Unifying estimation and inference for linear regression with stationary and integrated or near-integrated variables

Shaoxin Hong[a], Daniel J. Henderson[b], Jiancheng Jiang[c,*], Qingshan Ni[d]

[a]*Shandong University, Jinan 250100, China*
[b]*University of Alabama, Tuscaloosa, AL 35406, USA*
[c]*University of North Carolina at Charlotte, NC 28223, USA*
[d]*Hunan University, Changsha 410082, China*

Abstract

In linear time series regression, there is a discrepancy in the limiting distributions of least-squares estimators for stationary and integrated or near-integrated variables. This makes statistical inference difficult in practice as it must be decided which distribution should be used prior to constructing interval estimates and conducting hypothesis tests. This motivates us to develop a multiple linear regression model with stationary and integrated or near-integrated state variables to reduce this difficulty and propose a unifying inference procedure for the coefficient estimates. To facilitate this unifying inference, we propose a weighted estimation technique. The asymptotic distributions of the proposed estimators are developed. However, the asymptotic variance cannot be estimated well in practice due to erratic behavior of the residual-based variance estimate. As traditional bootstrap interval estimates do not work either (as bootstrap sampling from the residuals in this nonstationary setting can be riddled with many large values), a random weighting bootstrap method is proposed for constructing confidence regions. The proposed method works well (with time constant or time varying error variance) in our simulations and outperforms existing approaches. We use these methods to study the predictability of asset returns and further show how they can be implemented when some of our state variables are endogenous.

Keywords: Integrated, nearly integrated, random weighting, unit roots, weighted estimation equation.
JEL: C12, C58, G12

1. Introduction

In a simple linear time series model, we regress an outcome variable y_t on a state variable x_t

$$y_t = \beta_0 + x_t\beta_1 + v_t, \tag{1}$$

for $t = 1, 2, \ldots, n$, where β_0 and β_1 are unknown parameters and v_t is an error series. Our focus is on the estimation and inference of the parameter β_1 which depends on the order of integration of x_t. We know how to deal with x_t when it is I(0) (stationary). We know how to deal with x_t when it is I(1)

*Corresponding author
Email addresses: henryhong@sdu.edu.cn (Shaoxin Hong), djhenderson1@ua.edu (Daniel J. Henderson), jjiang1@uncc.edu (Jiancheng Jiang), nqs906@163.com (Qingshan Ni)

(nonstationary: unit root). But what happens when we don't know the order of integration or if the series is nearly integrated of order 1 (NI(1) – from above or below)?

In macroeconomics, many variables are considered to be I(1) and a common approach in a linear model is to perform first differences to make the process I(0). Estimation can be performed via ordinary least-squares (OLS), the asymptotic properties are well known for stationary data, and inference is straightforward. In many cases, we have a firm belief that a certain variable is stationary or nonstationary. However, suppose we are unsure of the order of integration. Common practice consists of using economic theory and/or performing statistical tests (e.g., Dickey and Fuller [31]). However, as Nelson and Plosser [63, pp. 152] note, we must be aware that "none of the tests presented, formal and informal, can have power against a [time series] alternative with an [autoregressive] root arbitrarily close to unity."

As expected, there are a host of variables whereby we are unsure whether they are stationary, have unit roots or are nearly integrated of order 1. For example, historically, it has been assumed that the unemployment rate fluctuates around some "natural rate" (I(0)). However, it has been argued (Blanchard and Summers [9, 10]) that changes in the unemployment rate can have long lasting impacts on the unemployment rate: the hysteresis hypothesis (I(1)). The literature shows mixed results (Clark [28], Nelson and Plosser [63], Song and Wu [88]). Similarly, evidence for stationarity of inflation is mixed, even within a given study (Ball [5], Culver and Papell [29], Ng and Perron [64]) and this is important as stationary prices are required for conventional sticky-price models (Dornbusch [32]). There is also a debate with respect to exchange rates (Roll [82], Adler and Lehmann [2], Mark [61], Chowdhury and Sdogati [27]). Many studies have been unable to reject the null hypothesis that the exchange rate is I(1). If the exchange rate between two nations is in fact a random walk, it would imply that purchasing power parity does not hold. Finally, as we will show in our application, there is a debate as to whether or not stock prices follow a random walk. While the efficient market hypothesis suggests that all information known about a stock is already factored into the price of that stock, the finance literature has found empirical puzzles where stock returns appear to be predictable (Phillips and Lee [74]). Recent research has modeled these series via a process with an autoregressive root that is close to, but exceeds, unity (Phillips and Yu [78]). In short, there is ample evidence to suggest that the order of integration of many variables is still in question (I(0), I(1) or NI(1)). In practice, we may be interested in using one or more of these variables in a regression. We can either make strong assumptions on the order of integration or we can switch to a methodology that does not rely on assuming a particular order of integration for estimation/inference. This is the approach we take.

More generally (and more formally), consider the multivariate version of our initial equation:

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + v_t, \tag{2}$$

where $E(v_t|\mathbf{x}_t) = 0$, and $\mathbf{x}_t = (1, x_{t,1}, \ldots, x_{t,k})^\top$ is a vector consisting of stationary and nonstationary variables and $\boldsymbol{\beta}$ is a $(k+1)\times 1$ parameter vector. Component $x_{t,i}$ $(i = 1, 2, \ldots, k)$ may be stationary or nonstationary, and if it is nonstationary, it follows that

$$x_{t,i} = \rho_i \, x_{t-1,i} + u_{t,i}, \tag{3}$$

where $\rho_i = 1 + \gamma_i/n$, and $u_{t,i}$ is a martingale difference. When $\gamma_i = 0$, $x_{t,i}$ is an I(1) process; when $\gamma_i < 0$, $x_{t,i}$ is called a nearly I(1) (NI(1), slightly stationary) process; when $\gamma_i > 0$, $x_{t,i}$ is called a nearly I(1) (NI(1), slightly explosive) process.

In our empirical application, the predictability of asset returns, y_t will represent excess stock returns in period t, while $\mathbf{x}_t$ will be a set of lagged stock characteristics (e.g., earnings-price ratio) and interest rate related variables (e.g., term spread). We find evidence that each type of state variable is present in our dataset (I(0), I(1) and NI(1)). Further, given that endogneity is known to exist in our dataset, we will show how our methodology can handle endogenous state variables (Section 5.1). After accounting for both endogeneity and acknowledging that the order of integration of our state variables is uncertain, we will be able to find some evidence of predictability of stock returns (when the estimate of a given element of $\boldsymbol{\beta}$ is significantly different from zero) and will compare our results with existing approaches.

It turns out, for model (2), ordinary least-squares can be used to estimate $\boldsymbol{\beta}$, but Theorem 1 in Section 2 shows that limiting distributions of least-squares estimators of $\boldsymbol{\beta}$ are different for I(0), I(1) and NI(1) cases, which makes inference for $\boldsymbol{\beta}$ difficult, as we have to decide which limiting distribution is used for inference (Basawa, Mallik, McCormick, Reeves and Taylor [6]). In particular, when $k = 1$ and ρ_1 is close to one, the variance estimate of the limiting distribution behaves erratically (Chan, Li and Peng [20]; Zhu, Cai and Peng [93]). Hence, it is desirable to develop a unifying inference tool which does not require us to *a priori* choose the limiting distribution. Several solutions have been proposed in literature: Phillips and Lee [74] and Kostakis, Magdalinos and Stamatogiannis [54] develop a new test for testing predictability by using an extended instrumental variable (dubbed as IVX) based inference; Choi, Jacewitz and Park [26] propose a testing procedure based on Cauchy estimation, which unifies the cases of nearly integrated and unit root; Phillips and Lee [75] propose a inference procedure for nonstationary variable regression that enables robust chi-square testing by using the mechanism of self-generated instruments called IVX instrumentation developed by Magdalinos and Phillips [60]; Ren, Tu and Yi [80] add an additional lag for highly persistent predictors to balance the predictive regression and propose simple testing procedures for univariate and multivariate predictive regressions; Zhu, Cai and Peng [93], Liu, Yang, Cai and Peng [59] and Hong, Jiang, Jiang and Xiao [48] consider empirical likelihood based approaches (among others).

In this paper, our solution is to propose a weighted estimation equation (WEE) method. This approach gives us a unifying limiting distribution whether the state variables are I(0), I(1), or NI(1). The asymptotic normal distribution of the proposed estimator is established. However, we cannot use this distribution to make inference directly, as the asymptotic variance depends on the variance of the error σ_v^2. This cannot be estimated well via the residuals due to the erratic behavior of residuals in the current nonstationary setting. Furthermore, standard bootstrap techniques do not provide an appropriate approximation to the distribution of our proposed estimator. This motivates us to propose a random weighting bootstrap estimation approach, based on the WEE. The proposed WEE estimation and random weighting bootstrap methods provide a necessary tool for unifying inference for model (2). We provide the necessary asymptotic theory and finite sample evidence to support the theory via simulations.

The remainder of this paper is organized as follows. In Section 2, we introduce multiple linear regression models with stationary and intergrated or nearly integrated variables and the corresponding estimation procedures. Unifying asymptotic distributions of the proposed estimators are established. Section 3 presents our random weighting bootstrap method for constructing confidence regions of the parameters and shows consistency of the bootstrap procedure. In Section 4, we compare the finite sample performance of our method relative to existing methods via simulations. In Section 5, we employ our methods to study the predicitability of stock returns. The final section (Section 6) concludes. Proofs of our theorems are relegated to the Appendix.

2. Multiple regression with stationary and integrated or nearly integrated variables

In this section, we begin by introducing linear regression models with both stationary and nonstationary state variables. We discuss least-squares estimation as well as the asymptotic distribution of these estimators. We offer an alternative estimation procedure and show that the estimators are asymptotically normal. However, we will show that both asymptotic and standard bootstrap procedures are insufficient for inference (which will be the motivation for our weighted bootstrap procedure in Section 3).

2.1. Least-squares estimation

Using OLS for model (2), we obtain the estimator $\widehat{\boldsymbol{\beta}}$ minimizing the least-squares objective function

$$\sum_{t=1}^{n}(y_t - \mathbf{x}_t^\top \boldsymbol{\beta})^2. \tag{4}$$

The solution, $\widehat{\boldsymbol{\beta}}$, has the closed form:

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{n} \mathbf{x}_t^{\otimes 2}\right)^{-1} \sum_{t=1}^{n} \mathbf{x}_t y_t, \tag{5}$$

where $a^{\otimes 2} = aa^\top$ for any matrix a. In the classical setting where $\mathbf{x}_t$ is stationary, under mixing conditions, it is known that $\widehat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent and asymptotically normal. However, as shown in Theorem 1, the limiting distributions of $\widehat{\boldsymbol{\beta}}$ are different for stationary and nonstationary cases.

Our focus here is on $\boldsymbol{\beta}$, the marginal impact of $\mathbf{x}_t$ on y_t. Without loss of generality, assume that $\mathbf{x}_t = (1, X_{t,1}^\top, X_{t,2}^\top)^\top$, where $X_{t,1}$ is a $d \times 1$ vector of I(1) or near I(1) variables (NI(1)), and $X_{t,2}$ is a $(k-d) \times 1$ vector of stationary I(0) variables[1] $(0 \le d \le k)$. In particular, let $X_{t,1} = (x_{t,1}, \ldots, x_{t,d})^\top$, and $X_{t,2} = (x_{t,d+1}, \ldots, x_{t,k})^\top$. As the intercept is included in the model, without loss of generality, we

[1]As we can write $\mathbf{x}_t$ as $\mathbf{x}_t = A(X_{t,1}^\top, X_{t,2}^\top)^\top$, with A being a $k \times k$ permutation matrix; our arguments apply in general.

assume each of the stationary components has zero mean, i.e. $E(X_{t,2}) = 0$. In these models, we will first assume that

(A_0) $X_{t,1} = AX_{t-1,1} + U_t$, where $A = I_d + n^{-1}C$, with $C = \text{diag}\,(\gamma_1, \ldots, \gamma_d)$, and $U_t = (u_{t,1}, \ldots, u_{t,d})^\top$ is a d-dimensional noise with mean zero and variance $\Sigma_0 = \text{Var}\,(U_t)$ being positive definite.

