Extensions to IVX Methods of Inference for Return Predictability

Robert Taylor, University of Essex

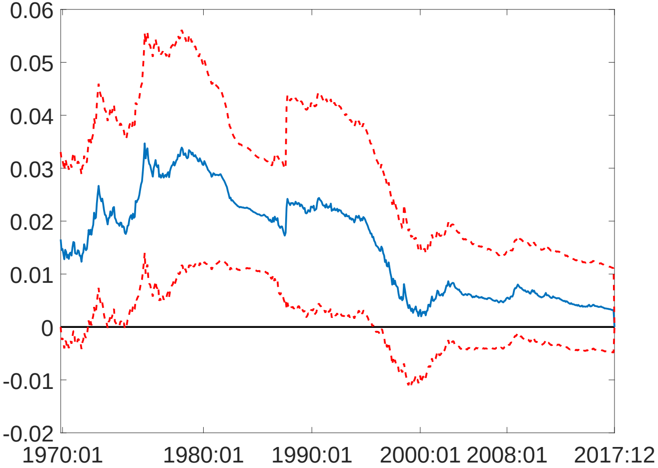
Funded by the Economic and Social Research Council of the United Kingdom under research grant ES/R00496X/1

https://rtaylor-essex.droppages.com/esrc2/default.htm

joint work with Matei Demetrescu, Iliyan Georgiev and Paulo Rodrigues

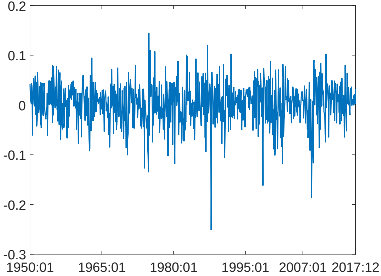
North American Meeting, Econometric Society
Montreal, June 10th 2021

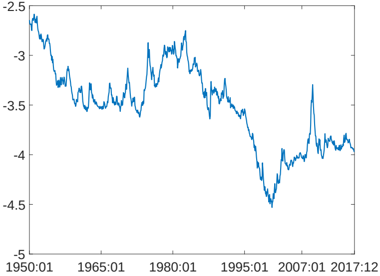
Is there any predictability in the equity premium?



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).

Outline

Moving on to ...

1. Background and Summary of Contributions
2. The (Episodic) Predictive Regression Model
3. IVX Predictability Tests
4. Finite Sample Simulations
5. Concluding Remarks

Trouble in paradise

Consider the predictive regression

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

where

$$x_t \quad = \quad \rho\, x_{t-1} + v_t,$$

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

$$\mathbf{\Sigma} = \mathrm{E}\left(\begin{pmatrix} u_t \\ v_t \end{pmatrix} \begin{pmatrix} u_t & v_t \end{pmatrix} \right) = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}.$$

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 (high) persistence

Should

- the shocks u_t and v_t correlate (so that $\phi := \sigma_{uv}/\sigma_u\sigma_v \neq 0$; for the EP-DY data above this correlation is estimated to be $\hat{\phi} = -0.98$), and
- the regressor x_t be autocorrelated,

one speaks of endogeneity. (A bit of a misnomer.)

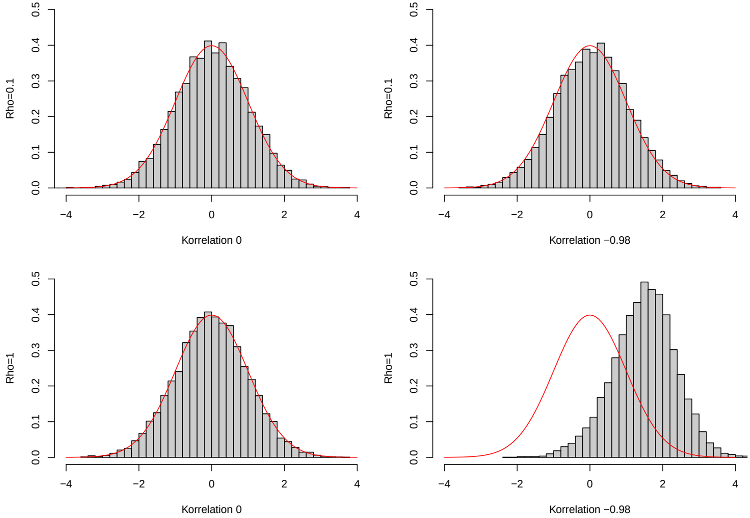
Under endogeneity and high persistence (near integration, $\rho = 1 - c/T$),

- the OLS estimator is 2nd order biased and
- the *t*-statistic has a non-normal limiting distribution.

