Methods on Testing Predictability of Asset Returns

Zongwu Cai

Department of Economics, University of Kansas and WISE, Xiamen University
E-mail: caiz@ku.edu

Present to Department of Economics, Vanderbilt University, April 4
A Note

Note that this talk is about my personal review on recent developments on testing predictability of asset returns. In particular, it includes my own work: two published papers, one paper revised and resubmitted, four papers submitted, and two working papers as well as several ongoing projects. In particular, I will summarize the recent results developed within the least 5 years and point out some future research in this area.


A Note


Motivation

The predictability of asset returns has been studied for recent three decades as a cornerstone research topic in economics and finance. It is widely examined in many financial applications such as the mutual fund performance, the conditional capital asset pricing, and the optimal asset allocations.

Predictability of asset returns has two major directions. The first one is to check whether the return as a time series is autocorrelated, or random walk, or martingale difference sequence (MDS). The second one is to use financial (state) variables as predictors to see if the financial (state) variables can predict asset returns.
Motivation

There is a vast amount of literature devoted to testing if asset returns are autocorrelated or random walk or MDS or other types of dependent structures; see, for example, the books by Campbell, Lo and MacKinlay (1997) and Jondeau, Poon and Rockinger (2010), and the references therein. I refer to the related books for this aspect. Alternatively, you can read my lecture notes “Econometric Analysis of Financial Data”.

In recent two decades, tremendous empirical studies have been devoted to testing predictability of asset returns using various lagged financial or state variables, such as the log dividend-price ratio, log earning-price ratio, the log book-to-market ratio, the dividend yield, the term spread and default premium, and the interest rates, as well as other economy variables.
Motivation

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

\[
\begin{align*}
y_t &= \alpha + \beta x_{t-1} + u_t, \\
x_t &= \theta + \phi x_{t-1} + v_t, \quad 1 \leq t \leq T,
\end{align*}
\]  

where \((u_t, v_t) \sim \mathcal{N}(0, \Sigma)\) are an independent and identically distributed (i.i.d.) series, and \(u_t\) and \(v_t\) can be correlated. Of course, the second equation in (1) can be in a higher order, say AR(p). For simplicity, the focus here is on AR(1).
Motivation

- Is model (1) easy? Answer: YES and NO!

- **Econometric/statistical issues:**
  
  (i) Correlation coefficient $\delta$ between $u_t$ and $v_t$ is non-zero in many real applications; see Table 4 in Campbell and Yogo (2006), which creates the so-called “embedded endogeneity”.

  (ii) Predicting regressor $x_t$ might be persistent. This means that $x_t$ 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 $c = 0$).

  (iii) When $\theta$ in the second equation of (1) is non-zero, it makes inferences for I(1) or NI(1) totally different from those for I(0) case.

- **WHY?** Let me show you some empirical studies in the literature.
Table 4
Estimates of the model parameters

