Appl. Math. J. Chinese Univ.
2018, 33(2): 127-144

A perspective on recent methods on testing predictability of asset returns

LIAO Xiao-sai[1] CAI Zong-wu[2,1,*] CHEN Hai-qiang[1]

Abstract. This paper highlights some recent developments in testing predictability of asset returns with focuses on linear mean regressions, quantile regressions and nonlinear regression models. For these models, when predictors are highly persistent and their innovations are contemporarily correlated with dependent variable, the ordinary least squares estimator has a finite-sample bias, and its limiting distribution relies on some unknown nuisance parameter, which is not consistently estimable. Without correcting these issues, conventional test statistics are subject to a serious size distortion and generate a misleading conclusion in testing predictability of asset returns in real applications. In the past two decades, sequential studies have contributed to this subject and proposed various kinds of solutions, including, but not limit to, the bias-correction procedures, the linear projection approach, the IVX filtering idea, the variable addition approaches, the weighted empirical likelihood method, and the double-weight robust approach. Particularly, to catch up with the fast-growing literature in the recent decade, we offer a selective overview of these methods. Finally, some future research topics, such as the econometric theory for predictive regressions with structural changes, and nonparametric predictive models, and predictive models under a more general data setting, are also discussed.

§1 Introduction

Testing predictability of asset returns has been studied for recent three decades as a cornerstone research topic in economics and finance. It not only attracts attention from financial practitioners as it is a key component to evaluate mutual fund managers' performance, examine the validity of asset pricing models, and improve asset allocation efficiency, but also has important implications in theoretical researches in finance. A large literature has been devoted to

Received: 2018-02-10. Revised: 2018-03-27.
MR Subject Classification: 62M20, 62M10, 62G35.
Keywords: asset returns, Heteroskedasticity, high persistency, nonlinearity, predictability, quantile regressions.
Digital Object Identifier (DOI): https://doi.org/10.1007/s11766-018-3590-0.
This work, in part, was supported by the National Natural Science Foundation of China (71631004, 71571152), and the Fundamental Research Funds for the Central Universities (20720171002, 20720170090), and the Fok Ying-Tong Education Foundation (151084).
* Correspondence author: CAI Zong-wu (zongwucai@gmail.com).

examining the validity of the efficient market hypothesis (Fama, 1965,1970) by testing whether asset returns are predictable or not.

There are two major directions in the area of testing asset return predictability. The first one is to check whether the return as a time series is a white noise process, or random walk process, or martingale difference sequence. The rejection of the null hypothesis implies that a certain type of dependent structure exists in return processes, which violates the definition of the weak-form efficient market hypothesis, so that one can predict future asset returns using past observations. A vast amount of literature has been devoted to this aspect; see, for example, the books by Campbell, Lo and MacKinlay (1997) and Jondeau, Poon and Rockinger (2007), and the references therein. However, as pointed out by Fama (1991), the past realized returns are noisy measures of expected return so that the test based on them lacks of power. Meanwhile, it is restrictive to use historical returns only as predictors since investors could observe other information. Therefore, the second stream of the literature is to expand predictive regressions to include other economic and financial variables, such as the dividend-price ratio, the earnings-price ratio, the book-to-market ratio, default spread, and interest rates as well as other economic variables; see, for example, Campbell (1987), Fama and French (1988), and Campbell and Shiller (1988). Empirical findings strongly suggest that the asset returns might be predictable by using these financial ratios and macroeconomic variables as in Lettau and Ludvigson (2001). But some critical analysis raises the concern that the predictability may be a result of data-snooping as addressed by Bossaerts and Hillion (1999) and Welch and Goyal (2008).

The typical econometric method used in aforementioned studies is an ordinary least squares (OLS) regression of returns versus the lag of the financial variables, and conventional t-statistics are used to check the significance of coefficients. However, a series of recent studies find that the statistical inference for predictive regressions, or more specifically, the limiting distribution of t-statistics, crucially relies on the time series properties of the regressors, i.e., their degree of persistence. Empirical analyses conclude evidently that most of predictors widely used in the literature are highly persistent with autoregressive roots extremely close to unity. Ignoring such persistence may lead to an over-rejection of the null hypothesis for conventional test statistics, and the problem is more serious if the innovation of the predictor is highly correlated with the error of return as studied by Campbell and Yogo (2006) and Torous, Valkanov and Yan (2004) and others. Acknowledging the invalidity of the standard approach, sequential studies have contributed to this subject and proposed various kinds of solutions, including, but not limit to, the bias-correction approaches, the linear projection approach, the IVX filtering method, the variable addition (or control function) ideas, the weighted empirical likelihood procedure, and the double-weight robust approach. To help the reader to catch up with the fast-growing literature, this paper offers a brief overview of these methods with some comments and suggestions with a further improvement. To this end, we first review the classical approaches and comment on their strengths and limitations, and then move to the recently proposed robust inferences regardless of the types of persistence, in the framework of linear mean regression models. Moreover, we extend the discussion into the framework of quantile regressions and nonlinear predictive models, and highlight recent developments on these predictive models. As for quantile regressions, we focus on the methods such as the IVX-QR idea by Lee (2016) and the weighted variable addition approach by Cai, Chen and Liao (2017b). For nonlinear predictive models, we mainly introduce the predictive regressions with varying-coefficients, in discontinuous or smoothing form. Nonparametric tests and their limiting distribution with nonstationary variables are discussed. Finally, we suggest several further research topics for future studies.

The rest of the paper is organized as follows. In Section 2 we discuss some econometric issues, classical methods and robust inference results for predictive mean regressions. Section 3 is devoted to the methods for quantile predictive regressions and Section 4 focuses on nonlinear predictive models. In particular, it elaborates some recent nonparametric tests for nonparametric predictive models. Section 5 concludes the paper with some discussions on future research directions.

§2 Mean predictive regression models

The classical predictive regression commonly considered in the literature is the following structural predictive linear model:

$$\begin{cases} y_t = \alpha + \beta x_{t-1} + u_t, \\ x_t = \theta + \phi x_{t-1} + v_t, \quad 1 \leq t \leq T, \end{cases} \tag{1}$$

where $(u_t, v_t) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ are independent and identically distributed (i.i.d.) series and $\mathbf{\Sigma} = \begin{pmatrix} \sigma_u^2 & \delta\, \sigma_u \sigma_v \\ \delta\, \sigma_u \sigma_v & \sigma_v^2 \end{pmatrix}$ so that u_t and v_t may be correlated if δ is non-zero. Of course, the second equation in (1) can be in a higher order in autoregressive (AR) model, say AR(p). For simplicity, the focus here is on AR(1). Given the above linear regression model, testing the predictability of x_{t-1} to y_t is equivalent to testing the null hypothesis $H_0 : \beta = 0$. However, it is not easy to estimate parameters and to make their statistical inferences for model (1) due to the following econometric issues:

- *Embedded endogeneity:* The correlation coefficient δ between u_t and v_t may be non-zero, inducing a finite sample bias in estimating β and the invalidity of traditional inference; see Torous, Valkanov and Yan (2004) and Campbell and Yogo (2006) for real examples.
- *Nonstationarity:* The predicting regressor x_t might be persistent. That is, x_t can be stationary if $|\phi| < 1$, denoted by I(0), or nonstationary if $\phi = 1 + c/T$, denoted by NI(1) or I(1) when $c = 0$. For the second case with nonstationary regressors, the limiting distribution of the estimator of β relies on the nuisance parameter c, which is not consistently estimable; see, for example, the theory developed in Cai and Wang (2014).
- *Drift term:* It is known that non-zero θ in the second equation of (1) does not make any difference in asymptotic theory when x_t is I(0). But, when θ is non-zero for I(1) or NI(1), it makes econometric inferences totally different from those for I(1) and NI(1) cases; see, for example, Theorem 2 in Cai and Wang (2014).

