BE990-8-SP - Research Methods in Financial Econometrics

Professor Robert Taylor
Room EBS.3.12, Essex Business School
University of Essex

e-mail: robert.taylor@essex.ac.uk

Spring Term 2023

Topic 3: Predictive Regressions for Returns

3.1 Introduction

- Testing for the predictability of asset returns has been the subject of numerous studies in the applied economics and finance literature, assessing the predictive strength of a range of candidate predictor variables, including valuation ratios, interest rates and other financial and macroeconomic variables.
- Fama (1981) examines the predictability of stock returns using various candidate predictors including interest rates, industrial production, GNP and capital stock and expenditure.
- Campbell and Yogo (2006) [CY, hereafter] consider candidate predictors that include the dividend and earnings price ratios, the three-month T-bill rate and the long-short yield spread.
- The standard approaches to determining whether returns are predictable are based on a simple linear predictive regression model with a constant and lagged putative predictor, which we denote as x_{t-1}, with slope coefficient β.

- A common finding in empirical studies into return predictability is that the putative predictor is often both strongly persistent and endogenous, with a non-zero (often strongly negative) correlation between the errors in the predictive regression and the innovations driving the predictor process; see, *inter alia*, CY and Welch and Goyal (2008).
- In this situation Cavanagh *et al.* (1995) [CES] show that the standard t test on the estimate of β suffers from severe size distortions that are a function of both the degree of persistence and the endogeneity of the predictor.
- Let's investigate this a little further ...

Is there any Predictability in the Equity Premium?

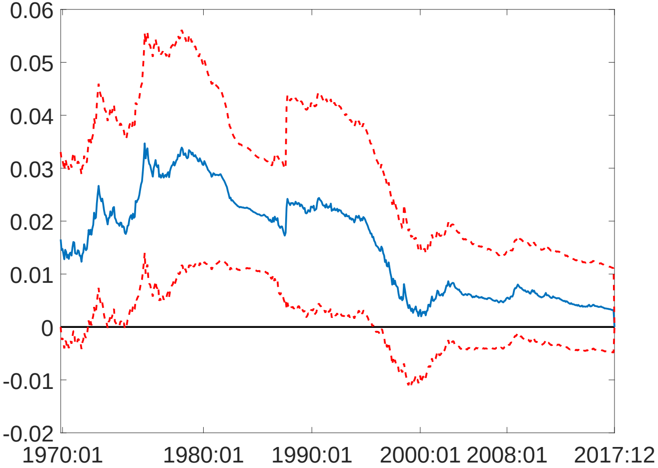
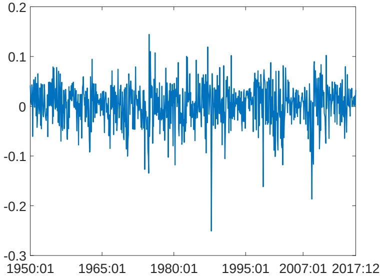
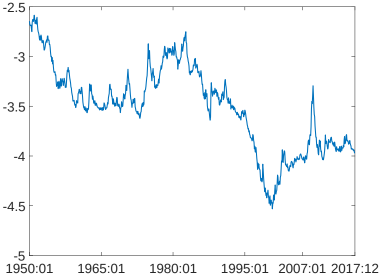


Figure 1 - Dividend yield: Forward Recursive IV regression estimates and pointwise CIs, 1950-2017 (Goyal-Welch 2008 updated monthly data).

... what about the persistence of the predictor?





The equity premium looks very mean reverting etc (almost noise), but the dividend yield looks strongly persistent (usual ADF test has p-value of 0.41).

The Basic Predictive Regression Set-up

Consider, for illustration, the very basic predictive regression

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

where

$$(x_t - \mu_x) \quad = \quad \rho(x_{t-1} - \mu_x) + e_t, \tag{2}$$

with $(u_t, e_t)' \sim iid(0, \mathbf{\Sigma})$ where

$$\mathbf{\Sigma} = \mathrm{E}\left(\begin{pmatrix} u_t \\ e_t \end{pmatrix}\begin{pmatrix} u_t & e_t \end{pmatrix}\right) = \begin{pmatrix} \sigma_u^2 & \sigma_{ue} \\ \sigma_{ue} & \sigma_e^2 \end{pmatrix}.$$

Null hypothesis: x_{t-1} does not predict y_t, i.e.

$$H_0: \quad \beta = 0\,.$$

Yet, even in this simplest setup...

Endogeneity and Strong Persistence

- The regression is subject to an endogeneity bias. This occurs where the shocks u_t and e_t are correlated. For the EP-DY data above the endogeneity correlation, $\delta := \sigma_{ue}/\sigma_u\sigma_e$, is estimated to be $\hat{\delta} = -0.98$.
- The regressor x_{t-1} is strongly autocorrelated (as we saw a DF test cannot reject a unit root in the predictor).

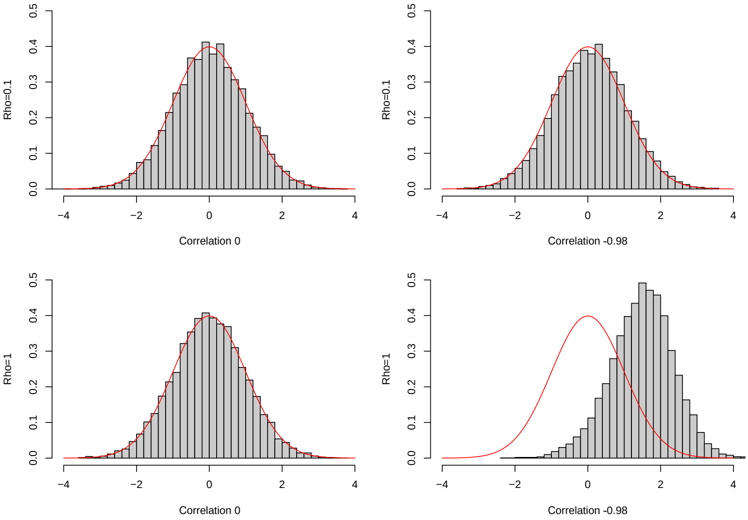
Under endogeneity and high persistence (near integration, $\rho = 1 - {}^{c}\!/_{T}$),

- the OLS estimator is biased and
- the *t*-statistic has a non-normal limiting distribution.