<table>
<thead>
<tr>
<th>Series</th>
<th>Obs.</th>
<th>Variable</th>
<th>$p$</th>
<th>$\delta$</th>
<th>DF-GLS</th>
<th>95% CI: $\rho$</th>
<th>95% CI: $c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P 500</td>
<td>123</td>
<td>$d-p$</td>
<td>3</td>
<td>-0.845</td>
<td>-0.855</td>
<td>[0.949, 1.033]</td>
<td>[-6.107, 4.020]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$e-p$</td>
<td>1</td>
<td>-0.962</td>
<td>-2.888</td>
<td>[0.768, 0.965]</td>
<td>[-28.262, -4.232]</td>
</tr>
<tr>
<td>Annual</td>
<td>77</td>
<td>$d-p$</td>
<td>1</td>
<td>-0.721</td>
<td>-1.033</td>
<td>[0.903, 1.050]</td>
<td>[-7.343, 3.781]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$e-p$</td>
<td>1</td>
<td>-0.957</td>
<td>-2.229</td>
<td>[0.748, 1.000]</td>
<td>[-19.132, -0.027]</td>
</tr>
<tr>
<td>Quarterly</td>
<td>305</td>
<td>$d-p$</td>
<td>1</td>
<td>-0.942</td>
<td>-1.696</td>
<td>[0.957, 1.007]</td>
<td>[-13.081, 2.218]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$e-p$</td>
<td>1</td>
<td>-0.986</td>
<td>-2.191</td>
<td>[0.939, 1.000]</td>
<td>[18.670, 0.145]</td>
</tr>
<tr>
<td>Monthly</td>
<td>913</td>
<td>$d-p$</td>
<td>2</td>
<td>-0.950</td>
<td>-1.657</td>
<td>[0.986, 1.003]</td>
<td>[-12.683, 2.377]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$e-p$</td>
<td>1</td>
<td>-0.987</td>
<td>-1.859</td>
<td>[0.984, 1.002]</td>
<td>[-14.797, 1.711]</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>115</td>
<td>$d-p$</td>
<td>3</td>
<td>-0.835</td>
<td>-2.002</td>
<td>[0.854, 1.010]</td>
<td>[-16.391, 1.079]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$e-p$</td>
<td>1</td>
<td>-0.958</td>
<td>-3.519</td>
<td>[0.663, 0.914]</td>
<td>[-38.471, -9.789]</td>
</tr>
<tr>
<td>Annual</td>
<td>69</td>
<td>$d-p$</td>
<td>1</td>
<td>-0.693</td>
<td>-2.081</td>
<td>[0.745, 1.010]</td>
<td>[-17.341, 0.690]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$e-p$</td>
<td>1</td>
<td>-0.959</td>
<td>-2.859</td>
<td>[0.591, 0.940]</td>
<td>[-27.808, -4.074]</td>
</tr>
<tr>
<td>Quarterly</td>
<td>273</td>
<td>$d-p$</td>
<td>1</td>
<td>-0.941</td>
<td>-2.635</td>
<td>[0.910, 0.991]</td>
<td>[-24.579, -2.470]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$e-p$</td>
<td>1</td>
<td>-0.988</td>
<td>-2.827</td>
<td>[0.900, 0.986]</td>
<td>[-27.322, -3.844]</td>
</tr>
<tr>
<td>Monthly</td>
<td>817</td>
<td>$d-p$</td>
<td>2</td>
<td>-0.948</td>
<td>-2.551</td>
<td>[0.971, 0.998]</td>
<td>[-23.419, -1.914]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$e-p$</td>
<td>2</td>
<td>-0.983</td>
<td>-2.600</td>
<td>[0.970, 0.997]</td>
<td>[-24.105, -2.240]</td>
</tr>
<tr>
<td><strong>Panel C: CRSP 1952–2002</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual</td>
<td>51</td>
<td>$d-p$</td>
<td>1</td>
<td>-0.749</td>
<td>-0.462</td>
<td>[0.917, 1.087]</td>
<td>[-4.131, 4.339]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$e-p$</td>
<td>1</td>
<td>-0.955</td>
<td>-1.522</td>
<td>[0.773, 1.056]</td>
<td>[-11.354, 2.811]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$r_3$</td>
<td>1</td>
<td>0.006</td>
<td>-1.762</td>
<td>[0.725, 1.040]</td>
<td>[-13.756, 1.984]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$y-r_1$</td>
<td>1</td>
<td>-0.243</td>
<td>-3.121</td>
<td>[0.363, 0.878]</td>
<td>[-31.870, -6.100]</td>
</tr>
<tr>
<td>Quarterly</td>
<td>204</td>
<td>$d-p$</td>
<td>1</td>
<td>-0.977</td>
<td>-0.392</td>
<td>[0.981, 1.022]</td>
<td>[-3.844, 4.381]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$e-p$</td>
<td>1</td>
<td>-0.980</td>
<td>-1.195</td>
<td>[0.958, 1.017]</td>
<td>[-8.478, 3.539]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$r_3$</td>
<td>4</td>
<td>-0.095</td>
<td>-1.572</td>
<td>[0.941, 1.013]</td>
<td>[-11.825, 2.669]</td>
</tr>
</tbody>
</table>
For each of our sampled stochastic explanatory variables, table 1 presents 95% confidence intervals for $U$. We provide results using the entire time series of data and, to investigate the robustness of our conclusions, the pre-1952 and post-1952 subsamples. In almost every case, these 95% confidence intervals include the unit root $U = 1$. The exceptions include the log dividend yield series over the 1926:12 to 1994:12 sample period whose upper limit of 0.996 is nearly indistinguishable from 1. While the 95% confidence interval for the term spread series based on the entire sample period does not contain 1, the interval based on the post-1952 subsample does.