(A_1) The series $(X_{t,2}, u_t)$ is strictly stationary and α-mixing with mixing coefficients $\alpha^*(s)$ satisfying $\sum_\ell \ell^a \{\alpha^*(\ell)\}^b < \infty$, for some $0 < b < 1$ and $a > b$. Denote by $\Gamma_2(h) = E\{[(X_{t+h,2} - E(X_{t+h,2})][X_{t,2} - E(X_{t,2})]^\top\}$ the autocovariance matrix function of $X_{t,2}$.

(A_2) The error series (u_t, v_t) is strictly stationary and α-mixing with the mixing coefficients satisfying the condition $\sum_{\ell=1}^{\infty} \ell^a \{\alpha(\ell)\}^b < \infty$, for some $0 < b < 1$ and $a > b$. Both u_t and v_t have the $(2+\delta)$th moment for some $\delta > 0$. Assume that $E(v_t|\mathbf{x}_t) = 0$. Denote by $\gamma_v(h)$ the autocovariance function of v_t. Let $\sigma_v^2 = \gamma_v(0) + 2\sum_{h=1}^{\infty} \gamma_v(h)$. Assume that v_t is independent of $\mathbf{x}_t$ at all leads and lags.

(A_3) If $x_{t,i}$ is stationary, $E|x_{t,i}|^4 < \infty$.

Similar to the literature (e.g., Phillips [71]; Phillips and Ouliaris [76]; Hansen [45]; Hansen [46]), in Assumption (A_0), we assume that Σ_0 is positive definite. Cointegration among the elements of $X_{t,1}$ is excluded with this assumption. Model (2), along with assumptions (A_0) and (A_1), contains many existing models well studied in the literature.[2] Conditions (A_2)–(A_3) are relatively mild. The finite fourth moment for stationarity (A_3) can be relaxed to a $2 + \delta$ moment for achieving robustness, but this leads to some efficiency loss as is common in robustness estimation.

For any unit root or nearly I(1) variable $x_{t,i}$, we let $U_{n,i}(r) = n^{-1/2} x_{[nr],i}$, where $r = t/n$ and $[w]$ denotes the integer part of w. Further, let $U_n(r) = (U_{n,1}(r), \ldots, U_{n,d}(r))^\top$ and $U_\gamma(r) = (U_{\gamma_1}(r), \ldots, U_{\gamma_d}(r))^\top$, where $U_{\gamma_i}(r) = \int_0^r \exp\{(r-s)\gamma_i\}\, dW_u^{(i)}(s)$ is a diffusion process, and $(W_u^{(1)}(s), \ldots, W_u^{(d)}(s))^\top$ is a d-dimensional Brownian motion with positive definite variance-covariance matrix Σ_u with $\Sigma_u = \text{Var}(U_1) + 2\sum_{s=2}^{\infty} E(U_1 U_s^\top)$. Under regularity conditions on $u_{t,i}$ (for example, the regularity conditions in Phillips [71]), we have that,

$$U_n(r) \Rightarrow U_\gamma(r) \tag{6}$$

as $n \to \infty$, where "$\Rightarrow$" represents weak convergence.[3] Lemma 1 in our appendix demostrates that the weak convergence result in (6) can be strengthened to strong convergence. This is important in order for us to derive our theoretical results. In addition, by Donsker's theorem (Billingsley [11], Theorem 19.2), $\sigma_v^{-1} n^{-1/2} \sum_{i=1}^{[nr]} v_i$ weakly converges to a standard Brownian motion $B_v(r)$ $(r \in [0,1])$ with mean $B_v(0) = 0$ and $var\{B_v(r)\} = r$.

[2]See Speckman [87], Carroll, Fan, Gijbels and Wand [17], Robinson [81], Chen, Liu and Tsay [25], Juhl and Xiao [52], Juhl and Xiao [53], Cai, Li and Park [14], Gao and Phillips [39], and Chen, Gao and Li [24], among others.

[3]Similarly, "$\xrightarrow{d}$" represents convergence in distribution.

For ease of exposition, we introduce the following additional notation. Let $D_n = \text{diag}(\sqrt{n}, nI_d, \sqrt{n}I_{k-d})$, where I_s represents an s dimensional identity matrix for any nonnegative integer s. Let $W_u^{(\ell)} = \int_0^1 \{U_\gamma(r)\}^{\otimes \ell}\, dr$ for $\ell = 1, 2$, $\Sigma_{2,v} = [\Gamma_2(0)\gamma_v(0) + 2\sum_{h=1}^{\infty} \Gamma_2(h)\gamma_v(h)]/\sigma_v^2$, and $\Sigma = \text{diag}\{\Sigma_u, \Sigma_{2,v}\}$, where $\Sigma_u = \begin{pmatrix} 1 & \{W_u^{(1)}\}^\top \\ W_u^{(1)} & W_u^{(2)} \end{pmatrix}$.

With each of the four assumptions in hand, the following theorem demonstrates that $\widehat{\boldsymbol{\beta}}_i$ is $\sqrt{n}$-consistent if $x_{t,i}$ is I(0), but n-consistent if $x_{t,i}$ is NI(1) or I(1).

Theorem 1. *Under conditions (A_0)–(A_3), with the probability going to 1, we have that*

$$D_n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sigma_v R^{-1}\Big(\int_0^1 dB_v(r), \int_0^1 U_\gamma^\top(r)\, dB_v(r), \int_0^1 dB_{ev}^\top(r)\Big)^\top + o_p(1), \tag{7}$$

where $B_{ev}(r)$ is a $(k-d)$-dimensional Brownian motion on $[0,1]$ with covariance matrix $\Sigma_{2,v}$. Further, $D_n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} MN(0, \sigma_v^2 \text{diag}\{\Sigma_u^{-1}, \Gamma_2^{-1}(0)\Sigma_{2,v}\Gamma_2^{-1}(0)\})$, given $\mathcal{F}_u = \{U_t, 1 \le t \le n\}$.

Corollary 1. *When $\{v_t\}_{t=1}^n$ are iid white noise, $R = \Sigma$ and the asymptotic variance-covariance matrix in Theorem 1 becomes* $\text{diag}\{\sigma_v^2\Sigma_u^{-1}, \sigma_v^2\{E(X_{t,2}^{\otimes 2})\}^{-1}\}$.

Theorem 1 indicates that the result holds for models with *multiple* local-to-unity predictors. To show this in practice, we will conduct simulations to include a non-zero intercept and four predictors in the new predictive regression model: one stationary process, one unit-root process, and two local-to-unity processes (see Section 4). Our results demonstrate that our method works well in all scenarios considered in the simulations.

The mixed normal distribution (MN) is defined in Phillips [72] and Park and Phillips [69]. When $E(X_{t,2}) = 0$ and v_t are iid white noise, the asymptotic conditional covariances between the coefficient estimator $\widehat{\boldsymbol{\beta}}_1$ of $X_{t,1}$ and $\widehat{\boldsymbol{\beta}}_2$ of $X_{t,2}$ and between the coefficient estimator $\widehat{\beta}_0$ of the intercept and $\widehat{\boldsymbol{\beta}}_2$ are both 0 (see Remark 1 of Cai and Wang [15]). If the stationary variables are centered, an oracle property arises, in that $\boldsymbol{\beta}_1$ can be estimated as if $\boldsymbol{\beta}_2$ is known, and vice versa. The result here is strikingly different from traditional linear regression models with stationary data. However, the limiting property of $\widehat{\boldsymbol{\beta}}$ depends on the stationarity of $\mathbf{x}_t$, which makes inference difficult for $\boldsymbol{\beta}$. It leads to different reference distributions for hypothesis testing. We therefore need to examine the stationarity of each component of $\mathbf{x}_t$ prior to conducting hypothesis testing. We could, instead, use a bootstrap procedure to obtain critical values, however, the full sample bootstrap is inconsistent for NI(1) or infinite variance settings (Datta [30]; Hall and Jing [43]). This motivates us to propose a weighted estimation procedure in Section 2.2.

2.2. Weighted score equation estimation

Given the issues with inference in least-squares model, we move to an estimator that leads to a parameter vector that converges to the normal distribution at the same rate ($\sqrt{n}$) for each component. Recall that $\widehat{\boldsymbol{\beta}}$ solves the estimation equation

$$\sum_{t=1}^{n} \mathbf{x}_t(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}) = 0. \tag{8}$$

Similar to Hong, Jiang, Jiang and Xiao [48], consider the following weighted estimation equations:

$$\sum_{t=1}^{n} \Omega_t \mathbf{x}_t (y_t - \mathbf{x}_t^\top \boldsymbol{\beta}) = 0, \tag{9}$$

where $\Omega_t = \operatorname{diag}(1, \omega_{t,1}, \ldots, \omega_{t,k})$ is a sequence of non-negative definite diagonal matrices. These are chosen and used to weigh down the contributions of data points to the score equations in (8). Our weighted score equation estimator of $\boldsymbol{\beta}$ is obtained by solving (9):

$$\widehat{\boldsymbol{\beta}}_\omega = \left(\sum_{t=1}^{n} \Omega_t \mathbf{x}_t^{\otimes 2} \right)^{-1} \sum_{t=1}^{n} \Omega_t \mathbf{x}_t y_t, \tag{10}$$

where

$$\omega_{t,i} = \begin{cases} 1 & \text{if} \quad i \in \mathcal{I} \\ (1 + \|\mathbf{x}_{t,\mathcal{I}^c}\|^2)^{-1/2} & \text{otherwise} \end{cases}$$

where $\mathcal{I} = \{i : n^{-\frac{1}{2}} \log(n) \max_{1 \le t \le n} |x_{t,i}| < c^*\}$, $\mathbf{x}_{t,\mathcal{I}^c}$ is the subvector of $\mathbf{x}_t$ with indexes not in $\mathcal{I}$, and c^* is a positive constant. The weight sequence involves the constant c^*. We can regard c^* as a tuning parameter and choose it by cross-validation or by maximizing the efficiency of $\widehat{\boldsymbol{\beta}}_\omega$ (see the paragraph under Theorem 2). The above weighting scheme is related to existing approaches. When $k = 1$, it reduces to the weight approach used in Chan, Li and Peng [20] and Zhu, Cai and Peng [93], but these approaches preassume the x-variable is I(1) or NI(1) while ours does not. For stationary cases with more than 2nd moments, given a positive constant c^*, we have

$$\Pr\left(\max_{1 \le t \le n} n^{-\frac{1}{2}} \log(n) |x_{t,i}| < c^* \right) \to 1, \text{ as } n \to \infty.$$

Hence, we set $\omega_{t,i}$ to be equal to one if $n^{-\frac{1}{2}} \log(n) \max_t |\tilde{x}_{t-1,i}| < c^*$, so that it leads to the most efficient unweighted least squares score equation with probability going to one.

The following theorem sets up a unifying limit for the proposed estimator.

Theorem 2. *Suppose model (2) holds with $x_{t,i}$ being stationary or an AR(1) process with $|\rho_i| < 1$ or $\rho_i = 1 + \gamma_i/n$ for some $\gamma_i \in \mathcal{R}$. Under the conditions (A_0)–(A_3), with probability going to 1,*

$$\left[\sum_{t=1}^{n} (\Omega_t \mathbf{x}_t)^{\otimes 2} \right]^{-1/2} \left(\sum_{t=1}^{n} \Omega_t \mathbf{x}_t^{\otimes 2} \right) \left(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta} \right) \xrightarrow{d} N(0, \sigma_v^2 I_k),$$

where I_k is a $k \times k$ identity matrix.