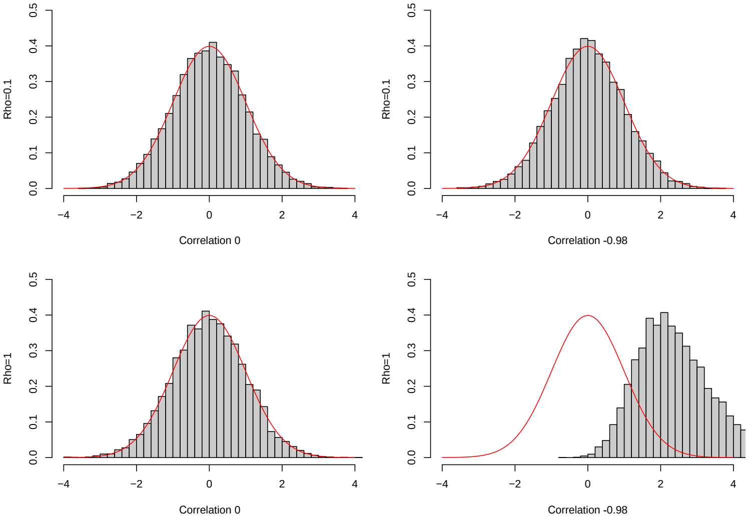
See, among others, CES, Stambaugh (1999), CY.

No problem when regressor is weakly persistent.

OLS t-statistics, $T = 305$



OLS t-statistics (White se's), $T = 305$. Volatility of both shocks 3 times higher in the first 20% of the sample



- In the near-unit root case, the limiting null distributions of the standard t-statistic for testing $\beta = 0$ in (1) depends on both δ and c, whenever neither is zero.
- If ρ were known, one could employ GLS estimation. For unknown ρ, there are a number of 'solutions' proposed in the literature:
  - Bonferroni - Cavanagh *et al.* (1995), Campbell and Yogo (2006)
  - Restricted log-likelihood - Jansson and Moreira (2006, Econometrica)
  - Almost optimal tests - Elliott *et al.* (2015, Econometrica)
  - Generic IV estimation - Breitung and Demetrescu (2015, Jnl Econometrics)
  - Extended Instrumental Variables [IVX] method of Kostakis *et al.* (2015, Review of Financial Studies) [KMS]

- Arguably the most commonly employed of these is the Bonferroni Q test proposed by CY. The method in CY is only valid for strongly persistent predictors. It can only be used in the case of a single predictor variable.
- We will review the CY method first before looking at approaches based on instrumental variable estimation proposed in *inter alia*, Kostakis *et al.* (2015) and Breitung and Demetrescu (2015). These tests are valid regardless of whether the predictor is weakly or strongly persistent and deliver standard limiting null distributions in either case. Unlike the CY method, these methods can also be implemented with multiple predictors.
- Monte Carlo simulations comparing these approaches and also the test of Elliott *et al.* (2015) can be found in Harvey *et al.* (2021). Some further predictability tests are also proposed in Harvey *et al.* (2021).

3.2 The Campbell-Yogo Q tests

- The Bonferroni Q test procedure of CY is based around computing a confidence interval for β using what is essentially a t-statistic obtained from the predictive regression augmented by the covariate $(x_t - \rho x_{t-1})$.
- When x_t is (near)-integrated, the local offset c in ρ is not consistently estimable, rendering the confidence interval calculation infeasible in practice.
- To overcome this problem, CY use a Bonferroni procedure, originally proposed in CES, whereby a confidence interval for ρ is first constructed by inverting the quasi-GLS demeaned Dickey-Fuller (ADF-GLS) unit root test of Elliott *et al.* (1996), that we reviewed in Topic 1, applied to the predictor, x_t.

- The bounds associated with this confidence interval for ρ are then used to deliver a feasible confidence interval for β. This can then be used to test the null hypothesis that β is equal to some specified value, usually zero.
- For strongly persistent predictors, CY show that the Bonferroni Q test procedure has well controlled size and good power properties regardless of the value of the non-centrality parameter c and the degree of endogeneity of the predictor.

CY's Predictive Regression Model

- CY consider the following predictive regression model

$$y_t = \alpha + \beta x_{t-1} + u_t, \qquad t = 1, ..., T \tag{3}$$

where y_t denotes the (excess) return in period t, and x_{t-1} denotes a (putative) predictor observed at time $t-1$.

- The DGP for x_t is given by

$$\begin{aligned} x_t &= \mu + w_t, \qquad t = 0, ..., T && (4) \\ w_t &= \rho w_{t-1} + v_t, \qquad t = 1, ..., T. && (5) \end{aligned}$$

- CY focus on testing the null hypothesis of no predictability, $H_0 : \beta = 0$, against the right-sided alternative (positive predictability) $H_1 : \beta > 0$.

Assumption 1

Assume that $\psi(L)v_t = e_t$ *where* $\psi(L) := \sum_{i=0}^{p-1} \psi_i L^i$ *with* $\psi_0 = 1$ *and* $\psi(1) \neq 0$, *with the roots of* $\psi(L)$ *assumed to be less than one in absolute value.*

Assume that $\xi_t := (u_t, e_t)'$ *is a bivariate martingale difference sequence [MDS] satisfying the following conditions: (i)* $E[\xi_t \xi_t'] = \begin{bmatrix} \sigma_u^2 & \sigma_{ue} \\ \sigma_{ue} & \sigma_e^2 \end{bmatrix}$, *(ii)* $\sup_t \, E[u_t^4] < \infty$, *and (iii)* $\sup_t \, E[e_t^4] < \infty$.

Define $\omega_v^2 := \lim_{T\to\infty} T^{-1} E(\sum_{t=1}^{T} v_t)^2 = \sigma_e^2/\psi(1)^2$ *to be the long run variance of the error process* $\{v_t\}$, *and* $\delta := \sigma_{ue}/\sigma_u\sigma_e$ *as the correlation between the innovations* $\{u_t\}$ *and* $\{e_t\}$.

- The assumptions placed on ξ_t allow the sequence of innovations to be conditionally heteroskedastic (e.g. GARCH). The MDS aspect of Assumption 1 implies the standard assumption made in the literature that the unpredictable component of returns, u_t, is serially uncorrelated.
- Assumption 1 allows the dynamics of the predictor variable to be captured by an $AR(p)$, with the degree of persistence of the predictor (strong or weak) controlled by the parameter ρ in (5).
- CY assume that the predictor $\{x_t\}$ is strongly persistent, with the autoregressive parameter ρ in (5) given by $\rho = 1 - c/T$ with c a finite non-zero constant. In the operational version of their statistic they impose that $c \in [-5, 50]$.
- The predictor is weakly persistent if $|\rho| < 1$, independent of T.
- CY assume that the initial condition is given by $w_0 = o_p(T^{1/2})$, so that it is asymptotically negligible.

