0.75$			$\delta = -0.95$		
		β_0	β_1	β_2	β_0	β_1	β_2
	$n = 100$	0.0540	0.0580	0.0940	0.0540	0.0580	0.0940
$c = 0$	$n = 250$	0.0530	0.0510	0.0770	0.0530	0.0510	0.0770
	$n = 500$	0.0540	0.0520	0.0660	0.0540	0.0510	0.0660
	$n = 100$	0.0280	0.0580	0.0280	0.0270	0.0580	0.0280
$c = -2$	$n = 250$	0.0400	0.0540	0.0520	0.0400	0.0540	0.0520
	$n = 500$	0.0480	0.0500	0.0510	0.0480	0.0500	0.0510

Table 7
Test powers in Example 2 at nominal size 5%.

	b	$\delta = -0.75$			$\delta = -0.95$		
		$n = 100$	$n = 250$	$n = 500$	$n = 100$	$n = 250$	$n = 500$
	0	0.0940	0.0770	0.0660	0.0940	0.0770	0.0660
	10	0.1710	0.1680	0.1640	0.5240	0.5230	0.5190
	20	0.4760	0.4780	0.4730	0.9590	0.9680	0.9720
	30	0.8220	0.8130	0.8110	0.9960	0.9980	1.0000
	40	0.9480	0.9540	0.9580	1.0000	1.0000	1.0000
$c = 0$	50	0.9820	0.9880	0.9940	1.0000	1.0000	1.0000
	60	0.9940	0.9960	0.9980	1.0000	1.0000	1.0000
	70	0.9980	0.9980	1.0000	1.0000	1.0000	1.0000
	80	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	90	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0	0.0280	0.0520	0.0510	0.0280	0.0520	0.0510
	10	0.0380	0.0620	0.0700	0.0790	0.1160	0.1300
	20	0.0720	0.1060	0.1200	0.3480	0.3820	0.3860
	30	0.1500	0.1920	0.2120	0.7900	0.7520	0.7360
	40	0.3000	0.3400	0.3500	0.9600	0.9380	0.9280
$c = -2$	50	0.5200	0.5190	0.5160	0.9940	0.9870	0.9860
	60	0.7300	0.7000	0.6850	1.0000	0.9980	0.9960
	70	0.8710	0.8320	0.8170	1.0000	1.0000	1.0000
	80	0.9420	0.9150	0.9010	1.0000	1.0000	1.0000
	90	0.9730	0.9560	0.9500	1.0000	1.0000	1.0000
	100	0.9880	0.9780	0.9760	1.0000	1.0000	1.0000

5. A real example

In the previous section, we conduct a Monte Carlo simulation study to illustrate the effectiveness of the two-step estimation method and also the validity of the proposed misspecification test statistics. In this section, we apply these methodologies to test the predictability of the equity return.

There are three time series, including the monthly S&P 500 excess returns, log dividend–price ratio and log earning–price ratio from 1932:12 to 2011:12. The monthly S&P 500 stock price, dividends and earnings are available on the web site.[1] The dividend–price ratio is calculated as the ratio of average dividends during the last year over the current stock price, and the earning–price ratio is computed as the average earnings over past ten years divided by the current price. The one-month T-bill from CRSP is selected to calculate the excess return. To check the predictability of the monthly return, we regress the excess return on the financial variable such as the log d–p ratio and the log e–p ratio, respectively. In order to catch different performances of these variables during various time periods, similar to Campbell and Yogo (2006), we investigate two time periods: 1932:12–1990:12 (the first period) and 1991:01–2011:12 (the second period). For the sake of simplicity, we use the log e–p ratio as the univariate predictor during the first period and the log d–p ratio as the predictor during the second period.

First, we check the persistency of the predictors and whether a non-zero drift should be included in their AR(1) models. During the first period, applying the augmented Dickey–Fuller (ADF) test on the log e–p ratio, the p-value is 0.0141 and the p-value to test the zero drift in the AR(1) model is 0.0011. This means the log e–p might not be the I(1) process during this period but the drift in its AR(1) model might not be zero. For the second period, the ADF test (p-value is 0.6098) fails to reject the null of unit root for the log d–p ratio and the p-value for the zero drift test is 0.3160. Therefore, the log d–p ratio should be the unit root process without drift during the second study period.