### TABLE 1 95% Confidence Intervals for the Largest Autoregressive Root of the Stochastic Explanatory Variables

<table>
<thead>
<tr>
<th>Series</th>
<th>Sample Period</th>
<th>$k$</th>
<th>ADF</th>
<th>95% Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dividend yield</td>
<td>1926:12–1994:12</td>
<td>5</td>
<td>-3.30</td>
<td>(.960, .996)</td>
</tr>
<tr>
<td></td>
<td>1926:12–1951:12</td>
<td>1</td>
<td>-2.84</td>
<td>(.915, 1.004)</td>
</tr>
<tr>
<td></td>
<td>1952:1–1994:12</td>
<td>1</td>
<td>-2.65</td>
<td>(.956, 1.004)</td>
</tr>
<tr>
<td>Default spread</td>
<td>1926:12–1994:12</td>
<td>2</td>
<td>-2.49</td>
<td>(.976, 1.003)</td>
</tr>
<tr>
<td></td>
<td>1926:12–1951:12</td>
<td>3</td>
<td>-0.90</td>
<td>(.984, 1.015)</td>
</tr>
<tr>
<td></td>
<td>1952:1–1994:12</td>
<td>2</td>
<td>-2.50</td>
<td>(.963, 1.004)</td>
</tr>
<tr>
<td>Book-to-market</td>
<td>1926:12–1994:08</td>
<td>6</td>
<td>-2.35</td>
<td>(.977, 1.003)</td>
</tr>
<tr>
<td></td>
<td>1926:12–1951:12</td>
<td>6</td>
<td>-1.60</td>
<td>(.967, 1.013)</td>
</tr>
<tr>
<td></td>
<td>1952:1–1994:08</td>
<td>6</td>
<td>-1.24</td>
<td>(.986, 1.008)</td>
</tr>
<tr>
<td>Term spread</td>
<td>1926:12–1994:12</td>
<td>6</td>
<td>-3.57</td>
<td>(.955, .992)</td>
</tr>
<tr>
<td></td>
<td>1926:12–1951:12</td>
<td>6</td>
<td>-3.11</td>
<td>(.943, .999)</td>
</tr>
<tr>
<td></td>
<td>1952:1–1994:12</td>
<td>2</td>
<td>-1.83</td>
<td>(.957, 1.012)</td>
</tr>
<tr>
<td>Short-term rate</td>
<td>1926:12–1994:12</td>
<td>8</td>
<td>-1.85</td>
<td>(.984, 1.004)</td>
</tr>
<tr>
<td></td>
<td>1926:12–1951:12</td>
<td>1</td>
<td>-1.90</td>
<td>(.955, 1.012)</td>
</tr>
<tr>
<td></td>
<td>1952:1–1994:12</td>
<td>7</td>
<td>-1.90</td>
<td>(.974, 1.007)</td>
</tr>
</tbody>
</table>

Note.—This table provides 95% confidence intervals for the largest autoregressive root $\rho$ of stochastic explanatory variables typically used in predictive regressions. The explanatory variables used are Dividend yield, Default spread, Book to market, Term spread, and Short-term rate. Dividend yield is the log real dividend yield, constructed as in Fama and French (1988). Default spread is the log of the difference between monthly averaged annualized yields of bonds rated Baa and Aaa by Moody’s. Book-to-market is the log of Pontiff and Schall’s (1998) Dow Jones Industrial Average (DJIA) book-to-market ratio. Term spread is the difference between annualized yields of Treasury bonds with maturity closest to 10 years at month end and 3-month Treasury bills. Short-term rate is the nominal 1-month Treasury bill rate. The augmented Dickey-Fuller statistic is denoted ADF, and we follow Ng and Perron (1995) in determining the maximum lag length $k$. We use the sequential pretesting method of Ng and Perron (1995) to determine the maximum lag length $k$. This method selects $k$ only after testing sequentially that the coefficients on additional lags, $k + 1$ and longer, are statistically insignificant.
Stambaugh (1999) showed that the least squares estimator for $\beta$ in (1) is biased in finite sample although it is asymptotically unbiased, due to the nonzero correlation between $u_t$ and $v_t$, under normality and with stationary regression ($|\phi| < 1$). How to improve the finite sample performance?

Indeed, Stambaugh (1999) derived the bias of the least squares estimator $\hat{\beta}$ as follows:

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

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

(i) First order bias correction (Stambaugh, 1999) as in (2). That is,

\[ \hat{\beta}_{bc} = \hat{\beta} - \hat{\rho}(1 + 3\hat{\phi}) / T, \]

where \( \hat{\rho} \) and \( \hat{\phi} \) are the estimator of \( \rho \) and \( \phi \), respectively.

(ii) Second-order bias-correction of Amihud and Hurvich (2004) based on a projection method \( (u_t = \rho v_t + \epsilon_t) \) and two stage augmented regression method as follows:
I: Bias-Correction

\[ y_t = \alpha + \beta x_{t-1} + \rho \hat{v}_t + \epsilon_t, \tag{3} \]

where \( \hat{v}_t \) is obtained from the second equation of (1). The OLS estimator of \( \beta \), denoted by \( \hat{\beta} \), is the so-called two-stage approach. By assuming that \( x_t \) is stationary, Amihud and Hurvich (2004) showed that the estimator is the second-order bias-correction and they also derived the asymptotic theory with a classical way so that the conventional t-type test can be used to test the predictability \( H_0 : \beta = 0 \). Moreover, Amihud and Hurvich (2004) and Amihud, Hurvich and Wang (2009) extended model (1) to multiple predictive regressions for stationary case.
Note that this approach in (3) is similar to the control function idea, which will be mentioned later.

(iii) When $\phi$ is close to 1, Lewellen (2004) suggested a conservative bias-adjusted estimator to $\beta$ as

$$\hat{\beta}_{bc_1} = \hat{\beta} + \hat{\rho}(0.9999 - \hat{\phi}).$$

Then, Lewellen (2004) argued that $\hat{\beta}_{bc_1}$ is the least biased estimator of $\beta$ when the true value of $\phi$ is very close to 1.
As mentioned above, the autoregressive parameter $\phi$ in $x_t$ might be very persistent.

