Journal of Econometrics 237 (2023) 105271

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Extensions to IVX methods of inference for return predictability☆

Matei Demetrescu [a], Iliyan Georgiev [b,c], Paulo M.M. Rodrigues [d,*], A.M. Robert Taylor [e]

[a] Department of Statistics, TU Dortmund University, Germany
[b] Department of Economics, University of Bologna, Italy
[c] Institut d'Anàlisi Econòmica, CSIC, Spain
[d] Banco de Portugal and NOVA School of Business and Economics, Portugal
[e] Essex Business School, University of Essex, United Kingdom

ARTICLE INFO

Article history:
Received 30 January 2021
Received in revised form 15 February 2022
Accepted 19 February 2022
Available online 18 April 2022

JEL classification:
C12
C22
G17

Keywords:
Predictive regression
IVX estimation
(Un)conditional heteroskedasticity
Subsample tests
Unknown regressor persistence
Endogeneity
Residual wild bootstrap

ABSTRACT

The contribution of this paper is threefold. First, we demonstrate that, provided either a suitable bootstrap implementation is employed or heteroskedasticity-consistent standard errors are used, the IVX-based predictability tests of Kostakis et al. (2015) retain asymptotically valid inference under the null hypothesis under considerably weaker assumptions on the innovations than are required by Kostakis et al. (2015). Second, under the same assumptions, we develop asymptotically valid bootstrap implementations of the IVX tests. Monte Carlo simulations show that the bootstrap tests deliver considerably more accurate finite sample inference than the asymptotic implementations of the tests under certain problematic parameter constellations, most notably for one-sided testing, and where multiple predictors are included. Third, we show how sub-sample implementations of the IVX approach can be used to develop asymptotically valid one-sided and two-sided tests for the presence of temporary windows of predictability.

1. Motivation

Predictive regression methods are a very important part of the statistical toolbox used in empirical finance, providing a framework to investigate whether a given series can be predicted by other lagged financial and macroeconomic variables. Two important applications are attempts to predict returns on financial assets, most notably equity returns (see, for example, the literature review in Campbell and Yogo, 2006), and developing regression-based tests for efficiency in foreign exchange markets (see Fama, 1984). In both of these applications the returns variable we wish to predict resembles a (near) martingale difference sequence (MDS), while the predictors used are often characterised by high persistence with

☆ The authors thank three anonymous referees, the Co-Editor (Torben Andersen), and Tassos Magdalinos for their helpful and constructive feedback on earlier versions of this paper. Rodrigues gratefully acknowledges financial support from the Portuguese Science Foundation (FCT) through project PTDC/EGE-ECO/28924/2017, and (UID/ECO/00124/2013 and Social Sciences DataLab, Project 22209), POR Lisboa, Portugal (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte, Portugal (Social Sciences DataLab, Project 22209). Taylor gratefully acknowledges financial support provided by the Economic and Social Research Council of the United Kingdom under research grant ES/R00496X/1.

* Correspondence to: Banco de Portugal, Economics and Research Department, Av. Almirante Reis, 71-6th Floor, 1150-012 Lisbon, Portugal.
E-mail address: pmrodrigues@bportugal.pt (P.M.M. Rodrigues).

https://doi.org/10.1016/j.jeconom.2022.02.007

a significant correlation existing between the predictive regression error and the innovations driving the predictors; see, among others, Campbell and Yogo (2006), Welch and Goyal (2008), Nelson and Kim (1993), Stambaugh (1999) and Pavlidis et al. (2017). In these circumstances, standard regression estimation and inference methods, including conventional regression t-tests, are rendered invalid; see, among others, Cavanagh et al. (1995), Campbell and Yogo (2006), Jansson and Moreira (2006) and Phillips and Magdalinos (2008).

As a result, a number of likelihood-based procedures have been developed for the case where the predictor is endogenous and displays strong persistence within the local-to-unity class of processes; see, in particular, Cavanagh et al. (1995), Campbell and Yogo (2006), Jansson and Moreira (2006) and Elliott et al. (2015). Excepting Elliott et al. (2015), a major practical drawback with these approaches is that they are invalid if the predictor is weakly persistent. An alternative approach is to base predictability tests on methods of estimating the predictive regression which are robust to the properties of the regressor. Various approaches have been considered, arguably the most successful is Kostakis et al. (2015) who estimate the predictive regression using the extended instrumental variable [IVX] procedure of Phillips and Magdalinos (2009); see also, Phillips and Lee (2013), Breitung and Demetrescu (2015), Lee (2016), Demetrescu and Hillmann (2022) and Demetrescu et al. (2022), among others. In the IVX approach each predictor in the predictive regression has an associated stochastic instrument formed by constructing a mildly integrated variable from the first differences of the predictor. The IVX instrument, by construction, has lower persistence than a near-integrated variable and, as a consequence, delivers predictability statistics with asymptotically pivotal limiting null distributions.

Kostakis et al. (2015) establish, under certain regularity conditions on the model innovations, an asymptotic mixed normality result for their IVX estimator and show that the associated IVX-based predictability statistics possess standard (pivotal) limiting null distributions, regardless of whether the predictor is local-to-unity or weakly dependent (stationary). The asymptotic theory for IVX predictability statistics can, however, provide a very poor approximation to their finite sample behaviour, particularly for highly persistent and endogenous predictors which, as noted above, is arguably the case of most practical relevance. To ameliorate these finite sample distortions from the asymptotic theory, Kostakis et al. (2015) suggest a finite sample modification to the standard errors used in computing the IVX statistics. While this correction appears to work well for tests against two-sided alternatives reported in the simulation study for the case of a single regressor in Kostakis et al. (2015), as we will show in this paper, tests against one-sided alternatives remain very badly size-distorted for highly persistent and endogenous regressors. Moreover, Xu and Guo (2021) present simulation evidence which suggests that the quality of the asymptotic distributional approximation under the null, even with the finite sample correction employed, also markedly deteriorates as the number of regressors in the predictive regression increases.

The regularity conditions adopted in Kostakis et al. (2015) include an assumption of unconditional homoskedasticity in the vector of innovations driving the predictive model. Kostakis et al. (2015) allow for some forms of conditional heteroskedasticity in the innovation vector (provided heteroskedasticity-consistent standard errors are used) although these conditions are rather restrictive in practice. In particular, while a relatively weak martingale difference assumption is placed on the innovations driving the regressors, the errors in the predictive regression equations are assumed to follow a finite-order parametric GARCH model. The latter precludes the conditional variance of the regression errors, as a function of the past, from involving any direct contributions of the lagged values of the innovations driving the predictors, a likely unrealistic restriction for many commonly posited predictors of stock returns. Moreover, while GARCH models are very widely used in empirical finance, their usefulness for returns data is not uncontentious; e.g., Carnero et al. (2004) argue that the class of autoregressive stochastic volatility [ARSV] models is better suited to capturing the main empirical properties of the volatility of financial returns series.

A major contribution of this paper is to address the foregoing issues with the practical implementation of the IVX tests. First, regarding the regularity conditions needed, we show that the IVX predictability statistics of Kostakis et al. (2015) (implemented with heteroskedasticity-consistent standard errors) continue to admit standard pivotal limiting null distributions (again regardless of the degree of persistence or endogeneity of the regressors) under essentially the same set of regularity conditions on the innovations as adopted by Demetrescu et al. (2022) for establishing that the (over-identified) two-stage least squares (2SLS) based predictability test statistics of Breitung and Demetrescu (2015) have standard pivotal limiting null distributions.[1] These conditions allow for quite general patterns of unconditional time heteroskedasticity in the innovations, allowing for time-varying innovation variances and the possibility of time-varying correlations between the innovations. The conditions also allow for a much larger martingale difference class of innovations than considered in Kostakis et al. (2015) with no need to exclude interdependence between the conditional variances of the innovations in the model. Moreover, no parametric model needs to be assumed for either the conditional or unconditional time-variation in the innovations.

The 2SLS tests of Breitung and Demetrescu (2015) are based on an (over-identified) regression where two instruments are used for each predictor: a Type-I instrument which by design has a lower degree of persistence than the predictor (an example is the IVX instrument), and a strictly exogenous Type-II instrument (such as a sine function of time). The Type-I (Type-II) instrument is asymptotically dominated under strong (weak) persistence by the type-II (Type-I) instrument; see Demetrescu et al. (2022). Consequently, to establish the limiting distribution of the 2SLS statistic under strong persistence one does not need to determine the large sample properties of the IVX instrument under strong persistence,

[1] The conditions we adopt are *higher-level* assumptions than those in Demetrescu et al. (2022) and, as such, avoid the *ad hoc* rate condition imposed on the fourth (mixed) moments by the latter; see Remark 5.

only the large sample behaviour of the Type-II instrument is required and this is a relatively easy task under the regularity conditions on the innovations considered in this paper; see Demetrescu et al. (2022). In contrast, to develop limiting distribution theory for the (just-identified) IVX statistics of Kostakis et al. (2015) one must establish invariance principles and asymptotic independence results (see Eq. (11)) for the IVX-filtered predictor in the strongly persistent case, which is a much more involved task. Showing that the allowable regularity conditions to establish these results can be weakened from those adopted by Kostakis et al. (2015) to essentially the same (albeit higher-level) conditions as Demetrescu et al. (2022) use for the 2SLS statistics constitutes a major technical contribution to the literature.

Establishing that the 2SLS tests of Breitung and Demetrescu (2015) and the IVX tests of Kostakis et al. (2015) are asymptotically valid under the same set of regularity conditions implies the practitioner can choose which of these tests to use, agnostic of the properties of the data. This has important practical ramifications. First, as shown in, among others, Harvey et al. (2021, pp. 208–212) for the univariate case, two-sided IVX tests are more powerful than the 2SLS tests with certain type-I instruments, and increasingly so as the persistence of the predictor decreases. Second, the randomness of the sign of the correlation between the Type-II instrument and the predictor entails that the 2SLS tests can only be validly implemented as two-sided tests (one reason for this is given in Remark 4 of Breitung and Demetrescu, 2015, p. 364), while the IVX tests can be validly implemented as either one-sided or two-sided tests. This forms an important practical distinction between the tests as several studies have found that imposing an economically motivated structure, such as a known slope sign, on the predictive regression model can lead to better and more accurate findings of return predictability. For example, Campbell and Thompson (2008) find that, among other things, imposing positive predictability (so that the sign of the predictor is imposed to be positive under the alternative) almost always improves the out-of-sample predictability obtained for the predictors considered for equity returns in Welch and Goyal (2008). Similarly, in a Bayesian setting, Pettenuzzo et al. (2014) find that imposing a non-negative prior on the Sharpe ratio aids identification of return predictability. Moreover, Pavlidis et al. (2017) consider an application of predictive regression to testing for bubbles in foreign exchange markets where a natural positive sign restriction applies under the alternative hypothesis.

Second, and associatedly, to improve on their finite sample performance we also discuss bootstrap implementations of the IVX tests which are shown to be asymptotically valid under the same set of regularity conditions on the innovations. Although there are papers already in the literature that consider the problem of bootstrapping mildly integrated variables, see Fan and Lee (2019) and Smeekes and Westerlund (2019), neither of these are capable of allowing for the generality of time-variation in the variance matrix of the vector of innovations we consider here. Moreover, neither of these approaches is concerned with partial-sums based statistics. More relevant to the IVX tests of Kostakis et al. (2015) considered in this paper, Demetrescu et al. (2022), develop subsample implementations of the 2SLS-based predictability tests of Breitung and Demetrescu (2015) and base inference on a fixed regressor wild bootstrap [FRWB] resampling scheme. In this approach the regressor (and instrument in the case of Breitung and Demetrescu, 2015) is treated as fixed in the resampling exercise, while the series being predicted is resampled using a wild bootstrap. Demetrescu et al. (2022) demonstrate that the FRWB approach correctly replicates the first-order limiting null distributions of the temporary predictability statistics they propose under both conditional and unconditional heteroskedasticity. The FRWB is also used by Georgiev et al. (2018, 2019) who develop bootstrap tests for structural change in the predictive regression model.

The FRWB can also be used to successfully replicate the first-order limiting null distribution of the full sample IVX statistics under the conditions on the innovations considered in this paper. However, in Monte Carlo simulations we find that it does not address the finite sample distortions with the asymptotic IVX tests discussed above, most notably the distortions that occur when the regressor is highly persistent and endogenous. This is perhaps unsurprising given that the FRWB does not replicate in the bootstrap data the contemporaneous correlation present between the model's innovations. We therefore also discuss an alternative residual wild bootstrap [RWB] resampling scheme which is designed to replicate this correlation. Here we jointly wild resample the residuals from the fitted predictive regression model and a parametric autoregressive model fitted to the predictor. We also investigate the conditions under which the RWB-based IVX predictability tests are first-order asymptotically valid, and show that these deliver substantial improvements in finite sample behaviour relative to the asymptotic IVX tests.

Although the main application of the IVX methodology has been to predictive regressions for forecasting stock returns, it has also recently been applied to Fama regressions in the context of detecting episodic bubble-type behaviour in foreign exchange markets by Pavlidis et al. (2017) who consider a rolling subsample-based implementation of one-sided (right-tailed) IVX tests of Kostakis et al. (2015) proposing a test which rejects the null hypothesis of no bubble if any of the subsample statistics in the rolling sequence exceeds a given critical value. To avoid multiple testing bias, Pavlidis et al. (2017) base their approach on a conservative critical value obtained using a Bonferroni correction which adjusts the nominal significance level for the number of statistics in the rolling sequence. Pavlidis et al. (2017) note that this approach is likely to deliver a conservative test and suggest that a bootstrap implementation might deliver more powerful size controlled tests.

Tests based on the suprema of rolling and recursive subsample sequences of the 2SLS statistics of Breitung and Demetrescu (2015) have also been implemented recently in the context of detecting temporary periods of stock return predictability (so-called *pockets of predictability*) by Demetrescu et al. (2022). As noted above, Demetrescu et al. (2022) use a FRWB to implement these tests. The final contribution of this paper is to 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 and that these are asymptotically valid under the same regularity conditions on the innovations as are required for the

corresponding bootstrap implementations of the full sample tests. As with the full sample tests, this allows practitioners to implement one-sided tests for temporary windows of predictability allowing for cases where the direction of predictability under the alternative is known, as with the foreign exchange rate bubble testing problem considered in Pavlidis et al. (2017), something not possible with the 2SLS-based tests of Demetrescu et al. (2022). This also ensures that a rejection of the null is only associated with a window of predictability with the anticipated sign. In the Fama-regression setting, as discussed by Pavlidis et al. (2017), the slope parameter is likely to be estimated to be negative over many subsamples of the data and so the tests of Demetrescu et al. (2022) are likely to reject in practice even when no bubble is present.

The remainder of the paper is organised as follows. Section 2, introduces the predictive regression model we consider together with the assumptions needed for our analysis. Section 3 reviews the full sample IV-based predictability tests of Kostakis et al. (2015) and details subsample implementations of these tests. Representations for the limiting distributions of the underlying test statistics under both the null and local alternatives are provided. These are shown to depend in general on any heteroskedasticity present. Moreover, the form of these limiting distributions depends on whether the predictor is weakly or strongly persistent, even under homoskedasticity. In the context of the full sample IVX statistic, however, the use of Eicker–White standard errors is shown to deliver a standard pivotal limiting null distribution regardless of the predictor's persistence. Section 4 discusses bootstrap implementations of the IVX tests and demonstrates the first-order asymptotic validity of these. Section 5 presents the results from a Monte Carlo analysis into the finite sample behaviour of both the full sample and subsample IVX tests, while empirical applications to stock returns and exchange rate data are reported in Section 6. Concluding comments including some suggestions for further research are provided in Section 7. Detailed proofs of the technical results given in the paper along with other supporting material appear in a supplementary appendix.

In what follows we use $\mathbb{I}(\cdot)$ to denote the indicator function, equal to one when its argument is true, zero otherwise, and $\|\cdot\|$ to denote the matrix norm $[Trace((\cdot)'(\cdot))]^{1/2}$. We denote by $\mathcal{D}^k$ the space of càdlàg real functions on $[0,1]^k$ equipped with the Skorokhod topology, and abbreviate $\mathcal{D}^1$ to $\mathcal{D}$. The weak convergence of probability measures on $\mathcal{D}^k$ and on $\mathbb{R}^k$ is denoted by $\Rightarrow$. We use the notation P, E, etc. for probability, expectation etc. with respect to the distribution of the original data and use P^*, E^*, etc. for probability, expectation, etc. induced by the data and the wild bootstrap multipliers (which we shall denote $\{R_t\}$) conditionally on the data. If w_T, w ($T \in \mathbb{N}$) are random elements of metric spaces, the weak-in-probability convergence $w_T \overset{w}{\Rightarrow}_p w$ means that $E^* f(w_T) \xrightarrow{p} Ef(w)$ for all continuous bounded real functions with matching domain. Finally, the O_p and o_p symbols have their usual meaning.

2. The predictive regression model and assumptions

Consider the predictive regression model for returns, y_t, allowing for time-variation in the slope coefficient on a lagged predictor, x_{t-1}, of the form

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

where x_t satisfies the additive component model

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

$$\xi_t = \rho \xi_{t-1} + w_t, \quad A(L)\, w_t = v_t, \qquad t = 1, \ldots, T, \tag{3}$$

in which u_t and v_t are serially uncorrelated (martingale difference [MD]) innovations, precise conditions on which are given in Assumption 3, and $A(L) := (1 - a_1 L - a_2 L^2 - \cdots - a_p L^p)$ is a stable autoregressive polynomial in the conventional lag operator, L. We define $\omega := 1/A(1)$ and, for the case where x_t also follows a stable autoregression, we let κ^2 denote the sum of the squared coefficients of the filter $((1 - \rho L)A(L))^{-1}$. In our exposition and technical analysis we follow the bulk of the literature and focus attention on the case of a single predictor, so that x_{t-1} in (1) is a scalar. Extensions to allow for multiple predictors will be discussed at various points in the text, although we leave a detailed treatment of this case for future research.

The DGP in (1) generalises the constant parameter predictive regression model considered in Kostakis et al. (2015) by allowing for the possibility that the slope coefficient on x_{t-1} varies over time, allowing for changes over time in the predictive content of the regressor x_{t-1}. The constant parameter predictive regression model obtains by setting a constant slope parameter such that $\beta_t = \beta$, for all $t = 1, \ldots, T$. The tests we consider in this paper are all for the null hypothesis, H_0, that $(y_t - \alpha)$ is a MD and, hence, that y_t is not predictable by x_{t-1}, which entails that $\beta_t = 0$, for all $t = 1, \ldots, T$, in (1).[2] The full-sample IVX tests of Kostakis et al. (2015) test the same null hypothesis, H_0, against the alternative that y_t is predictable by x_{t-1} with a constant slope parameter holding across the whole sample; that is, $\beta_t = \beta \neq 0$ for all $t = 1, \ldots, T$. The subsample implementations of IVX we discuss will be used to test against alternatives such that $\beta_t \neq 0$ for some t but without imposing constancy on β_t. In any case, some structure needs to be placed on the class of alternative hypotheses we may consider and this will be formalised below.