- *Nonlinearity:* The structural predictive linear model in (1) may not be appropriate in many applications due to the existence of nonlinearity. Also, as argued in Section 4.2 (see later), a nonparametric setting of model (1) may avoid the embedded endogeneity as mentioned earlier.
- *Heteroskedasticity:* It is well documented in the literature that heteroskedasticity issue exists for most of economic and financial data. How to characterize heteroskedasticity is of importance in real applications.

Indeed, it often happens to have the above econometric issues encountered in real applications; see, for example, Table 1 in Torous, Valkanov and Yan (2004) and Table 4 in Campbell and Yogo (2006), which show clearly that most of economic and financial variables applied to predict stock returns are highly persistent and/or δ is non-zero. Under these situations, conventional inferential techniques are invalid and might lead to misleading conclusions about the predictability. To overcome these issues, sequential studies have been devoted to this subject and proposed various solutions. In the following, we make a selective review on these studies.

2.1 Bias-correction methods under stationarity

The OLS estimator for α and β in (1) is given by

$$(\hat{\alpha}, \hat{\beta})^\top = \left[\sum_{t=2}^{T} X_{t-1} X_{t-1}^\top\right]^{-1} \sum_{t=2}^{T} X_{t-1} y_t,$$

where $X_{t-1} = (1, x_{t-1})$. Stambaugh (1999) shows that the OLS estimator $\hat{\beta}$ is biased in finite sample although it is asymptotically unbiased, due to the nonzero correlation between u_t and v_t under normality and stationary regression ($|\phi| < 1$). The bias can be represented as

$$\mathbb{E}[\hat{\beta} - \beta] = \rho \mathbb{E}[\hat{\phi} - \phi] = \rho[-(1 + 3\phi)/T] + O(T^{-2}), \tag{2}$$

where $\rho = \text{cov}(u_t, v_t)/\sigma_v^2 = \delta\sigma_u/\sigma_v$, based on the Kendall (1954)'s approximation under normality assumption.

Therefore, there are several bias correction methods proposed in the literature to make the estimation more accurate in finite sample. Particulally, the first is the so-called first-order bias-correction estimator $\hat{\beta}_{bc_1}$ proposed by Stambaugh (1999) from equation (2). That is, $\hat{\beta}_{bc_1} = \hat{\beta} - \hat{\rho}(1 + 3\hat{\phi})/T$, where $\hat{\rho}$ and $\hat{\phi}$ are the estimator of ρ and ϕ, respectively. The second one is the second-order bias-correction estimator $\hat{\beta}_{bc_2}$ of Amihud and Hurvich (2004) based on a linear projection of u_t onto v_t as $u_t = \rho v_t + \epsilon_t$. Therefore, model (1) can be rewritten as follows:

$$y_t = \alpha + \beta x_{t-1} + \rho\, v_t + \epsilon_t, \tag{3}$$

and the least squares estimator based on the estimated $\{\hat{v}_t\}$ is defined as

$$(\hat{\alpha}_{bc_2}, \hat{\beta}_{bc_2}, \hat{\rho}_{bc_2})^\top = \arg\min_{\alpha,\beta,\rho} \sum_{t=1}^{T} \left(y_t - \alpha - \beta x_{t-1} - \rho \hat{v}_t\right)^2, \tag{4}$$

where $\hat{v}_t$ is obtained from the second equation in (1). By assuming that x_t is stationary, Amihud and Hurvich (2004) argue that this estimator is indeed the second-order bias-correction and derive its asymptotic theory so that the conventional t-type test can be used to test the predictability $H_0 : \beta = 0$. Moreover, Amihud, Hurvich and Wang (2009) extend model (1) to multiple predictive regressions. Note that this approach in equation (3) is similar to the so-called control function idea as in Elliott (2011), which will be discussed later. Thirdly, when ϕ is close to 1, Lewellen (2004) suggests a conservative bias-adjusted estimator $\hat{\beta}_{bc_3}$ of β as $\hat{\beta}_{bc_3} = \hat{\beta} + \hat{\rho}(0.9999 - \hat{\phi})$, and argues that $\hat{\beta}_{bc_3}$ is actually the least biased estimator of β when the true value of ϕ is very close to 1.

2.2 Statistical inference under nonstationarity

Most of the bias-correction methods mentioned above assume $|\phi| < 1$. However, the autoregressive parameter ϕ in x_t might be very close to one. Therefore, to this end, a local-to-unity framework, that is, $\phi = 1 + c/T$, is pervasive in the literature; see, e.g., Lewellen (2004), Torous, Valkanov and Yan (2004), Campbell and Yogo (2006), Jansson and Moreira (2006), Cai and Wang (2014), and the references therein. In this case, the conventional test statistic, which converges to standard normal distribution with stationary x_t, is invalid due to the persistence of x_t and non-zero δ. In particular, given $\phi = 1 + c/T$ and $\delta \neq 0$, Campbell and Yogo (2006) show that the conventional t-test $t_{\hat{\beta}}$ for equation (1) has the following limit distribution instead of normal

$$t_{\hat{\beta}} \xrightarrow{d} \delta\, \tau_c / k_c + \left[1 - \delta^2\right]^{1/2} Z, \tag{5}$$

where $k_c^2 = \int_0^1 K_c^2(r)dr$, $\tau_c = \int_0^1 K_c(r)dW(r)$, and Z is a standard normal random variable independent of $K_c(\cdot)$ and $W(\cdot)$. Here, $W(\cdot)$ is the Brownian motion generated by $\{v_t\}$ and $K_c(\cdot)$ is a special case of the Ornstein-Uhlenbeck process satisfying the stochastic differential equation system (Black-Scholes model)

$$dK_c(r) = c\, K_c(r)dr + dW(r),$$

so that $K_c(\cdot)$ is a geometric Brownian motion. Then, it can be shown easily that $K_c(r) \sim N(0, \sigma_c^2(r))$, where $\sigma_c^2(r) = \sigma_u^2 \left[\exp(2cr) - 1\right]/2c$ and $\int_0^1 K_c(r)dr \sim N(0, \varsigma(c)^2)$ with $\varsigma(c)^2 = \sigma_u^2/c^2 + \sigma_u^2(e^{2c} - 4e^c + 3)/2c^3$. Also, it can be shown that $\lim_{c\to 0} \varsigma(c)^2 = \sigma_u^2/3$. Clearly, equation (5) shows that $t_{\hat{\beta}}$ contains the nuisance parameter c and the embedded endogeneity problem caused by non-zero δ. This implies that the asymptotic distribution of $t_{\hat{\beta}}$ deviates from the standard normal and it is impossible to construct a pivotal test statistic by self normalizing the OLS estimator. Therefore, conventional test fails with an over-rejection, which is verified by simulation studies in Campbell and Yogo (2006).

On the other hand, Cai and Wang (2014) extend the linear projection approach of Amihud and Hurvich (2004) to nonstationary case with $\phi = 1 + c/T$, and show that $\hat{\beta}_{bc_2}$ in (4) has the following asymptotic distribution

$$T(\hat{\beta}_{bc_2} - \beta) \xrightarrow{d} \xi_c,$$

where $\xi_c = \left[\tau_c - W(1)\int_0^1 K_c(r)dr\right]\left[k_c^2 - \tau_c^2\right]^{-1} + \delta\left[\tau_c + \Omega_1\right]/k_c^2$ with $\Omega_1 = \sum_{k=2}^{\infty} E(u_1 u_k)$. Based on this asymptotic theory, one can construct a test statistic and compute the critical values using a Monte Carlo simulation approach as suggested by Cai and Wang (2014) if $c = 0$.