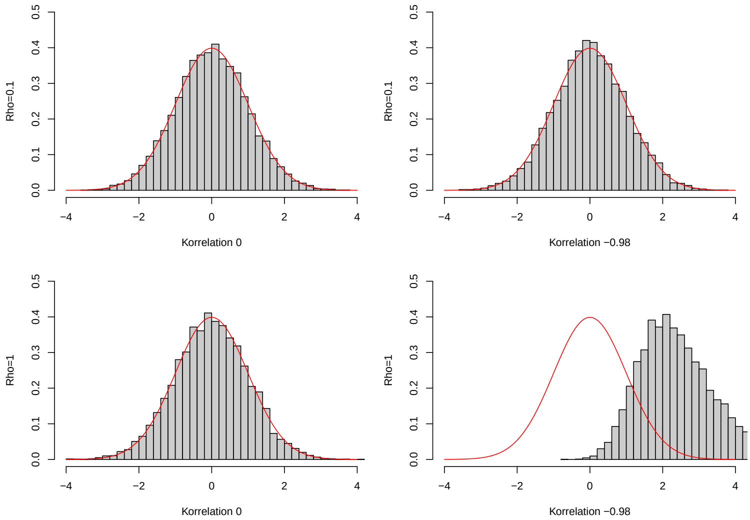
See Elliott/Stock (1994), Stambaugh (1999), Campbell/Yogo (2006) etc.

No problem when regressors are stationary or weakly persistent.

Trouble brewing - OLS t-statistics, $T = 305$



More trouble with variance breaks - volatility of both shocks 3 times higher in the first 20% of the sample



Feasible solutions

If ρ were known, one could employ GLS estimation. For unknown ρ:

- Bayes methods - Elliott/Stock (1994)
- Bonferroni - Campbell/Yogo (2006), but see Phillips (2012)
- Restricted log-likelihood - Jansson/Moreira (2006), Chen/Deo (2009)
- Almost optimal tests - Elliott *et al.* (2015)
- Variable addition - Toda/Yamamoto (1995), Dolado/Lütkepohl (1996)
- Generic IV estimation, including 2SLS methods - Breitung/Demetrescu (2015) [BD]
- Extended IV - or IVX - method of Kostakis *et al.* (2015) [KMS]

What we do in this paper

• Our focus in this paper is on IVX-based solutions which are becoming increasingly popular in the literature on return predictability.

• Under certain regularity conditions on the innovations, u_t and v_t, KMS establish that the IVX estimator is mixed normal and that the associated IVX-based predictability tests (t, F, and Wald type tests) have standard pivotal limiting null distributions, regardless of the degree of persistence or endogeneity of the predictor.

• In this paper we extend the IVX-based approach in three key directions:

Contribution I

• KMS assume that the innovations, u_t and v_t, are unconditionally homoskedastic. They allow for some conditional heteroskedasticity (provided White ses are used in the test statistics). In particular, although a relatively weak martingale difference assumption is placed on v_t, u_t is assumed to follow a finite-order parametric GARCH model. A consequence of this is that it imposes the absence of any dependence of the conditional variance of the regression errors on lagged values of the innovations driving the predictor. This is arguably unrealistic for many predictors used to predict stock returns.

• We show that the IVX statistics (with White ses) continue to admit standard pivotal limiting null distributions under considerably weaker assumptions on the innovations. In particular, we allow for quite general forms of conditional and unconditional heteroskedasticity in the joint innovation process, not tied to any parametric model.

Contribution II

• The asymptotic theory for IVX predictability statistics is known to provide a poor approximation to their finite sample behaviour, particularly for highly persistent and endogenous predictors which is of course the case of primary practical relevance. We develop asymptotically valid bootstrap implementations of the IVX tests. We investigate both fixed regressor wild bootstrap [FRWB] and residual wild bootstrap [RWB] resampling.