Next we apply the least squares method to estimate the coefficients in the AR(1) model: $x_t = \theta + \rho x_{t-1} + u_t$ for the log e–p ratio during the first period. The estimated values for parameters θ and ρ are $\hat{\theta} = -0.0645$ and $\hat{\rho} = 0.9777$. The 95% confidence interval for ρ is $[0.9772,\ 0.9781]$ and for c is $[-10.41,\ -10.01]$. Thus during this period, the log e–p ratio satisfies the NI(1) assumption on the regressor in the proposed model, since $\hat{\rho}$ is close to one but the 95% confidence interval for c does not contain zero. For the second period, $\hat{\rho} = 1.0003$ and the 95% confidence interval for ρ is $[0.9791,\ 1.0035]$, which implies that the log d–p ratio is a unit root process without drift during this period.

To establish the empirical relationship between the monthly excess return and the first lag of log e–p ratio (or log d–p ratio), now we consider the following linear regression of the response variable on a univariate regressor

$$r_t = \beta_0 + \beta_1 \hat{u}_t + \beta_2 x_{t-1} + v_t,$$

where $\hat{u}_t$ is the OLS estimated residual of the AR(1) model: $x_t = \theta + \rho x_{t-1} + u_t$, with $\theta \neq 0$ for the first period but zero for the second period. In the above model, r_t represents the monthly return of S&P 500 in month t, and the predictor x_{t-1} is the log e–p ratio (or the log d–p ratio) in month $t-1$. Applying the ordinary least squares method again, one can obtain the estimates $\hat{\beta}_i, i = 0, 1,$ and 2, over the two study periods, respectively. We also construct the confidence interval for each coefficient. The critical values are simulated according to the distributions given in (8)–(10) for the period 1932:12–1990:12 and in (11) for the period 1991:01–2011:12 respectively, with the population variances replaced by sample variances, and substituting c and θ by their estimates from the empirical data. The estimated coefficients, together with their 95% confidence intervals, are listed in Table 8. From the results in the table, one can observe that the confidence interval for coefficient β_2 covers zero during the period 1932:12–1990:12 for the log e–p ratio but does not do so for the second period for the log d–p ratio, which implies that the log d–p ratio has the predictability for the stock return during the second study period but the log e–p ratio might not be able to forecast stock return during the first study period. It is also interesting to notice that the stock return has a negative relation to the log d–p ratio since 1991.

Table 8
Two-step estimates (95% C.I.) in Section 5.

	1932:12–1990:12	1991:01–2011:12
β_0	0.0668 (−0.0552, 0.0785)	−0.0070 (−0.0078, −0.0062)
β_1	−0.9932 (−0.9936, −0.9927)	−0.9825 (−0.9830, −0.9821)
β_2	0.0225 (−0.0182, 0.0269)	−0.0021 (−0.0037, −0.0006)

6. Conclusion

In this paper, we study a predictive regression model which has the ability to include the regressor to be an I(1) or NI(1) process and allows the so-called embedded endogeneity in the model. By conducting a linear relationship between the two innovations in the predictive regression model, we develop a two-step estimation procedure for estimating the coefficients and study their asymptotic distributions. Then we investigate the usefulness of the estimation procedure using simulation examples and a real example, in which we illustrate how to construct the confidence interval for each coefficient based on the proposed asymptotic distribution, and also compare our estimation results with those obtained by directly applying the OLS method to a predictive regression with endogeneity to show that the estimates obtained by the proposed two-step method should be superior to the OLS method in the finite sample cases. However, as addressed by Viceira (1997) and Paye and Timmermann (2006), the stability of the predictability should also be a big concern in the reality. Thus, the predictive regression with time-varying coefficients deserves further discussion. The statistical inference based on the time-varying coefficient predictive regression model, such as the hypothesis testing of the stability of predictability, also deserves further investigation.