CY's Q Test

- CY propose testing for predictability based on Bonferroni procedures that make use of confidence intervals for the unknown autoregressive parameter $\rho = 1 - c/T$, with these confidence intervals constructed by inverting unit root tests.
- In the infeasible case where ρ is assumed known, CY derive an optimal test for H_0 against H_1 which rejects for large values of the likelihood ratio statistic:

$$Q(0,\rho) \quad := \quad \frac{\sum_{t=1}^{T} x_{t-1}^{\mu}\left[y_t - \frac{\sigma_{ue}}{\sigma_e \omega_v}(x_t - \rho x_{t-1})\right] + \frac{T}{2}\frac{\sigma_{ue}}{\sigma_e \omega_v}(\omega_v^2 - \sigma_v^2)}{\sqrt{\sigma_u^2(1-\delta^2)\sum_{t=1}^{T}(x_{t-1}^{\mu})^2}}$$

  where $x_{t-1}^{\mu} := x_{t-1} - T^{-1}\sum_{s=1}^{T} x_{s-1}$, and σ_v^2 denotes the short run variance of v_t.
- For known ρ, $Q(0,\rho)$ has a standard normal limiting null distribution.

- A $100(1-\alpha)\%$ confidence interval for β can then be derived based on the quantity $Q(0,\rho)$ and is given by $[\underline{\beta}(\rho,\alpha), \overline{\beta}(\rho,\alpha)]$ where

$$\begin{aligned} \underline{\beta}(\rho,\alpha) &= \{Q(0,\rho) + z_{\alpha/2}\} s\sqrt{1-\delta^2}, \\ \overline{\beta}(\rho,\alpha) &= \{Q(0,\rho) - z_{\alpha/2}\} s\sqrt{1-\delta^2} \end{aligned}$$

  where $s^2 := \sigma_u^2 / \sum_{t=1}^{T} (x_{t-1}^{\mu})^2$ and $z_{\alpha/2}$ denotes the $\alpha/2$ quantile of the standard normal distribution.
- But in practice $\rho = 1 - c/T$ is unknown and c cannot be consistently estimated, so this approach infeasible.
- In order to obtain an asymptotically size controlled test with good power across different values of c, CY propose using a confidence interval for ρ obtained by inverting the quasi-GLS demeaned ADF-GLS t-ratio based unit root test of Elliott *et al.* (1996) applied to x_t (allowing for $p-1$ lagged difference terms, as per Assumption 1), using pre-computed (asymptotic) confidence belts.

- To prevent the resulting confidence interval for β from suffering excess coverage, CY further propose a refinement whereby the significance level used to obtain the confidence interval for ρ is adapted to upper and lower bounds separately, and also according to the value of δ.
- Values of this significance level are chosen numerically to minimise over-coverage associated with the confidence interval for β, while ensuring that the overall Bonferroni test size does not exceed a chosen level across a specified range of c.
- Denoting the significance levels for the lower and upper confidence bounds for ρ by $\underline{\alpha}_1$ and $\overline{\alpha}_1$, respectively, the confidence interval for ρ can be written as $[\underline{\rho}(\underline{\alpha}_1), \overline{\rho}(\overline{\alpha}_1)]$, and the resulting $100(1-\alpha_2)\%$ confidence interval for β is obtained as $[\underline{\beta}(\overline{\rho}(\overline{\alpha}_1), \alpha_2), \overline{\beta}(\underline{\rho}(\underline{\alpha}_1), \alpha_2)]$ where

$$
\begin{aligned}
\underline{\beta}(\overline{\rho}(\overline{\alpha}_1), \alpha_2) &= \{Q(0, \overline{\rho}(\overline{\alpha}_1)) + z_{\alpha_2/2}\} s\sqrt{1-\delta^2}, \\
\overline{\beta}(\underline{\rho}(\underline{\alpha}_1), \alpha_2) &= \{Q(0, \underline{\rho}(\underline{\alpha}_1)) - z_{\alpha_2/2}\} s\sqrt{1-\delta^2}.
\end{aligned}
$$

- For a given value of δ, the resulting one-sided tests have an asymptotic size of exactly $\alpha_2/2$ for some value of c while remaining slightly undersized for other values of c. Consequently, two-sided tests will have size of at most α_2 across the specified range of c. CY calibrate this procedure by fixing $\alpha_2 = 0.1$ and consider $c \in [-5, 50]$ such that their resulting one-sided tests have a maximum (asymptotic) size of 5%. The appropriate values of $\underline{\alpha}_1$ and $\overline{\alpha}_1$ are given in Table 2 (p.40) of CY.
- These values of $\underline{\alpha}_1$ and $\overline{\alpha}_1$ are only provided for $\delta < 0$. For $\delta > 0$, CY note that replacing x_t in (3) with $-x_t$ flips the sign of both β and δ. Therefore, an equivalent right (left) tailed test for predictability when $\delta > 0$ can be performed as a left (right) tailed test for predictability based on (3) with x_t replaced by $-x_t$ using the values of $\underline{\alpha}_1$ and $\overline{\alpha}_1$ appropriate for a negative value of δ.
- In practice δ is unknown but can be consistently estimated. Use the values from Table 2 corresponding to this estimate.

- For full details on the practical implementation of the CY procedure, including consistent estimation of the parameters σ_e, σ_u, σ_v, σ_{ue}, ω_v and δ, implementation of the ADF-GLS unit root tests, and the pre-computed confidence belts, see CY and the corresponding supplementary material to CY available at

  https://scholar.harvard.edu/campbell/publications/implementing-econometric-methods-efficient-tests-stock-return-predictability-0

3.3 Instrumental Variable based Approaches

- We will next review two alternative approaches to the approach taken in CY. While CY use a Bonferroni-based correction of likelihood ratio-type statistics to control asymptotic size in the presence of endogenity and strong persistence, an alternative approach is to modify the estimation method used to estimate the parameters of the predictive regression in (3).
- As we have seen the problem is one of an endogenous strongly persistent predictor. From standard econometric theory we know that instrumental variable estimation is the classic solution to problems of endogeneity.
- Kostakis *et al.* (2015) [KMS] propose an exactly-identified IV solution, while Breitung and Demetrescu (2015) [BD] propose an over-identified two-stage least squares [2SLS] approach. Both are valid for both strongly and weakly persistent predictors, and both can be implemented with multiple predictors. We will review them in turn, starting with BD.

Uncertain Persistence

- For both the KMS and BD methods, we can now make the following much weaker assumption on ρ in (5).

Assumption 2

One of the following two conditions is assumed to hold:

1. ***Weakly persistent predictors:*** *The autoregressive parameter ρ in (5) is fixed and bounded away from unity, $|\rho| < 1$.*
2. ***Strongly persistent predictors:*** *The autoregressive parameter ρ in (5) is local-to-unity with $\rho := 1 - \frac{c}{T}$ where c is a fixed non-negative constant.*

- In fact the predictor can also be allowed to be *mildly integrated* which is a class of persistence between strongly and weakly persistent.