However, the test is infeasible if c is nonzero as the distribution of ξ_c depends on c which can not be estimated consistently. Moreover, it is interestingly to note that when θ in (1) is non-zero, i.e., there is a linear time trend in x_t, and Cai and Wang (2014) show that

$$T^{3/2}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma_{\theta,c}^2),$$

where $\sigma_{\theta,c}^2$ depends on θ, c and σ_v^2 with $\sigma_v^2 = \mathrm{Var}(v_t) + 2\sum_{t=2}^{\infty}\mathrm{Cov}(v_1, v_t)$. In particular, $\sigma_{\theta,0}^2 = 48\sigma_v^2/\theta^2$ for the case that $c = 0$; see Cai and Wang (2014) for details.

2.3 Robust inferences

As mentioned earlier, the asymptotic results in Campbell and Yogo (2006) and Cai and Wang (2014) depend on c, which is not consistently estimable although its estimate has a limiting distribution. Consequentially, the conventional test statistic is infeasible in applications. Recently, a series of researches investigate uniform (robust) inferences on predictive regressions in the sense that testing procedure for predictability is robust to general time series characteristics of the regressor and errors; see, for example, to name just a few, Campbell and Yogo (2006), Chen and Deo (2009), Phillips and Lee (2013), Zhu, Cai and Peng (2014), Demetrescu, Hanck and Tarcolea (2014), Kostakis, Magdalinos and Stamatogiannis (2015), Breitung and Demetrescu (2015), Rodrigues and Demetrescu (2016), Liu, Yang, Cai and Peng (2016), Cai, Chen and Liao (2017a, 2017b), and Yang, Liu, Cai and Peng (2017). In the following sections, u_t and v_t could be generalized to be a linear process as in Phillips and Solo (1992) and/or non-normally distributed.

2.3.1 Bonferroni method

Campbell and Yogo (2006) propose a new method (termed as Q-test) to construct a Bonferroni confidence interval for β. The procedure is as follows. The first step is to construct a $100(1-\alpha_1)\%$ confidence interval for ϕ, denoted as $CI_\phi(\alpha_1)$, while the second step is to construct $100(1-\alpha_2)\%$ confidence interval for β given ϕ, denoted as $CI_{\beta|\phi}(\alpha_2)$. Then, a confidence interval with the confidence level $100(1-\alpha^*)\%$, independ of ϕ, can be obtained by

$$CI_{\alpha^*}(\beta) = \cup_{\phi \in CI_{\alpha_1}(\phi)} CI_{\alpha_2}(\beta|\phi),$$

where $\alpha^* = \alpha_1 + \alpha_2$.

In particular, the confidence interval $CI_\phi(\alpha_1)$ given ϕ can be obtained by the procedures introduced by Stock (1991). One may choose any unit root test, such as the ADF test or the DF-GLS test, to construct the interval. It is noted that the DF-GLS test may generate a tighter confidence interval of ϕ as DF-GLS is more powerful than ADF test. Given the value of ϕ, the confidence interval $CI_{\alpha_2}(\beta|\phi)$ can be constructed using either the t-test or the Q-test. Again, the confidence interval based on Q-test is tighter than that based on t-test at the true value of ϕ. One may refer to Campbell and Yogo (2005) for more details.

As pointed out by Phillips and Lee (2013), Bonferroni approach might lead to an empirical size substantially lower than nominal size, resulting in a conservative test whose power is often negligible in nearly local alternative to the null of no predictability. Another critical limitation is the difficulty of extending the approach to multivariate regressions involving several predictors. Finally, the computing is very involved.

2.3.2 Weighted empirical likelihood

Zhu, Cai and Peng (2014), Liu, Yang, Cai and Peng (2016), and Yang, Liu, Cai and Peng (2017) propose an empirical likelihood approach together with a weighted least squares idea to construct confidence interval for β. If α in equation (1) is known, i.e., $\alpha = \alpha_0$, the empirical likelihood function for β is

$$L(\beta) = \sup\left\{\prod_{t=2}^{T} p_t : 0 \le p_t \le 1, \sum_{t=2}^{T} p_t = 1, \sum_{t=2}^{T} p_t H_t(\beta) = 0\right\},$$

where $H_t(\beta) = (y_t - \alpha_0 - \beta\, x_{t-1})x_{t-1}/\sqrt{1+x_{t-1}^2}$ and the weighted score equation is $\sum_{t=2}^{T} H_t(\beta) = 0$. After applying the Lagrange multiplier technique, we have

$$l(\beta) = -2\log L(\beta) = 2\sum_{t=2}^{T} \log\{1 + \lambda\, H_t(\beta)\},$$

where $\lambda = \lambda(\beta)$ satisfies $\sum_{t=2}^{T} H_t(\beta)/[1 + \lambda\, H_t(\beta)] = 0$. More importantly, Zhu, Cai and Peng (2014) show that under some regularity conditions, $l(\beta_0)$ converges in distribution to a chi-square distribution with one degree of freedom, where β_0 is the true value of β. Therefore, one can construct confidence interval with the significance level $100(1-b)\%$ as $\mathrm{CI}_b = \{\beta : l(\beta) \le \chi^2_{1,b}\}$. Rejecting $H_0 : \beta = 0$ is equivalent to checking if $0 \notin \mathrm{CI}_b$. The above result is the so called Wilks' theorem which holds regardless of the predicting variable being stationary or nonstationary. The key idea behind this approach is that the estimator of β defined from the empirical likelihood method could be associated to the weighted least squares as

$$\hat{\beta}_w = \arg\min_{\beta} \sum_{t=2}^{T} (y_t - \alpha_0 - \beta x_{t-1})^2 W_t,$$

where the weight is $W_t = 1/\sqrt{1+x_{t-1}^2}$ and it is easily to check that $x_{t-1}W_t = x_{t-1}/\sqrt{1+x_{t-1}^2}$ $= O_p(1)$ no matter that x_t is I(0) or I(1) or NI(1).

When α is unknown, Zhu, Cai and Peng (2014) show that Wilks' theorem for above empirical likelihood method fails when x_t is nonstationary and suggest using differencing idea as

$$y_t^* = \beta x_{t-1}^* + u_t^*,$$

where $y_t^* = y_{t+m} - y_t$ for $m = [T/2]$. Note that $x_t^* = O_p(\sqrt{T})$ if x_t is I(1) or NI(1). But, this differencing method is inefficient since it uses only half of sample. One may consider to take a difference as $x_{t-1}^* = x_{t-1} - x_{t-2}$ to remove α so that $x_t^* = O_p(1)$ for all cases of x_t but the convergence rate is $\sqrt{T}$ instead of T or $T^{3/2}$.

Recently, Liu, Yang, Cai and Peng (2016) extend this approach to bivariate regression but it still loses efficiency when α is unknown. Similar to the Bonferroni procedure, it is difficult to extend this approach to a general multivariate regression with mixed degree of persistence.