Appendix

In this appendix, we present briefly the derivations of the main results given in previous sections. Before embracing on the proofs, we define some notations and list a lemma. First, let C be a finite positive constant, which might be different in different appearances and $\|\cdot\|$ denotes the Euclidean norm. Define $\xi_{n,t} = x_t/\sqrt{n}$, $\zeta_t = \sum_{k=1}^{\infty} E_t(v_{t+k})$, and $W_v(r) = \sum_{t=1}^{[nr]} v_t/\sqrt{n}$. Let $\mathscr{F}_s = \sigma(\xi_{n,t}, v_t\ :\ t \le s)$ be the smallest sigma-field containing all the past histories of $(\xi_{n,s}, v_s)$ for all n. $E_t(X) = E(X|\mathscr{F}_t)$. In addition, for any $0 \le r \le 1$, $\Lambda_n^*(r) = \frac{1}{n}\sum_{t=1}^{[nr]}(x_t - x_{t-1})\zeta_t - x_{[nr]}\zeta_{[nr]+1}/n$.

Lemma A.1. *Under Assumptions* A1–A3, *we have* $\sup_{0\le r\le 1} |\Lambda_n^*(r)| = o_p(1)$.

Proof of Lemma A.1. By the definition of $\Lambda_n^*(r)$ and (2), one has

$$\sup_{0\le r\le 1} |\Lambda_n^*(r)| = \sup_{0\le r\le 1} \left| \frac{1}{n}\sum_{t=1}^{[nr]}[(\rho-1)x_{t-1} + u_t]\zeta_t - \frac{1}{n}x_{[nr]}\zeta_{[nr]+1} \right|$$
$$\le \frac{1-\rho}{n}\sup_{0\le r\le 1}\left|\sum_{t=1}^{[nr]} x_{t-1}\zeta_t\right| + \frac{1}{n}\sup_{0\le r\le 1}\left|\sum_{t=1}^{[nr]} u_t\zeta_t\right|$$

[1] See the home page of Professor Robert Shiller at http://www.econ.yale.edu/shiller/data.htm.

$$+\frac{1}{n}\sup_{0\le r\le 1}|x_{[nr]}\zeta_{[nr]+1}|$$
$$\equiv L_1+L_2+L_3.$$

We will check L_1, L_2 and L_3 respectively. First, based on the previous assumption, it is easy to see that $\|v_t\|_p = O(1)$ for any $p > 0$. Applying Minkowski's inequality for α-mixing sequence (McLeish, 1975), Davydov's inequality and Assumption A3, one obtains

$$\|\zeta_t\|_{\delta_1}\le\sum_{k=1}^{\infty}\|E_t(v_{t+k})\|_{\delta_1}\le C\sum_{k=1}^{\infty}\alpha^{1/\delta_1-1/\delta_2}(k)\|v_{t+k}\|_{\delta_2}\le C.$$

From Chebyshev's inequality, together with Assumption A3, one can show that for any $\epsilon > 0$,

$$P\left(\sup_{t\le n}|\zeta_t|>\sqrt{n}\epsilon\right)\le\epsilon^{-\delta_1}n^{1-\delta_1/2}E|\zeta_t|^{\delta_1}\le Cn^{1-\delta_1/2}\to 0$$

since $\delta_1 > 2$. Therefore, it follows that

$$L_3=\frac{1}{n}\sup_{0\le r\le 1}|x_{[nr]}\zeta_{[nr]+1}|$$
$$\le\sup_{0\le r\le 1}|x_{[nr]}|/\sqrt{n}\sup_{0\le r\le 1}|\zeta_{[nr]+1}|/\sqrt{n}=o_p(1)$$

by $\sup_{0\le r\le 1}|x_{[nr]}| = O_p(\sqrt{n})$ and the continuous mapping theorem (see, e.g., Theorem 2.7 in Billingsley (1999)). Similarly, since $\rho = 1 + c/n$,

$$L_1=\frac{1-\rho}{n}\sup_{0\le r\le 1}\left|\sum_{t=1}^{[nr]}x_{t-1}\zeta_t\right|\le\frac{-c}{n}\sup_{0\le r\le 1}|x_{[nr]}\zeta_{[nr]+1}|=o_p(1).$$