Theorem 2 shows that, under model (2), $\widehat{\boldsymbol{\beta}}_\omega$ is asymptotically unbiased and its variance can be approximated by

$$\Sigma_{\widehat{\boldsymbol{\beta}}_\omega} = \left[\left(\sum_{t=1}^{n} (\Omega_t \mathbf{x}_t)^{\otimes 2} \right)^{-1/2} \left(\sum_{t=1}^{n} \Omega_t \mathbf{x}_t^{\otimes 2} \right)^2 \left(\sum_{t=1}^{n} (\Omega_t \mathbf{x}_t)^{\otimes 2} \right)^{-1/2} \right]^{-1} \sigma_v^2, \tag{11}$$

where $\Sigma_{\widehat{\boldsymbol{\beta}}_\omega}$ is of the typical 'sandwich' form. Since σ_v^2 does not depend on c^*, we suggest choosing c^* to maximize the efficiency or equivalently the generalized variance of $\widehat{\boldsymbol{\beta}}_\omega$. In particular, we choose $\widehat{c}^* = \arg\max_{c^*} D_n(c^*)$, where

$$D_n(c^*) = \|\{\sum_{t=1}^n (\Omega_t \tilde{\mathbf{x}}_{t-1})^{\otimes 2}\}^{-1/2} (\sum_{t=1}^n \Omega_t \tilde{\mathbf{x}}_{t-1}^{\otimes 2})\|_F,$$

where $\|\cdot\|_F$ is the Hilbert norm of a matrix. c^* can also be chosen via cross-validation if we regard it as a tuning parameter [see Cai et al. (2000) and Jiang (2014)], but this is more computationally involved.

Using the limiting distribution in Theorem 2, we can construct a Wald type confidence region for $\boldsymbol{\beta}$:

$$\left(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}\right)^\top S_{\widehat{\boldsymbol{\beta}}_\omega}^{-1} \left(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}\right) \le \chi^2_{k,\alpha}, \tag{12}$$

where $\chi^2_{k,\alpha}$ is the α-quantile of $\chi^2(k)$ and $S_{\widehat{\boldsymbol{\beta}}_\omega}$ is defined in the same way as $\Sigma_{\widehat{\boldsymbol{\beta}}_\omega}$ but with σ_v^2 being replaced by $\widehat{\sigma}_v^2 = n^{-1}\sum_{t=1}^n \left(y_t - \mathbf{x}_t^\top \widehat{\boldsymbol{\beta}}_\omega\right)^2$. Similarly, we can construct the Wald type CI for β_j as

$$\left(\widehat{\beta}_{j,\omega} - \beta_j\right)^\top s_{\widehat{\boldsymbol{\beta}}_{j,\omega}}^{-1} \left(\widehat{\beta}_{j,\omega} - \beta_j\right) \le \chi^2_{1,\alpha},$$

for $j = 1, 2, \ldots, k$, where $s_{\widehat{\boldsymbol{\beta}}_{j,\omega}}$ is the estimate of the jth diagonal element of $\Sigma_{\widehat{\boldsymbol{\beta}}_\omega}$. However, in our simulations, and as noted in Chan, Li and Peng [20], such an interval estimate for $\boldsymbol{\beta}$ has unacceptable coverage probabilities for NI(1), I(1) and explosive cases as $\widehat{\sigma}_v^2$ behaves erratically. In addition, it assumes a constant variance for v_t.

Theorem 3. *Suppose model* (2) *is heteroskedastic such that*

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \sigma(\mathbf{x}_t) v_t^*, \tag{13}$$

where the scale $\sigma(\cdot)$ *is bounded away from* 0 *and* ∞ *and* v_t^* *satisfies* $E(v_t^*|\mathbf{x}_t) = 0$ *and* $E(v_t^{*2}|\mathbf{x}_t) = 1$, *Then, under the conditions of Theorem 2,*

$$\left[\sum_{t=1}^n (\Omega_t \mathbf{x}_t)^{\otimes 2} \sigma^2(\mathbf{x}_t)\right]^{-1/2} \left(\sum_{t=1}^n \Omega_t \mathbf{x}_t^{\otimes 2}\right) \left(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}\right) \xrightarrow{d} N(0, I_k).$$

If $\sigma(\cdot) = \sigma_v$, *this reduces to Theorem 2.*

Remark 1. *Theorem 3 indicates that, the estimator* $\widehat{\boldsymbol{\beta}}_\omega$ *of model* (2) *is consistent and asymptotically normal, even when the conditional variance of the error is misspecified. However, since* $\sigma(\cdot)$ *is unknown, it does not allow one to make inference for* $\boldsymbol{\beta}$ *in model* (13)*, unless one has a nonparametric estimator of* $\sigma(\cdot)$. *Even if such a nonparametric estimator is available, it will affect the convergence rate of* $\widehat{\boldsymbol{\beta}}_\omega$.

As pointed out in Zhu, Cai and Peng [93], a bootstrap method may be used for the interval estimate, but the full sample bootstrap method is inconsistent for NI(1) and infinite variance AR processes. This will be shown in more detail in our simulations (Section 4).

The above problem may be solved by employing a subsample bootstrap method, but the subsample size is difficult to choose in practice (Hall and Jing [43]; Datta [30]). For this approach, theoretical justification seems difficult, and the associated computational burden is heavy. This motivates us to propose the following random weighting bootstrap approach to circumvent this difficulty.

3. Random weighting bootstrap estimation

Given that the variance of $\widehat{\boldsymbol{\beta}}_\omega$ is difficult to estimate, we consider a random weighting bootstrap method to attack this problem. Random weighting bootstraps have been studied by Rubin [84] in a Bayesian setting, by Jin, Ying and Wei [51] for minimizing a general objective function with a U-process structure, and by Zheng, Zhu, Li and Xiao [92] for GARCH models (among others). This approach is robust against misspecification of the variance structure of the error in model (2).

3.1. Bootstrap estimator

Consider the random weighting estimation equation:

$$\sum_{t=1}^{n} \xi_t \Omega_t \mathbf{x}_t (y_t - \mathbf{x}_t^\top \boldsymbol{\beta}) = 0, \tag{14}$$

where ξ_t is a iid random variable with mean and variance equal to 1 (e.g., generated from a normal distribution).[4] The solution to the objective function in equation (14) is

$$\widehat{\boldsymbol{\beta}}_\omega^* = \left(\sum_{t=1}^{n} \xi_t \Omega_t \mathbf{x}_t^{\otimes 2} \right)^{-1} \sum_{t=1}^{n} \xi_t \Omega_t \mathbf{x}_t y_t. \tag{15}$$

For the estimator in Equation (10) and the bootstrap estimator in Equation (15), we present the following theorem:

Theorem 4. *Suppose $x_{t,i}$ is a stationary or AR(1) process with $|\rho_i| < 1$ or $\rho_i = 1 + \gamma_i/n$ for some $\gamma_i \in \mathcal{R}$. If model (13) holds, under the conditions (A_0)–(A_3), as $n \to \infty$, we have*

$$\sup_{x \in R} |P\{\sqrt{n}(\widehat{\beta}_{j,\omega}^* - \widehat{\beta}_{j,\omega}) < x_j,\ j = 1, \ldots, k | F_n\} - P\{\sqrt{n}(\widehat{\beta}_{j,\omega} - \beta_j) < x_j,\ j = 1, \ldots, k\}| \longrightarrow 0,$$

almost surely, where F_n is the empirical distribution of the sample $\{x_t, y_t\}_{t=1}^n$.

Remark 2. *As claimed in Remark 1, $\widehat{\boldsymbol{\beta}}_\omega$ cannot be used for making inference about $\boldsymbol{\beta}$ when there exists heteroskedasticity, but Theorem 4 demonstrates that the bootstrap estimator $\widehat{\beta}_{j,\omega}^*$ consistently estimates the distribution of $\widehat{\boldsymbol{\beta}}_\omega$ and thus can be employed for statistical inference about $\boldsymbol{\beta}$ in the heteroskedastic model* (13).

3.2. Bootstrap algorithm

By Theorem 4, the conditional distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_\omega^* - \widehat{\boldsymbol{\beta}}_\omega)$ given F_n is asymptotically the same as the unconditional distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta})$. Therefore, we propose the following algorithm to simulate the distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta})$:

(i) Use the original sample to calculate $\widehat{\boldsymbol{\beta}}_\omega$.

[4] We can also generate data from the exponential distribution with parameter $\lambda = 1$. The results are similar in our simulations and we exclude them for brevity.

(ii) Draw a random sample $\{\xi_t : t = 1, \ldots, n\}$ from the Normal distribution with mean 1 and standard deviation 1, and calculate $\widehat{\boldsymbol{\beta}}_\omega^*$.

(iii) Repeat step (ii) many (B) times and obtain a sample of $\widehat{\boldsymbol{\beta}}_\omega^*$, $\{\widehat{\boldsymbol{\beta}}_\omega^{*(b)}\}_{b=1}^B$. Compute $\sqrt{n}(\widehat{\boldsymbol{\beta}}_\omega^{*(b)} - \widehat{\boldsymbol{\beta}}_\omega)$.

(iv) Use the bootstrap sample $\{\widehat{\boldsymbol{\beta}}_\omega^{*(b)} - \widehat{\boldsymbol{\beta}}_\omega\}_{b=1}^B$ to determine the standard error of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta})$.

4. Simulation

To investigate the finite sample performance, we run 1000 simulations with data generated from the following predictive model:

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + v_t, \tag{16}$$

where $\mathbf{x}_t = (1, x_{t-1,1}, \ldots, x_{t-1,4})^\top$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_4)^\top = (2, 0.5, 1, 1.5, -1)^\top$, with a sample size $n = 200$ or 400. For the predicting variables, we consider

$$x_{t,i} = \rho_i \, x_{t-1,i} + u_{t,i}, \quad u_{t,i} \sim N(0,1), \quad i = 1, \ldots, 4,$$

where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_4)^\top = (0.6, 1, 1 - \frac{5}{n}, 1 - \frac{50}{n})^\top$. That is, $x_{t,1}$ is $I(0)$, $x_{t,2}$ is $I(1)$, and both $x_{t,3}$ and $x_{t,4}$ are $NI(1)$ (these coefficients converge to unity as $n \to \infty$). For each of the 1000 simulations, we run $B = 1000$ bootstraps with random weight $\xi_t \sim N(1,1)$.

We compare the following four inference methods:

(a) Wald-type confidence region (CR) based on the asymptotic normal distribution in Theorem 2, with σ_v being estimated via the standard error of the residuals;

(b) Conditional bootstrap CR based on the normal theory in Theorem 2:

  (i) Draw a bootstrap sample $\{\widehat{\varepsilon}_t^*\}$ from the residuals $\{\widehat{\varepsilon}_t\}$ after estimating $\boldsymbol{\beta}$ with $\widehat{\boldsymbol{\beta}}_\omega$;

  (ii) Obtain the conditional bootstrap sample $\{\mathbf{x}_t, \widehat{y}_{t,\omega}^*\}_{t=1}^n$, where $\widehat{y}_{t,\omega}^* = \mathbf{x}_t \widehat{\boldsymbol{\beta}}_\omega + \widehat{\varepsilon}_t^*$;

  (iii) Calculate the weighted score equation estimate $\widetilde{\boldsymbol{\beta}}_\omega^*$ of $\boldsymbol{\beta}$, based on the conditional bootstrap sample;

  (iv) Repeat (i)-(iii) $B = 1000$ times;

  (v) Estimate $\Sigma_{\widetilde{\boldsymbol{\beta}}_\omega}$ in (11) by the sample covariance $S_{\widetilde{\boldsymbol{\beta}}_\omega^*}$ of $\{\widetilde{\boldsymbol{\beta}}_\omega^{*(b)} - \widehat{\boldsymbol{\beta}}_\omega\}_{b=1}^B$.