2.3.3 IVX

Phillips and Magdalinos (2009) introduce a data-filtering approach, termed as IVX, to diminish the effect of embedded endogeneity in the predictive regression. The main idea of IVX is to construct an instrument variable z_t as

$$z_t = \phi_z z_{t-1} + u_{zt} \tag{6}$$

with $\phi_z = 1 + c_z/T^\eta$, $\eta \in (0, 1)$, and $c_z < 0$. Clearly, z_t is a mildly integrated process, which is less persistent than x_t if x_t is I(1) or NI(1). The reason of choosing the less persistent z_{t-1} is to ensure validity for chi-squared test limit theory at the cost of slow convergence rate. Using z_t as an instrumental variable, one can obtain an estimator of β, denoted by $\hat{\beta}_{IV}$. Phillips and Magdalinos (2009) show that the convergence rate for $\hat{\beta}_{IV}$ is $T^{(1+\eta)/2}$ instead of T; see Phillips and Magdalinos (2009) for details. Recently, Rodrigues and Demetrescu (2016) consider model (3) using the IVX approach.

In practice, how to choose the tuning parameters (c_z, η) and innovation u_{zt} is a challenging issue. To gain efficiency, one wishes η very close to 1. Indeed, Phillips and Lee (2013) and Kostakis, Magdalinos and Stamatogiannis (2015) suggest empirically choosing $\eta = 0.95$, $c_z = -I$, and $u_{zt} = \Delta x_{t-1}$, where I is the identity matrix. But, there is no theory on how to choose these tuning parameters and neither is there data-driven approach available in the literature.

In summary, for the IVX approach, one can observe the followings. First, the power of the test depends heavily on the choice of η. In other words, mis-choice of η might lose power; see, Kostakis, Magdalinos and Stamatogiannis (2015). Secondly, the IVX does not work well when x_t is I(0); see Phillips and Magdalinos (2009). Finally, the IVX approach requires the instrumental variable z_t to be less persistent than x_t, which sacrifices the convergence rate.

2.3.4 Dynamic approach

To improve the local power of the IVX based tests, Demetrescu, Hanck and Tarcolea (2014) and Yang, Liu, Cai and Peng (2017) consider adding the lagged variables into the model so that model (1) becomes dynamic as

$$y_t = \alpha + \gamma_1 y_{t-1} + \beta x_{t-1} + u_t.$$

Clearly, if $\beta = 0$, the above model is the AR(1) for y_t. Then, Demetrescu, Hanck and Tarcolea (2014) argue that the power losses of the IVX approach based test can be reduced to a minimum at the cost of loosening size control, while Yang, Liu, Cai and Peng (2017) employ the weighted empirical likelihood approach. Both show empirically that indeed, the dynamic approach can improve the local power than the IVX method.

2.3.5 Variable addition

Elliott (2011) and Breitung and Demetrescu (2015) initiate a variable addition (VA) or control function approach by adding an additional variable z_t into model (1) as

$$y_t = \alpha + \beta x_{t-1} + \gamma z_{t-1} + u_t, \tag{7}$$

where z_{t-1} is an augmented variable, and its generating mechanism will be specified later. The method is called as control function method in Elliott (2011) and Phillips and Lee (2013) as mentioned earlier. If z_{t-1} is taken to be $\hat{v}_t$ in equation (3), then model (7) becomes to the linear projection approach described in equation (3).

There have been some approaches proposed in recent years about choosing z_{t-1} in practice. For example, Elliott (2011) proposes adding a stationary variable (e.g., z_{t-1} is I(0) and possibly, it is orthogonal to u_t) into the predictive regression to help stabilize the limit theory. In simulations, this approach is shown to have better size control with higher local power than the method in Campbell and Yogo (2006). Recently, Liu, Yang, Cai and Peng (2016) consider taking z_{t-1} to be x_{t-2} so that the above model in (7) becomes to

$$y_t = \alpha + \beta x_{t-1} + \gamma\, x_{t-2} + u_t = \alpha + \beta\, \Delta x_{t-1} + \beta_1\, x_{t-2} + u_t, \tag{8}$$

where $\beta_1 = \beta + \gamma$, and then, they use the weighted empirical likelihood approach to construct confidence interval for β and β_1 regardless of ϕ. One of the nice properties of the above model in (8) is that even if $\beta_1 = 0$, y_t can be still predictable using Δx_{t-1}. Further, Breitung and Demetrescu (2015) give out some general conditions for z_t. That is, for some $0 \le \nu < 1/2$, z_t satisfies: (i) $V_{T,z} = \sum_{t=2}^T z_{t-1}^2/T^{1+2\nu} \to V_z$ and $V_{T,zu} = \sum_{t=2}^T z_{t-1}^2 u_t^2/T^{1+2\nu} \to V_{zu}$, where V_z and V_{zu} are positive and bounded; (ii) $\frac{1}{T^{1.5+\nu}}\sum_{t=2}^T z_{t-1}x_{t-1} \to 0$ and $\frac{1}{\sqrt{V_{zu}}}\frac{1}{T^{\nu+1/2}}\sum_{t=2}^T z_{t-1}\, u_t \to Z$, where Z is the standard normal random variable. Under the above conditions, there are a wide range of candidate variables, such as stationary processes, mild integrated processes, fractional integrated processes and long differences processes. For more details, one may refer to the paper by Breitung and Demetrescu (2015). However, the augmented variable z_t generated under the above conditions still maintains the property that it is less persistent than x_t, sacrificing the convergence rate.

2.3.6 Double-weighted robust

Curiously, one might ask why one can not take z_t to be nonstationary. To answer this question, recently, Cai, Chen and Liao (2017a) develop a new method by combining two coefficients estimators in the variable addition regression and allow for an exogenous but nearly integrated z_t. To this end, Cai, Chen and Liao (2017a) rewrite model (1) as follows:

$$y_t = \alpha + \beta x_{t-1} + u_t = \alpha + \beta_1 x_{t-1}^* + \beta_2\, z_{t-1} + u_t, \tag{9}$$

where $x_{t-1}^* = x_{t-1} - z_{t-1}$, $\beta_1 = \beta$, $\beta_2 = \beta$, and z_{t-1} is nearly integrated additional variable generated exogenously as $z_t = (1 + c_z/T)z_{t-1} + u_{zt}$, where $u_{zt} \sim N(0, 1)$. One can obtain the OLS estimator, denoted by $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)$. Since both $\hat{\beta}_1$ and $\hat{\beta}_2$ converge to the same value β, Cai, Chen and Liao (2017a) suggest using a weighted approach to estimate β as

$$\tilde{\beta}_w = \frac{W_1}{W_1 + W_2}\hat{\beta}_1 + \frac{W_2}{W_1 + W_2}\hat{\beta}_2,$$

where

$$W_1 = \sum_{t=2}^T x_{t-1}^* z_{t-1}/T^2 - \sum_{t=2}^T x_{t-1}^* \sum_{t=2}^T z_{t-1}/T^3 \text{ and } W_2 = \sum_{t=2}^T z_{t-1}^2/T^2 - \left(\sum_{t=2}^T z_{t-1}/T^{3/2}\right)^2.$$