IV Estimation

- BD discuss IV estimation using a vector $\boldsymbol{z}_t$ of instruments. They recommend basing tests on a 2SLS-based estimate, using two instruments per predictor (over-identified):

    1. a so-called type-I instrument, $z_{I,t}$, that is less persistent than x_t,
    2. a so-called type-II instrument, $z_{II,t}$, that is persistent yet exogenous.

- In contrast KMS's IV estimation uses a single instrument for each predictor (just-identified).

Instruments in practice

BD suggest using a vector of two instruments per predictor e.g.:

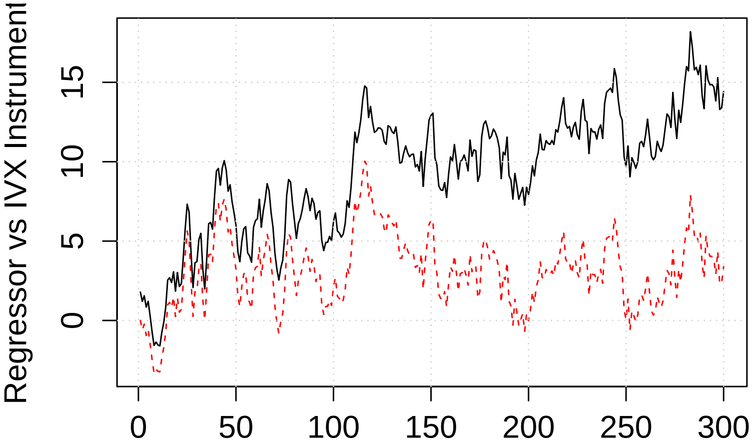
$$z_{I,0} = 0, \ \ z_{I,t} := \sum_{j=0}^{t-1} \varrho^j \Delta x_{t-j}, \ t = 1, ..., T$$

where $\varrho := 1 - \frac{a}{T^\gamma}$ with $\gamma \in (0, 1)$ and $a \geq 0$, and

$$z_{II,t} := \sin\left(\frac{\pi t}{2T}\right).$$

- $z_{I,t}$ is the IVX instrument of KMS.
- $z_{II,t}$ essentially delivers a spurious correlation with the strongly persistent x_t.
- Both satisfy generic high-level conditions for the instruments to be valid which are given in BD. BD discuss other possible choices.

The IVX trick applied to a random walk - using $a = 1$, $\gamma = 0.95$ as in Kostakis *et al.* (2015)



The 2SLS Test of BD

- For the case of a single predictor, the 2SLS t-ratio of BD (with Eicker-White standard errors) is given by:

$$t_\beta := \frac{\boldsymbol{A}_T' \mathbf{B}_T^{-1} \boldsymbol{C}_T}{\sqrt{\boldsymbol{A}_T' \mathbf{B}_T^{-1} \mathbf{D}_T \mathbf{B}_T^{-1} \boldsymbol{A}_T}} \tag{6}$$

where

$$\begin{aligned} \boldsymbol{A}_T &:= \sum_{t=1}^{T} \hat{x}_{t-1} \hat{\boldsymbol{z}}_{t-1}, \ \mathbf{B}_T := \sum_{t=1}^{T} \hat{\boldsymbol{z}}_{t-1} \hat{\boldsymbol{z}}_{t-1}', \\ \boldsymbol{C}_T &:= \sum_{t=1}^{T} \hat{\boldsymbol{z}}_{t-1} \hat{y}_t, \ \text{ and } \ \mathbf{D}_T := \sum_{t=1}^{T} \hat{\boldsymbol{z}}_{t-1} \hat{\boldsymbol{z}}_{t-1}' \hat{u}_t^2, \end{aligned}$$

and $\hat{y}_t$, $\hat{x}_{t-1}$ and $\hat{\boldsymbol{z}}_{t-1}$ are demeaned versions of y_t, x_{t-1} and $\boldsymbol{z}_{t-1}$.

- BD show that a necessary consequence of using over-identified IV inference with strictly exogenous instruments is that the 2SLS test cannot be used to test against one-sided alternatives. Consequently the appropriate form of the test is to reject for large values of the Wald statistic, $(t_{\beta_1})^2$.
- BD show that $(t_{\beta_1})^2$ has a limiting $\chi^2(1)$ distribution, even when ξ_t displays unconditional and/or conditional heteroskedasticity (regularity conditions assumed).

The IVX Test of KMS

In the case of a single predictor, the IVX-based t-ratio of KMS for testing $H_0 : \beta = 0$ in (3), instruments the endogenous predictor x_{t-1} with the IVX instrument $z_{I,t-1}$, and is given by

$$t_{zx} := \frac{\hat{\beta}_{zx}}{s.e.(\hat{\beta}_{zx})} \tag{7}$$

where $\hat{\beta}_{zx}$ is the IVX estimator of β,

$$\hat{\beta}_{zx} := \frac{\sum_{t=1}^{T} z_{I,t-1} \left(y_t - \bar{y} \right)}{\sum_{t=1}^{T} z_{I,t-1} \left(x_{t-1} - \bar{x}_{-1} \right)} \tag{8}$$

with $\bar{y} := T^{-1} \sum_{t=1}^{T} y_t$ and $\bar{x}_{-1} := T^{-1} \sum_{t=1}^{T} x_{t-1}$, and

$$s.e.(\hat{\beta}_{zx}) := \frac{\sqrt{\hat{\sigma}_u^2 \sum_{t=1}^{T} z_{I,t-1}^2}}{\sum_{t=1}^{T} z_{I,t-1} \left(x_{t-1} - \bar{x}_{-1} \right)} \tag{9}$$

with $\hat{\sigma}_u^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2$.

- A variety of choices for the residuals $\hat{u}_t$ in these procedures is possible. KMS recommend using the OLS residuals from estimating (3). One could also use residuals computed under the null; that is, $\hat{u}_t := y_t - \frac{1}{T}\sum_{s=1}^{T} y_s$, or use IV residuals.
- One-sided tests based on t_{zx} can be formed by rejecting against the right-sided alternative that $\beta > 0$ for large positive values of the statistics and against the left-sided alternative that $\beta < 0$ for large negative values of the statistics. The latter can be equivalently implemented as right-sided tests simply by replacing the predictor x_{t-1} by $-x_{t-1}$.
- Two-sided tests can be formed by rejecting against the alternative that $\beta \neq 0$ for large positive values of $(t_{zx})^2$.
- So the KMS tests have the advantage over the BD tests that they can be implemented as one-sided tests. This may therefore confer a power advantage in cases where finance theory predicts the sign of the slope parameter on x_{t-1} under predictability.