The degree of persistence of the regressor, x_t, is controlled via the parameter ρ. We allow x_t to be either weakly or strongly persistent through the following assumption.

[2] The methods which we outline in this paper could equally well be used to test the null hypothesis that $\beta_t = \beta_0$ for all $t = 1, \ldots, T$, but as the focus in both equity forecasting and Fama regressions is on testing the null hypothesis of a zero coefficient on the lagged predictor we will restrict our discussion to $\beta_0 = 0$.

Assumption 1. $A(L)$ is a finite-order ($p < \infty$) polynomial with all of the roots of $A(z) = 0$ lying outside the complex unit circle, $|z| = 1$. The initial condition, ξ_0, is a mean zero $O_p(1)$ variate. Moreover, exactly one of the two following conditions holds on the autoregressive parameter ρ in (3):

1. **Weakly persistent regressor:** ρ is fixed and bounded away from unity, $|\rho| < 1$.
2. **Strongly persistent regressor:** ρ is parameterised to be local-to-unity; that is, $\rho := 1 - cT^{-1}$, where c is a finite constant. This allows for pure $I(1)$ predictors ($c = 0$), locally stable predictors ($c > 0$), and locally explosive predictors ($c < 0$).

Remark 1. Assumption 1 imposes that the errors w_t in (3) follow a finite-order autoregression. This is imposed for the purposes of facilitating the RWB implementations of the full sample and subsample IVX tests in Section 4. Asymptotic versions of these tests (i.e. tests based on critical values from the limiting null distributions of the statistics) could equally well be based on a linear process assumption for w_t of the form considered in Assumption INNOV of Kostakis et al. (2015, p. 1512) or the slightly weaker Assumption M of Magdalinos (2020); in particular, Proposition 1 of this paper would remain valid. The FRWB implementations of the IVX tests discussed in Section 4 would also be asymptotically valid under a linear process assumption of this form. ◊

Remark 2. We follow the bulk of the predictive regression literature in considering regressors that follow either stable (weakly dependent) processes, see Amihud and Hurvich (2004), or are near-integrated, see Campbell and Yogo (2006), without assuming knowing which of these holds. As we shall see, the limiting behaviour of the IVX statistics can differ under the two types of persistence, but this can be consistently replicated (to asymptotic first order) by the bootstrap procedures we propose. An alternative framework, which we do not consider here, is to characterise the persistence of the predictors as lying in the class of fractionally integrated processes. Important contributions in predictability testing with fractionally integrated predictors include, Maynard and Phillips (2001), Maynard and Shimotsu (2009), Bauer and Maynard (2012), and Andersen and Varneskov (2021a, 2021b). The approach taken in Andersen and Varneskov (2021a, 2021b) is based on the concept of the local spectrum (LCM) methodology and allows for multivariate regressors with any mix of (fractional) integration degrees along with strong forms of endogeneity. The LCM methodology can therefore be viewed as complementary to the IVX approach. Andersen and Varneskov (2021a, pp. 227–228) demonstrate that the LCM approach remains asymptotically valid in the case where the predictors are observed with measurement error which is of a lower degree of persistence than the true predictor. A similar result holds in the set-up considered here; in particular, with an $I(0)$ measurement error, the large sample results given in this paper will continue to hold for strongly persistent predictors satisfying Assumption 1.2. Within the framework of Andersen and Varneskov (2021a), Andersen and Varneskov (2021c) consider the case of "imperfect" predictors, in the sense of Pastor and Stambaugh (2009), where a component of the conditional mean of the returns series exists that is not linearly spanned by the chosen predictor(s). Georgiev et al. (2019) consider essentially the same setting but in the context of standard linear predictive regression tests with predictors satisfying Assumption 1. In such cases the standard predictability tests should be interpreted not as tests for a perfect linear relation, but rather as tests of linear predictive power. Indeed, both Andersen and Varneskov (2021c) and Georgiev et al. (2019) develop tests that allow practitioners to distinguish between the "imperfect" and "perfect" regressor scenarios. ◊

The basic idea underlying the IVX procedure of Phillips and Magdalinos (2009) is to instrument the regressor x_{t-1} by a variable of controlled persistence, constructed as

$$z_0 = 0 \quad \text{and} \quad z_t = (1 - \varrho L)^{-1}_{+} \Delta x_t := \sum_{j=0}^{t-1} \varrho^j \Delta x_{t-j}, \, t = 1, \ldots, T, \tag{4}$$

and where $\varrho := 1 - aT^{-\eta}$ with $a > 0$ and $0 < \eta < 1$. The IVX scale and exponent parameters, a and η respectively, are tuning parameters set by the practitioner; Kostakis et al. (2015) recommend $a = 1$ and $\eta = 0.95$. Where x_t is near-integrated satisfying Assumption 1.2, z_t is approximately a mildly integrated process and therefore of lower persistence than x_t. Moreover, where x_t is weakly dependent satisfying Assumption 1.1, we have that $z_t \approx x_t$. As a result, Kostakis et al. (2015) demonstrate that the IVX full-sample estimator of the slope parameter in (1) is asymptotically (mixed) Gaussian under H_0 and under their Assumption INNOV regardless of whether Assumption 1.1 or 1.2 holds and that, consequently, the full-sample instrumental variable tests for H_0 have standard limiting null distributions regardless of the degree of persistence or endogeneity of x_t.

For our purposes we follow Demetrescu et al. (2022) and conduct our theoretical analysis of the large sample properties of both the full-sample and sub-sample IVX predictability statistics under local alternatives such that the slope parameter β_t is local-to-zero for an asymptotically non-vanishing set of the sample observations. This is an important generalisation of the large sample results presented for the full sample IVX-based tests in Kostakis et al. (2015) and Magdalinos (2020) which only apply under H_0. The localisation rate (or Pitman drift) is such that β_t is specified to lie in a neighbourhood of zero which shrinks with the sample size, T, at a rate which depends on which of Assumption 1.1 and Assumption 1.2 holds in (3). Specifically,[3]

[3] Notice that while the Pitman drift rate considered in Assumption 2 is required to obtain non-degenerate asymptotic (local) limiting distributions for the IVX predictability statistics (it ensures that the left and right hand sides of the predictive regression in (1) are always asymptotically "balanced",

Assumption 2. In (1)–(3), let $\beta_t := n_T^{-1} b(t/T)$, where $b(\cdot)$ is a piecewise Lipschitz-continuous real function on [0, 1], with $n_T = \sqrt{T}$ under Assumption 1.1, and $n_T = T^{1/2+\eta/2}$ under Assumption 1.2, recalling that η, $0 < \eta < 1$, is the IVX exponent used in the construction of z_t in (4).

Under the structure of Assumption 2, the null hypothesis H_0 that $\beta_t = 0$, for all $t = 1, \ldots, T$, can be expressed as

$$H_0 : \textit{The function } b(\cdot) \textit{ is identically zero on } [0, 1], \tag{5}$$

while the alternative hypothesis can be written as

$$H_{1,b(\cdot)} : \textit{The function } b(\cdot) \textit{ is non-zero over at least one non-empty open subinterval of } [0, 1]. \tag{6}$$

The latter entails that 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 exists for which $\beta_t \neq 0$, and where the size of this subset is proportional to the sample size T. One-sided alternatives that $\beta_t > 0$ ($\beta_t < 0$) in some subset(s) of the data can be considered simply by defining $b(\cdot)$ to be a non-negative (non-positive) function.

We conclude this section by detailing in Assumption 3 the conditions we will place on the innovations u_t and v_t in (1) and (3), respectively. Subsequently we will discuss these conditions relative to other sets of regularity conditions that have been adopted in the literature, before providing the key multivariate invariance principles [MIPs] that hold under these conditions.

Assumption 3. Let

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} := \mathbf{H}\left(\frac{t}{T}\right)\begin{pmatrix} a_t \\ e_t \end{pmatrix},$$

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 MD sequence satisfying $\mathrm{E}(\boldsymbol{\psi}_t \boldsymbol{\psi}_t') = \mathbf{I}_2$ and

$$\mathrm{E}\left\| \mathrm{E}_0 \sum_{t=1}^{T} (\boldsymbol{\psi}_t \boldsymbol{\psi}_t' - \mathbf{I}_2) \right\|^2 = O(T^{2\epsilon}) \tag{7}$$

for some $\epsilon < \frac{1}{2}$, with $\mathrm{E}_\tau(\cdot)$, denoting expectation conditional on the σ-algebra generated by $\{\boldsymbol{\psi}_{\tau-i}\}_{i=0}^{\infty}$, and where $\mathbf{I}_k$ denotes the $k \times k$ identity matrix.

Remark 3. As discussed in Demetrescu et al. (2022), Assumption 3.1 allows for unconditional time heteroskedasticity of quite general form in the innovations through the function $\mathbf{H}$, the unconditional covariance matrix of $(u_t, v_t)'$ being $\mathbf{H}(t/T)\mathbf{H}'(t/T)$. This allows u_t and v_t to display time-varying unconditional variances and both contemporaneous and time-varying (unconditional) correlation between u_t and v_t, including single or multiple (co-) variance shifts, (co-)variances following a broken trend, and smooth transition (co-) variance shifts. In contrast, Assumption INNOV of Kostakis et al. (2015, p.1512) and Assumption M of Magdalinos (2020) impose a constant unconditional variance matrix on $(u_t, v_t)'$, but do allow for conditional (stochastic) heteroskedasticity. Assumption 3.2 imposes a MD structure on $\boldsymbol{\psi}_t$ thereby also allowing for conditional heteroskedasticity. In common with Assumption INNOV of Kostakis et al. (2015) and Assumption M of Magdalinos (2020), Assumption 3.2 imposes finite fourth-order moments on $\boldsymbol{\psi}_t$. ◊

Remark 4. A crucial difference between the IVX tests of Kostakis et al. (2015) and the 2SLS tests of Breitung and Demetrescu (2015) is that in order to establish the large sample properties of the former in the strong persistence case we need to establish a weak convergence result for the partial sum process, $\frac{1}{\sqrt{T^{1+\eta}}} \sum_{t=1}^{[\tau T]} z_{t-1} u_t$ (see (11)). This is not required for the over-identified 2SLS statistics because, as discussed in Section 1, the Type-II instrument used in the case of these statistics asymptotically dominates the Type-I instrument (e.g., the IVX instrument) under strong persistence. For the case of full-sample sums, Kostakis et al. (2015) and Magdalinos (2020) need to 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 consequence to preclude the conditional variance of u_t, as a function of the past, to involve contributions of lagged v_t, an arguably unrealistic restriction for many potential predictors of stock returns; see Example 1 in the supplementary

possessing the same order of magnitude in probability), such local-to-zero magnitude alternative models should not necessarily be taken to have any inherent economic meaning. Indeed, in the case of excess returns, asset valuation theory stipulates that valuation ratios should have a fixed rather than shrinking magnitude coefficient in the predictive regression; see, for example, Campbell and Thompson (2008), and the references therein, who derive specific results for the case of the dividend price ratio based on the Gordon (1962) growth model. Moreover, dynamic asset pricing models often allow the equity premium to depend on persistent state variables, as for example in the long run risk model of Bansal and Yaron (2004).

appendix for further discussion. 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. 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 Merlevède et al. (2006). The exponent ϵ controls the degree of persistence permitted in the conditional variances of the innovations. 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 as, for example, are assumed in Johannes et al. (2014) also satisfy Assumption 3.2. ◊

Remark 5. It is instructive to compare the regularity conditions in Assumption 3 with those used in Demetrescu et al. (2022) in the context of establishing the large sample behaviour of the 2SLS-based predictability statistics of Breitung and Demetrescu (2015) for the case where the Type-I instrument used is set to be the IVX filter in (4). In doing so, it is important to note that the relevant regularity conditions in Demetrescu et al. (2022) are spread across Assumptions 3–6 of that paper; see, in particular, Lemma 1 of Demetrescu et al. (2022). Both sets of conditions impose the conditions in Assumption 3.1 and also require that the MD sequence $\boldsymbol{\psi}_t$ (in the notation of this paper) defined in Assumption 3.2 is strictly stationarity and ergodic. There are, however, some differences concerning the restrictions they place on the amount of serial dependence allowed in the conditional second order moments of the sequence $\boldsymbol{\psi}_t$. In particular, while the current paper assumes the single "high-level" condition in (7), Demetrescu et al. (2022) require that

$$\sup_{t\in\mathbb{N}}\mathrm{E}\|\mathrm{E}_{t-m}(\boldsymbol{\psi}_t\boldsymbol{\psi}_t'-\mathbf{I}_2)\|\to 0 \quad \text{as} \quad m\to\infty, \tag{8}$$

and that

$$\sup_{t=1,\ldots,T;T\in\mathbb{N}}\left|\mathrm{E}\left(\left(v_t^2-\mathrm{E}(v_t^2)\right)v_{t-k}v_{t-j}\right)\right| \le \frac{C}{(jk)^{1/2+\vartheta/2}} \tag{9}$$

for some $\vartheta > 0$ and any $k, j > 0$, where v_t are the shocks to the predictor (the second element of $\mathbf{H}(t/T)\boldsymbol{\psi}_t$, denoted by $\tilde{v}_t$ in Demetrescu et al., 2022). These conditions appear similar, and indeed coincide for many popular parametric models, such as for example finite-order stationary vector GARCH models or infinite-order ARCH models, for which if $\boldsymbol{\psi}_t$ satisfies (7) then it will also satisfy (8) and (9), and *vice versa*. Crucially, condition (7) replaces both condition (8) and the *ad hoc* rate of decay condition on the fourth order mixed moments stipulated under (9), and also allows us to make a strong connection with the probabilistic literature; see the discussion in Remark 4. Compared to (9), the condition in (7) is more easily interpretable, allowing us to state that the cumulated sum of $\boldsymbol{\psi}_t\boldsymbol{\psi}_t'-\mathbf{I}_2$ is 'almost' a martingale. This type of assumption is arguably more appealing in the context of modelling financial data where finance theory often predicts that the innovations in the model should have the MD property, rather than the conventional approach, typified by the assumptions in Demetrescu et al. (2022) and taken from the classical time series literature, which is to impose generic rate of decay conditions on higher-order moments. ◊

Under Assumption 1.1 (weak persistence), $\xi_t = (1-\rho L)_+^{-1}A(L)^{-1}v_t + \rho^t\xi_0$ (recall that $(1-\rho L)_+^{-1} := \sum_{j=0}^{t-1}\rho^j L^j$), which, given the exponential decay of the coefficients, is asymptotically equivalent to the process $(1-\rho L)^{-1}A(L)^{-1}v_t$, and with a slight abuse of notation, we will write $\xi_t = (1-\rho L)^{-1}A(L)^{-1}v_t$ in what follows, ignoring the asymptotically negligible term. Given Assumption 3, the normalised partial sums of $(u_t, v_t, \xi_{t-1}u_t)$ then satisfy the MIP,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor\tau T\rfloor}\begin{pmatrix}u_t\\ v_t\\ \xi_{t-1}u_t\end{pmatrix} \Rightarrow \int_0^\tau \mathbf{G}(s)\mathrm{d}\boldsymbol{B}(s) := \begin{pmatrix}M_u(\tau)\\ M_v(\tau)\\ M_{\xi u}(\tau)\end{pmatrix} \tag{10}$$

on $\mathcal{D}^3$, where $\mathbf{G}(\tau)$ is a 3×6 matrix of piecewise Lipschitz functions whose elements are formed from the elements of $\mathbf{H}(\tau)$, and where $\boldsymbol{B}(\tau)$ is a 6-dimensional Brownian motion. Explicit expressions for the covariance matrix of $\boldsymbol{B}(\tau)$ and for $\mathbf{G}(\tau)$ are provided in Lemma 4 in the supplement, where the result in (10) is formally established. Using the Phillips-Solo device, it is straightforwardly obtained from (10) that the normalised partial sums of ξ_t weakly converge to $\omega/(1-\rho)M_v$.

Remark 6. The MIP in (10) coincides with that given in Equation (2.6) and Lemma 1.1 of Demetrescu et al. (2022) for a weakly persistent predictor. The limiting processes M_u, M_v and $M_{\xi u}$ are individually variance-transformed Brownian motions; cf. Davidson (1994, section 29.4). They are, in general, correlated under Assumption 3, and indeed this correlation can be time-varying; see the supplementary appendix for precise expressions. Under conditional homoskedasticity, $M_{\xi u}$ can be seen to be uncorrelated with either M_u or M_v. Under conditional heteroskedasticity, however, M_v and $M_{\xi u}$ are in general dependent (as are M_u and $M_{\xi u}$), even where $\mathbf{H}(\tau)$ is constant, because $\mathrm{Cov}\,(\xi_{t-1}u_t\,, v_t)$ is not necessarily zero if the conditional correlation between u_t and v_t is nonzero. Where $\mathbf{H}(\tau)$ is constant, such that $(u_t, v_t)'$ is unconditionally homoskedastic, $\int_0^\tau \mathbf{G}(s)\mathrm{d}\boldsymbol{B}(s)$ reduces to a usual Brownian motion process. Where $\mathbf{H}(\tau)$ is non-constant the variance profiles of M_u, M_v and $M_{\xi u}$ will, in general, differ (we define the variance profile of a generic stochastic process $W(s)$ as $[W](s)/[W](1)$ where $[W](s)$ denotes the quadratic variation process of $W(s)$). Even in the special case where $\mathbf{H}(\tau)$ is a scalar multiple of the identity matrix, although M_u and M_v will share the same variance profile, this will not in general coincide with variance profile of $M_{\xi u}$ because the variance of its increments is a polynomial of degree four in the elements of $\mathbf{H}(\tau)$, while those of M_u and M_v are both polynomials of degree two (see the proof of Lemma 4 in the supplement). ◊

Under Assumption 1.2 (strong persistence), the normalised partial sums of (u_t, v_t) converge as previously to (M_u, M_v), where M_u and M_v are the same limiting processes as in (10). Moreover,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[\tau T]}\begin{pmatrix} v_t \\ \frac{1}{\sqrt{T^{\eta}}} z_{t-1}u_t \end{pmatrix} \Rightarrow \begin{pmatrix} M_v(\tau) \\ M_{zu}(\tau) \end{pmatrix} \tag{11}$$

on $\mathcal{D}^2$, with $M_{zu}(\tau) := \frac{\omega}{\sqrt{2a}}\int_0^\tau \sqrt{[M_v]'(s)[M_u]'(s)}\mathrm{d}B(s)$, where B is a standard Brownian motion independent of M_v, and where $[M_v]'(s)$ and $[M_u]'(s)$ denote the derivatives (with respect to s) of $[M_v](s)$ and $[M_u](s)$, respectively. These derivatives are well-defined at all but finitely many $s \in [0, 1]$, see Lemma 3 in the supplementary appendix. Convergence (11) is established in Lemma 5 in the supplementary appendix. Under strong persistence, the levels of ξ_t satisfy the weak convergence result $T^{-1/2}\xi_{\lfloor \tau T \rfloor} \Rightarrow \omega J_{c,\mathrm{H}}(\tau)$ on $\mathcal{D}$, where $J_{c,\mathrm{H}}(\tau)$ is an Ornstein–Uhlenbeck-type process driven by $M_v(\tau)$; that is, $J_{c,\mathrm{H}}(\tau) := \int_0^\tau e^{-c(\tau-s)}\mathrm{d}M_v(s)$.