Furthermore, Cai, Chen and Liao (2017a) show that

$$(W_1 + W_2)T\left(\tilde{\beta}_w - \beta\right)/\sqrt{W_2\hat{\sigma}_u^2} \xrightarrow{d} N(0,1),$$

where $\hat{\sigma}_u^2 = \sum_{t=1}^T \hat{u}_t^2/T$ and $\hat{u}_t$ is the residual from (9). Therefore, one can use this result to construct a test statistic for testing $H_0 : \beta = 0$, denoted by

$$Q_w = (W_1 + W_2)T\tilde{\beta}_w/\sqrt{W_2\hat{\sigma}_u^2}. \tag{10}$$

Also, Cai, Chen and Liao (2017a) show that if x_{t-1} is I(1) or NI(1), under $H_0 : \beta = 0$,

$$Q_w \xrightarrow{d} N(0,1)$$

and under the alternative

$$Q_w \xrightarrow{p} \infty,$$

which implies that Q_w is a consistent test. However, Cai, Chen and Liao (2017a) find out that similar to the IVX and VA approaches, Q_w in (10) might not be consistent when x_{t-1} is I(0). Therefore, to remedy this shortage, Cai, Chen and Liao (2017a) suggest constructing a test statistic to accommodate both stationary and nonstationary cases. Indeed, Cai, Chen and Liao (2017a) propose the following test statistic

$$Q_{RW} = \frac{W_*}{1 + W_*}Q_w + \frac{1}{1 + W_*}t_{\hat{\beta}_s}$$

for some W_* satisfying that $W_* \xrightarrow{p} 0$ if x_{t-1} is I(0) and $W_* \xrightarrow{p} \infty$ if x_{t-1} is I(1) or NI(1). For example, one can take W_* to be

$$W_* = \sum_{t=2}^T x_{t-1}^2/T^{1+\zeta}$$

for any $0 < \zeta < 1$, which can characterize the degree of nonstationarity. Here, $t_{\hat{\beta}}$ is the t-statistic and $\hat{\beta}$ is the OLS estimate of β for classical predictive regression. Finally, Cai, Chen and Liao (2017a) show that under H_0, $Q_{RW} = Q_w + o_p(1)$ for I(1) or NI(1) x_{t-1} and $Q_{RW} = t_{\hat{\beta}} + o_p(1)$ for I(0) x_{t-1}. Therefore, $Q_{RW} \xrightarrow{d} N(0,1)$ under H_0 for all cases. Also, they derive the asymptotic distribution of Q_{RW} under the local alternative. In all cases, $Q_{RW} \xrightarrow{p} \infty$ under the alternative hypothesis $H_a : \beta \neq 0$, which implies that Q_{RW} is consistent. Furthermore, Q_{RW} reaches its optimal convergence rate $\sqrt{T}$ with I(0) x_t and T with I(1) x_t, respectively.

§3 Quantile predictive regression models

In the preceding work, all errors in predictive regressions are assumed to be homoskedastic; that is, conditional variance is constant (do not change over x_{t-1} or time). However, it is well documented that economic or financial data are generally heteroskedastic. In the literature, there have been some papers devoted to predictive mean regression models with heteroskedasticity. For example, Han and Park (2012), Han and Zhang (2012) and Choi, Jacewitz and Park (2016) consider model (1) with heteroskedasticity as $u_t = \sigma_t e_t$, where $\sigma_t = \sigma(q_t/\sqrt{T})$ with some nonstationary regressor q_t or $\sigma_t = \sigma(t/\sqrt{T})$. Also, Choi, Jacewitz and Park (2016) propose a robust test based on the Cauchy estimator with no constant term in (1) and σ_t is stationary.

Another way to handle the heteroskedasticity issue is to use quantile regression. In addition to the reason for capturing heteroskedasticity automatically, there are still more reasons to model the predictive quantile regression with persistent predictors. First, quantile regression is wildly used in risk management to estimate VaR and CoVaR as in Adrian and Brunnermeier (2016). Secondly, quantile regression avoids the imbalance issue in the mean regression models, i.e., the regressors are highly persistent but the dependent variable might be close to be stationary. The imbalance issue does not only make equation (1) false but also make the OLS estimator of β always converging to zero even $\beta \neq 0$. Finally, quantile regression techniques can characterize well asymmetric and heavy-tailed distribution of the dependent variable.

In recent years, there have been several papers devoted to modeling the quantile regression with non-stationary explanatory variables. For example, Xiao (2009a) is the first work to consider the quantile regression with unit root regressors

$$q_\tau(x_t) = \alpha_\tau + \beta_\tau x_t, \tag{11}$$

where $P(y_t \leq q_\tau(x_t)|x_t) = \tau$ and x_t is unit root, whereas Lee (2016) generalizes (11) to the predictive setting and proposes using the IVX method (dubbed as IVX-QR) as mentioned above to estimate parameters in (11) as well as derives the asymptotic theory. In the following subsections, we review several methods to robust inferences under the framework of predictive quantile regressions.

3.1 IVX-QR

Given the following linear predictive quantile regression model,

$$q_\tau(x_{t-1}) = \alpha_\tau + \beta_\tau x_{t-1},\ 1 \leq t \leq T, \tag{12}$$

where $x_t = \phi x_{t-1} + u_t$ and $\phi = 1 + c/T^\kappa$. With different values of (c, κ), x_t could be stationary, or mild integrated, or nearly integrated or integrated, or mild explosive. Lee (2016) shows that the asymptotic distribution of the t-test statistic depends on the nuisance parameter c, given below (similar to (5))

$$t_{\hat{\beta}_\tau} = \frac{\hat{\beta}_\tau - \beta_\tau}{s.e.(\hat{\beta}_\tau)} \Rightarrow \sqrt{1-\delta_\tau^2}\, Z + \delta_\tau\, \tau_c/k_c,$$

where Z is the standard normal random variable, both τ_c and k_c are defined in (5), and δ_τ, which, similar to δ in a mean predictive model, is regarded as measuring the correlation coefficient between u_t and with $\psi_{t\tau} = \tau - I(y_t \leq \alpha_\tau + \beta_\tau x_{t-1})$. Here, $\hat{\alpha}_\tau$ and $\hat{\beta}_\tau$ are the quantile estimator of α_τ and β_τ, respectively, given by

$$(\hat{\alpha}_\tau, \hat{\beta}_\tau)^\top = \arg\min_{\alpha_\tau, \beta_\tau} \sum_{t=2}^{T} \rho_\tau\left(y_t - \alpha_\tau - \beta_\tau x_{t-1}\right),$$

where $\rho_\tau(v) = v\,(\tau - I(v \leq 0))$, the check function. Similar to Campbell and Yogo (2006), Lee (2016) demonstrates that δ_τ indeed is non-zero for some real applications, which implies that the so-called embedded endogeneity still exists under predictive quantile regressions.

To solve the above issues, Lee (2016) extends the IVX approach to quantile regression framework and proposes a new approach (IVX-QR), to estimate model (12) using an instrumental variable z_t generated by equation (6). The IVX-QR estimator is defined as:

$$(\hat{\alpha}_\tau^{\mathrm{IVX-QR}}, \hat{\beta}_\tau^{\mathrm{IVX-QR}})^\top = \arg\min_{\alpha_\tau,\beta_\tau} \sum_{t=2}^{T} \rho_\tau\left(y_t - \alpha_\tau - \beta_\tau z_{t-1}\right).$$

Under the null hypothesis $H_0 : \beta_\tau = 0$, Lee (2016) shows that

$$t_{\hat{\beta}_\tau} = \frac{\hat{\beta}_\tau^{\mathrm{IVX-QR}}}{s.e.(\hat{\beta}_\tau^{\mathrm{IVX-QR}})} \ \Rightarrow\ N(0,1).$$

However, there is still no theoretical criterion on how to choose turning parameters which might affect the finite-sample performance of inferences as elaborated in Phillips and Lee (2013). Moreover, the convergence rate of the IVX-QR estimator under the null hypothesis $H_0 : \beta_\tau = 0$ is less than the optimal rate T and thus it may suffer from a loss of power.