• Monte Carlo simulations show that, in particular the RWB, bootstrap methods we propose can deliver considerably more accurate finite sample inference than the asymptotic implementations of these tests under certain problematic parameter constellations, most notably for their implementation against one-sided alternatives, and where multiple predictors are included.

Contribution III

- The IVX methodology has recently been applied to Fama regressions in the context of detecting episodic bubble-type behaviour in Pavlidis *et al.* (2017). They use a rolling subsample-based implementation of one-sided IVX statistics and reject the no bubble null hypothesis if any of the subsample statistics in the rolling sequence exceeds a given critical value. To avoid the inherent multiple testing bias, they use a Bonferroni correction but note that this leads to highly conservative tests

- Tests based on the suprema of rolling and recursive subsample sequences of the 2SLS predictability statistics of BD are used to detect temporary periods of stock return predictability in Demetrescu *et al.* (2021). We show that both the RWB and FRWB approaches can also be implemented in the context of the corresponding tests from sequences of subsample IVX statistics under the same regularity conditions as for the full sample tests. Moreover, unlike the 2SLS-based tests, these can be implemented as either one-sided or two-sided tests for the presence of temporary windows of predictability.

Moving on to ...

Episodic Predictability

Consider the time-varying DGP

$$y_t = \alpha + \beta_t x_{t-1} + u_t, \qquad t = 1, \ldots, T \tag{1}$$

where the scalar predictor (extensions for multiple predictors are straightforward and given in the paper) x_t satisfies the DGP

$$x_t = \mu_x + \xi_t, \qquad t = 0, \ldots, T \tag{2a}$$

$$\xi_t = \rho\,\xi_{t-1} + w_t, \qquad t = 1, \ldots, T \tag{2b}$$

in which $\xi_0 = O_p(1)$ and where w_t is assumed to follow the pth order stable autoregression:

$$A(L)w_t = v_t, \;\; A(z) := (1 - a_1 z - a_2 z^2 - \cdots - a_p z^p).$$

Uncertain Persistence

Assumption 1

One of the following two conditions is assumed to hold:

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

The literature uses both equally often, but the resulting asymptotics differ.

Time-Varying Features I

Assumption 2

In the context of (1) *and* (2)*, let* $\beta_t := n_T^{-1} b\left(t/T\right)$, *where* $b(\cdot)$ *is a piecewise Lipschitz-continuous real function on* $[0, 1]$*, with* $n_T = \sqrt{T}$ *under Assumption 1.1, and* $n_T = T$ *under Assumption 1.2.*

We may reformulate the null hypothesis stated previously as

$$H_0 : \text{ The function } b(\tau) \text{ is identically zero for all } \tau \in [0, 1]. \tag{3}$$

We can now also formally specify the alternative hypothesis as,

$$H_{1,b(\cdot)} : b(\cdot) \text{ is non-zero over at least one non-empty subinterval of } [0, 1]. \tag{4}$$

Time-Varying Features II

- Under the null hypothesis, H_0, y_t is not predictable by x_{t-1} in any subsample.
- Under the alternative hypothesis, $H_{1,b(\cdot)}$, there exists at least one subset of the sample observations (this need not be a strict subset, so it could contain all of the sample observations) comprising contiguous observations and for which $\beta_t \neq 0$. A predictive episode, often termed a *pocket of predictability*. The size of this subset is proportional to the sample size T.

Time-Varying Features III

Assumption 3

Let $\begin{pmatrix} u_t \\ v_t \end{pmatrix} := \mathbf{H}(t/T) \begin{pmatrix} a_t \\ e_t \end{pmatrix}, \quad \begin{pmatrix} a_t \\ e_t \end{pmatrix} \sim \mathit{WN}\left(\mathbf{0}, \mathbf{I}_2\right)$, *where:*