  (vi) Compute the test statistic $(\widetilde{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta})^\top S_{\widetilde{\boldsymbol{\beta}}_\omega^*}^{-1} (\widetilde{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta})$.

(c) Random weighting bootstrap CR based on the normal theory in Theorem 4: the implementation of the algorithm in Section 3.2 can be summarized as
For each simulation,

  (i) **Input**: Use the sample to calculate $\widehat{\boldsymbol{\beta}}_\omega$,
  for $b = 1, \cdots, B$ **do**;

(ii) Draw a random weighting bootstrap estimators $\widehat{\boldsymbol{\beta}}_{\omega}^{*(b)}$ by (15) ;

(iii) Compute $\widehat{\boldsymbol{\beta}}_{\omega}^{*(b)} - \widehat{\boldsymbol{\beta}}_{\omega}$;

(iv) **End for**;

(v) Estimate $\Sigma_{\widehat{\boldsymbol{\beta}}_{\omega}}$ in (11) by the sample covariance $S_{\widehat{\boldsymbol{\beta}}^*_{\omega}}$ of $\{\widehat{\boldsymbol{\beta}}_{\omega}^{*(b)} - \widehat{\boldsymbol{\beta}}_{\omega}\}_{b=1}^{B}$;

(vi) Compute the test statistic $(\widehat{\boldsymbol{\beta}}_{\omega} - \boldsymbol{\beta})^{\top} S^{-1}_{\widehat{\boldsymbol{\beta}}^*_{\omega}} (\widehat{\boldsymbol{\beta}}_{\omega} - \boldsymbol{\beta})$.

(d) The IVX-Wald test in Kostakis, Magdalinos and Stamatogiannis [54].

4.1. Constant error variance

In the homoskedastic case, $v_t \in \{v_t^{(1)}, v_t^{(2)}\}$, $v_t^{(1)} \sim N(0,1)$ and $v_t^{(2)} \sim t(3)$. The simulation results are reported in the first two columns of Tables 1 and 2. The averages and standard deviations (in parentheses) of $\widehat{\boldsymbol{\beta}}_{\omega}$ are given in the first five rows. With nominal size 5%, the coverage probabilities of all four methods ((a)–(d)) are relatively close to 95% for each sample size.

4.2. Time-varying error variance

Our performance is strong with a constant error variance, but it is worth allowing the conditional variance of model (16) to vary with time: $var(v_t|x_t) = \sigma_t^2$. It is routinely assumed to be constant in the literature for testing the predictability of stock returns (for $H_0 : \beta = 0$). This assumption is common because it allows researchers to derive the asymptotic theory when x_t is nonstationary. However, in practice, σ_t^2 might not be constant and it is possibly a function of a predicting variable.[5] As stated above, the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{\omega}$ will be matched automatically by applying our random weighthing method. We argue that it will also work for the conditionally heteroskedastic case.

To illustrate the finite sample perfomance of the proposed method in the heteroskedastic case, we consider a data generating process with time-varying conditional variance. For ease of exposition, we re-write model (16) as,

$$y_t = \mathbf{x}_t^{\top} \boldsymbol{\beta} + \sigma_t \epsilon_t, \tag{17}$$

where $v_t = \sigma_t \epsilon_t$ with ϵ_t being a sequence of iid random variables from the standard normal distribution. Four types of conditional standard deviations are considered: $\sigma_t \in \{\sigma_{t,1}, \sigma_{t,2}, \sigma_{t,3}, \sigma_{t,4}\}$, where $\sigma_{t,1} = 1 + \frac{x_{t-1,1}^2}{10}$, $\sigma_{t,2} = 1 + \frac{x_{t-1,2}^2}{50}$, $\sigma_{t,3} = 1 + \frac{9t}{n} + |4sin(\frac{\pi t}{60})|$ and $\sigma_{t,4}^2 = 0.5 + 0.5v_{t-1}^2 + 0.49\sigma_{t-1,4}^2$. $\sigma_{t,1}$ is a function of the stationary predicting variable, $\sigma_{t,2}$ represents nonstationary nonlinear heteroskedasticity (Park [68]), $\sigma_{t,3}$ is periodically changing with a deterministic time trend and $\sigma_{t,4}$ represents a GARCH error setting with a high persistence level of volatility clustering. The simulation results are reported in last four columns of Tables 1 and 2. As expected, our proposed method still provides correct coverage probabilities for all examples whereas the other methods do not.

[5] For instance, Park [68] considered a time series with the conditional heteroskedasticities that are given by a nonlinear function of an integrated process.

Table 1: The simulation results of $n = 200$: $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_4)^\top = (2, 0.5, 1, 1.5, -1)^\top$, $v_t^{(1)}$ and $v_t^{(2)}$ are homoskedastic errors while $\sigma_{t,1}$–$\sigma_{t,4}$ represent the heteroskedastic errors, the four inference methods are represented by (a) Wald-type, (b) Conditional bootstrap, (c) Random weight and (d) IVX-Wald

$n = 200$	$v_t^{(1)}$	$v_t^{(2)}$	$\sigma_{t,1}$	$\sigma_{t,2}$	$\sigma_{t,3}$	$\sigma_{t,4}$
β_0	2.0007	2.0028	1.9998	2.0203	2.0127	1.9606
	(0.1444)	(0.2415)	(0.1682)	(0.6012)	(1.0219)	(0.8042)
β_1	0.4994	0.4909	0.4986	0.5023	0.5084	0.5081
	(0.0606)	(0.0992)	(0.0895)	(0.2733)	(0.4964)	(0.2634)
β_2	1.0010	0.9995	1.0011	0.9975	1.0020	0.9944
	(0.0180)	(0.0308)	(0.0209)	(0.0630)	(0.1558)	(0.0847)
β_3	1.5004	1.5005	1.5008	1.4992	1.4910	1.4927
	(0.0256)	(0.0413)	(0.0302)	(0.1369)	(0.2196)	(0.1573)
β_4	-0.9984	-0.9962	-0.9981	-0.9888	-0.9953	-1.0041
	(0.0483)	(0.0864)	(0.0567)	(0.2336)	(0.4291)	(0.3068)
(a)	0.9360	0.9360	0.9020	0.8970	0.9240	0.9170
(b)	0.9370	0.9360	0.9000	0.8890	0.9200	0.9180
(c)	0.9470	0.9480	0.9440	0.9570	0.9520	0.9560
(d)	0.9530	0.9520	0.9200	0.9180	0.9310	0.9280

For the IVX-Wald test, a bandwidth parameter M_n is involved in the Newey-West-type estimator which Kostakis, Magdalinos and Stamatogiannis [54] employ for the estimation of long-run covariance matrices (the related conditions that need to be satisfied are $M_n \to \infty$ and $M_n/\sqrt{n} \to 0$). We use $M_n = \lfloor n^{1/3} \rfloor$ which satisfies the above conditions, where $\lfloor C \rfloor$ is the integer not greater than C. For $\sigma_{t,1}$ and $\sigma_{t,2}$, the coverage probabilities of method (d) are much lower than the nominal level (95%). To investigate if this is due to a particular choice of M_n, we consider a large range of M_n, $M_n \in \{\lfloor n^{1/5} \rfloor, \lfloor n^{1/4} \rfloor, \lfloor 1.5n^{1/3} \rfloor, \lfloor 2n^{1/3} \rfloor, \lfloor 2.5n^{1/3} \rfloor\}$, for $\sigma_{t,1}$ and $n = 400$. The corresponding coverage probabilities are $\{0.916, 0.916, 0.918, 0.917, 0.917\}$, respectively. The IVX-Wald test is robust over M_n, but does not have great coverage probabilities (< 0.95).

5. Predictability of stock returns

For more than a century (Dow [33]), attempts have been made to predict equity returns. Historically, stock returns were regressed onto lags of financial variables (e.g., earnings price ratio), via OLS, and t-statistics were used to determine whether there is evidence (i.e., statistical significance) of predictability of returns (Goyal and Welch [40]). However, first-order distribution theory requires the autoregressive root on the predictor variables to be strictly less than unity in order to use t-distributions. As we have shown, when predictor variables are persistent, the null distributions of our t-statistics are no longer

Table 2: The simulation results of $n = 400$: $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_4)^\top = (2, 0.5, 1, 1.5, -1)^\top$, $v_t^{(1)}$ and $v_t^{(2)}$ are homoskedastic errors while $\sigma_{t,1}$–$\sigma_{t,4}$ represent the heteroskedastic errors, the four inference methods are represented by (a) Wald-type, (b) Conditional bootstrap, (c) Random weight and (d) IVX-Wald

$n = 400$	$v_t^{(1)}$	$v_t^{(2)}$	$\sigma_{t,1}$	$\sigma_{t,2}$	$\sigma_{t,3}$	$\sigma_{t,4}$
β_0	2.0032	1.9847	2.0035	1.9644	1.9915	2.0053
	(0.0974)	(0.1752)	(0.1140)	(0.8126)	(0.7749)	(0.4852)
β_1	0.4989	0.4982	0.4989	0.5185	0.5081	0.4974
	(0.0403)	(0.0657)	(0.0604)	(0.3277)	(0.3538)	(0.1543)
β_2	1.0001	1.0003	1.0002	1.0009	0.9984	0.9971
	(0.0090)	(0.0154)	(0.0105)	(0.0573)	(0.0829)	(0.0392)
β_3	1.4992	1.5008	1.4992	1.5016	1.4967	1.4994
	(0.0126)	(0.0208)	(0.0147)	(0.1128)	(0.1100)	(0.0648)
β_4	-0.9995	-0.9969	-0.9993	-0.9993	-1.0012	-1.0033
	(0.0263)	(0.0450)	(0.0309)	(0.2100)	(0.2253)	(0.1118)
(a)	0.9510	0.9490	0.9150	0.8790	0.9290	0.9240
(b)	0.9500	0.9450	0.9140	0.8770	0.9200	0.9200
(c)	0.9540	0.9580	0.9520	0.9550	0.9420	0.9510
(d)	0.9520	0.9560	0.9170	0.8950	0.9400	0.9370

standard normals in large samples.

This has led to different conclusions (regarding preditability) in the literature (Torous, Valkanov and Yan [89]) and many econometric attempts to reconcile these conflicting findings. As Campbell and Yogo [12] note, "A difficulty with understanding the rather large literature on predictability is the sheer variety of test procedures that have been proposed, which have led to different conclusions about the predictabiltiy of returns." Given that the degree of persistence of the various predictor variables are unknown, our unifying inference methods are a perfect fit for this scenario.

Looking at the most recent 50 years of data, we will find evidence that both stationary and nonstationary state variables exist in our model. Using our proposed inference procedure, we find evidence of predictibility of asset returns via observing confidence intervals that do not contain the point zero.

The remainder of this section proceeds as follows. Section 5.1 proposes our model and how we address the inherent endogeneity. Section 5.2 describes our data while Section 5.3 presents the results of our study.

5.1. Model

In order to address whether or not stock returns are predictable (e.g., Phillips, Shi and Yu [77]), we define y_t to be the (potentially) predictable variable (e.g., excess stock return), in period t, and x_{t-1} to be the given state variable (e.g., earnings price ratio), in period $t-1$. Using a similar framework as that above, our predictive regression model is

$$y_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_t, \qquad x_t = \rho x_{t-1} + u_t, \qquad 1 \le t \le n, \tag{18}$$

where $E(u_t|x_{t-1}) = 0$, but $E(\varepsilon_t|x_{t-1})$ may be nonzero. For ease of exposition and to compare to past studies, we consider a single x, but remind the reader that we can allow x to be a vector.