Recently, Fan and Lee (2017) provide a valid and easy-to-use inference procedure in predictive quantile framework with conditional heteroskedasticity innovations and suggest using the IVX-QR method and the moving block-wise Bootstrap of Bühlmann and Künsch (1999). Therefore, it inherits these limitations of the IVX-QR for quantile predictive regressions.

3.2 Weighted variable addition

Using a similar idea in equation (9), Cai, Chen and Liao (2017b) extend the weighted approach discussed in section (2.3.6) to quantile predictive regression framework. Similarly, one can rewrite the model (12) as follows:

$$q_\tau(x_{t-1}) = \alpha_\tau + \beta_\tau x_{t-1} = \alpha_\tau + \beta_{1\tau} x_{t-1}^* + \beta_{2\tau} z_{t-1},$$

where z_{t-1} is the additional variable, $x_{t-1}^* = x_{t-1} - z_{t-1}$, and $\beta_{1\tau} = \beta_{2\tau} = \beta_\tau$. The quantile estimator is defined as

$$(\hat{\alpha}_\tau, \hat{\beta}_{1\tau}, \hat{\beta}_{2\tau})^\top = \arg\min_{\alpha_\tau,\beta_{1\tau},\beta_{2\tau}} \sum_{t=2}^{T} \rho_\tau\left(y_t - \alpha_\tau - \beta_{1\tau} x_{t-1}^* - \beta_{2\tau} z_{t-1}\right).$$

To achieve the optimal convergence rate T, Cai, Chen and Liao (2017b) define the control function variable z_t as $z_{t-1} = x_{t-1}/\sqrt{1 + x_{t-1}^2}$, which is different from that in Cai, Chen and Liao (2017a), and they show that z_{t-1} is I(0) if x_{t-1} is I(0), and $z_{t-1} \Rightarrow \mathrm{sign}(K_c(r))$ as $t \to \infty$ if x_{t-1} is I(1) or NI(1), where sign$(\cdot)$ is the sign function and $K_c(\cdot)$ is defined in (5).

Define the weighted estimator $\tilde{\beta}_\tau^w$ as follows:

$$\tilde{\beta}_\tau^w = \frac{W_1}{W_1 + W_2}\hat{\beta}_{1\tau} + \frac{W_2}{W_1 + W_2}\hat{\beta}_{2\tau},$$

where

$$W_1 = \frac{1}{T^2}\sum_{t=2}^{T} x_{t-1}^* z_{t-1} - \frac{1}{T^{3/2}}\sum_{t=2}^{T} x_{t-1}^* \frac{1}{T^{3/2}}\sum_{t=2}^{T} z_{t-1} \text{ and } W_2 = \frac{1}{T^2}\sum_{t=2}^{T} z_{t-1}^2 - \left(\frac{1}{T^{3/2}}\sum_{t=2}^{T} z_{t-1}\right)^2.$$

Then, under the null hypothesis $H_0 : \beta_\tau = 0$, Cai, Chen and Liao (2017b) show that, the test statistic

$$Q_\tau^w = \hat{f}_{v_\tau}\ [W_2 \tau(1-\tau)]^{-1/2}\,(W_1 + W_2) T \tilde{\beta}_\tau^w \Rightarrow \mathrm{N}(0,1),$$

where $\hat{f}_{v_\tau}$ is a consistent estimate of a normalization constant f_{v_τ}; see Cai, Chen and Liao

(2017b) for details, and that under the alternative hypothesis $H_a : \beta_\tau \neq 0$,

$$Q_\tau^w \xrightarrow{p} \infty$$

no matter x_{t-1} is I(0), or I(1), or NI(1). Therefore, Q_τ^w is consistent for all cases. Furthermore, they show that Q_τ^w reaches the optimal rate $\sqrt{T}$ if x_{t-1} is I(0), and the optimal rate T if x_{t-1} is I(1) or NI(1), respectively.

§4 Nonlinear predictive models

4.1 Models for time-varying coefficients

So far, the parameters in both mean and quantile regressions, are assumed to be stable. That is, coefficients are constant (do not change over random variable or time). For time series data in economics and finance over a long time period, it is reasonable to expect that parameters in predictive regression models may experience structural changes at some unknown dates. Actually, Viceira (1997), Paye and Timmermann (2006), and Rapach and Wohar (2006) find strong evidence of instability in predictive regression models. However, they do not explain how to test predictability after detecting and estimating the break dates. Lettau and Nieuwerburgh (2008) focus on level shifts in the predictor variables and explain that the forecasting relationship is unstable unless such shifts are included in the analysis.

To model this instability in predictive regression, Cai, Wang and Wang (2015) consider a model with coefficients changing smoothly with time as

$$y_t = \alpha_t + \beta_t\, x_{t-1} + u_t, \tag{13}$$

where both α_t and β_t are smooth functions of time. Then, they propose a nonparametric testing procedure to test whether the time-varying coefficients are indeed changing with time. That is, the null hypothesis is $H_0 : \alpha_t = \alpha \;\&\; \beta_t = \beta$. They find that indeed, the coefficients are unstable for testing predictability of asset returns based on a real example.

Thereafter, the question arises is how to specify the form of time-varying coefficients. To solve this problem, Cai and Chang (2017) consider model (13), where both α_t and β_t are piecewise constants, with one break (easy to consider multiple breaks), specified as $\alpha_t = \alpha_1 I(t \leq \kappa_1) + \alpha_2 I(t > \kappa_1)$ and $\beta_t = \beta_1 I(t \leq \kappa_1) + \beta_2 I(t > \kappa_1)$, where the break point κ_1 is unknown. Also, x_{t-1} might be allowed to have a structural change at the same or different break point. They propose using the weighted empirical likelihood approach as mentioned earlier to test predictability in two time periods, i.e., $H_0 : \beta_1 = 0$ and/or $H_0 : \beta_2 = 0$. Of interest is to test $H_0 : \beta_1 = \beta_2$, which is the well known Chow test for the existence of structural changes.

Additionally, a threshold form of model in (1) is considered by Chen (2015) and Ganzalo and Pitarakis (2017) as

$$y_t = \alpha(q_{t-1}) + \beta(q_{t-1})x_{t-1} + u_t, \tag{14}$$

where q_t is an observable stationary variable (say, some variable proxying business cycles). Here, they assume that $\alpha(q_{t-1}) = \alpha_1 I(q_{t-1} \leq \gamma) + \alpha_2 I(q_{t-1} > \gamma)$ and $\beta(q_{t-1}) = \beta_1 I(q_{t-1} \leq \gamma) + \beta_2 I(q_{t-1} > \gamma)$, so that model (14) is a nonlinear model. They use the Wald-type test statistic to test predictability with the null hypothesis specified as $H_0 : \beta_1 = \beta_2 = 0$, or test the parameter stability with the null hypothesis defined as $H_0 : \alpha_1 = \alpha_2$. Note that if both $\alpha(\cdot)$ and $\beta(\cdot)$ are smoothing functions, model (14) reduces to the model in Cai, Li and Park (2009), Xiao (2009b) and Chen, Fang and Li (2015) when q_{t-1} is I(0) and x_{t-1} is I(1), and the model in Sun, Cai and Li (2013) when both q_{t-1} and x_{t-1} are I(1).

4.2 Nonparametric models and their tests

One of possible ways to avoid the embedded endongeneity is to consider a nonparametric model as

$$y_t = m(x_{t-1}) + u_t,$$

where $m(x_{t-1}) = E(y_t|x_{t-1})$ so that $E(u_t|x_{t-1}) = 0$, which is a more general model since there is no restriction on $m(\cdot)$. To estimate $m(\cdot)$, there is a vast amount of literature on this topic for both stationary and nonstationary cases; see, for example, to name just a few, the papers by Cai, Fan and Yao (2000) for stationary case, and Cai, Li and Park (2009) and Cai (2011) for nonstationary case. Note that as elaborated in Cai (2011), the local constant and local linear estimators share the exact same large sample behavior for nonstationary regressors.