- KMS implement a finite sample correction factor to correct for the finite sample effects of estimating the intercept term in (3). Details can be found in KMS and will be discussed in the class exercise.
- In the case where unconditional and/or unconditional heteroskedasticity is allowed for, the conventional standard error, $s.e.(\hat{\beta}_{zx})$, in (7) must be replaced by the corresponding Eicker-White (heteroskedasticity-robust) standard error. This is given by

$$t_{zx}^{EW} := \frac{\hat{\beta}_{zx}}{s.e.^{EW}(\hat{\beta}_{zx})}, \quad s.e.^{EW}(\hat{\beta}_{zx}) := \frac{\sqrt{\sum_{t=1}^{T} z_{t-1}^2 \hat{u}_t^2}}{\sum_{t=1}^{T} z_{t-1}\left(x_{t-1} - \bar{x}_{-1}\right)}.$$

- For a predictive regressions with multiple regressors, joint Wald (F) tests can be formed in the usual way with an IVX instrument formed for each regressor. However, the predictors all need to belong to the same persistence type.
- KMS show that their IVX statistics have standard limiting null distributions (standard normal/$\chi^2(1)$) for the case of one predictor, as above, and $\chi^2(K)$ for K predictors) regardless of whether x_t is strongly or weakly persistent.

3.4 Bootstrap IVX Tests

- Demetrescu *et al.* (2022a, Jnl Econometrics) [DGRT] show that, in the case of a single predictor, although the two-sided IVX tests proposed in KMS have good size control, one-sided tests do not. They also show that the finite sample size control of the IVX tests worsens as the number of predictors increases, other things equal.
- Consequently, DGRT explore two possible bootstrap implementations of KMS's IVX tests. The first, a residual wild bootstrap [RWB]. The second is a fixed regressor wild bootstrap [FRWB].
- DGRT show that both of these bootstrap methods are asymptotically valid and under weaker conditions on the errors, ξ_t, than are imposed by KMS.

A Residual Wild Bootstrap

1. Fit the predictive regression (3) to the sample data $(y_t, x_{t-1})'$ to obtain the residuals $\hat{u}_t$, $t = 1, ..., T$.
2. Fit by OLS an autoregression of order $p + 1$ to x_t; viz,

$$x_t = \hat{m} + \sum_{j=1}^{p+1} \hat{a}_j x_{t-j} + \hat{v}_t$$

   and compute the OLS residuals $\hat{v}_t$, $t = p + 1, \ldots, T$. Set $\hat{v}_t = 0$ for $t = 1, \ldots, p$.
3. Generate bootstrap innovations $(u_t^*, v_t^*)' = (D_t \hat{u}_t,\ D_t \hat{v}_t)'$, $t = 1 \ldots, T$, where D_t, $t = 1, ..., T$, is a scalar $i.i.d.(0, 1)$ sequence with $\mathrm{E}(D_t^4) < \infty$, which is independent of the sample data.

4 Define the bootstrap data $(y_t^*, x_{t-1}^*)'$ where $y_t^* = u_t^*$ (so that the null hypothesis is imposed on the bootstrap y_t^*) and where x_t^* is generated according to the recursion

$$x_t^* = \sum_{j=1}^{p+1} \hat{a}_j x_{t-j}^* + v_t^*, \quad t = 1, ..., T$$

with initial conditions $x_0^* = \ldots = x_{-p}^* = 0$. Create the associated bootstrap IVX instrument, z_t^*, as:

$$z_{I,0}^* = 0 \quad \text{and} \quad z_{I,t}^* = \sum_{j=0}^{t-1} \varrho^j \Delta x_{t-j}^*, \; t = 1, \ldots, T,$$

where ϱ is the same value as used in constructing the original IVX instrument, z_t.

5 Using the bootstrap sample data, $\left(y_t^*, x_{t-1}^*, z_{I,t-1}^*\right)'$, in place of the original sample data, $(y_t, x_{t-1}, z_{I,t-1})'$, construct the bootstrap analogues of the IVX statistics.

A Fixed-Regressor Wild Bootstrap

1. Construct the wild bootstrap innovations $y_t^* = \hat{y}_t D_t$, where $\hat{y}_t = y_t - \frac{1}{T}\sum_{t=1}^{T} y_t$ are the demeaned sample observations on y_t.
2. Using the bootstrap sample data $\left(y_t^*, x_{t-1}, \boldsymbol{z}_{I,t-1}'\right)'$, in place of the original sample data $\left(y_t, x_{t-1}, \boldsymbol{z}_{I,t-1}'\right)'$, construct the bootstrap analogues of the IVX statistics.

Key Differences?

- A key difference between the RWB and FRWB surrounds the generation of the bootstrap analogue data for x_t and $z_{I,t}$. While the RWB rebuilds into the bootstrap data (an estimate of) the correlation between the innovations u_t and v_t (it is crucial in doing so that the same D_t is used to multiply both $\hat{u}_t$ and $\hat{v}_t$), the FRWB does not. This is an important distinction because the finite sample behaviour of the IVX statistics is heavily dependent on the correlation between u_t and v_t when x_t is strongly persistent.
- A further difference is that because the RWB uses the bootstrap data x_t^* and $z_{I,t}^*$, one is implicitly using an estimate of ρ. Under strong persistence c, cannot be consistently estimated and so x_t^* will not be generated with the same local-to-unity parameter as x_t. However, the IVX statistics instrument x_{t-1} by $z_{I,t-1}$, and their bootstrap analogues instrument x_{t-1}^* by $z_{I,t-1}^*$. But both $z_{I,t}$ and $z_{I,t}^*$ are, by construction, mildly integrated processes, regardless of the value of c. There is therefore no necessity for the estimate of c to be consistent.

Monte Carlo Results from DGRT I

Case 1: Empirical Size: Scalar Predictor, IID errors

- DGP (3)-(5) with $\beta = 0$. Set $\alpha = \mu = 0$, w.n.l.o.g.
- $\rho := 1 - c/T$ with $c \in \{-0.5, -0.25, 0, 2.5, 5, 10, 25, ..., 250\}$
- $(u_t, v_t)'$ is zero-mean IID bivariate Gaussian with covariance matrix $\mathbf{\Sigma} := \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix}$ and $\delta = -0.95$
- IVX with $a = 1$, $\gamma = 0.95$, and KMS's finite-sample correction
- Report: $t_{zx}^{*,RWB}$ and $t_{zx}^{*,FRWB}$ (RWB and FRWB implementations of t_{zx}); t_{zx}^{EW} (asymptotic IVX test with conventional ses) and t_{zx} (asymptotic IVX test with White ses)
- $T = 250$, 10000 MC replications, 999 bootstrap replications with $D_t \sim NIID(0, 1)$. Nominal 5% level. In Step 2 of RWB p chosen by BIC over the search set $p \in \{0, ..., \lfloor 4(T/100)^{0.25} \rfloor\}$.