- 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 (2005, Table 1), Campbell and Yogo (2006, Table 4), Jansson and Moreira (2006), Cai and Wang (2014), and the references therein.

- Indeed, Cai and Wang (2014) showed that $\hat{\beta}$ in (3) has the following asymptotic distribution

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

where

$$\xi_c = \frac{\tau_c - W(1) \int K_c dr}{k_c^2 - (\int K_c dr)^2} + \rho [\tau_c + \Omega_1] / k_c^2$$

with $k_c^2 = \int K_c^2$, $\tau_c = \int K_c dW$, and $\Omega_1 = \sum_{k=2}^{\infty} E(u_1 u_k)$. 
II: Nonstationarity

Here, \( W(\cdot) \) is a 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)

\[
dK_c(r) = c K_c(r)dr + dW(r).
\]

Then, it can be shown easily that \( K_c(r) \sim N(0, \sigma_c^2(r)) \), where
\[
\sigma_c^2(r) = \sigma_u^2 [\exp(2cr) - 1] / 2c
\]
and
\[
\int_0^1 K_c(r)dr \sim N(0, \varsigma(c)^2)
\]
with
\[
\varsigma(c)^2 = \sigma_u^2/c^2 + \sigma_u^2(e^{2c} - 4e^c + 3)/2c^3.
\]
Clearly, it is easy to obtain that
\[
\lim_{c \to 0} \varsigma(c)^2 = \sigma_u^2/3.
\]
Also, when \( c = 0 \), \( K_c(r) \) becomes \( W(r) \). See Cai and Wang (2014) for details.

- Based on this asymptotic theory, one can construct a test on testing predictability using a Monte Carlo simulation approach as suggested by Cai and Wang (2014) if \( c = 0 \).
II: Nonstationarity

- Interestingly, when $\theta$ in (1) is non-zero, Cai and Wang (2014) showed that
  \[ T^{3/2}(\hat{\beta} - \beta) \rightsquigarrow N(0, \sigma_{\theta,c}^2), \]
  where $\sigma_{\theta,c}^2$ depends on only $\theta$, $c$ and $\sigma_v^2$ with
  \[ \sigma_v^2 = \text{Var}(v_t) + 2 \sum_{t=2}^{\infty} \text{Cov}(v_1, v_t). \]
  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.

- Campbell and Yogo (2006) showed that the classical t-test $t_{\hat{\beta}_s}$ has the following limit distribution instead of normal
  \[ t_{\hat{\beta}_s} \rightsquigarrow \delta(\tau_c/k_c) + \sqrt{1 - \delta^2} Z, \]  
  where $Z$ is a standard normal random variable independent of $K_c(\cdot)$ and $W(\cdot)$. Therefore, the classical t-test fails so that it is over-rejected, which is verified by simulation studies in Campbell and Yogo (2006).
Robust Inferences

But, the asymptotic results, in particular, the limiting distribution, depend on $c$, which is not consistently estimable although its estimate has a limiting distribution. Therefore, it is inconvenient and infeasible in applications.

Recently, a series of researches investigate uniformly (robust) statistical inferences on predictive regressions in the sense that testing procedure for predictability is robust to general time series characteristics on 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 (2014), Kostakis, Magdalinos and Stamatogiannis (2015), Breitung and Demetrescu (2015), Demetrescu and Rodrigues (2016), Liu, Yang, Cai and Peng (2016), Cai, Chen and Liao (2017a, 2017b), and Yang, Liu, Cai and Peng (2017).
In particular, Campbell and Yogo (2006) proposed a new method, the so-called Q-test, based on the Bonferroni idea to construct a confidence interval for $\beta$ for each $\phi$ as follows:

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

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

As pointed out by Phillips and Lee (2013), this approach might lead to that empirical size may be substantially lower than nominal size resulting in a conservative test whose power is often negligible in ear local alternative to the null of no predictability. Another critical limitation is the difficulty of extending this approach to multivariate regressions involving several predictors.
II: Weighted Empirical Likelihood Approach

Zhu, Cai and Peng (2014), Liu, Yang, Cai and Peng (2016), and Yang, Liu, Cai and Peng (2017) proposed an empirical likelihood approach together with a weighted least squares idea to construct confidence interval for $\beta$.

If $\alpha$ in known, the empirical likelihood function for $\beta$ is

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

where $H_t(\beta) = (y_t - \alpha - \beta x_{t-1})x_{t-1}/\sqrt{1 + x_{t-1}^2}$ and the weighted score equation is $\sum_{t=2}^{T} H_t(\beta) = 0$. After applying the Lagrange multiplier technique, the log of the empirical likelihood leads to

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

where $\lambda = \lambda(\beta)$ satisfies $\sum_{t=2}^{T} H_t(\beta)/[1 + \lambda H_t(\beta_1)] = 0$. 
II: Weighted Empirical Likelihood Approach

- The key idea behind this approach is the weighted least squares as

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

where $W_t = 1/\sqrt{1 + x_{t-1}^2}$ and $x_{t-1} W_t = x_{t-1}/\sqrt{1 + x_{t-1}^2} = O_p(1)$

no matter that $x_t$ is $I(0)$ or $NI(1)$.