Now it turns to L_2. At first, consider the term $u_t\zeta_t - E(u_t\zeta_t)$, for any $m \ge 1$,

$$\|E_{t-m}[u_t\zeta_t-E(u_t\zeta_t)]\|_{\delta_3}\le\sum_{k=1}^{\infty}\|E_{t-m}[u_tv_{t+k}-E(u_tv_{t+k})]\|_{\delta_3}$$
$$=\sum_{k=1}^{\infty}\|E_{t-m}(u_tv_{t+k})\|_{\delta_3}=\sum_{k=1}^{m}\|E_{t-m}(u_tv_{t+k})\|_{\delta_3}$$
$$+\sum_{k=m+1}^{\infty}\|E_{t-m}(u_tv_{t+k})\|_{\delta_3}\equiv N_1+N_2,$$

where $E(u_tv_{t+k}) = 0$ and $0 < \delta_3 < 2$ is defined in Assumption A3. By McLeish' inequality again and Assumption A3, it follows that

$$\|E_{t-m}(u_tv_{t+k})\|_{\delta_3}\le C\,\alpha^{1/\delta_3-1/2}(m)\|u_tv_{t+k}\|_2\le C\,\alpha^{1/\delta_3-1/2}(m).$$

Hence,

$$N_1=\sum_{k=1}^{m}\|E_{t-m}(u_tv_{t+k})\|_{\delta_3}\le Cm\,\alpha^{1/\delta_3-1/2}(m).$$

As for N_2, it follows by applying McLeish' inequality again that

$$\|E_{t-m}(u_tv_{t+k})\|_{\delta_3}\le\|u_tE_{t-m}(v_{t+k})\|_{\delta_3}$$
$$\le\|u_t\|_{\delta_4\delta_3/(\delta_4-\delta_3)}\|E_{t-m}(v_{t+k})\|_{\delta_4}$$
$$\le C\,\|u_t\|_{\delta_4\delta_3/(\delta_4-\delta_3)}\alpha^{1/\delta_4-1/2}(k)\|v_{t+k}\|_2$$
$$\le C\|u_t\|_{\delta_4\delta_3/(\delta_4-\delta_3)}\alpha^{1/\delta_4-1/2}(k)$$
$$\le C\alpha^{1/\delta_4-1/2}(k),$$

where $\delta_3 < \delta_4 < 2$ are also defined in Assumption A3. Then,

$$N_2=\sum_{k=m+1}^{\infty}\|E_{t-m}(u_tv_{t+k})\|_{\delta_3}\le C\sum_{k=m+1}^{\infty}\alpha^{1/\delta_4-1/2}(k).$$

Combining the above results, one obtains

$$\|E_{t-m}[u_t\zeta_t-E(u_t\zeta_t)]\|_{\delta_3}=N_1+N_2$$
$$\le C\left[m\alpha^{1/\delta_3-1/2}(m)+\sum_{k=m+1}^{\infty}\alpha^{1/\delta_4-1/2}(k)\right]\to 0.$$

It is easy to show by corollary to Theorem 3.3 in Hansen (1992) that

$$\frac{1}{n}\sup_{0\le r\le 1}\left|\sum_{t=1}^{[nr]}[u_t\zeta_t-E(u_t\zeta_t)]\right|\xrightarrow{p}0,$$

which, together with the fact that $|E(u_t\zeta_t)| = 0$, implies that

$$L_2=\frac{1}{n}\sup_{0\le r\le 1}\left|\sum_{t=1}^{[nr]}u_t\zeta_t\right|\xrightarrow{p}0.$$

This, in conjunction with the universe convergence of L_1 and L_3, implies that

$$\sup_{0\le r\le 1}|\Lambda_n^*(r)|\equiv L_1+L_2+L_3=o_p(1).$$

This proves the lemma. □

Proof of Theorem 1. First, based on the fact that $\sup_{2\le t\le n}x_{t-1}/\sqrt{n} = O(1)$, we consider the following transformation of the term $\hat{u}_t$:

$$\hat{u}_t=x_t-\hat{\rho}x_{t-1}=x_t-\rho x_{t-1}+(\rho-\hat{\rho})x_{t-1}$$
$$=u_t+\sqrt{n}(\rho-\hat{\rho})x_{t-1}/\sqrt{n}=u_t+O_p(n^{-1/2}),$$

where the last equation is easy to prove. Hence in the following proof, we will use u_t to replace $\hat{u}_t$ without further explanation. To prove the theorem, recall that

$$\hat{\beta}=\left(X^TX\right)^{-1}X^Ty=\left(X^TX\right)^{-1}X^T(X\beta+v)=\beta+\left(X^TX\right)^{-1}X^Tv.$$