Remark 7. The limiting process M_{zu} in (11) is also a variance-transformed Brownian motion. An important difference between the MIPs in (10) and (11) is that M_{zu} is independent of M_v irrespective of any conditional heteroskedasticity while, as discussed in Remark 6, $M_{\xi u}$ and M_v are in general dependent. Another important difference is that the processes $M_{\xi u}$ and M_{zu}, despite being determined by the same innovations, can have quite different behaviour depending on the pattern of conditional and unconditional heteroskedasticity present in $\boldsymbol{\psi}_t$. To illustrate, under unconditional heteroskedasticity the variance profiles of $M_{\xi u}$ and M_{zu} will in general differ where conditional heteroskedasticity is also present; see Example 2 in the supplementary appendix. ◊

Remark 8. The MIP in (11) generalises the corresponding convergence results in Kostakis et al. (2015) and Magdalinos (2020) in two ways. First, as discussed in Remarks 3 and 4, it establishes the existence of a MIP under much weaker conditions on the innovations than are allowed in Kostakis et al. (2015) and Magdalinos (2020). Second, Kostakis et al. (2015) and Magdalinos (2020) only provide a convergence result for the full sample quantity $\frac{1}{\sqrt{T^{1+\eta}}}\sum_{t=1}^{T} z_{t-1}u_t$, whereas the MIP in (11) establishes the joint limiting distribution of the corresponding sequence of quantities across all possible subsamples. The result in (11) also provides the necessary keystone for deriving the large sample properties of statistics arising in other settings involving IVX instrumentation of strongly persistent variables, under much weaker conditions than the extant literature allows. ◊

3. IVX-based predictability tests

3.1. Full-sample IVX tests

The full-sample IVX-based t-ratio, proposed in Kostakis et al. (2015), for testing the null hypothesis $H_0 : \beta_t = 0$ for all $t = 1, \ldots, T$ in (1) is given by[4]

$$t_{zx} := \frac{\hat{\beta}_{zx}}{s.e.(\hat{\beta}_{zx})}, \quad \hat{\beta}_{zx} := \frac{\sum_{t=1}^{T} z_{t-1}(y_t - \bar{y})}{\sum_{t=1}^{T} z_{t-1}(x_{t-1} - \bar{x}_{-1})} \tag{12}$$

$$s.e.(\hat{\beta}_{zx}) = \frac{\sqrt{\hat{\sigma}_u^2 \sum_{t=1}^{T} z_{t-1}^2}}{\sum_{t=1}^{T} z_{t-1}(x_{t-1} - \bar{x}_{-1})} \tag{13}$$

with $\bar{y} := T^{-1}\sum_{t=1}^{T} y_t$, $\bar{x}_{-1} := T^{-1}\sum_{t=1}^{T} x_{t-1}$, and $\hat{\sigma}_u^2 := T^{-1}\sum_{t=1}^{T}\hat{u}_t^2$.[5] A variety of choices for the residuals $\hat{u}_t$ is possible. Breitung and Demetrescu (2015) and Kostakis et al. (2015) recommend the OLS residuals from estimating (1) on the grounds that they come from the best linear projection of y_t on x_{t-1} regardless of the persistence of the putative predictor, and that their finite-sample behaviour appears to be more stable than that of the corresponding IV residuals. One could also use residuals computed under the null; that is, $\hat{u}_t := y_t - \frac{1}{T}\sum_{s=1}^{T} y_s$. Under the local alternatives considered in Assumption 2, these two possible choices are asymptotically equivalent in so far as the behaviour of the resulting IVX statistic is concerned. The IV residuals also have reduced convergence rates compared to the two possible choices above, so we will not consider them further.

[4] As discussed in Kostakis et al. (2015, p.1514), $\hat{\beta}_{zx}$ is invariant to whether z_{t-1} is demeaned or not.

[5] To ameliorate the finite sample effects of estimating the intercept term in (1), Kostakis et al. (2015, p. 1516) recommend the use of a finite-sample correction term. This entails replacing the numerator of (13) by $\sqrt{\hat{\sigma}_u^2\sum_{t=1}^{T} z_{t-1}^2 - \Xi}$ where $\Xi := T\bar{z}_{-1}^2(\hat{\sigma}_u^2 - \hat{\sigma}_{uw}^2\hat{\sigma}_w^{-2})$, with $\bar{z}_{-1} := T^{-1}\sum_{t=1}^{T} z_{t-1}$, and where $\hat{\sigma}_w^2$ and $\hat{\sigma}_{uw}$ are estimates of the long-run variance of w_t, and of the long-run covariance between u_t and w_t, respectively; a discussion on the practical choice of these estimators is provided in Kostakis et al. (2015, pp. 1513 and 1524). The inclusion of the correction term, Ξ, does not alter any of the large sample results that follow.

Kostakis et al. (2015) also consider a variant of the t_{zx} statistic based on the use of heteroskedasticity-robust (Eicker–White) standard errors. This is given by

$$t_{zx}^{EW} := \frac{\hat{\beta}_{zx}}{s.e.^{EW}(\hat{\beta}_{zx})}, \quad s.e.^{EW}(\hat{\beta}_{zx}) := \frac{\sqrt{\sum_{t=1}^{T} z_{t-1}^2 \hat{u}_t^2}}{\sum_{t=1}^{T} z_{t-1}\,(x_{t-1} - \bar{x}_{-1})}. \tag{14}$$

As we show in Section 3.3, t_{zx}^{EW} has a standard normal limiting null distribution under unconditional and/or conditional heteroskedasticity satisfying Assumption 3, regardless of whether x_t is strongly or weakly persistent. Kostakis et al. (2015) and Magdalinos (2020) have previously shown that this result holds under unconditional homoskedasticity and for the form of conditional heteroskedasticity they assume which, as discussed in Section 2, is a special case of our Assumption 3.2. The same result holds for t_{zx} in the strongly persistent case when the innovations are unconditionally homoskedastic, but does not hold in general otherwise. The finite sample correction term Ξ discussed in footnote 5 can also be applied to the numerator of $s.e.^{EW}(\hat{\beta}_{zx})$ in (14).

One-sided tests based on either t_{zx} or t_{zx}^{EW} can be formed by rejecting against the right-sided alternative that $\beta_t = \beta > 0$, for all $t = 1, \ldots, T$, for large positive values of the statistics and against the left-sided alternative that $\beta_t = \beta < 0$, for all $t = 1, \ldots, T$, 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 $t = 1, \ldots, T$, for large positive values of either $(t_{zx})^2$ or $(t_{zx}^{EW})^2$.

Remark 9. Kostakis et al. (2015) consider the more general set-up of multiple predictive regressions of the form $y_t = \alpha + \boldsymbol{\beta}'\boldsymbol{x}_{t-1} + u_t$, $t = 1, \ldots, T$, where $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_K)'$ and where $\boldsymbol{x}_t := (x_{1,t}, \ldots, x_{K,t})'$ is such that $\boldsymbol{x}_t = \boldsymbol{\mu}_{\boldsymbol{x}} + \boldsymbol{\xi}_t$ where $\boldsymbol{\xi}_t$ satisfies the K-dimensional generalisation of (3), $\boldsymbol{\xi}_t = \boldsymbol{\Gamma}\boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t$, $t = 1, \ldots, T$, and where $\boldsymbol{\mu}_{\boldsymbol{x}}$ is a K-vector of constants. In common with Kostakis et al. (2015), the dimension K of $\boldsymbol{x}_t$ is assumed to be fixed (does not increase with T). Kostakis et al. (2015) specify the matrix $\boldsymbol{\Gamma}$ to be diagonal with ith diagonal element ρ_i, $i = 1, \ldots, K$, and assume that the predictors all lie within the same persistence class; that is, the $x_{i,t}$, $i = 1, \ldots, K$, either all satisfy Assumption 1.1, or they all satisfy Assumption 1.2. Generating the set of K instruments, $\boldsymbol{z}_t := (z_{1,t}, \ldots, z_{K,t})'$, from the predictors $x_{i,t}$, $i = 1, \ldots, K$, each generated according to (4), a two-sided Wald-type IVX based test rejects the null $\mathbf{R}\boldsymbol{\beta} = 0$, where $\mathbf{R}$ is a known $q \times K$ matrix of full row rank, for large values of $W_{zx}^{\mathbf{R}} := \hat{\boldsymbol{\beta}}_{zx}'\mathbf{R}'(\mathbf{R}\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}_{zx})\mathbf{R}')^{-1}\mathbf{R}\hat{\boldsymbol{\beta}}_{zx}$ where $\hat{\boldsymbol{\beta}}_{zx} := \boldsymbol{A}_T^{-1}\boldsymbol{C}_T$ with $\boldsymbol{A}_T := \sum_{t=1}^{T} \boldsymbol{z}_{t-1}(\boldsymbol{x}_{t-1} - \bar{\boldsymbol{x}}_{-1})'$, $\boldsymbol{C}_T := \sum_{t=1}^{T} \boldsymbol{z}_{t-1}(y_t - \bar{y})$, $\bar{\boldsymbol{x}}_{-1} := T^{-1}\sum_{t=1}^{T} \boldsymbol{x}_{t-1}$, and where $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}_{zx}) := \hat{\sigma}_u^2 \boldsymbol{A}_T^{-1}\boldsymbol{B}_T(\boldsymbol{A}_T^{-1})'$ with $\boldsymbol{B}_T := \sum_{t=1}^{T} \boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'$, $\hat{\sigma}_u^2 := T^{-1}\sum_{t=1}^{2} \hat{u}_t^2$ and $\hat{u}_t$ being the OLS residuals of the estimated predictive regression. An Eicker–White version of $W_{zx}^{\mathbf{R}}$ can be formed by replacing $\hat{\sigma}_u^2\mathbf{B}_T$ in the expression of $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}_{zx})$ with $\mathbf{D}_T := \sum_{t=1}^{T} \boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\hat{u}_t^2$. A finite sample correction term can again be used; see Kostakis et al. (2015, p. 1515) for precise details. IVX (partial) t-type tests of the null hypothesis $\beta_i = 0$, $i \in \{1, \ldots, K\}$, can also be considered. ◊

3.2. Subsample IVX tests

As we will subsequently show in Proposition 1, the full-sample test based on t_{zx} has non-trivial asymptotic local power against $H_{1,b(\cdot)}$ of (6) for both weakly and strongly persistent regressors. However, these tests are clearly designed for the case where the function $b(\cdot)$ of Assumption 2 is such that $b(t/T) = b$, $t = 1, \ldots, T$. If it were known that a pocket of predictability might occur only over the particular subsample $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$, 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 $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$, viz,

$$t_{zx}(\tau_1, \tau_2) := \frac{\hat{\beta}_{zx}(\tau_1, \tau_2)}{s.e.(\hat{\beta}_{zx}(\tau_1, \tau_2))} \tag{15}$$

where

$$\hat{\beta}_{zx}(\tau_1, \tau_2) := \frac{\sum_{t=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} z_{t-1}\,(y_t - \bar{y}(\tau_1, \tau_2))}{\sum_{t=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} z_{t-1}\,(x_{t-1} - \bar{x}_{-1}(\tau_1, \tau_2))} \tag{16}$$

$$s.e.(\hat{\beta}_{zx}(\tau_1, \tau_2)) := \frac{\hat{\sigma}_u(\tau_1, \tau_2)\sqrt{\sum_{t=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} z_{t-1}^2}}{\sum_{t=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} z_{t-1}\,(x_{t-1} - \bar{x}_{-1}(\tau_1, \tau_2))} \tag{17}$$

with $\bar{y}(\tau_1, \tau_2) := (T^*)^{-1}\sum_{t=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} y_t$ and $\bar{x}_{-1}(\tau_1, \tau_2) := (T^*)^{-1}\sum_{t=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} x_{t-1}$, where $T^* := (\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor)$, and $\hat{\sigma}_u(\tau_1, \tau_2)^2$ is the analogue of $\hat{\sigma}_u^2$ in (13) computed for the subsample $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$. The corresponding subsample analogue of the full sample Eicker–White t_{zx}^{EW} statistic in (14) can be defined similarly and will be denoted $t_{zx}^{EW}(\tau_1, \tau_2)$.

In practice τ_1 and τ_2 are unknown and so, like in Demetrescu et al. (2022), we base tests on suitable functionals of sequences of subsample statistics. These need to be agnostic of the data to avoid any endogenous selection bias and any

test formed from them must be such that multiple testing issues are also avoided. Given we are testing the null of no predictability against the alternative of predictability in at least one subsample of the data, an approach based on the maximum (in the case of two-sided and right-tailed tests) or minimum (in the case of left-sided tests) of the sequence of subsample predictability statistics would seem appropriate. Common choices of such agnostic sequences of statistics include forward and reverse recursive sequences and rolling sequences. 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 are 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.

• The sequence of *forward recursive* statistics is given by $\{t_{zx}(0,\tau)\}_{\tau_L\le\tau\le 1}$, where the *warm-in* 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 and calculates the sequence of subsample predictive regression statistics $t_{zx}(0,\tau)$ for $t=1,\ldots,\lfloor\tau T\rfloor$, 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*,

$$\mathcal{T}_U^F := \max_{\tau_L\le\tau\le 1}\{t_{zx}(0,\tau)\}. \tag{18}$$

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 $\{t_{zx}(\tau,1)\}_{0\le\tau\le\tau_U}$ with $\tau_U\in(0,1)$ chosen by the user. Here one calculates the sequence of subsample predictive regression statistics $t_{zx}(\tau,1)$ for $t=\lfloor\tau T\rfloor+1,\ldots,T$, 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,

$$\mathcal{T}_U^B := \max_{0\le\tau\le\tau_U}\{t_{zx}(\tau,1)\} \tag{19}$$

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 $\{t_{zx}(\tau,\tau+\Delta\tau)\}_{0\le\tau\le 1-\Delta\tau}$ where the user-defined *window fraction* $\Delta\tau\in(0,1)$. Here one calculates the sequence of subsample statistics $t_{zx}(\tau,\tau+\Delta\tau)$ for $t=\lfloor\tau T\rfloor+1,\ldots,\lfloor\tau T\rfloor+\lfloor T\Delta\tau\rfloor$, where $\lfloor T\Delta\tau\rfloor$ is 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 sequence,

$$\mathcal{T}_U^R := \max_{0\le\tau\le 1-\Delta\tau}\{t_{zx}(\tau,\tau+\Delta\tau)\} \tag{20}$$

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.

Remark 10. The full sample IVX statistic t_{zx} of (12) is contained within the forward recursive, backward recursive, and rolling sequences of statistics, by setting $\tau=1$, $\tau=0$, and $\Delta\tau=1$. ◊

Remark 11. Subsample implementations of the multiple predictor IVX Wald tests discussed in Remark 9 can also be defined in an analogous fashion to $\mathcal{T}_U^F$, $\mathcal{T}_U^B$ and $\mathcal{T}_U^R$ of (18), (19) and (20), respectively. Here, defining the subsample analogue of the IVX Wald statistic $W_{zx}^{\mathbf{R}}$ computed over the data subsample $t=\lfloor\tau_1 T\rfloor+1,\ldots,\lfloor\tau_2 T\rfloor$, as $W_{zx}^{\mathbf{R}}(\tau_1,\tau_2)$, we can consider tests which reject for large values of the maxima from analogous forward recursive, backward recursive and rolling sequences of such subsample statistics, which we will denote $\mathcal{W}_F^{\mathbf{R}}$, $\mathcal{W}_B^{\mathbf{R}}$ and $\mathcal{W}_R^{\mathbf{R}}$, respectively. ◊

Tests based on sequences of subsample statistics have also been proposed in the literature on testing for episodic bubbles; see, for example, Phillips et al. (2015). Pavlidis et al. (2017) propose tests for episodic bubbles in foreign exchange markets based on the right-sided IVX t-ratios of Kostakis et al. (2015) applied to Fama regressions estimated over a rolling sequence of subsamples of the data. Their proposed test rejects the no bubble null hypothesis if any of the subsample statistics in the rolling sequence exceeds a given critical value. For size-controlled inference, they base their test on a conservative critical value obtained using a Bonferroni correction, adjusting the nominal significance level for the number of statistics in the sequence. Given that this number will generally be quite large (for given T, the number of statistics in the sequence will be larger the smaller the rolling window width, $\lfloor T\Delta\tau\rfloor$), Pavlidis et al. (2017) acknowledge that this approach will deliver a conservative test. The subsample maximum tests outlined above avoid the need for conservative testing methods and so would be expected to deliver more powerful bubble detection tests than those proposed in Pavlidis et al. (2017).

Demetrescu et al. (2022) also consider tests for episodic predictability based on the maxima from corresponding sequences of rolling and recursive subsample implementations of a 2SLS predictability statistic proposed in Breitung and Demetrescu (2015). It can be seen from Lemma S.6 of Demetrescu et al. (2022) and its proof that, under strong persistence and for any subsample of the type analysed there, the 2SLS t statistic with a heteroskedasticity-consistent standard error is distributed under the null in large samples as the product of a $\chi(1)$ variate[6] and a random sign that are statistically dependent in general, where the distribution of the random sign depends on the localisation parameter c. Moreover, neither the bootstrap studied by Demetrescu et al. (2022) nor the residual wild bootstrap studied in this paper can be validly applied to the 2SLS t statistic as the former fails to replicate the dependence structure of the $\chi(1)$ variate and the random sign, whereas the latter cannot mimic the distribution of the random sign as a function of c. Valid (asymptotic and bootstrap) tests can be based on the *squared* 2SLS t statistics (as this eliminates the random sign), but doing so precludes meaningful testing against one-sided predictability. In contrast, the subsample IVX-based tests and their bootstrap implementations proposed here can be validly used to test against either one-sided or two-sided alternatives as we show below, and so can be used to test against directed alternatives in cases where the predictor has a natural sign predicted by theory, as with the testing problem considered in Pavlidis et al. (2017) where a bubble implies a positive slope coefficient.