- Zhu, Cai and Peng (2014) showed that under some regularity conditions, $l(\beta_0)$ converges in distribution to a chi-square distribution with one degree of freedom, where $\beta_0$ is the true value of $\beta$. Therefore, one can construct confidence interval with the significance level $b$ as

$CI_b = \{\beta: l(\beta) \leq \chi^2_{1,b}\}$.

Rejecting $H_0: \beta = 0$ is equivalent to checking if $0 \notin CI_b$. 

II: Weighted Empirical Likelihood Approach

- Liu, Yang, Cai and Peng (2016) extended this approach to bivariate regression but it still loses efficiency.

- This approach loses efficiency when $\alpha$ is unknown. Also, like Bonferroni procedure, it is difficult to extend this approach to a general multivariate regression.
Data-Filtering (Instrumental Variable, termed as IVX) method of Phillips and Magdalinos (2009). To diminish the effect of endogeneity in the regression, the main idea of IVX is to construct an instrument variable which is less persistent than $x_{t-1}$ based on $\Delta x_{t-1}$. That is

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

with $\phi_z = 1 + c_z/T^\eta$, $\eta \in (0, 1)$, and $c_z < 0$. Clearly, if $x_t$ is I(1) or NI(1), $z_t$ is less persistent than $x_t$. The reason of choosing the less persistent $z_{t-1}$ is to ensure validity of chi-squared test limit theory at the cost of slow convergence rate. Then, one can use $z_t$ as an instrumental variable to estimate $\beta$ so that the IV estimation of $\beta$ can be obtained easily, denoted by $\hat{\beta}_{IV}$. Phillips and Magdalinos (2009) showed that the convergence rate for $\hat{\beta}_{IV}$ is $T^{(1+\eta)/2}$ instead of $T$; see Phillips and Magdalinos (2009) for details on theory.
Recently, Demetrescu and Rodrigues (2016) considered model (3) using the IVX approach.

How to choose the tuning parameters $c_z$ and $\eta$ and $u_{zt}$ in practice? To gain efficiency, one wishes $\eta$ very close to 1. Indeed, Phillips and Lee (2013), Kostakis, Magdalinos and Stamatogiannis (2015) suggested empirically choosing $\eta = 0.95$, $c_z = -1$, and $u_{zt} = \Delta x_{t-1}$. But, there is no theory on how to choose these tuning parameters.

For the IVX approach, we have the following observations:
Kostakis, Magdalinos and Stamatogiannis (2015) found that the power of the test depends on the choice of $\eta$. In other words, mis-choice of $\eta$ might lose power.

One can see that the choice of the tuning parameters in Kostakis et al. (2015) is not an easy task in practice especially when an intercept is included. Therefore, a data-driven method is preferred.

When $x_t$ is I(0), the IVX does not work well; see Phillips and Magdalinos (2009).

When $u_{zt} = \Delta x_{t-1}$, $z_{t-1}$ might not be orthogonal to $u_t$.

Is it possible to choose $z_t$ to be as the same persistency as $x_t$ if $x_t$ is persistent? ($\eta = 1$)?
To improve the local power of the IVX based tests, Demetrescu (2014) and Yang, Liu, Cai and Peng (2017) proposed adding the lagged variables into the model so that the model becomes dynamic, say

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

Clearly, if \( \beta = 0 \), the above model is the AR(1) for \( y_t \). Then, Demetrescu (2014) argued 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) employed the weighted empirical likelihood approach. Both showed empirically that indeed, the dynamic approach can improve the local power than the IVX method.
Variable Addition approach is called control function method in Elliott (2011) and Phillips and Lee (2013) as mentioned earlier. Elliott (2011) and Breitung and Demetrescu (2015) proposed adding an additional variable $z_t$ into the model as

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

where $z_{t-1}$ is an augmented variable, which is an augmented predictive regression. If $z_{t-1}$ is taken to be $v_t$ in model (3), then model (5) becomes to the linear projection approach in (3).

How to choose $z_{t-1}$ in practice? There have been some approaches proposed in recent years. For example,
V: Variable Addition (VA) Approach

- Elliott (2011) proposed 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 was shown to have better size control with higher local power than the method in Campbell and Yogo (2006).

- Liu, Yang, Cai and Peng (2016) proposed taking $z_{t-1}$ to be $x_{t-2}$ so that the above model in (5) becomes to

$$y_t = \alpha + \beta x_{t-1} + \gamma x_{t-2} + u_t = \alpha + \beta \Delta x_{t-1} + \beta_1 x_{t-2} + u_t, \quad (6)$$

where $\beta_1 = \beta + \gamma$, and then, they used the weighted empirical likelihood approach to construct confidence interval for $\beta$ and $\beta_1$ regardless of $\phi$. One of the nice properties of the above model in (6) is that even if $\beta_1 = 0$, $y_t$ can be still predictable using $\Delta x_{t-1}$.