In many applications, the correlation between innovations ε_t and u_t is nonzero (e.g., Table 1 in Torous, Valkanov and Yan [89] or Table 4 in Campbell and Yogo [12]), which causes a nonzero correlation between x_{t-1} and ε_t and creates "endogeneity". An implication of this is that directly regressing y_t on x_{t-1} may yield a biased OLS estimate of β_1. Again, ρ is the unknown degree of persistence of x_t: when $|\rho| < 1$, x_t is stationary (Viceira [91]; Amilud and Hurvich [3]; Amilud, Hurvich and Wang [4]); when $\rho = 1$, it is I(1); when $\rho = 1 + c/n$ with $c < 0$, it is nearly first-order integrated; when $\rho = 1 + c/n$ with $c > 0$, it is mildly explosive.[6] As our predictor x_t may be stationary or nonstationary and highly persistent, this makes modeling/inference difficult.

Several approaches have been designed to estimate model (18). The first one is the bias correction approach using information conveyed by the AR(1) process of x_t. For example, Kothari and Shanken [55] and Stambaugh [86] suggest a first-order bias-corrected OLS estimator, Amilud and Hurvich [3] propose a second-order bias-correction method, and Lewellen [57] study the conservative bias-correction vehicle which assumes the true autoregressive coefficient of AR(1) to be close to one. The second approach is the maximum likelihood estimation in Campbell and Yogo [12], who assume that innovations $\{\varepsilon_t, u_t\}$ are independently distributed bivariate normal $N(0, \Sigma)$. The third approach is based on linear projection and the least-squares method in Amilud and Hurvich [3], Amilud, Hurvich and Wang [4], and Cai and Wang [15]. Here the endogeneity may be removed from the model by the projection of ε_t onto u_t.

Formally, for $|\rho| < 1$, using the linear projection of ε_t onto u_t, $\varepsilon_t = \gamma u_t + v_t$, Amilud and Hurvich [3] re-express model (18) as

$$y_t = \beta_0 + \beta_1 x_{t-1} + \gamma u_t + v_t, \tag{19}$$

where v_t is a white noise sequence independent of x_t and u_t at all leads and lags. If u_t were known, the error in model (19) would satisfy the classical assumption of OLS without endogeneity. They applied a two-stage least-squares regression (2SLS). First, they obtained the OLS estimator $\widehat{\rho}$ of ρ. Then, they calculated the fitted residuals $\widehat{u}_t$. Finally, they regressed y_t on x_{t-1} and $\widehat{u}_t$ to obtain an estimate of β_1. For $\rho = 1 + c/n$ with $c \le 0$, Cai and Wang [15] investigated the 2SLS method and established its limiting distribution. Similar to above, these works show that the limiting distributions of the estimator of β_1 are different for I(0), I(1) and NI(1) cases, which again makes inference for β_1 difficult, as we have to decide which limiting distribution is used for inference. In addition, the above works cannot assess joint predictability of multiple state variables and thus may suffer from severe modeling bias problems due to

[6] A non-exhaustive set includes Elliott and Stock [34], Cavanagh, Elliott and Stock [18], Phillips [71], Torous, Valkanov and Yan [89], Campbell and Yogo [12], Polk, Thompson and Vuolteenaho [79], Rossi [83], and Cai and Wang [15].

missing important regressors.

Noting that $u_t = x_t - \rho x_{t-1} = \Delta x_t - (\rho - 1)x_{t-1}$, where $\Delta x_t = x_t - x_{t-1}$, we can rewrite model (19) in the following equivalent form:

$$y_t = \gamma_0 + \gamma_1 x_{t-1} + \gamma_2 \Delta x_t + v_t, \tag{20}$$

$$x_t = \rho x_{t-1} + u_t, \tag{21}$$

where $\gamma_0 = \beta_0$, $\gamma_1 = \beta_1 - \gamma(\rho - 1)$ and $\gamma_2 = \gamma$. That is,

$$\beta_1 = \gamma_1 + \gamma_2(\rho - 1).$$

Here the classical assumption of OLS holds, as the endogeneity disappears. The parameters γ_j and ρ can be estimated via OLS, which solve the equations:

$$\textstyle\sum_{t=1}^n \mathbf{z}_t(y_t - \mathbf{z}_t^\top \boldsymbol{\gamma}) = 0, \tag{22}$$

$$\textstyle\sum_{t=1}^n x_{t-1}(x_t - \rho x_{t-1}) = 0, \tag{23}$$

where $\mathbf{z}_t = (z_{t1}, z_{t2}, z_{t3})^\top = (1, x_{t-1}, \Delta x_t)^\top$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)^\top$. Since each of the above equations has the form of equation (8), the parameters in each equation can be estimated by the weighted estimation approach. We can use the random weighting bootstrap method to estimate the model parameters. Specifically, let

$$\mathcal{I}^* = \{i : n^{-\frac{1}{2}} \log(n) \max_{1 \le t \le n} |z_{t,i}| < c^*\},$$

then the estimation equations become

$$\textstyle\sum_{t=1}^n \xi_t^* \Omega_t^* \mathbf{z}_t(y_t - \mathbf{z}_t^\top \boldsymbol{\gamma}) = 0, \tag{24}$$

$$\textstyle\sum_{t=1}^n \xi_t w_t x_{t-1}(x_t - \rho x_{t-1}) = 0, \tag{25}$$

where $\Omega_t^* = \mathrm{diag}\{w_{t,1}, w_{t,2}, w_{t,3}\}$ with $w_{t,i} = 1$ if $i \in \mathcal{I}^*$ and $(1 + \|\mathbf{z}_{t,\mathcal{I}^{*c}}\|^2)^{-1/2}$ otherwise, $\mathbf{z}_{t,\mathcal{I}^{*c}}$ is the subvector of $\mathbf{z}_t$ with indexes not in $\mathcal{I}^*$, $w_t = 1$ if $n^{-1/2}(\log n)\max_{1\le t\le n}|x_{t-1}| < c^*$and $(1 + x_{t-1}^2)^{-1/2}$ otherwise, and $\{\xi_t^*\}$ and ξ_t are iid random weights with mean and variance equal to one. Once we estimate the random weighting estimators $\widehat{\boldsymbol{\gamma}}_\omega^*$ and $\widehat{\rho}_\omega^*$, we can obtain

$$\widehat{\beta}_{1,\omega}^* = \widehat{\gamma}_{1,\omega}^* + \widehat{\gamma}_{2,\omega}^*(\widehat{\rho}_\omega^* - 1), \tag{26}$$

which is our bootstrap estimator of β_1.

5.2. Data

Our data come directly from Welch and Goyal [90][7] and have been used in several other studies (e.g., Ceneizoglua and Timmermann [19], Kostakis, Magdalinos and Stamatogiannis [54], Ren, Tu and Yi

[7] The dataset updated until December 2021 is sourced from Amit Goyal's Web site: http://www.hec.unil.ch/agoyal/.

[80]). Our left-hand-side variable, monthly excess stock returns, is measured as the difference between the S&P 500 index, including dividends, and a one month Treasury bill rate. The month-end values come from the Center for Research in Security Press (CRSP). For the set of potential predictors of stock returns, we consider eleven variables and they can be divided into two groups.

The first set of predictors are **primarily stock characteristics**, they include: Dividend Price Ratio (dp), which is the difference between the log of dividends and the log of prices; Dividend Yield (dy), which is the difference between the log of dividends and the log of lagged prices; Earnings Price Ratio (ep); which is the difference between the log of earnings and the log of prices; Dividend Payout Ratio (de), which is the difference between the log of dividends and the log of earnings; Book-to-Market Ratio (bm), which is the ratio of book value to market value for the Dow Jones Industrial Average; Net Equity Expansion ($ntis$), which is the ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks.

The second set of predictors are **interest-rate related**, they include: Term Spread (tms), which is the difference between the long term yield on government bonds and the Treasury-bill; Default Yield Spread (dfy), which is the difference between BAA and AAA-rated corporate bond yields; Inflation ($infl$), which is the Consumer Price Index (All Urban Consumers) from the Bureau of Labor Statistics; Treasury Bill (tbl), which is the 3-Month Treasury Bill: Secondary Market Rate from the economic research data base at the Federal Reserve Bank at St. Louis (FRED); Long Term Yield (lty), which is from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook (see Welch and Goyal [90] and Lee [56]).

Most empirical researches include sample data observed in the early twentieth century. However, the empirical findings drawn from different research studies are not coincident. It has been argued in Campbell and Yogo [12] that the predictability of stock returns is much weaker over the postwar period (1952 to 2002) than the pre-war period (1926 to 2002), since the empirical results indicate that ep is significant at the 10% level in the pre-war sample but not significant in the post-war sample. Kostakis, Magdalinos and Stamatogiannis [54] extend the sample period to 2012, and their conclusions on significance of predictors are quite different between the postwar period (1952 to 2012) and the pre-war period (1926 to 2012), with the set of significant predictors being $\{dy, infl, dp, tbl, tms\}$ and $\{dy, ep, bm, ntis\}$ for monthly data, respectively. From the perspective of the time varying market structure, testing if a predictor has capacity on predicting stock returns over a long time period (for instance, 60 or 70 years) is "inconvincible" (it is much of interest to test the predictability of stock returns in different time periods, because the market structure is not consistent over time).

We concerntrate on the latest 50 years of stock returns (up to Dec. 2021, monthly data), and we divide the sample period into three segments to test predictability. In each segment we include two or three American business cycles. More precisely, the first segment starts from Dec. 1969 to Mar. 1987, and it consists of three business cycles: (i) Dec. 1969 to Mar. 1975, the characteristic of this period include the collapse of the Bretton Woods system of fixed exchange rates, the first oil crisis and hyperinflation; (ii) Apr. 1975 to Sep. 1982, the hyperinflation turned into deflation in this period; (iii) Oct. 1982 to Mar.

1987, characterised by large fiscal deficits and a stock market crash. The second segment we use is from Apr. 1987 to Dec. 2000, which consists of two business cycles: (1) Apr. 1987 to Mar. 1991, characterised by steady economic growth; (2) Apr. 1991 to Dec. 2000, this is a new economic period with high growth and low inflation. The third segment starts from Jan. 2001 to Dec. 2021, including two business cycles: (a) Jan. 2001 to Dec. 2009, characterised by the financial crisis triggered by the U.S. subprime mortgage crisis; (b) Jan. 2010 to Dec. 2021, the recovery of America's economy as the theme of the time.

5.3. Results

For each sample, we first check the correlation between x_{t-1} and the innovation of the prediction model v_t by computing the correlation coefficient r between the OLS estimates of u_t and v_t. A non-zero value of r implies the existence of embedded endogeneity which leads to a biased estimator of β_1 in (18), according to Nelson and Kim [62] and Stambaugh [86]. In empirical research, the value of r is usually different from zero. For example, Campbell and Yogo [12, Table 4] provide evidence of embedded endogeneity (empirically) when predicting stock returns via the earnings price ratio (ep) and dividend price ratio (dp). Once we have determined which variables are endogenous, we conduct our regression to determine whether or not the appropriate estimated coefficient has a confidence bound that overlaps zero.

Table 3 reports the estimated values of r and p-values of Pearson's correlation test corresponding to each of the eleven predictors above for each segment. At the 5% level, the existence of embedded endogeneity is confirmed and the corresponding predictors are bolded. It is seen from this table that $\{lty, dp, ep, bm\}$ always show the existence of embedded endogeneity in all segments, $\{de, tbl\}$ have exogeneity only in the second segment and $\{ntis, tms\}$ are only endogenous variables during the second segment. $infl$ and dfy are significant in the first and third segments, respectively. It is also instructive to contrast the estimated correlation coefficient $\widehat{r}$ during the three segments. The magnitude of $\widehat{r}$ for stock characteristics $\{ep, bm\}$ become smaller over time. The interest-rate related predictors $\{lty, tbl, dfy, infl\}$ change signs of correlation over time. This may due to the evolution of the financial market.