1. $\mathbf{H}(\cdot) := \begin{pmatrix} h_{11}(\cdot) & h_{12}(\cdot) \\ h_{21}(\cdot) & h_{22}(\cdot) \end{pmatrix}$ *is a matrix of piecewise Lipschitz-continuous bounded functions on* $(-\infty, 1]$, *which is of full rank at all but a finite number of points;*
2. $\boldsymbol{\psi}_t := (a_t, e_t)'$ *is a* L_4*-bounded stationary and ergodic martingale difference sequence satisfying* $\mathrm{E}(\boldsymbol{\psi}_t \boldsymbol{\psi}_t') = \mathbf{I}_2$ *and* $\mathrm{E}\|\mathrm{E}_0 \sum_{t=1}^{T} (\boldsymbol{\psi}_t \boldsymbol{\psi}_t' - \mathbf{I}_2)\|^2 = O(T^{2\epsilon})$ *for some* $\epsilon < \frac{1}{2}$, *with* $\mathrm{E}_0(\cdot)$ *denoting expectation conditional on* $\{\boldsymbol{\psi}_{-i}\}_{i=0}^{\infty}$ *and* $\mathbf{I}_k$ *denoting the* $k \times k$ *identity matrix.*

Key aspects of Assumption 3

- Assumption 3.1 allows for quite general unconditional time heteroskedasticity in the innovations through the function $\mathbf{H}$, whereby the unconditional covariance matrix of $(u_t, v_t)'$ is given by $\mathbf{H}(t/T)\mathbf{H}'(t/T)$. This allows both u_t and v_t to display time-varying unconditional variances and for both contemporaneous and time-varying (unconditional) correlation between u_t and v_t. Empirically plausible models of single or multiple (co-) variance shifts, (co-)variances which follow a broken trend, and smooth transition (co-) variance shifts are all permitted under this assumption. In contrast, KMS impose a constant unconditional variance matrix on $(u_t, v_t)'$.
- Assumption 3.2 imposes a martingale difference [MD] structure on $\boldsymbol{\psi}_t$ thereby allowing for conditional heteroskedasticity. In common with Assumption INNOV of KMS, Assumption 3.2 imposes finite fourth-order moments on $\boldsymbol{\psi}_t$.

Key aspects of Assumption 3

- To establish the large sample properties of the IVX tests of KMS in the *strong persistence* case we need to develop a weak convergence result for $\frac{1}{\sqrt{T^{1+\eta}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{t-1} u_t$.
- For the case of full-sample sums, in order to do so KMS make the parametric assumption that u_t is generated by a stationary finite-order GARCH(p, q) model with finite fourth moments. This assumption therefore has the unfortunate consequence that it imposes the absence of any dependence of the conditional variance of u_t on lags of v_t which is likely to be unrealistic for many predictors used to predict stock returns.
- Moreover, a number of authors, including Carnero *et al.* (2004) and Johannes *et al.* (2014) argue that ARSV models capture the main empirical properties of the volatility of financial returns series better than GARCH models.

Key aspects of Assumption 3

- To eliminate the need to choose a specific parametric volatility model, Assumption 3.2 instead adopts an explicit assumption of martingale approximability whereby

$$\mathrm{E}\|\mathrm{E}_0 \sum_{t=1}^{T} (\boldsymbol{\psi}_t \boldsymbol{\psi}_t' - \mathbf{I}_2)\|^2 = O(T^{2\epsilon})$$

  for some $\epsilon < \frac{1}{2}$, see Merlevede *et al.* (2006). The exponent ϵ controls the degree of persistence permitted in the conditional variances of the innovations.

Key aspects of Assumption 3

- Stationary vector GARCH processes with finite fourth-order moments satisfy Assumption 3.2 with $\epsilon = 0$, but the assumption is considerably more general as it also allows for asymmetric effects in the conditional variance.
- Stationary ARSV processes also satisfy Assumption 3.2.
- Assumption 3 is very similar to the assumptions made on $(u_t, v_t)'$ by Demetrescu *et al.* (2021) who develop subsample implementations of the 2SLS IV tests of Breitung and Demetrescu (2015). Notice, with 2SLS tests we don't have to worry about the asymptotic behaviour of $\frac{1}{\sqrt{T^{1+\eta}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{t-1} u_t$ in the strong persistent case as their Type-II instrument (eg a sine function of time) is chosen there not the IVX (Type-I) instrument.

Moving on to ...