- Breitung and Demetrescu (2015) suggested to define $z_{t-1}$ to be less persistent than $x_{t-1}$ to satisfy the following conditions:
Type I: $z_t$ is less persistent than $x_t$ so that $z_t$ satisfies the following assumptions:

$$V_{T,z} = \sum_{t=2}^{T} z_{t-1}^2 / T^{1+2\delta} \rightarrow V_z \text{ and } V_{T,zu} = \sum_{t=2}^{T} z_{t-1}^2 u_t^2 / T^{1+2\delta} \rightarrow V_{zu}$$

where $V_z$ and $V_{zu}$ are positive and bounded,

$$\frac{1}{T^{1.5+\delta}} \sum_{t=2}^{T} z_{t-1} x_{t-1} \rightarrow 0 \text{ and } \frac{1}{V_{zu}^{1/2}} \sum_{t=2}^{T} z_{t-1} u_t \rightarrow Z,$$

where $Z$ is the standard normal random variable.
V: VA Approach

Here are some examples:

- Stationary Variable: $z_t$ is iid $(0, \sigma^2_v)$ with $\delta = 1/2$ and $V_z = \sigma^2_v$.
- Bounded Deterministic Time Tread: $z_t = (t/T)^\kappa - 1/(1 + \kappa)$, where $\kappa > 0$, $\delta = 1/2$, and $V_z = \kappa^2/(2\kappa + 1)(\kappa + 1)^2$.

... 

**Type II:** Cai, Chen and Liao (2017a) argued that $z_t$ should be at the same order of persistency as $x_t$, which is type II. HOW?

For example, $z_t$ has the same degree of persistency as $x_t$. For example, nearly integrated variable: $z_t = (1 + \frac{d}{T})z_{t-1} + \epsilon_t$, where $\epsilon \sim N(0, 1)$, $\delta = 1$ and $V_z = \int K^2_d(r)dr$. 
Cai, Chen and Liao (2017a) proposed a new method as follows:

\[ y_t = \alpha + \beta x_{t-1} + u_t = \beta_0 + \beta_1 x^*_t + \beta_2 z_{t-1} + u_t = \beta^\top X_t + u_t, \]  

(7)

where \( x^*_t = x_{t-1} - z_{t-1} \), \( \beta_0 = \alpha \), \( \beta_1 = \beta \), \( \beta_2 = \beta \), \( \beta^\top = (\beta_0, \beta_1, \beta_2) \), \( X_t^\top = (1, x_{t-1}, z_{t-1}) \), and \( z_{t-1} \) is an additional variable generated exogenously or endogenously. One can obtain the OLS estimate of \( \beta \), denoted by \( \hat{\beta} \). Since both \( \beta_1 \) and \( \beta \) are same as \( \beta \), Cai, Chen and Liao (2017a) suggested using a weighted approach to estimate \( \beta \) as

\[ \hat{\beta}_w = \frac{W_1}{W_1 + W_2} \hat{\beta}_1 + \frac{W_2}{W_1 + W_2} \hat{\beta}_2, \]
VII: Double-Weighted Robust Approach

where

$$W_1 = \sum_{t=2}^{T} x_{t-1}^* z_{t-1} / T^2 - \sum_{t=2}^{T} x_{t-1}^* \sum_{t=2}^{T} z_{t-1} / T^3$$

and

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

Cai, Chen and Liao (2017a) showed that

$$\left( W_1 + W_2 \right) T \left( \hat{\beta}_w - \beta \right) / \sqrt{W_2 \hat{\sigma}_0^2} \xrightarrow{d} N(0, 1), \quad (8)$$

where $\hat{\sigma}_0^2 = \sum_{t=1}^{T} \hat{u}_t^2 / T$ and $\hat{u}_t$ is the residual from (7). So, one can use this result to construct a test statistic for testing $H_0 : \beta = 0$, denoted by

$$Q_w = (W_1 + W_2) T \hat{\beta}_w / \sqrt{W_2 \hat{\sigma}_0^2}. \quad (9)$$
Robust Inferences

VII: Double-Weighted Robust Approach

Also, Cai, Chen and Liao (2017a) showed 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 shows that $Q_{w}$ is a consistent test.

Furthermore, Cai, Chen and Liao (2017a) showed that similar to IVX and VA approaches, $Q_{w}$ in (9) might not be consistent when $x_{t-1}$ is I(0).

Can we construct a test statistic to accommodate both stationary and nonstationary cases?
The answer is YES! HOW?. Indeed, Cai, Chen and Liao (2017a) proposed the following test statistic

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

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

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

for any $0 < \zeta < 1$, which can characterize the degree of nonstationarity. Here, $t_{\hat{\beta}_s}$ is the t-statistic for the stationary $x_{t-1}$ and $\hat{\beta}_s$ is the OLS estimate of $\beta$ for classical predictive regression.
Cai, Chen and Liao (2017a) showed that under $H_0$, $Q_{RW} = Q_w + o_p(1)$ for I(1) or NI(1) $x_{t-1}$ and $Q_{RW} = t_{\hat{\beta}_s} + o_p(1)$ for I(0) $x_{t-1}$. Therefore, $Q_{RW} \xrightarrow{d} N(0,1)$ under $H_0$ for all cases.