Table 1: Size of Left-sided Tests
Gaussian IID innovations

c	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}
-5	0.046	0.004	0.004	0.003
-2.5	0.045	0.000	0.000	0.001
0	0.041	0.001	0.001	0.001
2.5	0.062	0.005	0.005	0.005
5	0.068	0.010	0.011	0.010
10	0.064	0.019	0.019	0.018
25	0.057	0.029	0.030	0.028
50	0.056	0.034	0.036	0.035
75	0.056	0.037	0.038	0.037
100	0.054	0.038	0.040	0.038
125	0.054	0.039	0.042	0.041
150	0.055	0.043	0.046	0.042
200	0.054	0.046	0.048	0.045
250	0.054	0.048	0.051	0.048

Table 2: Size of Right-sided Tests
Gaussian IID innovations

c	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}
-5	0.046	0.074	0.080	0.073
-2.5	0.041	0.094	0.097	0.093
0	0.053	0.105	0.114	0.110
2.5	0.064	0.112	0.116	0.115
5	0.062	0.107	0.116	0.112
10	0.062	0.097	0.102	0.099
25	0.057	0.078	0.084	0.080
50	0.052	0.067	0.072	0.067
75	0.053	0.064	0.068	0.065
100	0.053	0.061	0.065	0.062
125	0.052	0.060	0.063	0.060
150	0.053	0.056	0.060	0.059
200	0.050	0.054	0.056	0.053
250	0.051	0.051	0.055	0.053

Table 3: Size of Two-sided Tests
Gaussian IID innovations

c	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}
-5	0.048	0.038	0.044	0.039
-2.5	0.038	0.040	0.048	0.044
0	0.047	0.051	0.057	0.053
2.5	0.053	0.058	0.062	0.060
5	0.054	0.058	0.063	0.060
10	0.055	0.060	0.066	0.060
25	0.056	0.056	0.060	0.058
50	0.051	0.051	0.054	0.052
75	0.049	0.047	0.052	0.049
100	0.049	0.048	0.052	0.050
125	0.050	0.049	0.053	0.051
150	0.051	0.049	0.054	0.052
200	0.050	0.048	0.054	0.050
250	0.049	0.048	0.053	0.050

Monte Carlo Results from DGRT II

Case 2: Empirical Size: Multiple Predictors

- Multiple predictor simulation DGP:

$$\begin{aligned} y_t &= \alpha + \mathbf{x}'_{t-1}\boldsymbol{\beta} + u_t, \qquad t = 1, \ldots, T, \\ \mathbf{x}_t &= \boldsymbol{\rho}\mathbf{x}_{t-1} + \mathbf{v}_t, \qquad t = 0, \ldots, T, \end{aligned}$$

  where $\mathbf{x}_t := (x_{1,t}, ..., x_{K,t})'$ is a $K \times 1$ vector of predictor variables, $\boldsymbol{\beta}$ is a $K \times 1$ vector of parameters, $\alpha = 0.25$, $\boldsymbol{\rho}$ is a $K \times K$ diagonal matrix with common diagonal element ρ, i.e., $\boldsymbol{\rho} := \text{diag}(\rho, ..., \rho)$.
- The AR parameter ρ is again set equal to $1 - c/T$ with $c \in \{-5, -2.5, 0, 2.5, 5, 10, 25, ..., 250\}$

- The innovations are generated as $(u_t, \mathbf{v}_t')' \sim NIID(\mathbf{0}, \mathbf{\Sigma})$ where

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_u^2 & \sigma_{u,v_1} & 0 & \cdots & 0 \\ \sigma_{u,v_1} & \sigma_{v_1}^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_{v_2}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_{v_K}^2 \end{pmatrix} \quad (10)$$

  with $\sigma_u^2 = 0.037$, $\sigma_{u,v_1} = -0.035$, $\sigma_{v_1}^2 = ... = \sigma_{v_K}^2 = 0.045$.
- Notice, therefore, that the first predictor, $x_{1,t}$ is endogenous (with an endogeneity correlation parameter $\phi_1 = -0.83$), while the remaining predictors $x_{2,t}, ..., x_{K,t}$ are exogenous.
- Empirical sizes of the Wald tests for the joint significance of the K predictors. NB RWB uses obvious VAR generalisation of Step 2.

Table 4: Size of joint Wald Tests. $K = 3$ predictors.

c	$W_{zx}^{*,\text{RWB}}$	$W_{zx}^{*,\text{FRWB}}$	W_{zx}^{EW}	W_{zx}
-5	0.085	0.352	0.385	0.366
-2.5	0.097	0.176	0.193	0.177
0	0.075	0.105	0.117	0.104
2.5	0.067	0.086	0.103	0.090
5	0.059	0.077	0.095	0.083
10	0.054	0.066	0.083	0.071
25	0.052	0.061	0.075	0.066
50	0.053	0.057	0.070	0.061
75	0.053	0.053	0.069	0.058
100	0.051	0.053	0.069	0.057
125	0.052	0.054	0.070	0.058
150	0.052	0.054	0.069	0.058
200	0.052	0.055	0.071	0.059
250	0.053	0.055	0.071	0.060

Table 5: Size of joint Wald Tests. $K = 5$ predictors.

c	$W_{zx}^{*,\text{RWB}}$	$W_{zx}^{*,\text{FRWB}}$	W_{zx}^{EW}	W_{zx}
-5	0.074	0.402	0.466	0.421
-2.5	0.091	0.239	0.281	0.241
0	0.082	0.157	0.186	0.156
2.5	0.069	0.120	0.156	0.129
5	0.063	0.105	0.138	0.116
10	0.062	0.086	0.120	0.098
25	0.053	0.067	0.100	0.080
50	0.052	0.059	0.089	0.069
75	0.051	0.055	0.085	0.063
100	0.049	0.053	0.082	0.062
125	0.049	0.053	0.080	0.062
150	0.046	0.052	0.078	0.061
200	0.047	0.051	0.079	0.060
250	0.044	0.049	0.077	0.058

Table 6: Size of joint Wald Tests. $K = 10$ predictors.

c	$W_{zx}^{*,\text{RWB}}$	$W_{zx}^{*,\text{FRWB}}$	W_{zx}^{EW}	W_{zx}
-5	0.058	0.513	0.635	0.559
-2.5	0.072	0.398	0.505	0.425
0	0.087	0.306	0.406	0.324
2.5	0.075	0.238	0.342	0.262
5	0.067	0.191	0.301	0.225
10	0.060	0.141	0.244	0.175
25	0.050	0.089	0.174	0.118
50	0.048	0.067	0.142	0.091
75	0.046	0.060	0.129	0.081
100	0.046	0.056	0.120	0.077
125	0.043	0.053	0.117	0.074
150	0.042	0.052	0.116	0.071
200	0.039	0.049	0.116	0.070
250	0.036	0.050	0.116	0.072

3.5 Subsample Implementation of the KMS Test: Detecting *Pockets of Predictability*

- The testing approaches discussed so far are based on a maintained assumption that the slope coefficient β in (3) is constant over time. However, there are several reasons to suspect that if returns are predictable, then it is likely to be a time-varying phenomenon. The business cycle, time-varying risk aversion, rare disasters, structural breaks, speculative bubbles, investor's market sentiment, and regime changes in monetary policy have all be cited as possible reasons.
- Timmermann (2008) argues that for most time periods returns are not predictable but that there can be *pockets in time* where evidence of predictability is seen. If a variable begins to have predictive power for returns then a window of predictability might exist before investors learn about that relationship, but it will eventually disappear.