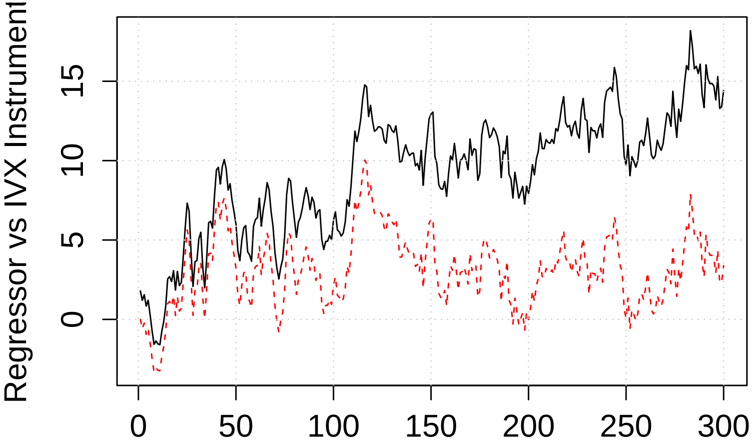
The IVX Instrument

KMS develop asymptotically valid methods of estimation and inference in the context of (1)-(2) based on the use of the mildly integrated IVX instrument

$$z_{I,t} := \sum_{j=0}^{t-1} \varrho^j \Delta x_{t-j} = (1 - \varrho L)_+^{-1} \Delta x_t$$

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

The IVX trick applied to a random walk - using $a = 1$, $\gamma = 0.95$ as in KMS



The Full Sample IVX Test of KMS I

The full-sample IVX-based t-ratio of KMS for testing $H_0 : \beta_t = \beta = 0$ for all $t = 1, ..., T$, instruments the endogenous predictor x_{t-1} with the IVX instrument $z_{I,t-1}$, and is given by

$$t_{zx} := \frac{\hat{\beta}_{zx}}{s.e.(\hat{\beta}_{zx})} \tag{5}$$

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

$$\hat{\beta}_{zx} := \frac{\sum_{t=1}^{T} z_{t-1} \left(y_t - \bar{y}\right)}{\sum_{t=1}^{T} z_{t-1} \left(x_{t-1} - \bar{x}_{-1}\right)} \tag{6}$$

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

$$s.e.(\hat{\beta}_{zx}) := \frac{\sqrt{\hat{\sigma}_u^2 \sum_{t=1}^{T} z_{t-1}^2}}{\sum_{t=1}^{T} z_{t-1} \left(x_{t-1} - \bar{x}_{-1}\right)} \tag{7}$$

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

The Full Sample IVX Test of KMS II

- A variety of choices for the residuals $\hat{u}_t$ is possible. KMS recommend using the OLS residuals from estimating (1) and we will use these also. 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 IV residuals.
- One-sided tests based on t_{zx} can be formed by rejecting against the right-sided alternative that $\beta_t = \beta > 0$, for all $t = 1, ..., T$, for large positive values of the statistics and against the left-sided alternative that $\beta_t = \beta < 0$, for all $t = 1, ..., T$, 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 $\beta_t = \beta \neq 0$, for all $t = 1, ..., T$, for large positive values of $(t_{zx})^2$.

The Full Sample IVX Test of KMS III

- KMS implement a finite sample correction factor to correct for the finite sample effects of estimating the intercept term in (1). Details can be found in KMS. We also implement this correction factor.
- In the case where conditional and/or unconditional heterskedasticity is allowed for under Assumption 3 the conventional standard error, $s.e.(\hat{\beta}_{zx})$, in (5) must be replaced by the corresponding Eicker-White standard error, and we denote the resulting statistic as t_{zx}^{EW}.
- We establish that t_{zx}^{EW} statistic has a standard normal limiting null distribution even under unconditional and/or conditional heteroskedasticity of the form specified in Assumption 3, regardless of whether x_t is strongly or weakly persistent. KMS have previously shown that this result holds under unconditional homoskedasticity and for the form of conditional heteroskedasticity they assume discussed above. Local limiting power functions also established.

Subsample Implementation of the KMS Test I

- If it were known that a *pocket of predictability* might occur only over the particular subsample $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$, such that $b(t/T) = b$ 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. We therefore base tests on forward and reverse recursive sequences and rolling sequences.

Subsample Implementation of the KMS Test II

- 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.
- The subsample IVX tests we propose based on these sequences of subsample statistics are then formally defined as follows. 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.

Subsample Implementation of the KMS Test III

- 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*,

$$\mathcal{T}_U^F := \max_{\tau_L \leq \tau \leq 1} \{t_{zx}(0,\tau)\}. \tag{8}$$

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 $\mathcal{T}_2^F$.