Also, they derived the asymptotic distribution of $Q_{RW}$ under local alternative.

Question: How to choose optimally the aforementioned three weights? $W_1$ and $W_2$ in $\hat{\beta}_w$ and $W_*$ in $Q_{RW}$.
In 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). It is well known in the literature that economic or financial data are not homoskedastic. Therefore, we need to accommodate heteroskedasticity.

Park (2012), Han and Zhang (2012) and Choi, Jacewitz and Park (2016) considered model (1) with heteroskedasticity as $u_t = \sigma_t e_t$, where $\sigma_t = \sigma(z_t/\sqrt{T})$ with some nonstationary regressor $z_t$ or $\sigma_t = \sigma(t/\sqrt{T})$. Also, Choi, Jacewitz and Park (2016) considered a robust test based on Cauchy estimator with no constant term in (1) with stationary $\sigma_t$.

It is well documented that a quantile regression (QR) approach can capture heteroskedasticity automatically.
Quantile Predictive Models

- Xiao (2009) considered the quantile regression with nonstationary regressors

\[ q_\tau(x_t) = \alpha_\tau + \beta_\tau x_t, \tag{10} \]

where \( P(y_t \leq q_\tau(x_t)|x_t) = \tau \) and \( x_t \) is nonstationary.

- First, Lee (2016) generalized (10) to the predictive quantile setting and showed that the t-test for (10) has a result similar to (4) for mean model as in Campbell and Yogo (2006). Then, Lee (2016) proposed an IVX-QR method using the IVX idea as mentioned above to estimate parameters in (10) and derived the asymptotic theory.

- Fan and Lee (2017) provided a valid and easy-to-use inference procedure in predictive quantile framework with conditional heteroskedasticity innovations and suggested using the IVX-QR method and the moving blockwise Bootstrap (MBB) of Künsch (1999).
Quantile Predictive Models

Recently, Cai, Chen and Liao (2017b) considered the predictive quantile model as in (10) and proposed using the weighted variable addition approach to testing predictability as follows. One can rewrite the model (10) 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 (control function 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

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

To achieve the optimal convergence rate \( T \), Cai, Chen and Liao (2017b) proposed using 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 showed that \( z_{t-1} \) is I(0) if \( x_{t-1} \) is I(0), and \( z_{t-1} \Rightarrow \text{sign}(K_c(r)) \) as \( t \to \infty \) if \( x_{t-1} \) is I(1) or NI(1), where \( \text{sign}(\cdot) \) is the sign function. Note that \( z_{t-1} \) here can be generalized to multivariate case as \( z_{t-1,j} = x_{t-1,j}/\sqrt{1 + x_{t-1,j}^2} \) for \( j \geq 1 \) if \( x_{t-1} \) is a multivariate variable.
Quantile Predictive Models

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

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

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

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

where $f_{v_\tau}(0) > 0$ is a normalization constant and unknown and $\hat{f}_{v_\tau}(0)$ is an estimate of $f_{v_\tau}(0)$, 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, $Q_\tau^w$ is consistent for all cases. Furthermore, they show that $Q_\tau^w$ 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.
In preceding work, parameters in predictive regressions are assumed to be stable; that is, coefficients are constant (do not change over time).

Because we are dealing with time series data in a long time period, it is reasonable to expect that parameters in predictive regression models may experience structural changes at some unknown dates.

Viceira (1997), Paye and Timmermann (2006), and Rapach and Wohar (2006) found strong evidence of instability in predictive regression models. However, they did not explain how to test predictability after detecting and estimating the break dates.

Lettau and Nieuwerburgh (2008) focused on level shifts in the predictor variables and explain that the forecasting relationship be instable unless such shifts are included in the analysis.
Cai, Wang and Wang (2015) considered a model with coefficients changing smoothly with time as
\[ y_t = \alpha_t + \beta_t x_{t-1} + u_t, \]
where both \( \alpha_t \) and \( \beta_t \) are smooth functions of time, and then they proposed a nonparametric testing procedure to test whether the time-varying coefficients are indeed changing with time. That is, \( H_0: \alpha_t = \alpha \) & \( \beta_t = \beta \). They found that indeed, the coefficients are instable for testing predictability of asset returns based on a real example.