3.3. Asymptotic theory

In this section we now provide limiting distribution theory for the IVX statistics from Sections 3.1 and 3.2.

Proposition 1. *Consider the model in (1)–(3) and let Assumptions 2 and 3 hold. Then under the local alternative $H_{1,b(\cdot)}$ of (6):*

(i) *Under Assumption 1.1, as $T \to \infty$*

$$
\begin{aligned}
t_{zx}(\tau_1,\tau_2) &\Rightarrow \frac{M_{\xi u}(\tau_2) - M_{\xi u}(\tau_1) + \kappa^2 \int_{\tau_1}^{\tau_2} b(s) d[M_v](s)}{\sqrt{\frac{\kappa^2}{\tau_2-\tau_1}\left([M_u](\tau_2) - [M_u](\tau_1)\right)\left([M_v](\tau_2) - [M_v](\tau_1)\right)}} := G_1(b,\tau_1,\tau_2);\\
\mathcal{T}_U^F &\Rightarrow \sup_{\tau\in[\tau_L,1]} \{G_1(b,0,\tau)\} := G_{1,U}^F(b);\\
\mathcal{T}_U^B &\Rightarrow \sup_{\tau\in[0,\tau_U]} \{G_1(b,\tau,1)\} := G_{1,U}^B(b);\\
\mathcal{T}_U^R &\Rightarrow \sup_{\tau\in[0,1-\Delta\tau]} \{G_1(b,\tau,\tau+\Delta\tau)\} := G_{1,U}^R(b).
\end{aligned}
$$

(ii) *Under Assumption 1.2, and with $\epsilon < \min\{1-\eta, \frac{1}{2}\eta\}$ in Assumption 3,*

$$
\begin{aligned}
t_{zx}(\tau_1,\tau_2) &\Rightarrow \frac{M_{zu}(\tau_2) - M_{zu}(\tau_1)}{\sqrt{\frac{1}{\tau_2-\tau_1}\left([M_u](\tau_2) - [M_u](\tau_1)\right)\left([M_v](\tau_2) - [M_v](\tau_1)\right)}}\\
&\quad + \sqrt{\frac{2\omega^2}{a}} \frac{\int_{\tau_1}^{\tau_2} b(s) d[M_v](s) + \int_{\tau_1}^{\tau_2} (J_{c,H}(s) - \bar{J}_{c,H}(\tau_1,\tau_2)) b(s) dJ_{c,H}(s)}{\sqrt{\frac{1}{\tau_2-\tau_1}\left([M_u](\tau_2) - [M_u](\tau_1)\right)\left([M_v](\tau_2) - [M_v](\tau_1)\right)}} := G_2(b,\tau_1,\tau_2);\\
\mathcal{T}_U^F &\Rightarrow \sup_{\tau\in[\tau_L,1]} \{G_2(b,0,\tau)\} := G_{2,U}^F(b);\\
\mathcal{T}_U^B &\Rightarrow \sup_{\tau\in[0,\tau_U]} \{G_2(b,\tau,1)\} := G_{2,U}^B(b);\\
\mathcal{T}_U^R &\Rightarrow \sup_{\tau\in[0,1-\Delta\tau]} \{G_2(b,\tau,\tau+\Delta\tau)\} := G_{2,U}^R(b),
\end{aligned}
$$

where a and η are the parameters defining the IVX filter in (4), ω and κ^2 are as defined in Section 2, and $\bar{J}_{c,H}(\tau_1,\tau_2) := \frac{1}{\tau_2-\tau_1}\int_{\tau_1}^{\tau_2} J_{c,H}(s)ds$. The results for $t_{zx}(\tau_1,\tau_2)$ hold for any given fixed values of τ_1 and τ_2, $0 \le \tau_1 < \tau_2 \le 1$.

Remark 12. Corresponding representations for the limiting distributions of the left-sided $\mathcal{T}_L^F$, $\mathcal{T}_L^B$ and $\mathcal{T}_L^R$ statistics under the conditions of Proposition 1 can be obtained by replacing the sup operator by the inf operator in the representations given in Proposition 1, and with an obvious notation we denote these limiting distributions as $G_{j,L}^F(b)$, $G_{j,L}^B(b)$ and $G_{j,L}^R(b)$, $j = 1, 2$, respectively. Similarly, representations for the limiting distributions of the two-sided statistics $\mathcal{T}_2^F$, $\mathcal{T}_2^B$ and $\mathcal{T}_2^R$, denoted $G_{j,2}^F(b)$, $G_{j,2}^B(b)$ and $G_{j,2}^R(b)$, $j = 1, 2$, respectively, can be obtained by squaring the limiting quantities over which the supremum is taken in the expressions in Proposition 1. ◊

[6] If a random variable is $\chi^2(1)$ distributed, then its positive square root obeys a $\chi(1)$ distribution.

Remark 13. Part (ii) of Proposition 1, which relates to the case where x_t is strongly dependent, imposes a further restriction on the degree of persistence permitted in the conditional variances via the additional requirement that $\epsilon < \min\{1-\eta, \frac{1}{2}\eta\}$. This restriction therefore entails that $\epsilon < 1/3$ (with this maximum upper bound for ϵ corresponding to the use of an IVX filter with $\eta = 2/3$). Recalling, for example, that parametric GARCH models are such that $\epsilon = 0$, it seems likely that this additional restriction would not be restrictive in practice. ◊

Remark 14. The results in Proposition 1 establish the asymptotic local power functions of the tests based on the subsample and full sample IVX-based statistics (the latter obtained by setting $\tau_2 = 1$ and $\tau_1 = 0$ in the limiting representations for $t_{zx}(\tau_1, \tau_2)$) from Sections 3.1 and 3.2, respectively, under the local alternative $H_{1,b(\cdot)}$. These local power functions depend, in general, on any heteroskedasticity and/or weak autocorrelation (short-run dynamics) present in the errors and differ according to whether x_t is weakly or strongly persistent. In the strongly persistent case they also depend on the parameter a used in the IVX filter and on the local-to-unity parameter, c. For the full sample t_{zx} test these results therefore complement those provided in Kostakis et al. (2015) and Magdalinos (2020) which apply only under the null hypothesis. From Proposition 1 it can be seen that the full sample t_{zx} test exhibits non-trivial power against the class of time-varying local alternatives we consider in this paper; that is, it has power to detect predictive episodes. In the case where, for some constant $b \neq 0$, $b(s) = b$ for all s, the results in Proposition 1 provide the asymptotic local power functions of the tests against the alternative of (local) predictability with a fixed slope coefficient across the full sample. ◊

Remark 15. The limiting null distributions of the statistics obtain from the results in Proposition 1 on setting $b(s) = 0$ for all s. Doing so, the limiting null distributions of the individual statistics $t_{zx}(\tau_1, \tau_2)$ can be seen to be (pointwise) normal. For example, under strong persistence, we have for the full-sample statistic that

$$t_{zx} \Rightarrow \frac{M_{zu}(1)}{\sqrt{[M_u](1)[M_v](1)}} = \frac{\int_0^1 \sqrt{[M_u]'(s)[M_v]'(s)}\mathrm{d}B(s)}{\sqrt{[M_u](1)[M_v](1)}} \overset{d}{=} N\left(0, \frac{\int_0^1 [M_u]'(s)[M_v]'(s)\mathrm{d}s}{\int_0^1 [M_u]'(s)\mathrm{d}s \int_0^1 [M_v]'(s)\mathrm{d}s}\right).$$

It can then be seen that in the unconditionally homoskedastic case where $\mathbf{H}$ is constant, the limiting null distribution of t_{zx} is standard normal under strong persistence, and hence that of $(t_{zx})^2$ is χ^2_1. This holds regardless of any conditional heteroskedasticity present in the innovations. In the weakly persistent case, however, we have that

$$t_{zx} \Rightarrow \frac{M_{\xi u}(1)}{\sqrt{\kappa^2 [M_u](1)[M_v](1)}} \overset{d}{=} N\left(0, \frac{[M_{\xi u}](1)}{\kappa^2 [M_u](1)[M_v](1)}\right)$$

whereby it follows that the variance of the limiting distribution of t_{zx} will in general depend on any conditional heteroskedasticity and/or short-run dynamics (the latter through the parameter κ^2) present, even where $\mathbf{H}$ is constant. On the other hand, κ^2 drops out of this expression under conditional homoskedasticity of $\boldsymbol{\psi}_t$, even if $\mathbf{H}$ is time-varying. For further details see the proof of Lemma 4 in the supplementary appendix. ◊

Remark 16. The limiting null distributions of the subsample-based statistics, $\mathcal{T}_j^F$, $\mathcal{T}_j^B$ and $\mathcal{T}_j^R$, $j \in \{U, L, 2\}$, all depend, in general, in a highly complicated way on nuisance parameters arising from any heteroskedasticity and (in the weakly dependent case) serial correlation present in $(u_t, v_t)'$ and on whether x_t is strongly or weakly persistent. While, as we show below in Proposition 2, these dependencies can be removed from the limiting null distribution of the full sample statistic by using Eicker–White standard errors, this is not true of the subsample-based statistics. ◊

As discussed in Remark 15, the standard t_{zx} statistic, while having a limiting null distribution that is free of nuisance parameters when x_t is strongly persistent and the innovations are unconditionally homoskedastic, does not in general have a pivotal limiting null distribution when x_t is weakly persistent. The non-pivotal nature of the limiting null distribution of t_{zx} under conditional heteroskedasticity in the case of a weakly persistent predictor motivated Kostakis et al. (2015) to also consider the Eicker–White statistic t_{zx}^{EW} in (14). In Proposition 2 we demonstrate that the limiting (marginal) null distribution of the subsample Eicker–White statistic $t_{zx}^{EW}(\tau_1, \tau_2)$ is standard normal under the conditions of Proposition 1 and regardless of whether x_t is weakly dependent or near-integrated.

Proposition 2. *Under the conditions of Proposition 1, and for any given fixed values of τ_1 and τ_2, $t_{zx}^{EW}(\tau_1, \tau_2) \Rightarrow N(0,1)$, and hence $(t_{zx}^{EW}(\tau_1, \tau_2))^2 \Rightarrow \chi^2_1$, under the null hypothesis, H_0, regardless of whether Assumption 1.1 or Assumption 1.2 holds.*

Remark 17. As a consequence of Proposition 2 the full-sample t_{zx}^{EW} statistic of (14) is seen to have a standard normal limiting null distribution under H_0 regardless of whether x_t is weakly or strongly persistent. The standard normality of the limiting null distribution of t_{zx}^{EW} has previously been shown to hold by Kostakis et al. (2015) under their Assumption INNOV and by Magdalinos (2020) under his Assumption M, both of which assume unconditional homoskedasticity. The result in Proposition 2 therefore establishes that this result holds under the much more general conditions of Assumption 3, which includes: (i) the case where $\mathbf{H}$ is non-constant such that the innovations are unconditionally heteroskedastic, and (ii) the case where the sequence $\boldsymbol{\psi}_t$ exhibits conditional heteroskedasticity of very general form; see again the discussion in Remarks 4 and 5. ◊

Remark 18. Provided the vector $(u_t, \boldsymbol{v}_t')'$ satisfies an obvious $(k+1)$-dimensional generalisation of Assumption 3, then the multiple predictor full sample Wald statistic, $W_{zx}^{\mathbf{R}}$ of Remark 9, when implemented with Eicker–White standard errors, can be shown to have a χ_q^2 limiting null distribution regardless of whether $\boldsymbol{x}_t$ is strongly or weakly persistent. The limiting null distributions of the corresponding subsample-based statistics, $\mathcal{W}_F^{\mathbf{R}}$, $\mathcal{W}_B^{\mathbf{R}}$ and $\mathcal{W}_R^{\mathbf{R}}$, of Remark 11 will, like the corresponding subsample-based tests for a scalar predictor, x_t, discussed in this section, have limiting null distributions which will, in general, depend in a highly complicated way on nuisance parameters arising from any heteroskedasticity and (in the weakly dependent case) serial correlation present in $(u_t, \boldsymbol{v}_t')'$ and on whether $\boldsymbol{x}_t$ is strongly or weakly persistent. ◊

4. Bootstrap IVX tests

As the results in Section 3.3 show, implementing tests based on either the full sample t_{zx} statistic from Section 3.1 or the subsample-based $\mathcal{T}_j^F$, $\mathcal{T}_j^B$ and $\mathcal{T}_j^R$, $j = U, L, 2$, statistics from Section 3.2 will require us to address the fact that their limiting null distributions will, in general, depend on nuisance parameters arising from heteroskedasticity and/or serial correlation present in the data, and on whether the predictor x_{t-1} is weakly dependent or near-integrated.

To that end, we will consider two bootstrap resampling schemes for t_{zx}, $\mathcal{T}^F$, $\mathcal{T}^B$ and $\mathcal{T}^R$. The first, a residual wild bootstrap [RWB], is outlined in Algorithm 1. In Algorithm 2 we then outline how the fixed regressor wild bootstrap [FRWB] employed by Demetrescu et al. (2022) can also be used with the full sample and subsample IVX statistics discussed in this paper.[7] Although we will not formally establish large sample validity here, the RWB could also be validly employed in connection with the corresponding 2SLS tests of Demetrescu et al. (2022).

Algorithm 1 (*Residual Wild Bootstrap*)**.**

Step 1: Fit the predictive regression to the sample data $(y_t, x_{t-1})'$ to obtain the residuals $\hat{u}_t$, $t = 1, \ldots, T$, using any of the two choices outlined below (13).

Step 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 $t = 1, \ldots, p$.

Step 3: Generate bootstrap innovations $\left(u_t^*, v_t^*\right)' := \left(R_t \hat{u}_t,\ R_t \hat{v}_t\right)'$, $t = 1 \ldots, T$, where R_t, $t = 1, \ldots, T$, is a scalar *i.i.d.*$(0, 1)$ sequence with $\mathrm{E}(R_t^4) < \infty$, which is independent of the sample data.

Step 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

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

with initial conditions $x_0^* = \cdots = x_{-p}^* = 0$. Create the associated bootstrap IVX instrument, z_t^*, via $z_0^* = 0$ and $z_t^* = \sum_{j=0}^{t-1} \varrho^j \Delta x_{t-j}^*$, $t = 1, \ldots, T$, where ϱ is the same value used in constructing the original IVX instrument, z_t.

Step 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 $t_{zx}(\tau_1, \tau_2)$, $\mathcal{T}_j^F$, $\mathcal{T}_j^B$ and $\mathcal{T}_j^R$, $j = U, L, 2$, statistics from Section 3.2. Denote these bootstrap statistics as $t_{zx}^*(\tau_1, \tau_2)$, $\mathcal{T}_j^{*,F}$, $\mathcal{T}_j^{*,B}$ and $\mathcal{T}_j^{*,R}$, $j = U, L, 2$.

Step 6: Taking $\mathcal{T}_U^F$ to illustrate, a bootstrap p-value is then computed as $p_{1,T}^* := 1 - G_{1,T}^*(\mathcal{T}_U^F)$, where $G_{1,T}^*(\cdot)$ denotes the conditional (on the original sample data) cumulative distribution function (cdf) of $\mathcal{T}_U^{*,F}$. The bootstrap test, run at the λ significance level, based on $\mathcal{T}_U^F$ is therefore defined such that it rejects H_0 if $p_{1,T}^* < \lambda$. Bootstrap p-values for the other tests are similarly obtained.

Algorithm 2 (*Fixed Regressor Wild Bootstrap*)**.**

Step 1: As Step 1 in Algorithm 1.

Step 2: Generate bootstrap innovations $u_t^* := R_t \hat{u}_t$, $t = 1 \ldots, T$, where R_t satisfies the same conditions as given in Step 3 of Algorithm 1.

Step 3: For $t = 1, \ldots, T$, define the bootstrap data $y_t^* = u_t^*$ (so that the null hypothesis is imposed on the bootstrap y_t^*).

[7] To save space we outline our proposed bootstrap procedures for the case where conventional standard errors are used and where the finite sample correction term of Kostakis et al. (2015) is not employed; cf. footnote 5. Bootstrap implementations of the tests with the finite sample correction term can instead be used without altering any of the large sample properties given in this section. Moreover, bootstrap implementations of the IVX tests based around Eicker–White standard errors may also be considered and share the same asymptotic validity properties as the bootstrap tests based on conventional standard errors.

Step 4: As detailed in Step 5 of Algorithm 1, but where the original sample data, $(y_t, x_{t-1}, z_{t-1})'$ are instead replaced by the fixed regressor bootstrap sample data, $\left(y_t^*, x_{t-1}, z_{t-1}\right)'$.

Step 5: As Step 6 of Algorithm 1.