Therefore,

$$D_n(\hat{\beta}-\beta)=\left(n^{-1}D_n^{-1}X^TXD_n^{-1}\right)^{-1}n^{-1}D_n^{-1}X^Tv\equiv S_n^{-1}T_n,\qquad\text{(A.1)}$$

where $S_n = n^{-1}D_n^{-1}X^TXD_n^{-1}$ and $T_n = n^{-1}D_n^{-1}X^Tv$. We will check the two terms respectively. At first,

$$S_n=\frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix}1&&\\&1&\\&&1/\sqrt{n}\end{pmatrix}\begin{pmatrix}1&u_t&x_{t-1}\\u_t&u_t^2&u_tx_{t-1}\\x_{t-1}&u_tx_{t-1}&x_{t-1}^2\end{pmatrix}\begin{pmatrix}1&&\\&1&\\&&1/\sqrt{n}\end{pmatrix}$$
$$=\frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix}1&u_t&x_{t-1}/\sqrt{n}\\u_t&u_t^2&u_tx_{t-1}/\sqrt{n}\\x_{t-1}/\sqrt{n}&u_tx_{t-1}/\sqrt{n}&x_{t-1}^2/n\end{pmatrix}.\qquad\text{(A.2)}$$

For an α-mixing process satisfying (1.3) of Rio (1995), one obtains the limiting distribution of the partial sum of u_t: $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_t\xrightarrow{d}N(0,\sigma_u^2)$, where $\sigma_u^2 = Var(u_1) + 2\sum_{k=2}^{\infty}Cov(u_1, u_k)$. Thus, we conclude

$$\frac{1}{n}\sum_{t=1}^{n}u_t=O(n^{-1})=o(1).$$

Similarly, based on the results $E\left(n^{-1}\sum_{t=1}^{n}u_t^2\right) = \sigma_u^2$ and $E\left(n^{-1}\sum_{t=1}^{n}u_t^2\right)^2 = o(1)$, we obtain

$$\frac{1}{n}\sum_{t=1}^{n}u_t^2=\sigma_u^2+o(1).$$

These results, together with the asymptotic properties for NI(1) processes in (5), imply that

$$S_n = \begin{pmatrix} 1 & 0 & \int_0^1 K_c(r)dr \\ 0 & \sigma_u^2 & 0 \\ \int_0^1 K_c(r)dr & 0 & \int_0^1 K_c^2(r)dr \end{pmatrix} + o_p(1) \equiv S_0 + o_p(1). \tag{A.3}$$

Second,

$$T_n = n^{-1}D_n^{-1}X^T v = \frac{1}{n}\sum_{t=1}^n \left(1, u_t, x_{t-1}/\sqrt{n}\right)^T v_t. \tag{A.4}$$

Applying Theorem 1 in Rio (1995) again, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n v_t \xrightarrow{d} N(0, \sigma_v^2) = W_v(1).$$

Similarly, by recalling that u_t and v_t are independent, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n u_t v_t \xrightarrow{d} W_{uv}(1),$$

where $W_{uv}(s)$ is a one-dimensional Brownian motion with variance σ_{uv}^2. Finally, applying Theorem 4.4(b) in Hansen (1992) and combining with Lemma A.1, we obtain

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n x_{t-1}v_t \xrightarrow{d} \int_0^1 K_c(r)dW_v(r). \tag{A.5}$$

These results, together with (A.4), imply that

$$\sqrt{n}T_n = \left(W_v(1), W_{uv}(1), \int_0^1 K_c(r)dW_v(r)\right)^T + o_p(1). \tag{A.6}$$

Hence,

$$\sqrt{n}S_n^{-1}(s)T_n(s) \xrightarrow{d} S_0^{-1}\left(W_v(1), W_{uv}(1), \int_0^1 K_c(r)dW_v(r)\right)^T. \tag{A.7}$$

Since x_{t-1} and v_t are uncorrelated, $K_c(r)$ and $W_v(r)$ are uncorrelated and $\int_0^1 K_c(r)dW_v(r)$ follows a mixed normal distribution. The conditional covariance of

$$\left(W_v(1), W_{uv}(1), \int_0^1 K_c(r)dW_v(r)\right)^T$$

is diag $\left(\sigma_v^2, \sigma_{uv}^2, \sigma_v^2\int_0^1 K_c^2(r)dr\right)$, where σ_{uv}^2 denotes the variance of $W_{uv}(1)$. By combining (A.1) and (A.7), we establish the asymptotic distribution in Theorem 1. That is,

$$\sqrt{n}\,D_n\left(\hat\beta - \beta\right) \xrightarrow{d} \text{MN}(0, \Sigma_\beta).$$