Subsample Implementation of the KMS Test IV

- 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,

$$\mathcal{T}_U^B := \max_{0 \leq \tau \leq \tau_U} \{t_{zx}(\tau, 1)\} \tag{9}$$

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.

Subsample Implementation of the KMS Test V

- 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,

$$\mathcal{T}_U^R := \max_{0 \leq \tau \leq 1-\Delta\tau} \{t_{zx}(\tau, \tau + \Delta\tau)\} \tag{10}$$

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.

Subsample Implementation of the KMS Test VI

Demetrescu *et al.* (2021) also consider tests for episodic predictability based on the maxima from corresponding sequences of rolling and recursive subsample implementations of a 2SLS predictability statistic as discussed by Breitung and Demetrescu (2015). As a necessary consequence of over-identified IV inference with strictly exogenous instruments, the approach proposed in Demetrescu *et al.* (2021) can only be used to test against two-sided alternatives, while as we have seen the subsample IVX-based tests considered in this paper can be used to test against either one-sided or two-sided alternatives. Where, as is often the case, theory predicts the sign of the slope parameter on x_{t-1} under predictability, being able to consider one-sided tests will clearly deliver tests with greater power relative to two-sided testing.

Subsample Implementation of the KMS Test VII

Limiting distribution theory is developed in the paper for the subsample IVX tests under the same set of assumptions as for the full sample IVX tests. The main take-aways from these limiting results are that the limiting null distributions of the subsample tests defined above:

1. Do not depend on the magnitude of ρ under Assumption 1.1 or the mean-reversion parameter c under Assumption 1.2 (in both cases asymptotic local power does though).
2. Depend in general on any heteroskedasticity present, despite being based on White standard errors. They also depend on any serial correlation present in $(u_t, v_t)'$ in the weakly persistent case.
3. Have different functional forms depending on whether x_t is near-integrated or stable.

The last two features pose significant problems for conducting inference that are not encountered with tests based on the full sample IVX test. However, these issues can all be solved by using bootstrap methods.

Bootstrap IVX Tests I

- We will explore two bootstrap resampling schemes. The first, a residual wild bootstrap [RWB]. The second is the fixed regressor wild bootstrap [FRWB] employed for 2SLS IV tests by Demetrescu *et al.* (2021). We show that both of these are first-order asymptotically valid.
- The approaches can be applied to any of the full sample and subsample-based IVX statistics discussed above. They can be implemented with either regular standard errors or White standard errors (the wild bootstrap obviates the need for White ses when heteroskedasticity is present).

A Residual Wild Bootstrap

1. Fit the predictive regression 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 $(u_t^*, v_t^*)' := (R_t \hat{u}_t,\ R_t \hat{v}_t)'$, $t = 1 \ldots, T$, where R_t, $t = 1, ..., T$, is a scalar $i.i.d.(0, 1)$ sequence with $\mathrm{E}(R_t^4) < \infty$, 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:

$$z_0^* = 0 \quad \text{and} \quad z_t^* = \sum_{j=0}^{t-1} \varrho^j \Delta x_{t-j}^*, \; t = 1, \ldots, T,$$

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

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

A Fixed-Regressor Wild Bootstrap

1. Construct the wild bootstrap innovations $y_t^* := \hat{y}_t R_t$, 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 $\left(y_t^*, x_{t-1}, \boldsymbol{z}_{t-1}'\right)'$, in place of the original sample data $\left(y_t, x_{t-1}, \boldsymbol{z}_{t-1}'\right)'$, construct the bootstrap analogues of the IVX statistics.

Bootstrap IVX Tests II

- A key difference between the RWB and FRWB surrounds the generation of the bootstrap analogue data for x_t and z_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 R_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_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_{t-1}, and their bootstrap analogues instrument x_{t-1}^* by z_{t-1}^*. But both z_t and z_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.

Bootstrap IVX Tests III

- In practice the autoregressive lag truncation order used in the second step of the RWB will be unknown. This can be selected in the usual way using a consistent information criterion such as the Bayes Information Criterion (BIC) or Hannan-Quinn [HQ] information criterion. A less parsimonious information criterion, such as the Akaike Information Criterion [AIC] could also be used, or even a deterministic truncation lag chosen according to, for example, the popular Schwert (1989) rule where the lag truncation is set equal to $\lfloor \kappa(T/100)^{1/4} \rfloor$, for some positive constant κ.
- The lag length fitted in the second step of the RWB turns out to have rather little bearing on the power of the resulting bootstrap tests. No choice of p is required in connection with the FRWB.