Remark 19. A key difference between the RWB and FRWB outlined in Algorithms 1 and 2, respectively, surrounds the generation of the bootstrap analogue data for x_t and, hence, z_t. In the FRWB scheme one calculates the bootstrap statistics in Step 4 using the data $\left(y_t^*, x_{t-1}, z_{t-1}\right)'$; that is, y_t^* is generated exactly as in Algorithm 1, but the observed outcomes on $\mathbf{x} := [x_0, x_1, \ldots, x_T]'$ and $\mathbf{z} := [z_0, z_1, \ldots, z_T]'$ are treated as a fixed regressor and fixed instrument vector, respectively. As such, while the RWB rebuilds into the bootstrap data (an estimate of) the correlation between the innovations u_t and v_t through Step 3 of Algorithm 1 (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, as the simulation results we report in Section 5 will show, the finite sample behaviour of the IVX statistics is heavily dependent on the correlation between u_t and v_t in the case where x_t is strongly persistent. As a result we find that the RWB delivers considerably better finite sample performance than the FRWB in the case where x_t is strongly persistent. ◊

Remark 20. A further difference between the RWB and the FRWB is that because one creates bootstrap analogues of x_t and z_t, x_t^* and z_t^* respectively, one implicitly has to use an estimate of ρ in doing so. Under Assumption 1.2 (strong persistence) it is well known that the associated local-to-unity parameter, c, cannot be consistently estimated. Consequently, when x_t is strongly persistent the bootstrap data on x_t^* will not be generated with the same local-to-unity parameter as the original data x_t. For the FRWB this issue does not arise as the original data on x_t are used in calculating the bootstrap statistics. However, the IVX statistics instrument x_{t-1} by z_{t-1}, and their bootstrap analogue statistics instrument x_{t-1}^* by z_{t-1}^*, where z_t and z_t^* are, by construction, both mildly integrated processes regardless of the value of c under Assumption 1.2. There is therefore no necessity for the estimate of c from Step 2 to be consistent in order to validly implement the RWB in Algorithm 1. Notice that this would not be true under Assumption 1.2 if we were bootstrapping the standard OLS t-statistic from (1) because this statistic does not instrument x_t by a variable of lower persistence and, as result, has a limiting null distribution which depends on c. ◊

Remark 21. It could also be possible to implement a moving block bootstrap [MBB] based scheme, similar to that used in Fan and Lee (2019), for the IVX-based tests considered here. An outline of this algorithm can be found in the supplementary appendix. We conjecture that this MBB procedure is asymptotically valid provided **H** is constant such that the innovations were unconditionally homoskedastic. To account for unconditional heteroskedasticity a block wild adaptation of this bootstrap could be employed and this is also outlined in the Supplementary Appendix. We will not pursue either of these methods further here as in unreported simulations we found them to perform poorly in finite samples relative to the RWB-based tests. ◊

Remark 22. With simple modifications, the RWB of Algorithm 1 can be implemented for the multiple regressor full sample Wald statistic, $W_{zx}^{\mathbf{R}}$ of Remark 9, and the corresponding subsample-based statistics, $\mathcal{W}_F^{\mathbf{R}}$, $\mathcal{W}_B^{\mathbf{R}}$ and $\mathcal{W}_R^{\mathbf{R}}$, of Remark 11. In Step 2 of Algorithm 1 a vector autoregression of order $p+1$ is fitted to $\boldsymbol{x}_t$ to obtain the residuals $\hat{\boldsymbol{v}}_t$ with the residuals from these collected into $\hat{\mathbf{v}}_t$. In Step 3 one then calculates the bootstrap innovations $\left(u_t^*, \boldsymbol{v}_t^{*\prime}\right)' = \left(R_t\hat{u}_t, R_t\hat{\boldsymbol{v}}_t'\right)'$, $t = 1, \ldots, T$. In Step 4 one generates the bootstrap data $y_t^* = u_t^*$ imposing the null, together with the bootstrap predictor vector, $\boldsymbol{x}_t^*$, by the recursion based on the coefficient estimates obtained in Step 2. The bootstrap instruments, $\boldsymbol{z}_t^*$, are derived from $\boldsymbol{x}_t^*$ according to the same IVX filter used to obtain $\boldsymbol{z}_t$ from $\boldsymbol{x}_t$. The RWB statistics are then computed from the bootstrap sample data, $\left(y_t^*, \boldsymbol{x}_{t-1}^*, \boldsymbol{z}_{t-1}^*\right)'$. The FRWB of Algorithm 2 can also be modified to allow for multiple regressors by using the bootstrap sample data, $\left(y_t^*, \boldsymbol{x}_{t-1}, \boldsymbol{z}_{t-1}\right)'$ in Step 4. Provided the conditions outlined in Remark 18 hold, the FRWB and RWB-based tests for multiple regressors will share analogous asymptotic validity properties to the bootstrap tests in the case of a single regressor established below. ◊

In Proposition 3 we demonstrate the large sample validity of the RWB and FRWB implementations of the IVX tests from Algorithms 1 and 2, respectively. In particular, these are shown to correctly replicate the first order asymptotic null distributions of the IVX statistics under both the null and local alternatives. For the RWB-based tests this result requires a further restriction to hold on the fourth moments of the innovations in the case where x_t is weakly persistent. This additional restriction is not required for the asymptotic validity of the FRWB-based tests.

Proposition 3. *Consider the model in (1)–(3) and let Assumptions 2 and 3 hold. Then under the local alternative $H_{1,b(\cdot)}$ of (6):*

(i) *Under Assumption 1.1,*

(a) *For the bootstrap statistics generated according to the RWB scheme in Algorithm 1, provided $E[(\boldsymbol{\psi}_1\boldsymbol{\psi}_1') \otimes (\boldsymbol{\psi}_{-i}\boldsymbol{\psi}_{-j}')] = 0$ for all natural $i \neq j$, it holds that $t_{zx}^*(\tau_1, \tau_2) \overset{w}{\Rightarrow}_p G_1(0, \tau_1, \tau_2)$ for fixed $0 \leq \tau_1 < \tau_2 \leq 1$, and $\mathcal{T}_j^{*,F} \overset{w}{\Rightarrow}_p G_{1,j}^F(0)$, $\mathcal{T}_j^{*,B} \overset{w}{\Rightarrow}_p G_{1,j}^B(0)$ and $\mathcal{T}_j^{*,R} \overset{w}{\Rightarrow}_p G_{1,j}^R(0)$, in each case for $j = U, L, 2$.*

(b) *For the bootstrap statistics generated according to the FRWB scheme in Algorithm 2,* $t^*_{zx}(\tau_1,\tau_2) \overset{w}{\Rightarrow}_p G_1(0,\tau_1,\tau_2)$ *for fixed* $0 \le \tau_1 < \tau_2 \le 1$, *and* $\mathcal{T}_j^{*,F} \overset{w}{\Rightarrow}_p G^F_{1,j}(0)$, $\mathcal{T}_j^{*,B} \overset{w}{\Rightarrow}_p G^B_{1,j}(0)$ *and* $\mathcal{T}_j^{*,R} \overset{w}{\Rightarrow}_p G^R_{1,j}(0)$, *in each case for* $j = U, L, 2$.

(ii) *Under Assumption 1.2, and with* $\epsilon < \min\{\eta, \frac{1}{2}\}$ *in Assumption 3, and regardless of whether the bootstrap statistics are generated according to the RWB scheme in Algorithm 1 or the FRWB scheme in Algorithm 2,* $t^*_{zx}(\tau_1,\tau_2) \overset{w}{\Rightarrow}_p G_2(0,\tau_1,\tau_2)$ *for fixed* $0 \le \tau_1 < \tau_2 \le 1$, *and* $\mathcal{T}_j^{*,F} \overset{w}{\Rightarrow}_p G^F_{2,j}(0)$, $\mathcal{T}_j^{*,B} \overset{w}{\Rightarrow}_p G^B_{2,j}(0)$ *and* $\mathcal{T}_j^{*,R} \overset{w}{\Rightarrow}_p G^R_{2,j}(0)$, *in each case for* $j = U, L, 2$.

Remark 23. A comparison of the limiting results for the bootstrap statistics in Proposition 3 with those given for the corresponding statistics in Proposition 1 demonstrates the usefulness of the RWB and FRWB procedures from Algorithms 1 and 2, respectively; as the number of observations increases, the bootstrapped statistics have the same first-order limiting null distributions as the corresponding original test statistic.[8] For this result to hold for the RWB statistics, however, it is seen that fourth moments of the form $\mathrm{E}[(\boldsymbol{\psi}_1\boldsymbol{\psi}_1') \otimes (\boldsymbol{\psi}_{-i}\boldsymbol{\psi}_{-j}')]$ for natural $i \neq j$ should not contribute to the quadratic variation of the process $M_{\xi u}$. The reason is that in the RWB world the mixed fourth moments $\mathrm{E}^*[(R_t^2\boldsymbol{\psi}_t\boldsymbol{\psi}_t') \otimes (R_{t-i}R_{t-j}\boldsymbol{\psi}_{t-i}\boldsymbol{\psi}_{t-j}')] = 0$ by construction for all positive natural $i \neq j$, and hence, these do not contribute to the quadratic variation of the RWB analogue of $M_{\xi u}$. As with the conditions placed on $\{\boldsymbol{\psi}_t\}$ by Assumption 3.2, this assumption is not tied to any specific parametric model. Even where this condition is violated, the impact on the (asymptotic) size of the resulting RWB test might still be relatively small, given that the quantities $\mathrm{E}[(\boldsymbol{\psi}_1\boldsymbol{\psi}_1') \otimes (\boldsymbol{\psi}_{-i}\boldsymbol{\psi}_{-j}')]$, for all natural $i \neq j$, only constitute part of the quadratic variation of $M_{\xi u}$ and it is this latter quantity which the bootstrap limit needs to reproduce. A well known class of models which violate this condition are GARCH models with non-zero leverage effects. We will explore the impact of such a model on the finite sample size behaviour of the RWB tests in Section 5. ◊

Remark 24. In the case of the RWB, the asymptotic validity result in Proposition 3 requires knowledge of the true autoregressive lag length, p, used in Step 2 of Algorithm 1. In practice p, 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 without affecting the stated asymptotic validity results for the RWB. A less parsimonious information criterion, such as the Akaike Information Criterion [AIC] could also be used. Furthermore, we conjecture that the RWB tests would still be asymptotically valid for more general linear process innovations of the form discussed in Remark 1, provided a sieve-type device is used in Step 2 whereby the truncation lag for the fitted autoregression is allowed to increase at a suitable rate with the sample size, e.g. $\lfloor \kappa(T/100)^{1/4} \rfloor$, for a positive constant κ. A formal proof of this conjecture is beyond the scope of the present paper but constitutes an interesting topic for future research. Along these lines, in unreported simulations we found that the lag length fitted in Step 2 has rather little bearing on the power of the resulting bootstrap tests. It should also be stressed that no choice of truncation lag is required in connection with the FRWB outlined in Algorithm 2. ◊

Remark 25. Although, as Proposition 3 shows, the RWB and FRWB are asymptotically equivalent, to first-order, to each other and to the limiting null distributions of the corresponding asymptotic statistics, they can be shown to differ in higher-order terms. In particular, Demetrescu and Hosseinkouchack (2021) demonstrate that in the strongly persistent case the second-order term in a Taylor expansion of the full-sample IVX statistic is a function of c. In preliminary work we have found that the FRWB fails to replicate this second-order term entirely, while the RWB, conditional on the data, replicates a similar functional to the second-order term but with $\hat{c}$ (the implied estimate of c obtained from Step 2 of Algorithm 1) replacing the true c. So although the RWB does not correctly replicate the second-order term from the limiting null distribution of the IVX statistic it replicates part of it and this would be anticipated to effect a reduced sensitivity to c in the finite sample size properties of the RWB test relative to the FRWB test, a prediction borne out by the simulations results in Section 5. A full treatment of this issue is beyond the scope of the present paper, but constitutes a useful topic for further research. ◊

Remark 26. A consequence of the results in Proposition 3, using the same arguments as in the proof of Theorem 5 in Hansen (2000), is that for each of the tests the bootstrap p-values are (asymptotically) uniformly distributed under the unit root null hypothesis, H_0, leading to tests with (asymptotically) correct size, thereby establishing the asymptotic validity of the bootstrap tests. In the case of the FRWB, this validity result is achieved without the practitioner needing to have knowledge of whether x_t is weakly or strongly persistent and holds regardless of any autocorrelation or heteroskedasticity present in u_t and v_t satisfying Assumption 3. For the RWB this is also true, provided the condition $\mathrm{E}[(\boldsymbol{\psi}_1\boldsymbol{\psi}_1') \otimes (\boldsymbol{\psi}_{-i}\boldsymbol{\psi}_{-j}')] = 0$ for all natural $i \neq j$ holds. A further consequence of the result in Proposition 3 for $t^*_{zx}(\tau_1,\tau_2)$, setting $\tau_1 = 0$ and $\tau_2 = 1$, is therefore that under the null the RWB and FRWB bootstrap implementations of the full sample t_{zx} test deliver asymptotically valid (by which we mean that the bootstrap p-values are asymptotically invariant to any nuisance parameters under the null) inference under Assumption 3 (or the restricted version thereof in the case of the RWB scheme) without the need for Eicker–White standard errors. ◊

[8] Observe that the condition placed on ϵ in part (ii) of Proposition 3 is less restrictive than that imposed for part (ii) of Proposition 1 regardless of the value of η used in the IVX filter and therefore this result holds for all DGPs such that Proposition 1 holds.

Remark 27. A further implication of Proposition 3 is that the bootstrap IVX tests from Algorithms 1 and 2 will admit the same asymptotic local power functions under $H_{1,b(\cdot)}$ as the corresponding (infeasible) size-adjusted tests based on the corresponding original IVX statistic. ◊

Remark 28. As discussed in Remark 23, a key difference between the large sample properties of the RWB and FRWB is that the former can only be validly applied in the case where x_t is weakly persistent if the mixed fourth moments $\mathrm{E}[(\boldsymbol{\psi}_1\boldsymbol{\psi}_1')\otimes(\boldsymbol{\psi}_{-i}\boldsymbol{\psi}_{-j}')]$ with $i \neq j$ do not contribute to the quadratic variation of the process $M_{\xi u}$. However, as we will see in the simulations in Section 5, the RWB delivers considerably better finite sample performance than the FRWB when x_t is strongly persistent, while the two display similar performance when the degree of persistence in x_t is weaker. In principle one might use the sample data on x_t to decide which of the RWB and FRWB to use. In particular, one could adopt the RWB of Algorithm 1 unless the sample data suggested the persistence in x_t was relatively weak. This idea has previously been advocated in the predictability testing literature by Elliott et al. (2015) who propose a procedure which switches between a weighted average power test where x_t is strongly persistent and the standard OLS t-test from (1) when x_t is weakly persistent. The switching mechanism they adopt is to use the OLS t-test when $\hat{c} \geq 130$ and the weighted average power test otherwise, where $\hat{c}$ is an estimate of the local-to-unity parameter, c. A similar rule could be used here, whereby we use the RWB unless $\hat{c}$ exceeds some specified value in which case we use the FRWB. An obvious estimate of c, based on the autoregressive estimates from Step 2 of Algorithm 1, is $\hat{c} := T(1 - \sum_{j=1}^{p} \hat{a}_j)$. This rule ensures that, with probability approaching one, the RWB would not be chosen in large samples when x_t was weakly dependent, and so this hybrid bootstrap will share the asymptotic validity result enjoyed by the FRWB in the weak persistence case. ◊

Remark 29. In practice the cdf $G^*_{1,T}(\cdot)$ of the bootstrap $\mathcal{T}_U^{*,F}$ statistic, and the corresponding cdfs for the other statistics, required in Step 6 of Algorithm 1 and Step 5 of Algorithm 2 will be unknown but can be approximated in the usual way through numerical simulation. To illustrate, again for the case of the $\mathcal{T}_U^F$ statistic, this is achieved by generating B bootstrap (conditionally) independent statistics, say $\mathcal{T}_{U,b}^{*,F}$, $b = 1, \ldots, B$, each computed as in Algorithm 1 above. The simulated bootstrap p-value for the test is then computed as $\tilde{p}^*_{1,T} = B^{-1}\sum_{b=1}^{B} \mathbb{I}\left(\mathcal{T}_{U,b}^{*,F} > \mathcal{T}_U^F\right)$ and is such that $\tilde{p}^*_{1,T} \xrightarrow{a.s.} p^*_{1,T}$ as $B \to \infty$, where $\xrightarrow{a.s.}$ denotes almost sure convergence. An approximate standard error for $\tilde{p}^*_{1,T}$ is given by $(\tilde{p}^*_{1,T}(1 - \tilde{p}^*_{1,T})/B)^{1/2}$. Simulated bootstrap critical values can also be obtained; e.g. for the $\mathcal{T}_U^F$ statistic, a λ level bootstrap critical value, $cv_{\lambda,B}$ say, can be calculated as the upper tail λ percentile from the order statistic formed from the B bootstrap statistics, $\mathcal{T}_{U,b}^{*,F}$, $b = 1, \ldots, B$. The resulting bootstrap test, which rejects H_0 if $\mathcal{T}_U^F > cv_{\lambda,B}$, will have asymptotic size that for sufficiently large B will be as close as desired to λ. ◊

5. Monte Carlo simulations

We now discuss the results from a detailed Monte Carlo study into the finite sample properties of IVX-based predictability tests. In Section 5.1 for the case of a single predictor and Section 5.2 for multiple predictors, we compare the properties of the full sample tests of Kostakis et al. (2015) based on asymptotic critical values with their RWB and FRWB bootstrap implementations developed in this paper. A comparison of the subsample bootstrap IVX tests proposed in this paper with their 2SLS counterparts from Demetrescu et al. (2022) is made in Section 5.3. For all statistics, OLS residuals are used in computing the standard errors. All simulations are preformed in MATLAB, versions R2018b and R2020a, using the Mersenne Twister random number generator. Results are reported for tests run at the 5% nominal significance level. Unless otherwise stated, the results are based on $B = 999$ bootstrap replications, and 10,000 Monte Carlo replications.

5.1. Single predictor regressions — full sample tests

We first consider the case where a single predictor, x_{t-1}, is included in the predictive regression. Results are reported for the IVX test of Kostakis et al. (2015) both with and without Eicker–White corrected standard errors, t_{zx}^{EW} and t_{zx}, respectively; these statistics were computed exactly as detailed in Section 3.1 with the finite sample correction term, Ξ, (see footnote 5) included using a Bartlett kernel with bandwidth $T^{1/3}$ as recommended by Kostakis et al. (2015) — this choice was made in all of the numerical experiments and empirical applications reported in this paper. We will compare these with their RWB and FRWB bootstrap analogues, denoted $t_{zx}^{*,RWB}$ and $t_{zx}^{*,FRWB}$, described in Algorithms 1 and 2, respectively. In the context of the RWB the autoregressive lag length used in Step 2 of Algorithm 1 was chosen applying the BIC over $p \in \{0, \ldots, \lfloor 4(T/100)^{0.25} \rfloor\}$. The bootstrap statistics are all based on conventional standard errors and all include the finite sample correction term. Our analysis consists of testing the null hypothesis of no predictability, $H_0 : \beta = 0$, in (1) in the context of a constant parameter prediction model, so that $\beta_t = \beta$, for all $t = 1, \ldots, T$. We will consider tests directed against both one-sided alternatives, left-tailed tests for $H_1 : \beta < 0$, and right-tailed tests for $H_1 : \beta > 0$, together with two-sided tests for $H_1 : \beta \neq 0$.

5.1.1. Empirical size

To investigate the finite sample size properties of t_{zx}, t_{zx}^{EW}, $t_{zx}^{*,RWB}$ and $t_{zx}^{*,FRWB}$ under the null hypothesis of no predictability, we generate data according to (1)–(3) with $\beta_t = \beta = 0$ for all $t = 1, \ldots, T$. We initialised the autoregressive

process characterising the dynamics of the putative predictor, x_t, in (3) at $\xi_0 = 0$, and considered a wide range of values for the autoregressive parameter ρ in (3) covering stationary, near-integrated and mildly explosive predictors; in particular, we set $\rho = 1 - c/T$ with $c \in \{-5, -2.5, 0, 2.5, 5, 10, 25, 50, 75, 100, 125, 150, 200, 250\}$. All results reported, both in the main text and in the supplementary appendix, are for sample sizes $T = 250$ and $T = 1000$. In total, for the single predictor case, we consider 11 distinct classes of DGP. For the sake of space we will present Tables of results for two of these DGPs in this section. A summary of the results for the other 9 DGPs will also be given, with the full details of these DGPs and the associated tables of results for these cases given in the supplementary appendix.