Now, the question is how to specify the form of time-varying coefficients? That is, how do the coefficients change with time?
Cai and Chang (2018) assumed that the coefficients are piecewise constant: structural changes. Therefore, they considered the following model

\[ y_t = \alpha_t + \beta_t x_{t-1} + u_t, \]

where both \( \alpha_t \) and \( \beta_t \) are piecewise constant, say with one break (easy to consider multiple breaks) \( \alpha_t = \alpha_1 I(t < T_y) + \alpha_2 I(t \geq T_y) \) and \( \beta_t = \beta_1 I(t < T_y) + \beta_2 I(t \geq T_y) \) with unknown \( T_y \). Also, \( x_{t-1} \) might be allowed to have a structural break with/without the same change point. They used the weighted empirical likelihood approach as mentioned earlier to testing predictability in two time periods. That is, \( H_0 : \beta_1 = 0 \) and/or \( H_0 : \beta_2 = 0 \).

Of interest here is to test \( H_0 : \beta_1 = \beta_2 \) (the well known Chow test).
Nonlinear Models

A threshold form of model in (1) was considered by Ganzalo and Pitarakis (2017) as

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

(11)

where $q_t$ is a stationary variable (say, some variable for proxying business cycle), $\alpha(q_{t-1}) = \alpha_1 I(q_{t-1} \leq \gamma) + \alpha_2 I(q_{t-1} < \gamma)$ and $\beta(q_{t-1}) = \beta_1 I(q_{t-1} \leq \gamma) + \beta_2 I(q_{t-1} < \gamma)$, which is a nonlinear model. They used the Wald type test to testing predictability as $H_0 : \beta_1 = \beta_2 = 0$ or $H_0 : \alpha_1 = \alpha_2$.

Note that if both $\alpha(\cdot)$ and $\beta(\cdot)$ are smoothing function, model (11) reduces to the model in Cai, Li and Park (2009) and Xiao (2009) 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 $I(1)$.
Nonparametric Tests

- One of the possible ways to remove the embedded endogeneity is to consider a nonparametric model as

\[ y_t = m(x_{t-1}) + u_t, \]  

(12)

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.
Nonparametric Tests

- **Question:** How to test $H_0 : m(x) = m_0(x, \beta)$? where $m_0(\cdot, \cdot)$ is a known function, say a threshold function or other types, to see if some financial theory holds.

- Cai and Wu (2013) considered this kind of test and proposed a $L_2$ type test statistic. That is,

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

for some weighting function to avoid a random denominator, which can be simplified as a **U-statistic** as

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

where $h$ is the bandwidth and $K(\cdot)$ is the kernel function used to obtain $\hat{m}(x)$. 

(Cai, Zongwu (Department of Economics, University of Kansas and WISE, Xiamen University E-mail: caiz@ku.edu) Predictive Regressions Present to Department of Economics, Vanderbilt University, April 4)
Nonparametric Models

- Cai and Wu (2013) derived the asymptotic distribution of the proposed test statistic under $H_0$ and showed that the test statistic $J_T$ diverges to $\infty$ under the alternative hypothesis. In particular, of interest is that they derived the limiting result for a U-statistic involving nonstationary variables. This result is new.

- Recently, Kasparis, Andreou and Phillips (2015) considered 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 proposed a naive test as follows:

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

where $A(\cdot)$ is a self-normalized function. They showed that $\hat{F}_{\text{sum}} \xrightarrow{d} \chi^2_m$ and $\hat{F}_{\text{max}} \xrightarrow{d} Y$ for some random variable $Y$. But, the proposed test depends on the choice of $\{z_j\}$. 

Cai, Zongwu (Department of Economics, University of Kansas and WISE, Xiamen University E-mail: caiz@ku.edu)
Further, by assuming that $\mu_m = 0$, Juhl (2014) considered 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 (13) involving $\{\hat{u}_t\}$, the test statistic proposed in Juhl (2014) is given as follows:

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

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

Future Work

1. How about the model with $x_t$ having structural changes too? Easy? Indeed, Cai and Chang (2018) did consider this case with known break point.

2. How about a test similar to the well known Chow test? That is to use the whole sample to form a test statistic for testing $\beta_1 = \beta_2$ when regressors are nonstationary.

3. How to develop a new method to avoid using a part of sample when the model contains an intercept to improve the efficiency?

4. How about nonparametric or semiparametric models or other types of models, say quantile models? How about models with capturing heteroskedasticity?
5. How to generalize predictive model to a more general setting for predicting global asset returns (panel or longitudinal data)? See Hjalmarsson (2010, JFQA). The key issue here is how to deal with the cross-sectional dependence or/and structural changes or/and time-varying setting.

6. For nonparametric models, how to conduct nonparametric tests, especially under nonparametric quantile framework? How about using nonparametric quantile to construct directly prediction interval instead of conducting tests?

7. How about the model with $x_t$ satisfying $\phi = 1 + c/T^\zeta$ with $0 \leq \zeta < 1$?

8. More econometric/statistical issues ....?
9. Finally, I would like to mention the related issues in other fields in economics. For example, how about this kind of models applied to other fields such as macroeconomics? to study bubbles by assuming that $\phi = 1 + c/T^\zeta$ with $c > 0$ and $0 \leq \zeta \leq 1$ so that $x_t$ is explosive; see several papers by Philipps, Shi and Yu (2015) on characterizing the bubbles in macroeconomics.
Thank You!