For testing predictability, when the predictors are endogenous, we should use model (19) and estimate β_1 by $\widehat{\beta}^*_{1,\omega}$ in (26); otherwise, model (2) with $k = 1$ is employed, and β_1 is estimated by $\widehat{\beta}_{1,\omega}$. The 90% and 95% confidence intervals are constructed by employing the proposed random weighting method. The empirical results of testing predictability for eleven predictors are reported in Table 4. Bold values indicate rejection of the null hypothesis of no predictability at the 10% or 5% level.

The IVX estimator $\widehat{\beta}_{IVX}$ and IVX-Wald statistics W_{IVX} proposed by Kostakis, Magdalinos and Stamatogiannis [54] are shown in the last two columns of the table. For the period December 1969-March 1987, we find that the null of no predictability can be rejected at the 5% level when dfy, $ntis$ and tms are predicting variables (de is significant at the 10% level). Comparing our findings with the IVX-Wald test, the differences are the set of significant predictors and at which level they become significant. Further, calculating the 99% CI of dfy, $ntis$ and tms ($[0.0841, 3.6468]$, $[-0.7741, 0.0621]$ and

Table 3: The estimate values of r and p-values of Pearson's correlation test

	Dec. 1969-Mar. 1987		Apr. 1987-Dec. 2000		Jan. 2001-Dec. 2021	
	$\widehat{r}$	p-value	$\widehat{r}$	p-value	$\widehat{r}$	p-value
de	**0.1461**	0.0361	0.0439	0.5780	**-0.2008**	0.0014
lty	**-0.3724**	<0.0001	**-0.2211**	0.0046	**0.2730**	<0.0001
dy	-0.0096	0.8915	-0.0032	0.9673	-0.0844	0.1835
dp	**-0.9926**	<0.0001	**-0.9979**	<0.0001	**-0.9778**	<0.0001
tbl	**-0.2702**	0.0001	-0.0402	0.6103	**0.1897**	0.0026
ep	**-0.9667**	<0.0001	**-0.9231**	<0.0001	**-0.3038**	<0.0001
bm	**-0.8894**	<0.0001	**-0.7887**	<0.0001	**-0.5308**	<0.0001
dfy	0.1109	0.1125	-0.0473	0.5491	**-0.2541**	<0.0001
ntis	-0.0065	0.9260	**-0.2066**	0.0081	0.0210	0.7411
tms	0.0913	0.1917	**-0.1687**	0.0313	0.1121	0.0769
infl	**-0.1807**	0.0093	-0.1118	0.1554	0.0358	0.5736

$[0.0269, 1.0272]$, respectively) with our method, we find that dfy is significant at the 1% level for both methods but de, $ntis$ and tms are significant at the 10% and 5% levels for the IVX-Wald test. *tbl* and $infl$ are identified to be significant predictors by IVX-Wald statistics, which are insignificant according to our method. The predictability evidence entirely disappears in the period April 1987-December 2000 for both methods. For the period January 2001-December 2021, the IVX-Wald test indicates that only lty is significant at the 1% level but fails to report the significance of *tbl* even at the 10% level. Our method finds evidence of predictability at the 5% level. Moreover, by calculating the 99% CI of lty and *tbl*, we find that lty is still significant at the 1% level (with CI: [-1.1485,-0.0745]) as well by our method. While the empirical results of both methods show that the predictability of stock returns has weakened in the process of time, our method was generally able to find stronger evidence of predictability.

6. Conclusion

In this paper, we have provided a unified approach for inference in multiple linear regression models with stationary and integrated or near-integrated variables. Since OLS with a standard bootstrap does not work well, we proposed the WEE method. We provide the asymptotic distribution, but unfortunately, the asymptotic variance cannot be estimated well in finite samples given the erratic behavior of the residual based variance estimate. The standard bootstrap does not consistently estimate the sampling distributions of our proposed WEE estimators either. We therefore propose a random weighting bootstrap method (which does not rely on resampling residuals) to construct confidence regions for the coefficients of our state variables. Our asymptotic theory is supported in finite samples via simulations.

In the application, we contributed to the well studied analysis of asset return prediction. We were

Table 4: Empirical results: confidence intervals in bold imply rejection of the null hypothesis at the 10% or 5% level as the *, **, and *** imply rejection of the null hypothesis at 10%, 5%, and 1% level, respectively.

	$\widehat{\beta}_{1,\omega}$	90% CI		95% CI		$\widehat{\beta}_{IVX}$	W_{IVX}
Dec. 1969-Mar. 1987							
de	0.0471	**0.0069**	**0.0872**	-0.0008	0.0949	0.0648	6.5232^{**}
lty	0.0132	-0.2035	0.2298	-0.2450	0.2713	-0.1392	0.9559
dy	0.0156	-0.0085	0.0396	-0.0131	0.0443	0.0090	0.2554
dp	0.0170	-0.0082	0.0421	-0.0130	0.0469	0.0098	0.3562
tbl	-0.1202	-0.2984	0.0581	-0.3326	0.0922	-0.2212	4.0070^{**}
ep	-0.0015	-0.0195	0.0166	-0.0230	0.0200	-0.0086	0.4861
bm	0.0053	-0.0284	0.0390	-0.0348	0.0454	-0.0036	0.0348
dfy	1.8655	**0.7279**	**3.0030**	**0.5100**	**3.2209**	1.7892	6.6923^{***}
$ntis$	-0.3560	**-0.6230**	**-0.0890**	**-0.6742**	**-0.0379**	-0.3452	3.6256^{*}
tms	0.5270	**0.2076**	**0.8464**	**0.1465**	**0.9076**	0.4967	5.8661^{**}
$infl$	-1.4426	-3.7563	0.8711	-4.1996	1.3144	-2.0909	6.0484^{**}
Apr. 1987-Dec. 2000							
de	-0.0112	-0.0343	0.0119	-0.0387	0.0163	-0.0111	0.5306
lty	-0.2815	-0.9305	0.3675	-1.0548	0.4918	-0.4127	1.7866
dy	0.0016	-0.0144	0.0177	-0.0175	0.0207	0.0024	0.0502
dp	0.0028	-0.0136	0.0192	-0.0168	0.0224	0.0023	0.0432
tbl	-0.1402	-0.5113	0.2309	-0.5823	0.3020	-0.1182	0.2328
ep	0.0125	-0.0070	0.0321	-0.0108	0.0359	0.0103	0.6846
bm	-0.0012	-0.0516	0.0492	-0.0612	0.0588	0.0013	0.0012
dfy	-0.4978	-3.8861	2.8906	-4.5353	3.5397	0.3873	0.0313
$ntis$	0.0812	-0.2086	0.3710	-0.2641	0.4265	-0.0164	0.0070
tms	-0.1337	-0.6046	0.3373	-0.6948	0.4275	-0.1159	0.1595
$infl$	-2.3958	-5.3225	0.5309	-5.8832	1.0915	-2.3331	1.8159
Jan. 2001-Dec. 2021							
de	0.0045	-0.0133	0.0223	-0.0167	0.0257	-0.0012	0.0346
lty	-0.6115	**-0.9544**	**-0.2685**	**-1.0201**	**-0.2028**	-0.7145	7.9664^{***}
dy	0.0283	-0.0061	0.0626	-0.0127	0.0692	0.0016	0.0021
dp	0.0287	-0.0084	0.0658	-0.0155	0.0729	-0.0008	0.0009
tbl	-0.3736	**-0.6472**	**-0.0999**	**-0.6996**	**-0.0475**	-0.1720	0.4588
ep	-0.0002	-0.0165	0.0162	-0.0197	0.0193	0.0014	0.0405
bm	0.0869	-0.1218	0.2955	-0.1618	0.3355	-0.0062	0.0119
dfy	-0.4336	-2.5881	1.7209	-3.0008	2.1336	-0.9198	1.8775
$ntis$	0.1885	-0.1364	0.5133	-0.1986	0.5756	0.1967	1.6638
tms	-0.1743	-0.5353	0.1867	-0.6044	0.2558	-0.2931	1.9593
$infl$	0.5166	-0.7333	1.7665	-0.9727	2.0059	0.7672	1.1075

able to address issues of endogeneity and different orders of integration. Accounting for each of these led to evidence of prediction. We believe that the methods that we have developed here will work well in other settings where the order of integration of state variables are uncertain.

Appendix. Proofs of theorems

To facilitate our arguments for our proofs, we first introduce the following lemma.

Lemma 1. *(Lemma A.1 of Cai, Wang and Wang [16]) Assume $u_{t,i}$ is a stationary α-mixing process with mixing coefficient $\alpha_i(n)$ satisfying that $E(|u_{t,i}|^r) < \infty$ and $\sum_{n=1}^{\infty} \alpha_i^s(n) < \infty$, where $s = 1/(2+\delta_*) - 1/r$, $r > 2 + \delta_*$, and $0 < \delta_* \leq 2$. Let $\theta_* = 1/2 - 1/(2+\delta_*)$ and $\lambda_* > 0$ is a function of δ_*. Then the NI(1) or I(1) process $U_{n,i}(r) = n^{-1/2} x_{[nr],i}$ for $0 \leq r \leq 1$ admits the following strong approximation*

$$\sup_{0 \leq r \leq 1} |U_{n,i}(r) - U_{\gamma_i}(r)| = O[n^{-\theta_*}\{\log(n)\}^{\lambda_*}]$$

holds almost surely, where $U_{\gamma_i}(\cdot)$ is the diffusion process in (6).

Proof of Theorem 1. By (5), we have

$$D_n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = R_n^{-1} V_n \tag{27}$$

where $R_n = D_n^{-1} \sum_{t=1}^n \mathbf{x}_t^{\otimes 2} D_n^{-1}$ and $V_n = D_n^{-1} \sum_{t=1}^n \mathbf{x}_t v_t$. For $(t-1)/n \leq r \leq t/n$ we define $U_n(r) \equiv U_{n,t} \equiv n^{-1/2} X_{t,1}$. Then, by Lemma 1,

$$\sum_{t=1}^n \|X_{t,1}\|^2 = O_p(n^2). \tag{28}$$

It follows from Theorem 1.2 of Berkes and Horváth [7] and Lemma 1 that, for $\ell = 1, 2$,

$$n^{-1-\ell/2} \sum_{t=1}^n X_{t,1}^{\otimes \ell} = n^{-1} \sum_{t=1}^n U_{n,t}^{\otimes \ell} \xrightarrow{d} \int_0^1 \{U_\gamma(r)\}^{\otimes \ell} \, dr.$$

Therefore,

$$\sum_{t=1}^n X_{t,1}^{\otimes \ell} = n^{1+\ell/2} W_u^{(\ell)} + o_p(n^{1+\ell/2}), \tag{29}$$

where $W_u^{(\ell)} = \int_0^1 \{U_\gamma(r)\}^{\otimes \ell} \, dr$. By the stationarity mixing property of $X_{t,2}$, we have

$$\sum_{t=1}^n X_{t,2}^{\otimes 2} = nE(X_{t,2}^{\otimes 2})\{1 + o_p(1)\} = n\Sigma_2\{1 + o_p(1)\}. \tag{30}$$

Then, by (29)-(30) and simple matrix operation,

$$R_n = \begin{pmatrix} R_{n,11} & R_{n,12} & R_{n,13} \\ R_{n,21} & R_{n,22} & R_{n,23} \\ R_{n,31} & R_{n,32} & R_{n,33} \end{pmatrix}, \tag{31}$$

where $R_{n,11} = 1$, $R_{n,12} = n^{-1} \sum_{t=1}^n U_{n,t}^\top = \{W_u^{(1)}\}^\top + o_p(1)$, $R_{n,13} = n^{-1} \sum_{t=1}^n X_{t,2}^\top = E(X_{t,2}^\top) + o_p(1)$, $R_{n,21} = R_{n,12}^\top$, $R_{n,22} = n^{-1} \sum_{t=1}^n U_{n,t}^{\otimes 2} = W_u^{(2)} + o_p(1)$, $R_{n,23} = n^{-1} \sum_{t=1}^n U_{n,t} X_{t,2}^\top$, $R_{n,31} = R_{n,13}^\top$, $R_{n,32} = R_{n,23}^\top$, and $R_{n,33} = n^{-1} \sum_{t=1}^n X_{t,2}^{\otimes 2} = E(X_{t-1,2}^{\otimes 2}) + o_p(1)$. Since $X_{t,2}$ has zero mean, we have