Main results

The first DGP (DGP1) we will consider corresponds to (1)–(3) with the innovation vector $(u_t, v_t)'$ drawn from an i.i.d. bivariate Gaussian distribution with mean vector zero and covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}$, where ϕ corresponds to the correlation between u_t and v_t. Results from DGP1 for $\phi = -0.95, -0.90$, and -0.50 are reported in Table 1.[9] Results for tests run at the 1% and 10% significance levels are qualitatively similar and can be found in Tables D.1–D.3 of the supplementary appendix. Additional results for $\phi = 0$ can also be found in Table D.4.

The second DGP (DGP2) we will consider is one designed to be such that the regularity conditions needed for the validity of the RWB when x_t is weakly persistent are violated. The DGP we consider is a well known model where the conditional variance of the innovations $(u_t, v_t)'$ follows a stationary ARCH model with leverage effects and is of the form

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0 \\ \rho x_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} 0 \\ \rho x_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \boldsymbol{\psi}_t \tag{21}$$

with

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

and $(\varepsilon_{1t}, \varepsilon_{2t})' \sim NIID(0, \mathbf{I}_2)$. The AR parameter ρ is again set equal to $1 - c/T$.

DGP2 satisfies our assumptions of finite fourth moments of $\boldsymbol{\psi}_t$ and martingale approximability of $\boldsymbol{\psi}_t \boldsymbol{\psi}_t'$ (with $\epsilon = 0$). However, and crucially, the quadratic variation of $M_{\xi u}$ depends on,

$$\begin{aligned} h_{11}^2 h_{21}^2 b_1 b_2 \, \mathrm{E}(a_t^2 a_{t-1} a_{t-2}) &= \rho^3 \, \mathrm{E}(a_t^2 a_{t-1} a_{t-2}) \\ &= \frac{\rho^3}{8} \, \mathrm{E} \, |\varepsilon_1|^3 \, \mathrm{E} \left\{ |a_1| \left[\sqrt{\left(1 + \frac{1}{2} a_1^2\right)^3} - 1 \right] \right\} > 0 \end{aligned} \tag{22}$$

for $\rho > 0$; see the proof of Lemma 4. This model therefore violates the limiting condition that $M_{\xi u}^* \overset{d}{=} M_{\xi u}$ which is necessary and sufficient for the validity of the RWB in the case where x_t is weakly persistent. Specifically, the non-zero term in (22) is absent from the quadratic variation of $M_{\xi u}^*$ in the limiting distribution of the RWB bootstrap statistic when x_t is weakly persistent; cf. Remark 23. Therefore, results will be reported only for $c \in \{5, 10, 25, 50, 75, 100, 125, 150, 200, 250\}$. Recall, however, that this limiting condition is not required for the asymptotic validity of the FRWB statistic. Results from DGP2 are reported in Table 2; additional results for tests run at the 1% and 10% significance levels can be found in Table D.5 of the supplementary appendix.

Consider first the results for the homoskedastic DGP1. A comparison of the results in Table 1 for $\phi = -0.95, -0.90$, and -0.50, respectively, shows that when the innovations are homoskedastic the endogeneity correlation parameter, ϕ, has relatively little impact on the size properties of the two-sided tests, at least for cases where the autoregressive parameter c is positive and not close to zero. Here there is relatively little difference between the tests based on asymptotic critical values and the corresponding RWB and FRWB bootstrap tests. For all of these cases the two-sided tests display finite sample size close to the nominal level. However, where x_t is mildly explosive with $c = -5$ there is a tendency to undersize in t_{zx}, t_{zx}^{EW} and $t_{zx}^{*,FRWB}$ for $\phi = -0.95$ and $\phi = -0.90$ which is largely redressed by $t_{zx}^{*,RWB}$. For $0 \leq c \leq 10$ slight oversizing is also seen for both $\phi = -0.95$ and $\phi = -0.90$ with t_{zx}, t_{zx}^{EW} and $t_{zx}^{*,FRWB}$ which is again largely eliminated by $t_{zx}^{*,RWB}$.

A different picture emerges for the one-sided implementations of the tests. The one-sided t_{zx}, t_{zx}^{EW} and $t_{zx}^{*,FRWB}$ tests display severe size distortions for $c < 50$ when $\phi = -0.95$. Specifically, for $\phi = -0.95$ the left-tailed t_{zx}, t_{zx}^{EW} and $t_{zx}^{*,FRWB}$ tests display very significant undersizing, while their right-tailed counterparts are severely oversized (for instance when $c < 10$ empirical size is in most cases more than double the nominal size). The size distortions observed for these one-sided tests decrease, other things equal, as $|\phi|$ decreases, but significant size distortions are still observed even for $\phi = -0.5$. We also observe that the empirical rejection frequencies of the one-sided t_{zx}, t_{zx}^{EW} and $t_{zx}^{*,FRWB}$ tests under DGP1 are all very similar to each other for given values of ϕ and c. Consequently, the FRWB based implementations of the one-sided IVX tests do not appear to offer any tangible improvement on the finite sample size properties of the

[9] Notice that, because we report results for both left-tailed, right-tailed and two-tailed tests, it is not necessary to report results for positive values of ϕ; cf. Campbell and Yogo (2006, p. 30).

Table 1
Empirical rejection frequencies at 5% significance level of one-sided (left and right tail) and two-sided predictability tests, for sample sizes $T = 250$ and $T = 1000$. DGP1 (**homoskedastic IID innovations**): $y_t = \beta x_{t-1} + u_t, x_t = \rho x_{t-1} + w_t$ and$w_t = \psi w_{t-1} + v_t$, where $\beta = 0$, $\rho = 1 - c/T$, $\psi = 0$ and $(u_t, v_t)' \sim NIID(\mathbf{0}, \Sigma)$, with $\Sigma = [1 \quad \phi; \quad \phi \quad 1]$.

ϕ	c	Left-sided tests								Right-sided tests								Two-sided tests							
		$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}
		$T = 250$				$T = 1000$				$T = 250$				$T = 1000$				$T = 250$				$T = 1000$			
−0.95	−5	0.046	0.004	0.004	0.003	0.045	0.002	0.003	0.003	0.046	0.074	0.080	0.073	0.039	0.064	0.065	0.064	0.048	0.038	0.044	0.039	0.040	0.030	0.032	0.031
	−2.5	0.045	0.000	0.000	0.001	0.046	0.000	0.000	0.000	0.041	0.094	0.097	0.093	0.041	0.092	0.092	0.091	0.038	0.040	0.048	0.044	0.037	0.042	0.043	0.042
	0	0.041	0.001	0.001	0.001	0.042	0.001	0.001	0.001	0.053	0.105	0.114	0.110	0.050	0.103	0.104	0.102	0.047	0.051	0.057	0.053	0.041	0.050	0.050	0.049
	2.5	0.062	0.005	0.005	0.005	0.060	0.006	0.006	0.006	0.064	0.112	0.116	0.115	0.059	0.107	0.108	0.107	0.053	0.058	0.062	0.060	0.050	0.056	0.057	0.058
	5	0.068	0.010	0.011	0.010	0.068	0.013	0.014	0.013	0.062	0.107	0.116	0.112	0.059	0.105	0.106	0.106	0.054	0.058	0.063	0.060	0.052	0.056	0.058	0.058
	10	0.064	0.019	0.019	0.018	0.064	0.022	0.021	0.021	0.062	0.097	0.102	0.099	0.059	0.097	0.098	0.098	0.055	0.060	0.066	0.060	0.055	0.061	0.063	0.062
	25	0.057	0.029	0.030	0.028	0.061	0.030	0.030	0.031	0.057	0.078	0.084	0.080	0.055	0.081	0.082	0.081	0.056	0.056	0.060	0.058	0.052	0.056	0.057	0.056
	50	0.056	0.034	0.036	0.035	0.057	0.035	0.035	0.035	0.052	0.067	0.072	0.067	0.051	0.070	0.071	0.070	0.051	0.051	0.054	0.052	0.051	0.053	0.054	0.053
	75	0.056	0.037	0.038	0.037	0.055	0.039	0.039	0.038	0.053	0.064	0.068	0.065	0.051	0.067	0.068	0.067	0.049	0.047	0.052	0.049	0.051	0.052	0.054	0.052
	100	0.054	0.038	0.040	0.038	0.055	0.040	0.040	0.040	0.053	0.061	0.065	0.062	0.051	0.065	0.066	0.065	0.049	0.048	0.052	0.050	0.051	0.051	0.053	0.052
	125	0.054	0.039	0.042	0.041	0.054	0.041	0.042	0.041	0.052	0.060	0.063	0.060	0.051	0.063	0.063	0.063	0.050	0.049	0.053	0.051	0.050	0.051	0.052	0.052
	150	0.055	0.043	0.046	0.042	0.053	0.043	0.043	0.043	0.053	0.056	0.060	0.059	0.051	0.061	0.061	0.061	0.051	0.049	0.054	0.052	0.050	0.052	0.053	0.051
	200	0.054	0.046	0.048	0.045	0.053	0.044	0.044	0.042	0.050	0.054	0.056	0.053	0.050	0.059	0.060	0.059	0.050	0.048	0.054	0.050	0.052	0.051	0.052	0.051
	250	0.054	0.048	0.051	0.048	0.053	0.044	0.044	0.044	0.051	0.051	0.055	0.053	0.050	0.059	0.059	0.059	0.049	0.048	0.053	0.050	0.050	0.051	0.051	0.051
−0.90	−5	0.049	0.004	0.005	0.004	0.047	0.003	0.003	0.003	0.044	0.074	0.079	0.073	0.041	0.066	0.067	0.064	0.045	0.036	0.044	0.038	0.041	0.032	0.034	0.033
	−2.5	0.047	0.000	0.001	0.001	0.048	0.000	0.000	0.000	0.042	0.093	0.100	0.094	0.040	0.092	0.091	0.090	0.037	0.042	0.047	0.044	0.037	0.040	0.043	0.041
	0	0.039	0.001	0.001	0.001	0.040	0.002	0.002	0.002	0.054	0.104	0.112	0.108	0.051	0.098	0.101	0.100	0.048	0.052	0.057	0.053	0.043	0.048	0.049	0.048
	2.5	0.059	0.005	0.005	0.005	0.058	0.007	0.007	0.007	0.063	0.110	0.114	0.112	0.058	0.103	0.106	0.105	0.055	0.059	0.062	0.061	0.050	0.056	0.057	0.057
	5	0.066	0.010	0.011	0.010	0.065	0.014	0.014	0.014	0.062	0.106	0.114	0.109	0.059	0.100	0.102	0.102	0.054	0.059	0.064	0.060	0.051	0.058	0.059	0.058
	10	0.063	0.019	0.020	0.020	0.063	0.023	0.022	0.022	0.061	0.094	0.100	0.096	0.059	0.094	0.095	0.094	0.055	0.057	0.065	0.061	0.054	0.059	0.061	0.060
	25	0.055	0.030	0.032	0.030	0.059	0.031	0.031	0.031	0.058	0.078	0.082	0.079	0.055	0.080	0.080	0.078	0.054	0.055	0.060	0.057	0.052	0.056	0.056	0.055
	50	0.054	0.033	0.036	0.034	0.056	0.036	0.036	0.036	0.053	0.067	0.071	0.068	0.051	0.069	0.070	0.069	0.049	0.050	0.054	0.052	0.051	0.053	0.054	0.052
	75	0.054	0.038	0.038	0.037	0.054	0.039	0.040	0.039	0.052	0.064	0.067	0.064	0.052	0.066	0.066	0.066	0.048	0.049	0.053	0.049	0.051	0.053	0.053	0.053
	100	0.050	0.037	0.040	0.039	0.054	0.040	0.041	0.040	0.052	0.062	0.065	0.062	0.053	0.065	0.064	0.064	0.049	0.048	0.052	0.050	0.052	0.052	0.052	0.053
	125	0.053	0.041	0.043	0.041	0.055	0.044	0.043	0.043	0.053	0.059	0.062	0.059	0.051	0.060	0.062	0.062	0.052	0.050	0.054	0.052	0.051	0.052	0.053	0.053
	150	0.054	0.043	0.045	0.044	0.055	0.045	0.044	0.044	0.052	0.056	0.059	0.058	0.051	0.061	0.060	0.060	0.049	0.050	0.054	0.052	0.051	0.052	0.054	0.052
	200	0.054	0.046	0.049	0.047	0.054	0.045	0.046	0.046	0.051	0.055	0.058	0.055	0.053	0.059	0.061	0.061	0.049	0.048	0.052	0.050	0.049	0.050	0.053	0.051
	250	0.055	0.048	0.051	0.048	0.052	0.046	0.046	0.046	0.048	0.049	0.053	0.051	0.051	0.058	0.059	0.060	0.049	0.049	0.054	0.049	0.050	0.050	0.051	0.050
−0.50	−5	0.053	0.019	0.024	0.019	0.050	0.018	0.018	0.018	0.046	0.072	0.079	0.072	0.047	0.070	0.072	0.070	0.048	0.043	0.055	0.042	0.047	0.042	0.043	0.041
	−2.5	0.050	0.006	0.007	0.005	0.048	0.004	0.004	0.004	0.053	0.101	0.107	0.101	0.047	0.095	0.096	0.094	0.049	0.050	0.059	0.051	0.043	0.047	0.047	0.045
	0	0.030	0.005	0.006	0.006	0.032	0.009	0.008	0.008	0.061	0.097	0.102	0.096	0.056	0.091	0.090	0.091	0.048	0.051	0.057	0.053	0.045	0.049	0.050	0.048
	2.5	0.045	0.016	0.017	0.016	0.049	0.018	0.018	0.018	0.061	0.090	0.096	0.090	0.058	0.087	0.087	0.086	0.052	0.054	0.059	0.055	0.050	0.052	0.053	0.052
	5	0.050	0.023	0.024	0.023	0.054	0.024	0.024	0.025	0.061	0.081	0.087	0.085	0.057	0.082	0.081	0.080	0.052	0.055	0.059	0.057	0.050	0.053	0.054	0.053
	10	0.052	0.030	0.031	0.030	0.053	0.031	0.031	0.031	0.056	0.074	0.078	0.076	0.053	0.073	0.074	0.074	0.052	0.053	0.057	0.055	0.050	0.052	0.054	0.054
	25	0.050	0.036	0.038	0.036	0.053	0.039	0.039	0.039	0.053	0.065	0.069	0.067	0.052	0.064	0.065	0.063	0.049	0.050	0.054	0.052	0.049	0.050	0.052	0.051
	50	0.048	0.039	0.040	0.039	0.051	0.042	0.042	0.043	0.054	0.063	0.066	0.064	0.049	0.057	0.059	0.058	0.049	0.050	0.054	0.050	0.054	0.052	0.053	0.053
	75	0.050	0.042	0.045	0.043	0.053	0.045	0.045	0.045	0.055	0.060	0.065	0.062	0.049	0.056	0.057	0.057	0.049	0.049	0.054	0.052	0.053	0.053	0.054	0.053
	100	0.049	0.043	0.045	0.043	0.052	0.047	0.046	0.045	0.054	0.059	0.063	0.061	0.049	0.056	0.059	0.057	0.050	0.048	0.054	0.052	0.052	0.052	0.053	0.052
	125	0.051	0.044	0.046	0.045	0.051	0.045	0.046	0.045	0.055	0.059	0.061	0.059	0.049	0.054	0.056	0.055	0.051	0.050	0.055	0.050	0.052	0.052	0.052	0.053
	150	0.051	0.046	0.048	0.047	0.052	0.046	0.046	0.046	0.055	0.057	0.061	0.058	0.050	0.055	0.055	0.054	0.051	0.049	0.054	0.052	0.052	0.052	0.053	0.054
	200	0.052	0.047	0.050	0.048	0.053	0.047	0.046	0.046	0.053	0.054	0.057	0.054	0.050	0.054	0.054	0.054	0.049	0.049	0.055	0.052	0.052	0.052	0.053	0.053
	250	0.053	0.049	0.052	0.051	0.052	0.048	0.047	0.046	0.051	0.052	0.055	0.051	0.050	0.053	0.054	0.055	0.049	0.049	0.053	0.050	0.050	0.050	0.051	0.051

Note: t_{zx} and t_{zx}^{EW} correspond to the statistics presented in (12) and (14) of the main text, and $t_{zx}^{*,RWB}$ and $t_{zx}^{*,FRWB}$ are the corresponding residual Wild bootstrap (RWB) and fixed regressor wild bootstrap (FRWB) implementations of (12) computed as described in Algorithms 1 and 2 of Section 4.