This proves the theorem. □

Proof of Proposition 1. By our previous analysis, when variable x_t satisfies the autoregressive model in (6) with nonzero drift and nonzero value of c, it follows

$$x_t = \theta\sum_{s=0}^{t-1}\rho^s + \sum_{r=0}^{t-1}\rho^r u_{t-r} = \frac{n\theta}{c}(e^{ct/n} - 1) + o(1),$$

where the last equation is due to $\rho = 1 + c/n = \exp(n^{-1}c) + o(n^{-1})$, which implies that

$$n^{-1}x_{[nr]} \xrightarrow{p} \theta(e^{cr} - 1)/c + o(1), \quad (t-1)/n \le r < t/n$$

where $[nr]$ denotes the integer part of nr. Next, we have

$$n^{-2}\sum_{t=1}^n x_{t-1} = n^{-2}\sum_{t=1}^n \frac{n\theta}{c}(\rho^{t-1} - 1) + o(1) = \frac{\theta}{c^2}(e^c - 1 - c) + o(1) \xrightarrow{p} \theta(e^c - 1 - c)/c^2.$$

Similarly,

$$n^{-3}\sum_{t=1}^n x_{t-1}^2 = \frac{\theta^2}{c^2}n^{-1}\sum_{t=1}^n (1 - \rho^{t-1})^2 + o(1) \xrightarrow{p} \theta^2[(e^c - 2)^2 + 2c - 1]/2c^3.$$

Finally, by Theorem 1 in Rio (1995), we have $n^{-1/2}\sum_{t=1}^n v_t \rightarrow N(0, \sigma_v^2)$, which implies that

$$n^{-3/2}\sum_{t=1}^n x_{t-1}v_t = n^{-1/2}\sum_{t=1}^n \frac{\theta}{c}(1 - \rho^{t-1})v_t + o(1)$$

is also normally distributed with mean zero and variance $\frac{\theta^2\sigma_v^2}{c^2}\left[(e^c - 2)^2 + 2c - 1\right]$. Thus, part (d) is proved. □

Proof of Theorem 2. When the drift in the AR model of the regressor is nonzero, we have the properties: $x_{[nr]}/n = O_p(1)$ and

$$\begin{bmatrix} n^{1/2}(\hat\theta - \theta) \\ n^{3/2}(\hat\rho - \rho) \end{bmatrix} \xrightarrow{d} \text{N}(0, \sigma_u^2 Q^{-1}),$$

where $Q = \begin{pmatrix} 1 & \theta/2 \\ \theta/2 & \theta^2/3 \end{pmatrix}$. Thus, the estimated residual of the AR(1) model for x_{t-1} follows:

$$\hat u_t = x_t - \hat\theta - \hat\rho x_{t-1} = x_t - \theta - \rho x_{t-1} + (\theta - \hat\theta) + (\rho - \hat\rho)x_{t-1} = u_t + O_p(n^{-1/2}),$$

which has the same order as if the drift is zero. The estimated β still follows (A.1), but with D_n replaced by D_n^*:

$$D_n^*(\hat\beta - \beta) = \left(n^{-1}D_n^{*-1}X^TXD_n^{*-1}\right)^{-1} n^{-1}D_n^{*-1}X^T v \equiv S_n^{*-1}T_n^*,$$

where the definitions of $S_n^*(s)$ and $T_n^*(s)$ are very similar to $S_n(s)$ and $T_n(s)$ in the proof of Theorem 1, but with D_n^* instead of D_n. To find the asymptotic property for S_n^*, notice

$$S_n^* = \frac{1}{n}\sum_{t=1}^n \begin{pmatrix} 1 & u_t & x_{t-1}/n \\ u_t & u_t^2 & u_t x_{t-1}/n \\ x_{t-1}/n & u_t x_{t-1}/n & x_{t-1}^2/n^2 \end{pmatrix}.$$