$R_{n,13} = \mathbf{0}_{k-d}^\top + o_p(1)$ and $R_{n,33} = \Gamma_{X_2}(0) + o_p(1)$. Using an argument similar to that in Cai, Li and Park [14, pp. 108], we obtain that

$$n^{-1}\sum_{t=1}^{n} U_{n,t}\left[X_{t,2}^\top - E(X_{t,2}^\top)\right] = o_p(1).$$

This, combined with (29), leads to

$$\begin{aligned} R_{n,23} &= n^{-1}\sum_{t=1}^{n} U_{n,t}E(X_{t,2}^\top) + n^{-1}\sum_{t=1}^{n} U_{n,t}\{X_{t,2}^\top - E(X_{t,2}^\top)\} + o_p(1) \\ &= \mathbf{0}_{d\times(k-d)} + o_p(1). \end{aligned}$$

Let $\Sigma_u = \begin{pmatrix} 1 & \{W_u^{(1)}\}^\top \\ W_u^{(1)} & W_u^{(2)} \end{pmatrix}$. Then it is straightforward to verify that

$$R_n = R + o_p(1) \;\text{ and }\; R_n^{-1} = R^{-1} + o_p(1), \tag{32}$$

where $R = \text{diag}\{\Sigma_u, \Gamma_2(0)\}$. Let $W_n(r) = \sum_{s=1}^{[nr]} n^{-1/2}v_s$ and $U_n(r) = U_{n,[nr]} = n^{-1/2}X_{[nr],1}$. It is easy to see that, for the mixing process $\{v_t\}$, $\sigma_v^{-1}W_n(r)$ converges weakly to a standard Brownian motion $B_v(r)$ with $B_v(0) = 0$ and $\text{var}\{B_v(r)\} = r$, where $\sigma_v^2 = \gamma_v(0) + 2\sum_{h=1}^{\infty}\gamma_v(h)$. Hence, $W_n(r)$ converges weakly to the Brownian motion

$$W_v(r) = \sigma_v B_v(r) \tag{33}$$

on $r \in [0,1]$. By Lemma A.1 of Cai, Wang and Wang [16], $U_n(r) = U_\gamma(r) + o_p(1)$ uniformly for $r \in [0,1]$. Let $V_{n0} = n^{-1/2}\sum_{t=1}^{n} v_t$, $V_{n1} = n^{-1}\sum_{t=1}^{n} X_{t,1}v_t$ and $V_{n2} = n^{-1/2}\sum_{t=1}^{n} X_{t,2}v_t$. Then $V_n = (V_{n0}, V_{n1}^\top, V_{n2}^\top)^\top$. Note that $V_{n0} = W_n(1) \stackrel{d}{\to} \sigma_v B_v(1)$. Applying Lemma 1 and (33), we obtain that

$$\begin{aligned} V_{n1} &= \sum_{t=1}^{n} U_{n,t}n^{-1/2}v_t = \int_0^1 U_n(r)\,dW_n(r) + o_p(1) \\ &= \sigma_v\int_0^1 U_\gamma(r)\,dB_v(r) + o_p(1), \end{aligned} \tag{34}$$

where $U_\gamma(r)$ and $B_v(r)$ are uncorrelated processes, because v_t and $\mathbf{x}_{t-1}$ are uncorrelated. Since $X_{t-1,2}$ and v_t involve only stationary mixing variables and $E(V_{t,2}) = 0$, it is easy to show that

$$\begin{aligned} \text{Var}(V_{n2}) &= n^{-1}\sum_{s,t=1}^{n} E(X_{s,2}X_{t,2}^\top v_s v_t) \\ &= \Gamma_2(0)\gamma_v(0) + 2\sum_{h=1}^{n}(1-\frac{h}{n})\Gamma_2(h)\gamma_v(h) \\ &\to \Gamma_2(0)\gamma_v(0) + 2\sum_{h=1}^{\infty}\Gamma_2(h)\gamma_v(h) = \sigma_v^2\Sigma_{2,v}, \end{aligned}$$

where $\Sigma_{2,v} = \{\Gamma_2(0)\gamma_v(0) + 2\sum_{h=1}^{\infty}\Gamma_2(h)\gamma_v(h)\}/\sigma_v^2$, and

$$V_{n2} \xrightarrow{d} N(0, \sigma_v^2\Sigma_{2,v}) = \sigma_v B_{ev}(1), \tag{35}$$

where $B_{ev}(r)$ is a $(k-d)$-dimensional Brownian motion on $[0,1]$ with covariance matrix $\Sigma_{2,v}$. Therefore, by (34)-(35) and the definition of V_n,

$$V_n = \left(\sigma_v\int_0^1 dB_v(r), \sigma_v\int_0^1 U_\gamma^\top(r)\,dB_v(r), \sigma_v\int_0^1 dB_{ev}^\top(r)\right)^\top + o_p(1) \equiv (V_{n0}^*, V_{n1}^{*\top}, V_{n2}^{*\top})^\top + o_p(1).$$

Since v_t is independent of u_t, conditional on $\mathcal{F}_u$, V_{n0}^* is asymptotically normally distributed with zero mean and variance σ_v^2, and stochastic integral V_{n1}^* is asymptotically normal with mean zero and variance $\sigma_v^2 W_u^{(2)}$. Note that $\text{cov}(V_{n0}^*, V_{n1}^*|\mathcal{F}_u) = \sigma_v^2 W_u^{(1)}$, $\text{cov}(V_{n0}^*, V_{n2}^*|\mathcal{F}_u) = \mathbf{0}_{k-d} + o_p(1)$, and $\text{cov}(V_{n1}^*, V_{n2}^{*\top}|\mathcal{F}_u) = \mathbf{0}_{d\times(k-d)} + o_p(1)$, since $X_{t,2}$ is centered. Using the Cramér-Wold device, we obtain that, V_n is asymptotically normal with mean zero and variance-covariance matrix $\sigma_v^2\Sigma$, where $\Sigma = \text{diag}\{\Sigma_u, \Sigma_{2,v}\}$. Further, by (27) and (32), $D_n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically mixed normal with mean zero and variance-covariance matrix equal to $R^{-1}\Sigma R^{-1} = \sigma_v^2 \text{diag}\{\Sigma_u^{-1}, \Gamma_2^{-1}(0)\Sigma_{2,v}\Gamma_2^{-1}(0)\}$, which is $\sigma_v^2\text{diag}\{\Sigma_u^{-1}, \{E(X_{t,2}^{\otimes 2})\}^{-1}\}$ if $\{v_t\}$ are iid white noise.

$\diamond$

Proof of Theorem 2. Let $A_n = n^{-1}\sum_{t=1}^n(\Omega_t\mathbf{x}_t)^{\otimes 2}$ and $B_n = n^{-1}\sum_{t=1}^n \Omega_t\mathbf{x}_t^{\otimes 2}$. By the definition of $\widehat{\boldsymbol{\beta}}_\omega$, we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}) = \left(n^{-1}\sum_{t=1}^n \Omega_t\mathbf{x}_t^{\otimes 2}\right)^{-1} n^{-1/2}\sum_{t=1}^n \Omega_t\mathbf{x}_t v_t. \tag{36}$$

Therefore,

$$\sqrt{n}A_n^{-1/2}B_n(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}) = A_n^{-1/2}n^{-1/2}\sum_{t=1}^n \Omega_t\mathbf{x}_t v_t = S_n, \tag{37}$$

where $S_n = A_n^{-1/2}n^{-1/2}C_n v$ with $C_n = (\Omega_1\mathbf{x}_1, \ldots, \Omega_n\mathbf{x}_n)$ and $v = (v_1, \ldots, v_n)^\top$. Let $\mathcal{F}_u = \sigma(u_t, t \le n)$. Then $\Omega_t\mathbf{x}_t$ is $\mathcal{F}_u$-measurable. It can be shown that $E(S_n|\mathcal{F}_u) = 0$ and

$$E(S_n^{\otimes 2}|\mathcal{F}_u) = \sigma_v^2 A_n^{-1/2} n^{-1}\sum_{t=1}^n (\Omega_t\mathbf{x}_t)^{\otimes 2} A_n^{-1/2} = \sigma_v^2 I_k,$$

where I_k is a $k\times k$ identity matrix. Using the martingale limit theorem (Hall and Heyde [42]), we can show that, for any $k\times 1$ vector a,

$$a^\top S_n \xrightarrow{d} N(0, \sigma_v^2 a^\top a).$$

Then, by the Wald device, we have $S_n \xrightarrow{d} N(0, \sigma_v^2 I_k)$. By (37), we conclude the result of the theorem. $\diamond$

Proof of Theorem 3. Let $A_n^* = n^{-1}\sum_{t=1}^n(\Omega_t\mathbf{x}_t)^{\otimes 2}\sigma^2(\mathbf{x}_t)$. Then, by the same argument as that for (36),

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}) = \left(n^{-1}\sum_{t=1}^n \Omega_t\mathbf{x}_t^{\otimes 2}\right)^{-1} n^{-1/2}\sum_{t=1}^n \Omega_t\mathbf{x}_t\sigma(\mathbf{x}_t)v_t^*. \tag{38}$$

Hence,

$$\sqrt{n}A_n^{*-1/2}B_n(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}) = A_n^{*-1/2}n^{-1/2}\sum_{t=1}^n \Omega_t\mathbf{x}_t\sigma(\mathbf{x}_t)v_t^* \equiv S_n^*. \tag{39}$$

Note that $E(S_n^*|\mathcal{F}_u) = 0$ and $E(S_n^{*\otimes 2}|\mathcal{F}_u) = I_k$. The result holds from the same argument as for Theorem 2.

Proof of Theorem 4. By (15), we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_\omega^* - \boldsymbol{\beta}) = \left(n^{-1}\sum_{t=1}^{n}\xi_t\Omega_t\mathbf{x}_t^{\otimes 2}\right)^{-1} n^{-1/2}\sum_{t=1}^{n}\xi_t\Omega_t\mathbf{x}_t\sigma(\mathbf{x}_t)v_t^* \equiv B_n^{*-1}\eta_n^*, \tag{40}$$

where $\eta_n^* = n^{-1/2}\sum_{t=1}^n \xi_t\Omega_t\mathbf{x}_t\sigma(\mathbf{x}_t)v_t^*$. By independence between $\{\xi_t\}$ and F_n, we have $E(B_n^* - B_n|F_n) = n^{-1}\sum_{t=1}^n(\xi_t - 1)\Omega_t\mathbf{x}_t^{\otimes 2} = 0$ and

$$E\{(B_n^* - B_n)^{\otimes 2}|F_n\} = n^{-2}\sum_{t=1}^{n}\left(\Omega_t\mathbf{x}_t^{\otimes 2}\right)^{\otimes 2} = O_p(1/n).$$

Then $B_n^* = B_n\{1 + o_p(1)\}$. It is straightforward to verify that $A_n = O_p(1)$, $B_n = O_p(1)$, $B_n^{-1} = O_p(1)$, and $\eta_n^* = O_p(1)$. Then

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_\omega^* - \boldsymbol{\beta}) = B_n^{-1}\eta_n^* + o_p(1).$$

This, together with (39), leads to

$$\sqrt{n}A_n^{*-1/2}B_n(\widehat{\boldsymbol{\beta}}_\omega^* - \widehat{\boldsymbol{\beta}}_\omega) = A_n^{*-1/2}n^{-1/2}\sum_{t=1}^{n}(\xi_t - 1)\Omega_t\mathbf{x}_t\sigma(\mathbf{x}_t)v_t^* + o_p(1) \equiv A_n^{*-1/2}\kappa_n.$$

Conditional on F_n, κ_n has mean zero and variance equal to $n^{-1}\sum_{t=1}^n(\Omega_t\mathbf{x}_t)^{\otimes 2}\sigma^2(\mathbf{x}_t) = A_n^*$. Therefore, conditional on F_n, using the martingale central limit theorem (Hall and Heyde [42]), we can show that $\sqrt{n}A_n^{*-1/2}B_n(\widehat{\boldsymbol{\beta}}_\omega^* - \widehat{\boldsymbol{\beta}}_\omega)$ is asymptotically normal with mean zero and variance I_k. Applying the Polya Theorem and Theorem 3, we complete the proof of the theorem. ◇

References

[1] Abuaf, N. and Jorion, P. (1990). Purchasing power parity in the long run. *Journal of Finance* **45**, 157–174.