The key point is how to test $H_0 : m(x) = m_0(x, \beta)$, where $m_0(\cdot, \cdot)$ is a known function but β is an unknown parameter. By this way, it could be applied to check if $m(x)$ is a threshold function or other types of parametric forms to see if some financial/economic theory holds. To this end, Cai and Wu (2013) consider this kind of test and propose a L_2-type test statistic. That is,

$$||\hat{m}(\cdot) - m_0(\cdot, \hat{\beta})||_2^2 = \int \left(\hat{m}(x) - m_0(x, \hat{\beta})\right)^2 D(x)dx$$

for some weighting function $D(\cdot)$ to avoid a random denominator and $\hat{\beta}$ is the estimate of β under the null hypothesis, which can be simplified as a U-statistic as

$$J_T = \frac{1}{T^{3/4}\, h^{1/2}} \sum_{t \neq s} \hat{u}_t \hat{u}_s K_{ts}, \tag{15}$$

where h is the bandwidth and $K_{ts} = K((x_{t-1} - x_{s-1})/h)$ with $K(\cdot)$ being the kernel function used to obtain $\hat{m}(x)$. Also, Cai and Wu (2013) derive the asymptotic distribution of the proposed test statistic J_T under H_0 and show that it diverges to ∞ under the alternative hypothesis. In particular, of interest is that they derive the limiting result for a U-statistic involving nonstationary variables.

Recently, by assuming that $m_0(\cdot) = \mu_m$ but unknown, Kasparis, Andreou and Phillips (2015) consider a special case of the above hypothesis test by specifying $H_0 : m(z) = \mu_m$ for all z. Then, based on the asymptotic distribution of the nonparametric estimation of $m(z_j)$, denoted by $\hat{m}(z_j)$ for some grid points $\{z_j\}_{j=1}^m$, they propose a naive test as follows:

$$\hat{F}_{sum} = \sum_{j=1}^{m} A(z_j) \left(\hat{m}(z_j) - \hat{\mu}_m\right)^2 \text{ or } \hat{F}_{max} = \max_{1 \leq j \leq m} A(z_j) \left(\hat{m}(z_j) - \hat{\mu}_m\right)^2,$$

where $A(\cdot)$ is a self-normalized function and $\hat{\mu}_m$ is the estimate of μ_m under the null hypothesis. They argue that $\hat{F}_{sum} \stackrel{d}{\to} \chi_m^2$ and $\hat{F}_{max} \stackrel{d}{\to} Z_*$ for some random variable Z_*. But, the proposed test depends on the choice of $\{z_j\}_{j=1}^m$ and m.

Further, by assuming that $\mu_m = 0$, Juhl (2014) considers testing the hypothesis $H_0 : m(z) = 0$ for all z, and proposes using the conditional moment testing approach as in Zheng (1996) and Fan and Li (1999) to construct the test statistic. Different from J_T in (15) involving $\{\hat{u}_t\}$, the test statistic proposed in Juhl (2014) is given as follows:

$$U_T = \frac{1}{T^2\,h}\sum_{t\neq s} y_t\, y_s K_{ts},$$

which is a function of $\{y_t\}$ instead of $\{\hat{u}_t\}$, and its limiting distribution under H_0 is derived by Juhl (2014); see Juhl (2014) for details.

§5 Conclusions

Although the literature on modeling the predictive regression with stationary or nonstationary predictor is tremendously and rapidly growing, there are still some open issues left to be addressed. First, how to deal with the predictive regression model with x_t having structural changes is still unclear, although Cai and Chang (2017) consider the case with a known break point in x_t. Secondly, it deserves to develop a test similar to the Chow test for testing the existence of structural changes in predictive regression models, i.e., test $\beta_1 = \beta_2$, when regressors are nonstationary. The third issue is how to develop a new method, unlike the weighted empirical likelihood approach, to avoid losing half of the sample, when the model contains an unknown intercept. The fourth one is how to generalize predictive models to a more general setting for predicting global asset returns (panel data); see, for example, Hjalmarsson (2010) for an example on panel data with/without cross-sectional dependence. The fifth issue is how to conduct nonparametric tests, especially under nonparametric quantile predictive models, which can provide a good way to construct the prediction interval to directly predict the asset returns. Finally, the interesting topic is how to develop predictive regression models applied to other fields such as macroeconomics. For instance, it is worth studying bubbles by assuming that $\phi = 1 + c/T^{\zeta}$ with $c > 0$ and $0 \leq \zeta < 1$ so that x_t is mildly explosive; see the papers by Philipps, Shi and Yu (2015a, 2015b) and others on characterizing the bubbles in macroeconomics. Clearly, the aforementioned issues and the related topics are very interesting and they are warranted as future research topics.

References

[1] T Adrian, M K Brunnermeier. *CoVaR*, Am Econ Rev, 2016, 106(7): 1705-1741.

[2] Y Amihud, C M Hurvich. *Predictive regressions: A reduced-bias estimation method*, J Financ Quant Anal, 2004, 39(4): 813-841.

[3] Y Amihud, C M Hurvich, Y Wang. *Multiple-predictor regressions: Hypothesis testing*, Rev Financ Stud, 2009, 22(1): 413-434.

[4] P Bossaerts, P Hillion. *Implementing statistical criteria to select return forecasting models: what do we learn?*, Rev Financ Stud, 1999, 12(2): 405-428.

[5] J Breitung, M Demetrescu. *Instrumental variable and variable addition based inference in predictive regressions*, J Econometrics, 2015, 187(1): 358-375.

[6] P Bühlmann, H R Künsch. *Block length selection in the bootstrap for time series*, Comput Statist Data Anal, 1999, 31(3): 295-310.

[7] Z Cai. *Nonparametric regression models with integrated covariates*, In: *Non-parametric Statistical Methods and Related Topics: A Festschrift in Honor of Professor P.K. Bhattacharya on his 80th Birthday*, J Jiang, G G Roussas and F J Samaniego Eds, 2011: 257-275.

[8] Z Cai and S Y Chang. *A new test on asset return predictability with structural breaks*, Working paper, Xiamen University, 2017.

[9] Z Cai, H Chen, X Liao. *A double weighted robust approach to test predictability with persistent predictors*, Working Paper, Xiamen University, 2017a.

[10] Z Cai, H Chen, X Liao. *A weighted variable additon approach to test predictability in quantile regression with persistent predictors*, Working Paper, Xiamen University, 2017b.

[11] Z Cai, J Fan, Q Yao. *Functional-coefficient regression models for nonlinear time series*, J Amer Statist Assoc, 2000, 95(451): 941-956.

[12] Z Cai, Q Li, J Y Park. *Functional-coefficient models for nonstationary time series data*, J Econometrics, 2009, 148(2): 101-113.

[13] Z Cai, Y Wang. *Testing predictive regression models with nonstationary regressors*, J Econometrics, 2014, 178: 4-14.

[14] Z Cai, Y Wang , Y Wang. *Testing instability in a predictive regression model with nonstationary regressors*, Econometric Theory, 2015, 31(5): 953-980.

[15] Z Cai, L Wu. *A consistent nonparametric test when regressors are nonstationary*, Working paper, University of Kansas, 2013.

[16] J Y Campbell. *Stock returns and the term structure*, J Financ Econ, 1987, 18(2): 373-399.