Bootstrap IVX Tests III

- We show that both bootstraps correctly replicate the first order asymptotic null distributions of the IVX statistics under both the null hypothesis and local alternatives. However, *in the case where x_t is weakly persistent*, for the RWB-based tests this result requires a further restriction to hold on the fourth moments of the innovations. This additional restriction is not required for the asymptotic validity of the FRWB tests.
- Specifically, we require that fourth moments of the form $\mathrm{E}[(\boldsymbol{\psi}_1\boldsymbol{\psi}_1') \otimes (\boldsymbol{\psi}_{-i}\boldsymbol{\psi}_{-j}')]$ are zero for all natural $i \neq j$. Although this condition is not tied to any specific parameteric model, a well known class of models which violate this condition are GARCH models with non-zero leverage effects.
- The RWB does not replicate the contribution of these moment terms to the quadratic variation of a key limiting process in the weakly dependent case. The FRWB implicitly replicates these terms.

Moving on to ...

1. Background and Summary of Contributions
2. The (Episodic) Predictive Regression Model
3. IVX Predictability Tests
4. Finite Sample Simulations
5. Concluding Remarks

Monte Carlo Results I

Case 1: Empirical Size: Scalar Predictor, IID errors

- DGP (1)-(2) with $\beta_t = 0$ for all t. Set $\alpha = \mu_x = 0$ 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 $\boldsymbol{\Sigma} := \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}$ and $\phi = -0.95$
- 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. 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 II

Case 2: Empirical Size: Scalar Predictor, ARCH with leverage

- Next a case which violates the regularity conditions for validity of the RWB when x_t is weakly persistent, where the conditional variance of $(u_t, v_t)'$ follows a stationary ARCH model with leverage effects:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0 \\ \rho x_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} 0 \\ \rho x_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \psi_t$$

$$\psi_t = \begin{pmatrix} a_t \\ e_t \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t}\sqrt{1 + \frac{1}{2}a_{t-1}^2 \mathbb{I}_{\{a_{t-1}<0\}}} \\ \varepsilon_{2t} \end{pmatrix}$$

  and $(\varepsilon_{1t}, \varepsilon_{2t})' \sim NIID(\mathbf{0}, \mathbf{I_2})$.
- The AR parameter ρ is again set equal to $1 - c/T$ with $c \in \{5, 10, 25, ..., 250\}$
- All computational aspects as for Case 1.

Table 4: Size of left-sided Tests. ARCH innovations with leverage.

c	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}
5	0.062	0.099	0.106	0.107
10	0.059	0.088	0.096	0.099
25	0.059	0.076	0.081	0.092
50	0.058	0.067	0.074	0.089
75	0.060	0.062	0.069	0.090
100	0.060	0.061	0.067	0.089
125	0.059	0.060	0.066	0.088
150	0.059	0.058	0.063	0.085
200	0.057	0.056	0.060	0.082
250	0.053	0.053	0.057	0.078

Table 5: Size of right-sided Tests. ARCH innovations with leverage.

c	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}
5	0.058	0.014	0.015	0.015
10	0.059	0.021	0.022	0.024
25	0.061	0.030	0.030	0.039
50	0.062	0.037	0.038	0.053
75	0.061	0.039	0.041	0.061
100	0.059	0.037	0.040	0.065
125	0.057	0.039	0.041	0.067
150	0.059	0.040	0.042	0.071
200	0.056	0.041	0.045	0.072
250	0.054	0.043	0.047	0.075

Table 6: Size of two-sided Tests. ARCH innovations with leverage.

c	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}
5	0.053	0.055	0.065	0.063
10	0.051	0.053	0.060	0.062
25	0.056	0.052	0.057	0.069
50	0.059	0.050	0.056	0.081
75	0.061	0.051	0.057	0.090
100	0.063	0.053	0.058	0.095
125	0.062	0.051	0.059	0.098
150	0.060	0.051	0.057	0.097
200	0.058	0.050	0.056	0.096
250	0.053	0.051	0.054	0.092

Monte Carlo Results III

Case 3: Empirical Size: Multiple Predictors

- The multiple predictor simulation DGP we use is as in Xu and Guo (2020):

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

  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

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_u^2 & \sigma_{u,v_1} & 0 & \cdots & 0 \\ \sigma_{u,v_1} & \sigma_{v_1}^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_{v_2}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_{v_K}^2 \end{pmatrix} \quad (11)$$

  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.
- We report the empirical sizes of the Wald tests for the joint significance of the K predictors. NB RWB uses obvious VAR generalisation of Step 2.