- It therefore seems reasonable to consider the possibility that the predictive relationship might change over time, so that over a long span of data one may observe some temporary window(s) of time during which predictability occurs.
- The assumption in (3) that β is constant across the sample implies that under the alternative hypothesis x_{t-1} is predictive for y_t across the whole sample. Look again at Figure 1. Here we would be able to reject the null of no predictability if the data ended at, for example, 1990M1 (because here the confidence interval for the slope does not include zero) but we cannot reject the null based on the whole sample (because zero is now in the confidence interval). So full sample tests may not be well suited to detecting pockets of predictability.
- So can we do any better than the full sample tests if we allow the slope parameter β in (3) to be time-varying? Let's call it β_t.

- If it were known that a *pocket of predictability* might occur only over the particular subsample $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$, $0 \leq \tau_1 < \tau_2 \leq 1$, such that $\beta_t = \beta \neq 0$ for $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$ but was zero elsewhere, then it would be more logical to base a test for this on the IVX statistic computed only on the subsample $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$. With an obvious notation denote this statistic as $t_{zx}(\tau_1, \tau_2)$, and the corresponding subsample analogue of the full sample Eicker-White t_{zx}^{EW} statistic denoted $t_{zx}^{EW}(\tau_1, \tau_2)$.
- In practice it is unlikely to be known which specific subsample(s) of the data might admit predictive regimes. Tests based on forward and reverse recursive sequences and rolling sequences might therefore be useful, as they were with bubble detection methods in Topic 2.

- Tests based on the forward recursive sequence of statistics are designed to detect pockets of predictability which begin at or near the start of the full sample period, while those based on the reverse recursive sequence are designed to detect end-of-sample pockets of predictability. For a given window width, tests based on a rolling sequence of statistics are designed to pick up a window of predictability, of (roughly) the same length, within the data.
- DGRT propose tests based on these sequences of subsample statistics. We will formally define these on the next slide. We will outline these for the case of IVX statistics computed with conventional standard errors, but these can also be implemented with Eicker-White standard errors by replacing $t_{zx}(\cdot,\cdot)$ with $t_{zx}^{EW}(\cdot,\cdot)$ throughout.

- The sequence of *forward recursive* statistics is given by $\{t_{zx}(0,\tau)\}_{\tau_L \leq \tau \leq 1}$, where the parameter $\tau_L \in (0,1)$ is chosen by the user. The forward recursive regression approach uses $\lfloor T\tau_L \rfloor$ start-up observations, where τ_L is the *warm-in* fraction, and then calculates the sequence of subsample predictive regression statistics $t_{zx}(0,\tau)$ for $t = 1, ..., \lfloor \tau T \rfloor$, with τ travelling across the interval $[\tau_L, 1]$. An upper-tailed test can then be based on the maximum taken across this sequence, *viz*,

$$\mathcal{T}_U^F := \max_{\tau_L \leq \tau \leq 1} \{t_{zx}(0,\tau)\}. \tag{11}$$

The corresponding left-tailed test can be based on the minimum across this sequence, denoted $\mathcal{T}_L^F$, and a two-tailed test can be based on the corresponding maximum taken over the sequence of $(t_{zx}(0,\tau))^2$ statistics, denoted $\mathcal{T}_2^F$.

- The sequence of *backward recursive* statistics is given by $\{t_{zx}(\tau, 1)\}_{0 \leq \tau \leq \tau_U}$ with $\tau_U \in (0, 1)$ again chosen by the user. Here one calculates the sequence of subsample predictive regression statistics $t_{zx}(\tau, 1)$ for $t = \lfloor \tau T \rfloor + 1, ..., T$, with τ travelling across the interval $[0, \tau_U]$. Analogously to the forward recursive case, an upper-tailed test can again be based on the maximum from this sequence,

$$\mathcal{T}_U^B := \max_{0 \leq \tau \leq \tau_U} \{t_{zx}(\tau, 1)\} \tag{12}$$

while corresponding lower-tailed tests and two-sided tests can be formed from the statistics $\mathcal{T}_L^B$ and $\mathcal{T}_2^B$, defined analogously to the forward recursive case.

- The sequence of *rolling* statistics is given by $\{t_{zx}(\tau, \tau + \Delta\tau)\}_{0 \leq \tau \leq 1-\Delta\tau}$ where the user-defined parameter $\Delta\tau \in (0, 1)$. Here one calculates the sequence of subsample statistics $t_{zx}(\tau, \tau + \Delta\tau)$ for $t = \lfloor \tau T \rfloor + 1, ..., \lfloor \tau T \rfloor + \lfloor T\Delta\tau \rfloor$, where $\Delta\tau$ is the window fraction with $\lfloor T\Delta\tau \rfloor$ the window width, with τ travelling across the interval $[0, 1 - \Delta\tau]$. An upper-tailed test can again be based on the maximum from this rolling sequence,

$$\mathcal{T}_U^R := \max_{0 \leq \tau \leq 1-\Delta\tau} \{t_{zx}(\tau, \tau + \Delta\tau)\} \tag{13}$$

while corresponding lower-tailed tests and two-sided tests can again be formed from the statistics $\mathcal{T}_L^R$ and $\mathcal{T}_2^R$, defined analogously to the recursive cases.

- DGRT show that these subsample predictability tests can only be validly implemented (i.e. with controlled asymptotic size) using either the RWB or the FRWB discussed previously for the full sample IVX tests.
- As discussed in Pavlidis *et al* (2017), subsample implementations of KMS's right-tailed IVX predictability tests can be used to test for the presence of a speculative bubble in exchange rates. They do not find evidence of a bubble. DGRT revisit an application in Pavlidis *et al* (2017) with an updated dataset and using their proposed bootstrap subsample maximum tests and find evidence of a bubble in the UK/US exchange rate which crashed in 2016.
- Demetrescu *et al.* (2022b) also consider tests for episodic predictability based on the maxima from corresponding sequences of rolling and recursive subsample implementations of the 2SLS predictability statistics of BD discussed in section 3.3. Again these have the disadvantage that they cannot be used for one-tailed testing.