Table 2
Empirical rejection frequencies at 5% significance level of one-sided (left and right tail) and two-sided predictability tests, for sample sizes $T = 250$ and $T = 1000$. **DGP2 (ARCH with Leverage Effects):** $\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0 \\ \rho x_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} 0 \\ \rho x_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \boldsymbol{\psi}_t$ with $\boldsymbol{\psi}_t = (a_t; e_t)' = (\varepsilon_{1t}\sqrt{1 + \frac{1}{2}a_{t-1}^2 \mathbb{I}_{\{a_{t-1}<0\}}}; \varepsilon_{2t})'$ and $(\varepsilon_{1t}, \varepsilon_{2t})' \sim NIID(0, \mathbf{I}_2)$.

c	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}
	$T = 250$				$T = 1000$			
	Left-sided tests							
5	0.062	0.099	0.106	0.107	0.060	0.097	0.101	0.102
10	0.059	0.088	0.096	0.099	0.059	0.090	0.092	0.093
25	0.059	0.076	0.081	0.092	0.055	0.076	0.075	0.080
50	0.058	0.067	0.074	0.089	0.053	0.066	0.068	0.074
75	0.060	0.062	0.069	0.090	0.055	0.064	0.066	0.075
100	0.060	0.061	0.067	0.089	0.057	0.061	0.063	0.076
125	0.059	0.060	0.066	0.088	0.058	0.061	0.062	0.078
150	0.059	0.058	0.063	0.085	0.059	0.061	0.063	0.080
200	0.057	0.056	0.060	0.082	0.061	0.060	0.063	0.085
250	0.053	0.053	0.057	0.078	0.062	0.062	0.063	0.088
	Right-sided tests							
5	0.058	0.014	0.015	0.015	0.057	0.013	0.013	0.013
10	0.059	0.021	0.022	0.024	0.057	0.021	0.020	0.021
25	0.061	0.030	0.030	0.039	0.057	0.030	0.030	0.034
50	0.062	0.037	0.038	0.053	0.058	0.037	0.037	0.045
75	0.061	0.039	0.041	0.061	0.061	0.041	0.040	0.052
100	0.059	0.037	0.040	0.065	0.060	0.041	0.041	0.055
125	0.057	0.039	0.041	0.067	0.061	0.042	0.042	0.058
150	0.059	0.040	0.042	0.071	0.060	0.042	0.042	0.062
200	0.056	0.041	0.045	0.072	0.062	0.042	0.042	0.067
250	0.054	0.043	0.047	0.075	0.062	0.042	0.044	0.071
	Two-sided tests							
5	0.053	0.055	0.065	0.063	0.051	0.056	0.058	0.058
10	0.051	0.053	0.060	0.062	0.053	0.056	0.058	0.059
25	0.056	0.052	0.057	0.069	0.053	0.053	0.055	0.059
50	0.059	0.050	0.056	0.081	0.054	0.051	0.052	0.062
75	0.061	0.051	0.057	0.090	0.059	0.052	0.054	0.069
100	0.063	0.053	0.058	0.095	0.061	0.055	0.054	0.074
125	0.062	0.051	0.059	0.098	0.064	0.054	0.055	0.079
150	0.060	0.051	0.057	0.097	0.064	0.054	0.056	0.084
200	0.058	0.050	0.056	0.096	0.068	0.053	0.056	0.089
250	0.053	0.051	0.054	0.092	0.068	0.052	0.055	0.097

Note: t_{zx} and t_{zx}^{EW} correspond to the statistics presented in (12) and (14) of the main text, and $t_{zx}^{*,RWB}$ and $t_{zx}^{*,FRWB}$ are the corresponding residual Wild bootstrap (RWB) and fixed regressor wild bootstrap (FRWB) implementations of (12) computed as described in Algorithms 1 and 2 of Section 4.

asymptotic tests, as might be expected in the light of Remark 19. In contrast, both the left-sided and right-sided tests implemented with the RWB offer empirical size properties close to the nominal level throughout.

Consider next the results in Table 2 for DGP2 where the conditional variance of $(u_t, v_t)'$ follows an ARCH model with leverage effects. The results show that in general the two-sided versions of the t_{zx}^{EW}, $t_{zx}^{*,RWB}$ and $t_{zx}^{*,FRWB}$ tests all display reasonable size control throughout. In contrast, significant size distortions are seen for the two-sided t_{zx} test. The latter finding is consistent with our discussion in Remark 15 on the non-pivotal nature of the limiting null distribution of t_{zx} under conditional heteroskedasticity when x_t is weakly dependent. Large size distortions are also seen for the one-sided t_{zx} tests. Moreover, and as observed with DGP1, although the two-sided t_{zx}^{EW} and $t_{zx}^{*,FRWB}$ tests show decent finite sample size control the same is not true of the one-sided versions of these tests. In contrast the one-sided $t_{zx}^{*,RWB}$ tests deliver decent finite sample size control for all values of c and regardless of the sample size. Consequently, although the limiting condition $M_{\xi u}^{*} \overset{d}{=} M_{\xi u}$ formally required for the asymptotic validity of the RWB tests is not met by DGP2, the results in Table 2 suggest that $t_{zx}^{*,RWB}$ nonetheless displays arguably the most reliable finite sample size control among the tests considered for data generated according to DGP2.

Summary of additional results

We also investigated the impact on the finite sample performance of the IVX statistics and their bootstrap implementations from a variety of additional empirically relevant models. Full details of the simulation DGPs considered and the tabulated results (which appear in Tables D.7–D.42) are given in the supplementary appendix. In what follows we provide a summary of these results:

(1). We repeated the experiments in Table 1 for the case where v_t follows either a positively autocorrelated (DGP3) or a negatively autocorrelated (DGP4) stationary AR(1) process. These results, which can be found in Tables D.7–D.14, were qualitatively very similar to those reported for serially uncorrelated v_t in Table 1.

(2). We consider two DGPs which include a contemporaneous one-time break of equal magnitude in the unconditional variances of u_t and v_t, as in Georgiev et al. (2018) and Demetrescu et al. (2022). The first, labelled DGP5, contains an upward change in the unconditional variances of u_t and v_t at the sample midpoint (Tables D.15–D.18), while the second, labelled DGP6, contains a corresponding downward change in the unconditional variances of u_t and v_t (Tables D.19–D.22).

The results reported in Tables D.15 to D.22 reveal that, as expected, the two-sided IVX test with conventional standard errors, t_{zx}, displays significant size distortions. For example, for a 5% significance level and $\phi = -0.95$ the rejection frequencies observed across all values of c considered, when an upward change in variance occurs (Table D.15) are in the range $[0.064, 0.095]$ for $T = 250$ and $[0.066, 0.097]$ for $T = 1000$. For a downward change in variance (Table D.19) results are similar ($[0.017, 0.098]$ for $T = 250$ and $[0.018, 0.091]$ for $T = 1000$), except for cases where $c < 0$ (mildly explosive predictors) in which case some undersizing is observed. The magnitude of these size distortions are relatively stable across ϕ.

In contrast, for the one-sided versions of t_{zx} the empirical size distortions for the former worsen, other things equal, as $|\phi|$ increases. For example, for DGP5 with $T = 250$ and $\phi = -0.95$ the range of empirical rejection frequencies for the left-sided tests is $[0.003, 0.075]$ and for the right-sided tests $[0.085, 0.151]$; see Table D.15. On the other hand, for $\phi = 0$ the left and right-sided tests rejection frequencies' range is $[0.064, 0.081]$; see Table D.18.

The size distortions seen in the two-sided t_{zx} test for DGP5 and DGP6 are significantly ameliorated by using Eicker–White standard errors (t_{zx}^{EW}) when $c \geq -2.5$. However, the one-sided (left and right-sided) t_{zx}^{EW} tests do not improve much relative to t_{zx} when $c \leq 25$; see Tables D.15 to D.22.

The RWB and FRWB bootstrap implementations of the two-sided t_{zx} test both do a very good job of controlling finite sample size in the presence of unconditional heteroskedasticity. For the one-sided tests, $t_{zx}^{*,RWB}$ displays empirical rejection frequencies which are again in general close to the nominal significance level considered, regardless of the values of c and ϕ. In contrast, the one-sided $t_{zx}^{*,FRWB}$ test displays significant size distortions for values of $c \leq 25$; these improve as $|\phi|$ decreases, as anticipated by the discussion in Remark 19.

(3). To further evaluate the impact of conditional heteroskedasticity we considered three further volatility specifications: (i) a GARCH(1,1) model with either $N(0, 1)$ (DGP7) or Student-t distributed innovations with 5 degrees of freedom [t_5] (DGP8); (ii) a GoGARCH(1,1) model, see for example (Van der Weide, 2002), allowing for either $N(0, 1)$ (DGP9) or t_5 innovations (DGP10); and (iii) an autoregressive stochastic volatility process (DGP11).

As observed earlier, the non-pivotal nature of the t_{zx} statistic's limiting null distribution under GARCH type conditional heteroskedasticity is also apparent in the results in Tables D.23 to D.26 and D.27 to D.30 corresponding to DGP7 and DGP8, respectively. These results highlight that the size distortion of the two-sided t_{zx} statistic increases as $|\phi|$ increases regardless of whether $N(0, 1)$ (Tables D.23 to D.26) or Student-t innovations (Tables D.27 to D.30) are used in generating the data. The magnitude of the size distortions is, however, considerably exacerbated when the innovations are heavy tailed (DGP8). For instance, for $N(0, 1)$ innovations, $T = 250$, $\phi = -0.95$ and for a 5% significance level the range of the empirical rejection frequencies for t_{zx} is $[0.042, 0.082]$, whereas for t_5 innovations the range is $[0.081, 0.167]$. The Eicker–White correction does a good job in correcting the size distortion of the two-sided t_{zx} test regardless of whether the innovations are $N(0, 1)$ or Student-t distributed. In the previous example, the ranges of the rejection frequencies of t_{zx}^{EW} when the innovations are $N(0, 1)$ and t_5 distributed is $[0.047, 0.066]$ and $[0.062, 0.068]$, respectively. The results also show that the RWB and FRWB both display good empirical size properties in a two-sided hypothesis testing context. However, for one-sided testing $t_{zx}^{*,RWB}$ delivers significantly better finite sample size control than $t_{zx}^{*,FRWB}$ when x_t is strongly persistent, while they display similar performance for weaker levels of persistence in x_t. Overall $t_{zx}^{*,RWB}$ is the best performing test regardless of the nominal significance levels used and regardless of the underlying distribution of the innovations. All of the other one-sided tests display serious size distortions when the predictor is strongly persistent ($c < 25$), for both $N(0, 1)$ or t_5 distributed innovations.

For the GoGARCH models (DGP9 and DGP10 in Tables D.31 to D.34 and Tables D.35 to D.38, respectively), qualitatively similar conclusions can be drawn to those discussed above for the GARCH(1,1) case albeit the magnitude of the size distortions observed for the $t_{zx}^{*,FRWB}$, t_{zx}^{EW} and t_{zx} tests are generally smaller.

Finally, for stochastic volatility (DGP11), the results in Tables D.39 to D.42 suggest that all of the two-sided tests display adequate finite sample size control, with the exception of t_{zx}^{EW} which is oversized for $T = 250$, although its size properties are improved for $T = 1000$. For the one-sided tests, similar conclusions are drawn as for the GARCH and GoGARCH specifications. Specifically, $t_{zx}^{*,FRWB}$, t_{zx}^{EW} and t_{zx} are considerably oversized when the predictor is strongly persistent and $\phi = -0.95$, but $t_{zx}^{*,RWB}$ displays reliable empirical rejection frequencies, across c.

5.1.2. Finite sample local power

We next provide a brief analysis of the finite sample local power properties of one-sided and two-sided implementations of the IVX tests from Section 3.1 together with their bootstrap analogues from Section 4 and compare these with the 2SLS predictability tests of Breitung and Demetrescu (2015). To that end, we simulate data from DGP1 under a variety of local alternatives. For the sake of space, we only report results for $\phi = \{-0.95, -0.50\}$, for a sample of size $T = 250$ and

for four values of the persistence parameter, c, associated with x_t; specifically, $c = \{0, 10, 25, 50\}$. The slope parameter β is parameterised in (1) as $\beta = b/T$, with results reported for $b \in \{-20, -19, \ldots, 19, 20\}$.

Due to the large finite sample size distortions seen with the one-sided t_{zx}, t_{zx}^{EW} and $t_{zx}^{*,FRWB}$ tests discussed in Section 5.1.1 for these combinations of c and ϕ, we only report local power results for the two-sided t_{zx}, $t_{zx}^{*,RWB}$ and $t_{zx}^{*,FRWB}$ tests and for the one-sided $t_{zx}^{*,RWB}$ test all of which have well controlled empirical size under DGP1. Results are also reported for the two-sided test of Breitung and Demetrescu (2015), denoted $t_{zx}^{*,2SLS}$, implemented with a fixed regressor wild bootstrap (the asymptotic validity of which is established in Demetrescu et al., 2022) using the choice of instruments recommended by Breitung and Demetrescu (2015), namely the type-I fractionally integrated instrument $z_{1t} := (1-L)_+^{0.51} x_t$ and type-II sine instrument $z_{2t} := \sin(\frac{\pi}{2} t/T)$. The finite sample local power curves of these tests are graphed in Fig. 1. Recalling from Remark 27 that the RWB and FRWB implementations of the IVX tests share the same asymptotic local power functions as the corresponding (size-adjusted) asymptotic IVX test, Fig. 1 shows that this prediction from the limiting theory is borne out well even for a sample of size $T = 250$ with the power curves of the bootstrap and asymptotic two-sided tests being almost indistinguishable from each other for all of the values of c considered. The power dominance of the IVX tests, both one-sided and two-sided, over the two-sided 2SLS $t_{zx}^{*,2SLS}$ test is clearly seen in Fig. 1, confirming the findings of Harvey et al. (2021). For alternatives where $\beta < 0$ the gains from using the left-tailed IVX tests over the two-sided IVX tests are also clearly seen in Fig. 1, with the magnitude of the power gains from using the one-sided tests generally larger for $\phi = -0.95$ *vis-à-vis* $\phi = -0.5$, and greater the larger is c. For alternatives where $\beta > 0$, the gains from using a right-tailed IVX test over the two-tailed IVX test are much less obvious than for testing against $\beta < 0$, but are nonetheless still apparent for $c \geq 10$ when $\phi = -0.50$ and for $c \geq 25$ when $\phi = -0.95$.

5.2. Multiple predictors — full sample tests

We now investigate the finite sample behaviour of the asymptotic IVX test and its RWB and FRWB bootstrap counterparts in cases where multiple predictors are included in the predictive regression. For our analysis we use the same DGP as is considered in Xu and Guo (2021); that is,

$$y_t = \alpha + \boldsymbol{x}_{t-1}'\boldsymbol{\beta} + u_t, \qquad t = 1, \ldots, T, \tag{23}$$

$$\boldsymbol{x}_t = \boldsymbol{\Gamma}\boldsymbol{x}_{t-1} + \boldsymbol{v}_t, \qquad t = 0, \ldots, T, \tag{24}$$

where $\boldsymbol{x}_t := (x_{1,t}, \ldots, x_{K,t})'$ is a $K \times 1$ vector of predictor variables, $\boldsymbol{\beta}$ is a $K \times 1$ vector of parameters, $\alpha = 0.25$, $\boldsymbol{\Gamma}$ is the $K \times K$ diagonal matrix $\boldsymbol{\Gamma} := \mathrm{diag}(\rho, \ldots, \rho)$, and $(u_t, \boldsymbol{v}_t')' \sim NIID(\boldsymbol{0}, \boldsymbol{\Sigma})$ where

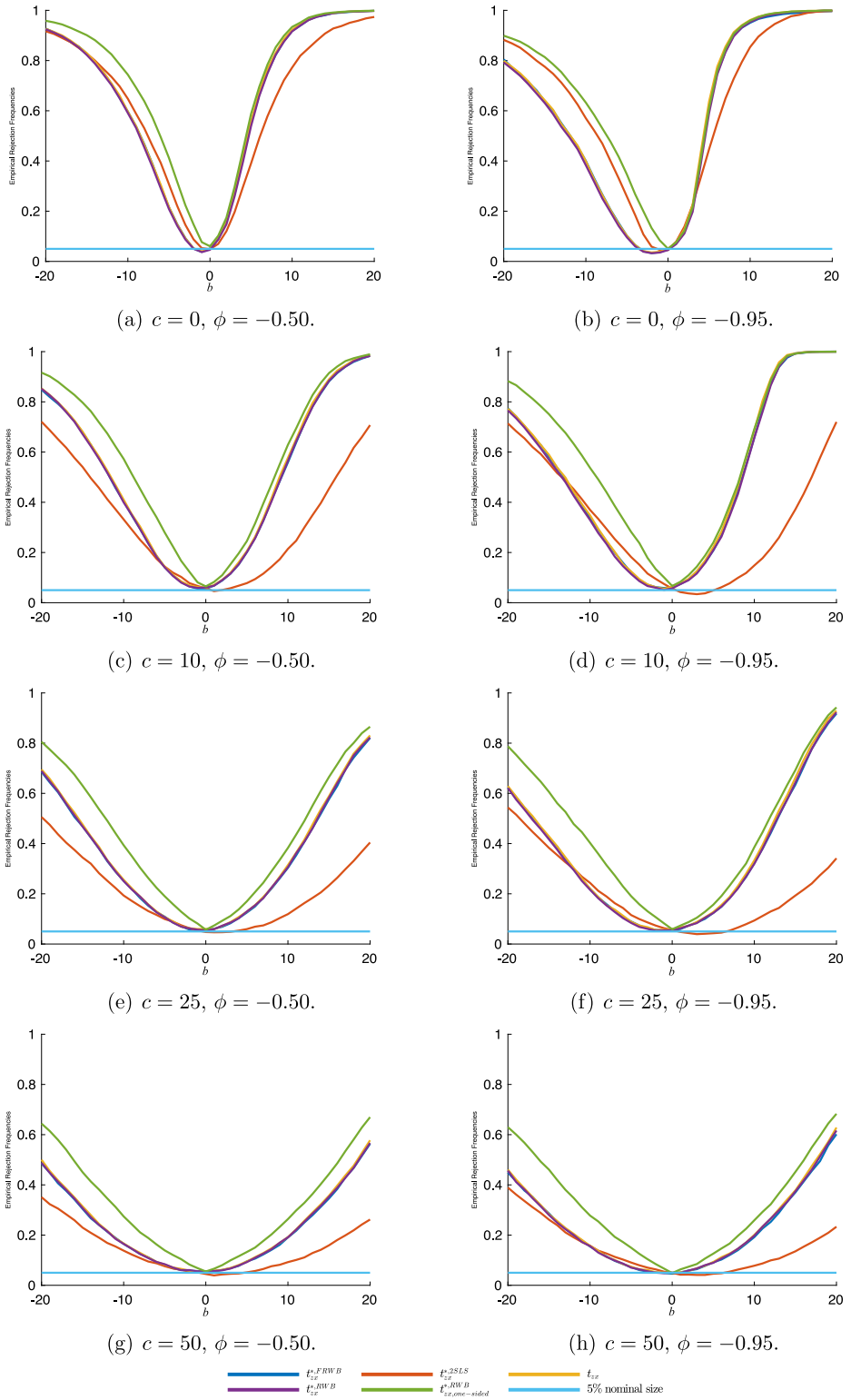
$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_u^2 & \sigma_{u,v_1} & 0 & \cdots & 0 \\ \sigma_{u,v_1} & \sigma_{v_1}^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_{v_2}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_{v_K}^2 \end{pmatrix} \tag{25}$$

with $\sigma_u^2 = 0.037$, $\sigma_{u,v_1} = -0.035$, $\sigma_{v_1}^2 = \cdots = \sigma_{v_K}^2 = 0.045$. Notice, therefore, that the first predictor, $x_{1,t}$ is endogenous (with an endogeneity correlation parameter $\phi_1 = -0.83$), while the remaining predictors $x_{2,t}, \ldots, x_{K,t}$ are exogenous. For the autoregressive parameter we again consider $\rho = 1 - c/T$ with $c \in \{-5, 2.5, 0, 2.5, 5, 10, 25, 50, 75, 100, 125, 150, 200, 250\}$.