Table 7: 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 8: 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 9: Size of joint Wald Tests. $K = 10$ predictors.

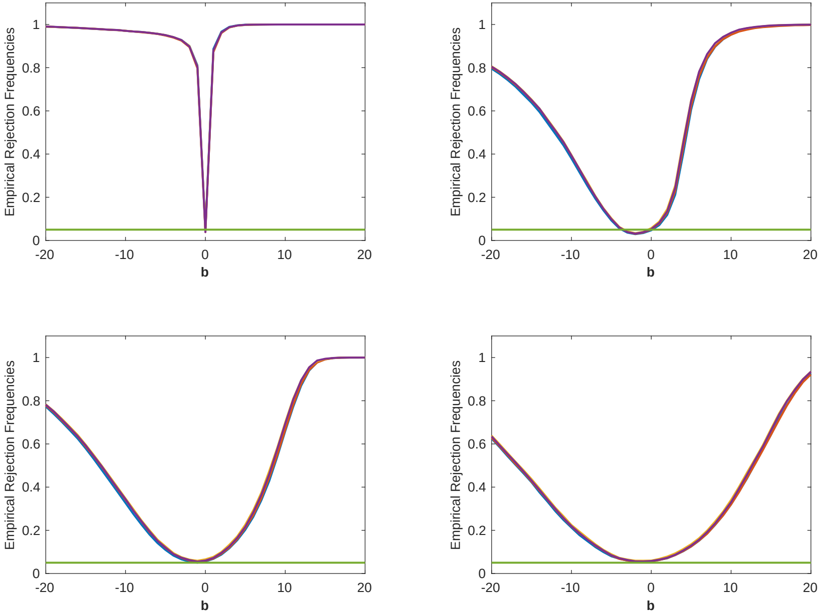
c	$W_{zx}^{*,\mathsf{RWB}}$	$W_{zx}^{*,\mathsf{FRWB}}$	W_{zx}^{EW}	W_{zx}
-5	0.058	0.513	0.635	0.559
-2.5	0.072	0.398	0.505	0.425
0	0.087	0.306	0.406	0.324
2.5	0.075	0.238	0.342	0.262
5	0.067	0.191	0.301	0.225
10	0.060	0.141	0.244	0.175
25	0.050	0.089	0.174	0.118
50	0.048	0.067	0.142	0.091
75	0.046	0.060	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

Monte Carlo Results IV

Case 4: Empirical Power: Scalar Predictor, IID errors

- DGP (1)-(2) with $\beta = b/T$, with the following values considered for the Pitman drift parameter, $b \in \{-20, -19, ..., 19, 20\}$.
- All other aspects as for Case 1.
- Results reported for four values of the persistence parameter, c, associated with x_t; specifically, $c = \{-5, 0, 10, 20\}$.
- All computational aspects as for Case 1.

Figure 1: Empirical Power of two-sided tests, $c = -5, 0, 10, 25$



Moving on to ...

1. Background and Summary of Contributions
2. The (Episodic) Predictive Regression Model
3. IVX Predictability Tests
4. Finite Sample Simulations
5. Concluding Remarks

Summing up

- Inference in predictive regressions with regressors of uncertain persistence is highly challenging!
- IVX tests of KMS are very popular but make overly restrictive assumptions on the innovations (notably that the predictive regression error follows a GARCH process) and can have poor finite sample size control when the predictor is strongly persistent and endogenous.
- We show that the IVX tests with White standard errors are valid under much weaker conditions (no need to assume a GARCH model for the errors).
- Bootstrap implementations are shown to considerably improve finite sample properties.
- Subsample IVX tests developed with bootstrap implementation.