Based on the asymptotic properties given in (7), we can easily obtain

$$S_n^* = \begin{pmatrix} 1 & 0 & \theta(e^c - 1 - c)/c^2 \\ 0 & \sigma_u^2 & 0 \\ \theta(e^c - 1 - c)/c^2 & 0 & \theta^2[(e^c - 2)^2 + 2c - 1]/2c^3 \end{pmatrix} + o_p(1) \equiv S_0^* + o_p(1).$$

Further,

$$T_n^* = n^{-1}D_n^{*^{-1}}X^T v = \frac{1}{n}\sum_{t=1}^n (1, u_t, x_{t-1}/n)^T v_t.$$

Applying the above asymptotic properties again, we can show that

$$\sqrt{n}T_n^* = \left(W_v(1), W_{uv}(1), |\theta|W_v(1)\sqrt{[(e^c - 2)^2 + 2c - 1]/2c^3}\right)^T + o_p(1).$$

Therefore,

$$\sqrt{n}\, S_n^{*-1}(s) T_n^*(s) \xrightarrow{d} S_0^{*-1} \left(W_v(1), W_{uv}(1), |\theta| W_v(1) \sqrt{[(e^c-2)^2+2c-1]/2c^3} \right)^T.$$

The covariance matrix of

$$\left(W_v(1), W_{uv}(1), |\theta| W_v(1) \sqrt{[(e^c-2)^2+2c-1]/2c^3} \right)^T$$

is diag $\left(\sigma_v^2, \sigma_{uv}^2, \theta^2\sigma_v^2[(e^c-2)^2+2c-1]/2c^3\right)$. By Slusky's theorem again, we obtain the asymptotic distribution in Theorem 2. □

References

Amihud, Y., Hurvich, C., 2004. Predictive regression: A reduced-bias estimation method. Journal of Financial and Quantitative Analysis 39, 813–841.

Amihud, Y., Hurvich, C., Wang, Y., 2009. Multiple-predictor regressions: hypothesis testing. Review of Financial Studies 22, 413–434.

Billingsley, P., 1999. Convergence of Probability Measures, second ed. Wiley, New York.

Cai, Z., 2002. A two-stage approach to additive time series models. Statistica Neerlandica 56, 415–433.

Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association 95, 941–956.

Campbell, J., Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Econometrics 81, 27–60.

Cavanagh, C.L., Elliott, G., Stock, J.H., 1995. Inference in models with nearly integrated regressors. Econometric Theory 11, 1131–1147.

Dangl, T., Halling, M., 2012. Predictive regressions with time-varying coefficients. Journal of Financial Economics 106, 157–181.

Elliott, G., Stock, J.H., 1994. Inference in time series regression when the order of integration of a regressor is unknown. Econometric Theory 10, 672–700.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton.

Hansen, B.E., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 489–500.

Lewellen, J., 2004. Predicting returns with financial ratios. Journal of Financial Economics 74, 209–235.

Mankiw, N.G., Shapiro, M., 1986. Do we reject too often? Small sample properties of tests of rational expectation models. Economics Letters 20, 139–145.

McLeish, D.L., 1975. A maximal inequality and dependent strong laws. The Annuals of Probability 3, 829–839.

Nelson, C.R., Kim, M.J., 1993. Predictable stock returns: the role of small sample bias. Journal of Finance 48, 641–661.

Paye, B.S., Timmermann, A., 2006. Instability of return prediction models. Journal of Empirical Finance 13, 274–315.

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

Polk, C., Thompson, S., Vuolteenaho, T., 2006. Cross-sectional forecasts of the equity premium. Journal of Financial Economics 81, 101–141.

Rio, E., 1995. The functional law of the iterated logarithm for stationary strongly mixing sequences. The Annals of Probability 23, 1188–1203.

Rossi, B., 2007. Expectation hypothesis tests and predictive regressions at long horizons. Econometrics Journal 10, 1–26.

Stambaugh, R., 1986. Bias in regressions with lagged stochastic regressors. Working Paper, University of Chicago.

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

Torous, W., Valkanov, R., Yan, S., 2004. On predicting stock returns with nearly integrated explanatory variables. Journal of Business 77, 937–966.

Viceira, L.M., 1997. Testing for structural change in the predictability of asset returns. Working Paper, Harvard University.