[17] J Y Campbell, A W C Lo, A C MacKinlay. *The Econometrics of Financial Markets*, Princeton University Press, 1997.

[18] J Y Campbell, R J Shiller. *The dividend-price ratio and expectations of future dividends and discount factors*, Rev Financ Stud, 1988, 1(3): 195-228.

[19] J Y Campbell, M Yogo. *Implementing the econometric methods in "Efficient tests of stock return predictability"*, Working paper, University of Pennsylvania, 2005.

[20] J Y Campbell, M Yogo. *Efficient tests of stock return predictability*, J Financ Econ, 2006, 81(1): 27-60.

[21] H Chen. *Robust estimation and inference for threshold models with integrated regressors*, Econometric Theory, 2015, 31(4): 778-810.

[22] H Chen, Y Fang, Y Li. *Estimation and inference for varying-coefficient models with nonstationary regressors using penalized splines*, Econometric Theory, 2015, 31(1): 753-777.

[23] W W Chen, R S Deo. *Bias reduction and likelihood-based almost exactly sized hypothesis testing in predictive regressions using the restricted likelihood*, Econometric Theory, 2009, 25(5): 1143-1179.

[24] Y Choi, S Jacewitz, J Y Park. *A reexamination of stock return predictability*, J Econometrics, 2016, 192(1): 168-189.

[25] M Demetrescu, C Hanck, A I Tarcolea. *IV-based cointegration testing in dependent panels with time-varying variance*, J Time Ser Anal, 2014, 35(5): 393-406.

[26] G Elliott. *A control function approach for testing the usefulness of trending variables in forecast models and linear regression*, J Econometrics, 2011, 164(1): 79-91.

[27] E F Fama. *The behavior of stock-market prices*, J Business, 1965, 38(1): 34-105.

[28] E F Fama. *Efficient capital markets: A review of theory and empirical work*, J Financ, 1970, 25(2): 383-417.

[29] E F Fama. *Efficient capital markets: II*, J Financ, 1991, 46(5): 1575-1617.

[30] E F Fama, K R French. *Dividend yields and expected stock returns*, J Financ Econ, 1988, 22(1): 3-25.

[31] R Fan, J H Lee. *Predictive quantile regressions under persistence and conditional heteroskedasticity*, Working paper, University of Illinois at Champion and Urbana, 2017.

[32] Y Fan, Q Li. *Consistent model specification tests: Omitted variables and semi-parametric functional forms*, Econometrica, 1999, 64(4): 865-890.

[33] J Gonzalo, J Y Pitarakis. *Inferring the predictability induced by a persistent regressor in a predictive threshold model*, J Bus Econom Statist, 2017, 35(2): 202-217.

[34] H Han, J Y Park. *ARCH/GARCH with persistent covariate: Asymptotic theory of MLE*, J Econometrics, 2012, 167(1): 95-112.

[35] H Han, S Zhang. *Non-stationary non-parametric volatility model*, Econom J, 2012, 15(2): 204-225.

[36] E Hjalmarsson. *Predicting global stock returns*, J Financ Quant Anal, 2010, 45(1):49-80.

[37] T Juhl. *Nonparametric test of the predictive regression model*, J Bus Econom Statist, 2014, 32(3): 387-394.

[38] M Jansson, M J Moreira. *Optimal inference in regression models with nearly integrated regressors*, Econometrica, 2006, 74(3): 681-714.

[39] E Jondeau, S H Poon, M Rockinger. *Financial Modeling Under Non-Gaussian Distributions*, Springer Science & Business Media, 2007.

[40] I Kasparis, E Andreou, P C Phillips. *Nonparametric predictive regression*, J Econometrics, 2015, 185(2): 468-494.

[41] A Kostakis, T Magdalinos, M P Stamatogiannis. *Robust econometric inference for stock return predictability*, Rev Financ Stud, 2015, 28(5): 1506-1553.

[42] J H Lee. *Predictive quantile regression with persistent covariates: IVX-QR approach*, J Econometrics, 2016, 192(1): 105-118.

[43] M Lettau, S Ludvigson. *Consumption, aggregate wealth, and expected stock returns*, J Financ, 2001, 56(3): 815-849.

[44] M Lettau, S Van Nieuwerburgh. *Reconciling the return predictability evidence*, Rev Financ Stud, 2008, 21(4): 1607-1652.

[45] J Lewellen. *Predicting returns with financial ratios*, J Financ Econ, 2004, 74(2): 209-235.

[46] X Liu, B Yang, Z Cai and L Peng. *A unified test for predictability of asset returns regardless of properties of predicting variables*, forthcoming: J Econometrics, 2016.

[47] B S Paye, A Timmermann. *Instability of return prediction models*, J Empir Financ, 2006, 13(3): 274-315.

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

[49] P C Phillips, T Magdalinos. *Econometric inference in the vicinity of unity*, Working paper, Yale University, 2009.

[50] P C Phillips, S Shi, J Yu. *Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500*, Int Econ Rev, 2015a, 56(4): 1043-1078.

[51] P C Phillips, S Shi, J Yu. *Testing for multiple bubbles: Limit theory of real-time detectors*, Int Econ Rev, 2015b, 56(4): 1079-1134.

[52] P C Phillips, V Solo. *Asymptotics for linear processes*, Ann Statist, 1992, 20(2): 971-1001.

[53] D E Rapach, M E Wohar. *In-sample vs. out-of-sample tests of stock return predictability in the context of data mining*, J Empir Financ, 2006, 13(2): 231-247.

[54] P M M Rodrigues, M Demetrescu. *Residual-augmented IVX predictive regression*, Working paper, Goethe University Frankfurt, 2016.

[55] R F Stambaugh. *Predictive regressions*, J Financ Econ, 1999, 54(3): 375-421.

[56] J H Stock. *Confidence intervals for the largest autoregressive root in US macroeconomic time series*, J Monetary Econ, 1991, 28(3): 435-459.

[57] Y Sun, Z Cai, Q Li. *Semiparametric functional coefficient models with integrated covariates*, E-conometric Theory, 2013, 29(3): 659-672.

[58] W Torous, R Valkanov, S Yan. *On predicting stock returns with nearly integrated explanatory variables*, J Business, 2004, 77(4): 937-966.

[59] L M Viceira. *Testing for structural change in the predictability of asset returns*, Manuscript, Harvard University, 1997.

[60] I Welch, A Goyal. *A Comprehensive look at the empirical performance of equity premium prediction*, Rev Financ Stud, 2008, 21(4): 1455-1508.

[61] Z Xiao. *Quantile cointegrating regression*, J Econometrics, 2009a, 150(2): 248-260.

[62] Z Xiao. *Functional-coefficient cointegrating models*, J Econometrics, 2009b, 152(2): 81-92.

[63] B Yang, X Liu, Z Cai and L Peng. *New robust econometric tests for a dynamic predictive regression*, Working paper, University of Kansas, 2017.

[64] J X Zheng. *A consistent test of functional form via nonparametric estimation techniques*, J E-conometrics, 1996, 75(2): 263-289.

[65] F Zhu, Z Cai, L Peng. *Predictive regressions for macroeconomic data*, Ann Appl Stat, 2014, 8(1): 577-594.

[1] The Wang Yanan Institute for Studies in Economics, MOE Key Laboratory of Econometrics (Xiamen University), Fujian Provincial Key Laboratory of Statistical Science, Xiamen University, Xiamen 361005, China.

Email: liaoxiaosai@126.com
hc335@xmu.edu.cn

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

Email: zongwucai@gmail.com