[2] Adler, M. and Lehmann, B. (1983). Deviations from purchasing power parity in the long run. *Journal of Finance* **38**, 1471–1487.

[3] Amilud, Y. and Hurvich, C. M. (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis* **39**, 813–841.

[4] Amilud, Y., Hurvich, C. M. and Wang, Y. (2009). Multiple-predictor regressions: Hypothesis testing. *Review of Financial Studies* **22**, 413–434.

[5] Ball, L. (1993). The dynamics of high inflation. *NBER Working Paper* No. 4578.

[6] Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H. & Taylor, R. L.(1991) Bootstrapping unstable first-order autoregressive processes. *Annals of Statistics* **19**, 1098–1101.

[7] Berkes, I. and Horváth, L. (2006). Convergence of integral functionals of stochastic processes. *Econometric Theory* **22**, 304–322.

[8] Bickel, P. J. (1975). One-step Huber estimates in linear models. *Journal of the American Statistics Association* **70**, 428–433.

[9] Blanchard, O. J. and Summers, L. H. (1986). Hysteresis and the European unemployment problem. Fischer Stanley (Ed.), *NBER Macroeconomics Annual* Vol. 1.

[10] Blanchard, O. J. and Summers, L. H. (1987). Hysteresis in unemployment. *European Economic Review* **31**, 288–295.

[11] Billingsley, P. (1999). Convergence of Probability Measures. *2nd ed.* Wiley, New York.

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

[13] Cai, Z. W., Fan, J. Q. & Yao, Q. W. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* **95**, 941–56.

[14] Cai, Z., Li, Q. and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics* **148**, 101–113.

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

[16] Cai, Z., Wang, Y. and Wang, Y. (2015). Testing instability in a predictive regression model with nonstationary regressions. *Econometric Theory* **31**, 953–980.

[17] Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially single-index models. *Journal of the American Statistical Association* **92**,477–489.

[18] Cavanagh, C. L., Elliott, G. and Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric Theory* **11**, 1131–1147.

[19] Cenesizoglua, T. and Timmermann, A. (2012). Do return prediction models add economic value?. *Journal of Banking and Finance* **36**, 2974–2987.

[20] Chan, N. H., Li, D. and Peng, L. (2012). Toward a unified interval estimation of autoregressions. *Econometric Theory* **28**, 705-717.

[21] Chan, N. H. and Peng, L. (2005). Weighted least absolute deviations estimation for an AR(1) process with ARCH(1) errors. *Biometrika* **92**, 477–484.

[22] Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) process. *Annals of Statistics* **15**, 1050-1063.

[23] Chan, N. H. and Zhang, R. M. (2009). Inference for nearly nonstationary processes under strong dependence and infinite variance. *Statistica Sinica* **19**, 925-947.

[24] Chen, J., J. Gao and D. Li (2014). Estimation in semi-parametric regression with non-stationary regressors. *Bernoulli* **18**, 678–702.

[25] Chen, R., J. S. Liu and R. S. Tsay (1995). Additivity tests for nonlinear autoregressions. *Biometrika* **82**, 369–383.

[26] Choi, Y., Jacewitz, S. and Park, J. Y. (2016). A reexamination of stock return predictability. *Journal of Econometrics* **192**, 168–189.

[27] Chowdhury, A. R. and Sdogati, F. (1993). Purchasing power parity in the major EMS countries: The role of price and exchange rate adjustment. *Journal of Macroeconomics* **15**, 25–45.

[28] Clark, P. K. (1989). Trend reversion in real output and unemployment. *Journal of Econometrics* **40**, 15–32.

[29] Culver, S. E. and Papell, D. H. (1989). Is there a unit root in the inflation rate? Evidence from sequential break and panel data models. *Journal of Applied Econometrics* **12**, 435–444.

[30] Datta, S. (1996). On asymptotic properties of bootstrap for AR(1) processes. *Journal of Statistical Planning and Inference* **53**, 361–374.

[31] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**, 427–431.

[32] Dornbusch, R. (1976). Expectations and exchange rate dynamics. *Journal of Political Economy* **84**, 1161–1176.

[33] Dow, C. H. (1920). Scientific stock speculation. *The Magazine of Wall Street.*

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

[35] Engle, R. F. and Granger, C.W. J. (1987). Co-integration and error-correction: Representation, estimation and testing. *Econometrica* **55**, 251–76.

[36] Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications.* CRC Press.

[37] Fan, J. and Jiang, J. (2005). Nonparametric inference for additive models. *Journal of the American Statistical Association* **100**, 890–907.

[38] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Berlin: Springer-Verlag.

[39] Gao, J. and P. C. B. Phillips (2013). Functional coefficient nonstationary regression with non- and semi-parametric cointegration. *Monash Econometrics and Business Statistics Working Papers, Monash University,* Department of Econometrics and Business Statistics.

[40] Goyal, A. and Welch, I. (2003) Predicting the equity premium with dividend ratios. *Management Science,* **49**, 639–654.

[41] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* New York: Springer.

[42] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press.

[43] Hall, P. and Jing, B. Y. (1998). Comparison of bootstrap and asymptotic approximations to the distribution of a heavy-tailed mean. *Statistica Sinica* **8**, 887–906.

[44] Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review* **58**, 109–127.

[45] Hansen, B. E. (1992a). Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* **60**, 967–972.

[46] Hansen, B. E. (1992b). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* **8**, 489–500.

[47] Hong, S. and Jiang, J. (2018). Unifying inference for multiple linear predictive regression models with stationary and nonstationary data. Manuscript.

[48] Hong, S., Jiang, J., Jiang, X. and Xiao, Z. (2021). Unifying inference for semiparametric regression. *Econometrics Journal* **24**, 482–501.

[49] Jiang, J. (2014). Multivariate Functional-coefficient Regression Models for Multivariate Nonlinear Times Series. *Biometrika* **101**, 689-702.

[50] Jiang, J. and Mack, Y. P. (2001). Robust local polynomial regression for dependent data. *Statistica Sinica* **11**, 705–722.

[51] Jin, Z., Ying, Z., and Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381-390.

[52] Juhl, T. and Xiao, Z. (2005a). Testing for cointegration using partially linear models. *Journal of Econometrics* **124**, 363–394.

[53] Juhl, T. and Xiao, Z. (2005b). Partially linear models with unit roots. *Econometric Theory* **21**, 877–906.

[54] Kostakis, A., Magdalinos, T. and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies*, **28**, 1506–1553.

[55] Kothari, S. P. and Shanken (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics* **44**, 169–203.

[56] Lee, J. (2016). Predictive quantile regression with persistent covariates: IVX-QR approach. *Journal of Econometrics*, **192**, 105–118.

[57] Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics* **74**, 209–235.

[58] Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society.* Series B **67**, 381-393.

[59] Liu, X., Yang, B., Cai, Z. and Peng, L. (2019). A unified test for predictability of asset returns regardless of properties of predicting variables. *Journal of Financial Economics* **208**, 141–159.

[60] Magdalinos, T. and Phillips, P. C. B (2011). Econometric inference in the vicinity of unity. *Singapore Management University*, Singapore.

[61] Mark, N.C. (1990). Real and nominal exchange rates in the long run: An empirical investigation. *Journal of International Economics* **28**, 115–136.

[62] Nelson, C.R. and Kim, M.J. (1993). Predictable stock returns: the role of small sample bias. *Journal of Finance*, **48**, 641-–661.

[63] Nelson, C. R. and Plosser, C. R. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics* **10**, 139–162.

[64] Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica* **69**, 1519–1554.

[65] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

[66] Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics* **18**, 90–120.

[67] Owen, A. B. (2001). Empirical Likelihood. *CRC Press.*

[68] Park, J. Y. (2002). Nonstationary nonlinear heteroskedasticity. *Journal of Econometrics,* **110**, 383–415.

[69] Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica* **69**, 117–161.

[70] Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance* **13**, 274–315.

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

[72] Phillips, P. C. B. (1989). Partially identified econometric models. *Econometric Theory* **5**, 181–240.

[73] Phillips, P. C. B. and Han, C. (2008). Gaussian inference in AR(1) times series with or without a unit root. *Econometric Theory* **63**, 1023-1078.

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

[75] Phillips, P. C. B. and Lee, J. H. (2016). Robust econometric inference with mixed integrated and mildly explosive regressors. *Journal of Econometrics* **192**, 433–450.

[76] Phillips, P. C. B. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica* **58**, 165-193.

[77] Phillips, P. C. B., Shi, S. and Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review* **56**, 1043–1077.

[78] Phillips, P. C. B. and Yu, J. (2011). Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics* **2**, 455–491.

[79] Polk, C., Thompson, S. and Vuolteenaho, T. (2006). Cross-sectional forecasts of the equity premium. *Journal of Financial Economics* **81**, 101–141.

[80] Ren, Yu and Tu, Yundong and Yi, Yanping (2019). Balanced predictive regressions. *Journal of Empirical Finance* **54**, 118–142.

[81] Robinson, P. M. (1988). Semiparametric econometrics: A survey. *Journal of Applied Econometrics* **3**, 35-51.

[82] Roll, R. (1979). Violations of purchasing power parity and their implications for efficient international commodity markets. in M. Sarnat and G.P. Szego (eds) *International Finance and Trade*, Vol. 1, Chapter 6, Cambridge, MA.

[83] Rossi, B. (2007). Expectations hypotheses tests at Long Horizons. *Econometrics Journal* **10**, 1–26.

[84] Rubin, D.B. (1981). The Bayesian bootstrap. *Annals of Statistics* **9**, 130–134.

[85] So, B. S. and Shin, D. W.(1999). Cauchy estimators for autoregressive processes. *Econometric Theory* **15**, 165–176.

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

[87] Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B* **50**, 413–436.

[88] Song, F. M. and Wu, Y. (2007). Hysteresis in unemployment: Evidence from 48 U.S. States. *Economic Inquiry* **35**, 235–243.

[89] Torous, W., Valkanov, R. and Yan, S. (2004). On predicting stock returns with nearly integrated explanatory variables. *Journal of Business* **77**, 937–966.

[90] Welch, I. and Goyal A. (2008) A comprehensive look at the empirical performance of equity premium prediction *Review of Financial Studies,* **21**, 1455–1508.

[91] Viceira, L. M. (1997). Testing for structural change in the predictability of asset returns. *Working Paper, Harvard University.*

[92] Zheng, Y., Zhu, Q., Li, G. and Xiao, Z. (2018). Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. *Journal of the Royal Statistical Society, Series B* **80**, 975-993.

[93] Zhu, F., Cai, Z. and Peng, L.(2014). Predictive regressions for macroeconomic data. *Annals of Applied Statistics* **8**, 577–594.