3.6 Long-Horizon Predictive Regressions

- There has been an increasing interest in long-horizon predictive regressions, because a number of studies using long-horizon variables seem to find significant results where previous short-run predictive regression models, like those we have considered so far, fail to do. Some of the most important among these studies are listed in the literature review given in section 1 of Valkanov (2003).
- The results in the aforementioned studies are based on long-horizon variables, where the long-horizon variable is a rolling sum of the original series. In the literature, it is heuristically argued that long-run regressions produce more accurate results by strengthening the signal coming from the data while eliminating the noise.
- We will very briefly review long-horizon methods.

- In the long-horizon literature you'll see that they end to write the short-run predictive regression that we have encountered so far with the time index shifted one-period forward, i.e., the short-run (one period) predictive recursive system,

$$\begin{aligned} y_{t+1} &= \alpha_1 + \beta_1 x_t + u_{t+1}, \qquad t = 1, \ldots, T-1, && (14) \\ x_{t+1} &= \mu + w_{t+1}, \qquad \text{and} \qquad w_{t+1} = \rho w_t + v_{t+1} && (15) \end{aligned}$$

  where y_{t+1} is, for example, a continuously compounded excess return of an asset or the variation of a nominal exchange rate from t to $t+1$ and x_{t+1} is some (putative) predictor variable.
- This is just a cosmetic shift in the time-index, and α and β relabeled as α_1 and β_1 respectively; it's essentially the same DGP as we had before. This is termed the *short-horizon* or short-run predictive regression model in the literature.

- The most common long-horizon predictive regression specification used in empirical analysis results from the h-period, $h \geq 1$, temporal aggregation of (14) and is given by

$$y_{t+h}^{(h)} = \alpha_h + \beta_h x_t + error_{t+h}, \ t = 1, \ldots, T - h \tag{16}$$

  where $y_{t+h}^{(h)} := \sum_{j=1}^{h} y_{t+j}$ is the h-period cumulative variable to be predicted.
- Notice that for $h = 1$, (16) is simply the short-horizon predictive regression in (14). To gain further insight into the specific features of (16), let us examine the h-horizon cumulated dependent variable $y_{t+h}^{(h)}$ more closely.

- From (14), the long-horizon predictive model can be written as,

$$y_{t+h}^{(h)} = h\alpha_1 + \beta_1 \sum_{j=0}^{h-1} x_{t+j} + u_{t+h}^{(h)}, \tag{17}$$

  where $u_{t+h}^{(h)} := \sum_{j=1}^{h} u_{t+j}$.
- Recall that the short-horizon predictability testing methods we have discussed are all based on the assumption that the error term in the predictive regression is serially uncorrelated.
- It is clear from (17) that this condition is violated in long-horizon case because serial correlation is induced into the error term in the long-horizon predictive regression, arising from the temporal aggregation of the dependent variable.
- The methods developed for short-horizon predictability testing cannot therefore be validly used for $h > 1$.

- To address this issue, Valkanov (2003) and Hjalmarsson (2011) propose using the conventional OLS t-statistic from (16) but scaled by a constant to reflect the inflation of the standard errors as the prediction horizon increases. Both approaches are valid only for strongly persistent predictors.
- Tests for multiple-horizon predictability designed to be asymptotically valid regardless of whether the predictors are strongly or weakly persistent and for handling the issues arising from temporal aggregation are proposed in Phillips and Lee (2013) who develop tests from a reversed predictive regression framework, estimated by IVX.
- Other more recent approaches are considered in Xu (2020) and Demetrescu *et al.* (2022c). The latter provides a review of existing long-horizon methods in the literature.

Some Useful Resources

- The confidence belts and code for the procedures needed for running the CY tests are available from Motohiro Yogo's website:

  https://sites.google.com/site/motohiroyogo/research/asset-pricing
- You will develop MATLAB code in the computer class to be able to run the IVX tests of KMS.
- The Welch-Goyal (2008) dataset (updated) can obtained from

  https://sites.google.com/view/agoyal145/
- The Campbell-Yogo dataset (updated) can be obtained from

  https://sites.google.com/site/motohiroyogo/research/asset-pricing

References

- Breitung, J. and Demetrescu, M. (2015). Instrumental variable and variable addition based inference in predictive regressions. Journal of Econometrics 187, 358-375.
- Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. Journal of Financial Economics 81, 27-60.
- Cavanagh, C. L., Elliott, G. and Stock, J.H. (1995). Inference in models with nearly integrated regressors. Econometric Theory 11, 1131-1147.
- Demetrescu, M., I. Georgiev, P. M. M. Rodrigues, and A. M. R. Taylor (2022a). Extensions to IVX methods of inference for return predictability. Journal of Econometrics (forthcoming). Downloadable from https://doi.org/10.1016/j.jeconom.2022.02.007
- Demetrescu, M., I. Georgiev, P. M. M. Rodrigues, and A. M. R. Taylor (2022b). Testing for episodic predictability in stock returns. Journal of Econometrics 227, 85-113.

- Demetrescu, M., P.M.M. Rodrigues, and A.M.R. Taylor (2022c). Transformed regression-based long-horizon predictability tests, Journal of Econometrics (forthcoming). Downloadable from https://doi.org/10.1016/j.jeconom.2022.06.006
- Elliott, G., Müller, U.K. and Watson, M.W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. Econometrica 83, 771-811.
- Fama, E.F. (1981). Stock returns, real activity, inflation, and money. American Economic Review 71, 545-565.
- Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2021). Simple tests for stock return predictability with good size and power properties. Journal of Econometrics 224, 198-214.
- Hjalmarsson, E. (2011). New methods for inference in long-horizon regressions. Journal of Financial and Quantitative Analysis 46, 815–839.

- Jansson, M. and M. J. Moreira (2006). Optimal inference in regression models with nearly integrated regressors. Econometrica 74, 681–714.
- Kostakis, A., Magdalinos, T. and Stamatogiannis, M.P. (2015). Robust econometric inference for stock return predictability. Review of Financial Studies 28, 1506-1553.
- Pavlidis, E. G., I. Paya, and D. A. Peel (2017). Testing for speculative bubbles using spot and forward prices. International Economic Review 58, 1191–1226.
- Phillips, P. C. B. and J. H. Lee (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. Journal of Econometrics 177, 250–264.
- Stambaugh, R. F. (1999). Predictive regressions. Journal of Financial Economics 54, 375–421.

- Timmermann, A. (2008). Elusive return predictability. International Journal of Forecasting 24, 1–18.
- Valkanov, R. (2003). Long-horizon regressions: theoretical results and applications. Journal of Financial Economics 68, 201–232.
- Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455-1508.
- Xu, K.-L. (2020). Testing for multiple-horizon predictability: Direct regression based versus implication based. Review of Financial Studies 33, 4403–4443.