Table 3 reports the empirical rejection frequencies at a 5% significance level, for $T = 250$ and $T = 1000$ and for $K \in \{1, 3, 5, 10\}$, for the Wald-type IVX tests W_{zx} and W_{zx}^{EW} discussed in Remark 9, together with the RWB and FRWB bootstrap implementations of W_{zx}, denoted $W_{zx}^{*,RWB}$ and $W_{zx}^{*,FRWB}$, respectively, computed as described in Remark 22 (results for 1% and 10% significance levels are reported in Table D.6 of the supplementary appendix). In the context of $W_{zx}^{*,RWB}$, in Step 2 of the multivariate version of Algorithm 1 autoregressions of length $p+1$ were fitted to each element of $\boldsymbol{x}_t$ with p selected in each case by BIC using the same range of values of p as were used in the simulations for a single predictor.

For $K = 1$ (the single predictor case), and in line with what was observed in Section 5.1.1 for the two-sided tests based under DGP1, all of the Wald-based IVX statistics display empirical rejection frequencies close to the nominal level. Again, $W_{zx}^{*,RWB}$ displays the smallest size distortions among the tests considered. For instance, for a 5% significance level the rejection frequencies of $W_{zx}^{*,RWB}$ are in the range [0.042, 0.056] for $T = 250$ and [0.038, 0.056] for $T = 1000$, whereas for $W_{zx}^{*,FRWB}$, W_{zx}^{EW} and W_{zx} these are [0.037, 0.058], [0.045, 0.064], and [0.040, 0.060], respectively, when $T = 250$ and [0.034, 0.060], [0.036, 0.060] and [0.035, 0.059], respectively, when $T = 1000$.

However, it is as K increases that the significant advantage of the RWB becomes clear, particularly in the case where the predictors are strongly persistent. It is clear from the results that the $W_{zx}^{*,FRWB}$, W_{zx}^{EW} and W_{zx} tests are not reliable when the predictors are strongly persistent. The rejection frequencies we observe for W_{zx} are in line with those reported in Xu and Guo (2021) who also show that the quality of the prediction from the asymptotic theory deteriorates as the number of regressors, K, specified in the predictive regression increases. For instance, for $K = 3$ and $c < 0$; for $K = 5$ and $c < 2.5$; and for $K = 10$ and $c < 25$, even for $T = 1000$ all three of these tests display rejection frequencies larger than



(a) $c = 0$, $\phi = -0.50$. (b) $c = 0$, $\phi = -0.95$.

(c) $c = 10$, $\phi = -0.50$. (d) $c = 10$, $\phi = -0.95$.

(e) $c = 25$, $\phi = -0.50$. (f) $c = 25$, $\phi = -0.95$.

(g) $c = 50$, $\phi = -0.50$. (h) $c = 50$, $\phi = -0.95$.

Fig. 1. Power plots for two-sided $t_{zx}^{*,FRWB}$, $t_{zx}^{*,2SLS}$, t_{zx}, $t_{zx}^{*,RWB}$ tests and the one-sided $t_{zx}^{*,RWB}$ tests for predictability. Data generated from DGP1 with $\phi = \{-0.95, -0.50\}$ and for $T = 250$.
Note: The green line corresponds to the rejection frequencies of the left-sided and right-sided RWB-based t-tests in the relevant tail: i.e., for $b < 0$ the line corresponds to the left-sided test, while for $b > 0$ it corresponds to the right-sided test.

Table 3
Empirical rejection frequencies at 5% significance level of Wald-type IVX-based tests for predictability in a multiple predictive regression context with $K \in \{1, 3, 5, 10\}$ predictors, for sample sizes $T = 250$ and $T = 1000$.

K	c	$W_{zx}^{*,\mathrm{RWB}}$	$W_{zx}^{*,\mathrm{FRWB}}$	W_{zx}^{EW}	W_{zx}	$W_{zx}^{*,\mathrm{RWB}}$	$W_{zx}^{*,\mathrm{FRWB}}$	W_{zx}^{EW}	W_{zx}
		$T = 250$				$T = 1000$			
1	−5	0.048	0.037	0.045	0.040	0.043	0.034	0.036	0.035
	−2.5	0.042	0.045	0.051	0.048	0.038	0.042	0.043	0.043
	0	0.050	0.053	0.059	0.056	0.045	0.051	0.051	0.050
	2.5	0.054	0.058	0.062	0.059	0.051	0.054	0.056	0.056
	5	0.056	0.058	0.064	0.060	0.054	0.059	0.060	0.058
	10	0.054	0.056	0.061	0.056	0.055	0.060	0.060	0.059
	25	0.055	0.057	0.060	0.057	0.056	0.059	0.059	0.059
	50	0.054	0.055	0.059	0.056	0.055	0.059	0.059	0.058
	75	0.055	0.055	0.060	0.056	0.056	0.057	0.058	0.058
	100	0.055	0.054	0.059	0.055	0.056	0.056	0.057	0.058
	125	0.055	0.054	0.058	0.054	0.056	0.056	0.057	0.056
	150	0.053	0.051	0.056	0.052	0.055	0.053	0.055	0.055
	200	0.052	0.051	0.056	0.052	0.054	0.054	0.053	0.052
	250	0.051	0.049	0.055	0.051	0.053	0.052	0.053	0.053
3	−5	0.085	0.352	0.385	0.366	0.083	0.346	0.354	0.346
	−2.5	0.097	0.176	0.193	0.177	0.092	0.162	0.159	0.155
	0	0.075	0.105	0.117	0.104	0.071	0.096	0.096	0.095
	2.5	0.067	0.086	0.103	0.090	0.063	0.081	0.084	0.083
	5	0.059	0.077	0.095	0.083	0.060	0.076	0.079	0.077
	10	0.054	0.066	0.083	0.071	0.057	0.072	0.078	0.075
	25	0.052	0.061	0.075	0.066	0.056	0.064	0.067	0.065
	50	0.053	0.057	0.070	0.061	0.053	0.058	0.061	0.059
	75	0.053	0.053	0.069	0.058	0.050	0.055	0.058	0.055
	100	0.051	0.053	0.069	0.057	0.048	0.052	0.054	0.051
	125	0.052	0.054	0.070	0.058	0.048	0.049	0.054	0.050
	150	0.052	0.054	0.069	0.058	0.047	0.049	0.052	0.049
	200	0.052	0.055	0.071	0.059	0.046	0.048	0.051	0.048
	250	0.053	0.055	0.071	0.060	0.046	0.048	0.050	0.048
5	−5	0.074	0.402	0.466	0.421	0.074	0.398	0.408	0.403
	−2.5	0.091	0.239	0.281	0.241	0.091	0.237	0.238	0.230
	0	0.082	0.157	0.186	0.156	0.085	0.152	0.154	0.148
	2.5	0.069	0.120	0.156	0.129	0.069	0.118	0.126	0.117
	5	0.063	0.105	0.138	0.116	0.063	0.104	0.110	0.105
	10	0.062	0.086	0.120	0.098	0.058	0.089	0.096	0.092
	25	0.053	0.067	0.100	0.080	0.052	0.069	0.077	0.071
	50	0.052	0.059	0.089	0.069	0.049	0.057	0.064	0.059
	75	0.051	0.055	0.085	0.063	0.050	0.055	0.062	0.057
	100	0.049	0.053	0.082	0.062	0.051	0.056	0.060	0.057
	125	0.049	0.053	0.080	0.062	0.052	0.055	0.060	0.057
	150	0.046	0.052	0.078	0.061	0.051	0.055	0.059	0.056
	200	0.047	0.051	0.079	0.060	0.051	0.052	0.057	0.054
	250	0.044	0.049	0.077	0.058	0.051	0.052	0.058	0.054
10	−5	0.058	0.513	0.635	0.559	0.060	0.502	0.526	0.501
	−2.5	0.072	0.398	0.505	0.425	0.076	0.384	0.394	0.371
	0	0.087	0.306	0.406	0.324	0.091	0.295	0.300	0.280
	2.5	0.075	0.238	0.342	0.262	0.078	0.229	0.244	0.224
	5	0.067	0.191	0.301	0.225	0.068	0.188	0.211	0.191
	10	0.060	0.141	0.244	0.175	0.061	0.147	0.166	0.151
	25	0.050	0.089	0.174	0.118	0.057	0.101	0.119	0.108
	50	0.048	0.067	0.142	0.091	0.057	0.081	0.096	0.085
	75	0.046	0.060	0.129	0.081	0.055	0.071	0.085	0.077
	100	0.046	0.056	0.120	0.077	0.055	0.066	0.080	0.071
	125	0.043	0.053	0.117	0.074	0.055	0.061	0.077	0.067
	150	0.042	0.052	0.116	0.071	0.052	0.058	0.075	0.064
	200	0.039	0.049	0.116	0.070	0.053	0.057	0.070	0.063
	250	0.036	0.050	0.116	0.072	0.051	0.055	0.070	0.061

Note: W_{zx} and W_{zx}^{EW} are the Wald-type IVX-based statistics discussed in Remark 9 of the main text, and $W_{zx}^{*,RWB}$ and $W_{zx}^{*,FRWB}$ are the corresponding residual Wild bootstrap (RWB) and fixed regressor wild bootstrap (FRWB) versions of W_{zx} computed as described in Algorithms 1 and 2 of Section 4 of the main text.

15% at a 5% nominal level. For the smaller sample, $T = 250$, qualitatively similar size behaviour is observed (but with distortions of larger magnitude) for $W_{zx}^{*,FRWB}$ and W_{zx}. However, W_{zx}^{EW} becomes severely oversized as K increases, for all values of c. For instance, for $K = 10$, $T = 250$ and a 5% significance level, the smallest empirical rejection frequencies seen for this statistic is more than double the significance level considered. To illustrate the severity of the size distortions,

observe from Table 3 that, for $K = 10$ unit root predictors ($c = 0$) and a 5% significance level, the empirical rejection frequencies of $W_{zx}^{*,FRWB}$, W_{zx}^{EW} and W_{zx} are 30.6%, 40.6% and 32.4%, respectively, for $T = 250$, and 29.5%, 30.0% and 28.0%, respectively for $T = 1000$. For mildly explosive predictors, the situation is even worse with empirical size in the region of 70% for each of $W_{zx}^{*,FRWB}$, W_{zx}^{EW} and W_{zx} when $K = 10$ and $c = -5$.

In contrast, the residual wild bootstrap based test, $W_{zx}^{*,RWB}$, controls empirical size much better than the other tests with empirical rejection frequencies acceptably close to the nominal level for all of the values of K considered. Some size distortions remain for values of $c \leq 5$, albeit unlike with the other tests these do not get appreciably worse as K increases. Moreover, in those cases where size distortions are seen with the $W_{zx}^{*,RWB}$ test, these are very much smaller than those seen for those cases with the other tests. Indeed, there are no entries in Table 3 where $W_{zx}^{*,RWB}$ displays an empirical size in excess of 10%, which compares very favourably with the other tests.

Finally, although not reported here, we also investigated the finite sample behaviour of the partial IVX t-type tests discussed in Remark 9. To summarise our findings, for both one-sided and two-sided implementations, the t-type tests associated with the exogenous predictors, $x_{2,t}, \ldots, x_{K,t}$, all displayed qualitatively similar finite sample size properties to those which were observed for the single predictive regression case for DGP1 with $\phi = 0$ (see Table D.6 of the supplementary appendix), i.e., the t-type tests display empirical rejection frequencies which are close to the nominal significance levels considered. For the t-type tests associated with the endogenous predictor, $x_{1,t}$, both one-sided and two-sided versions of the RWB implementation of the tests continued to display good finite sample size control, regardless of the number of predictors, K, and the value of c. In contrast, the empirical sizes of the other implementations of the tests, including the FRWB tests, deteriorated very badly as K increased, rendering these tests highly unreliable in practice.

5.3. Subsample tests

We next summarise the findings from a Monte Carlo study of the finite sample performance of the subsample IVX-based predictability tests proposed in Section 3.2. The full set of results can be found in the supplementary appendix: the empirical size results in Tables D.43–D.48, and the empirical local power results in Figures D.1–D.3. Results are reported for two-sided and one-sided tests, implemented with either the RWB or FRWB, together with the corresponding rolling, forward and backward recursive two-sided 2SLS-based tests of Demetrescu et al. (2022) which use a type-I instrument constructed as in (4) and the type-II sine instrument, $z_{2t} := \sin(\frac{\pi}{2}t/T)$, and are implemented using the FRWB. For the recursive sequences we set $\tau_L = 1/3$ and $\tau_U = 2/3$, respectively, and for the rolling sequences we set $\Delta\tau = 1/3$. The empirical size and local power simulations are based on 3000 and 1000 Monte Carlo replications, respectively, and $B = 399$ bootstrap replications. All other computational aspects are as outlined previously.

5.3.1. Empirical size

The results in Tables D.43–D.48 pertain to data generated from DGP1, as described in Section 5.1.1, setting $\beta_t = 0$ for all t and $\rho = 1 - c/T$, with $c \in \{-5, -2.5, 0, 2.5, 5, 10, 25, 50, 75, 100, 125, 150, 200, 250\}$, and $T \in \{250, 1000\}$.

Consider first the rolling tests. The results in Tables D.43–D.44 suggest that the correlation parameter ϕ has relatively little impact on the empirical size properties of the two-sided rolling tests. For $c \geq 0$ there is relatively little difference between the two-sided 2SLS test and the two-sided IVX-based tests, regardless of whether a RWB or FRWB implementation is used with the latter, with all of the tests displaying good finite sample size control. For $c < 0$ (locally explosive predictors) the two-sided FRWB IVX-based test is rather conservative while the 2SLS test is slightly over-sized. Turning to the one-sided rolling tests, both the lower- and upper-tail RWB based IVX tests display decent finite sample size control (the empirical sizes of the nominal 5% level lower-tailed and upper-tailed statistics across all the values of c considered lie in the range [0.026, 0.063] and [0.029, 0.064], respectively). In contrast, the rolling upper-tail FRWB IVX-based test displays a tendency to over-sizing, most notably when $c \geq 0$ and $\phi = -0.95$ (rejection frequencies at a nominal 5% level when $T = 250$ are between [0.065, 0.117] and between [0.090, 0.129] for $T = 1000$). This over-sizing moderates for smaller ϕ; however, Table D.43 shows that for c close to zero (regardless of the sample size) significant over-sizing is still observed in the FRWB IVX-based test. The lower-tailed FRWB IVX-based rolling test also displays notable under-sizing when the predictor is strongly persistent, which is severe for locally explosive processes. Notice also that these patterns of size-distortion in the one-sided FRWB IVX-based tests closely mirror the full sample FRWB (and asymptotic) IVX-based tests under DGP1 reported in Section 5.1.1.

Turning to the recursive tests, the results in Tables D.45–D.48 show similar patterns to those observed for the rolling tests. Specifically, for $c \geq 0$ there is little to choose between the two-sided 2SLS and IVX-based tests, with all of these tests displaying decent size control throughout. Differences again surface between the one-sided IVX-based tests: the RWB-based implementations of the tests displaying good size control across c, while the lower- and upper-tailed forward and backward FRWB IVX-based tests display size distortions. To illustrate the latter, the upper-tailed forward and backward recursive FRWB IVX-based tests tend to be over-sized, particularly when for a strongly persistent predictor with a high endogeneity correlation (e.g., for $\phi = -0.95$, the forward recursive test displays size between [0.071, 0.088] for $0 \leq c \leq 25$ and the backward recursive test between [0.074, 0.091] for $0 \leq c \leq 10$). The lower-tailed FRWB IVX-based recursive tests are again correspondingly undersized. As with the full sample IVX tests, our simulation results strongly support using RWB implementations of the subsample IVX tests rather than the FRWB.

5.3.2. Finite sample local power

Figure D.1 (rolling tests) and Figures D.2 and D.3 (forward and backward recursive tests) graph finite sample local power functions of the subsample tests for data generated from DGP (1)–(3) under local alternatives satisfying $H_{1,b(\cdot)}$ of (6). We report results for $\phi = -0.95$ and $\phi = -0.5$, for the homoskedastic case, $\sigma_{ut}^2 = \sigma_{vt}^2 = 1$, for $T = 250$ and for $c = \{0, 10, 25, 50\}$. The slope parameter in (1) is parameterised as $\beta_t = b_t/T$, with $b_t = b$ with $b \in \{-40, -39, \ldots, -1, 1, \ldots, 39, 40\}$ for $t = 1, \ldots, \lfloor T/3 \rfloor$ and $b_t = 0$ for $t = \lfloor T/3 \rfloor + 1, \ldots, T$, such that a window of predictability occurs in the first third of the sample. We would therefore expect the rolling and forward recursive tests to display more power against this DGP than the backward recursive tests.

Consider the results for the rolling tests in Figure D.1. For testing against right-tailed alternatives (or equally, against left-tailed alternatives when ϕ is positive) there is little difference in general between those tests which are size-controlled. Exceptions occur for the case where $c = 0$ and $\phi = -0.95$ where the (FRWB) two-sided 2SLS-based test is slightly more powerful than the other tests, and where $c = 50$ and $\phi = -0.50$ where the one-sided RWB IVX-based test is slightly more powerful than the two-sided 2SLS and IVX tests. The pattern is very different when testing against left-tailed alternatives (right-tailed alternatives when ϕ is positive) where clear power gains over the two-sided tests are achieved by the one-sided IVX tests, except for $c = 0$ where the one-sided IVX test has almost identical power to the two-sided 2SLS-based test. The power gains for the left-sided IVX tests over the two-sided tests are larger for $\phi = -0.95$ *vis-à-vis* $\phi = -0.50$.

Turning to the forward and backward recursive tests in Figures D.2–D.3, as expected the forward recursive tests are the more powerful against this DGP and so we will focus our discussion on those tests. In contrast to the rolling tests discussed above, the gains to using the one-tailed IVX tests are most obvious when testing against right-sided alternatives, and again these power gains are larger for $\phi = -0.95$ *vis-à-vis* $\phi = -0.50$. The one-sided RWB IVX-based test clearly dominates the other tests against right-sided alternatives (noting that the FRWB IVX-based test is not size-controlled). For left-sided alternatives the one-sided IVX tests are significantly more powerful than the two-sided IVX tests but roughly as powerful as the two-sided 2SLS tests; albeit for $c = 0$ and $\phi = -0.95$ the 2SLS test is somewhat more powerful than the one-sided IVX test, although it should be recalled from Table D.47 that the 2SLS test is a little over-sized in this case.

6. Empirical applications

6.1. Testing for predictability in the equity premium

We first re-evaluate the predictability of the equity premium based on the predictors considered in the empirical case studies of Welch and Goyal (2008) and Campbell and Yogo (2006), using both one-sided and two-sided bootstrap and asymptotic tests. Specifically, we will first test for predictability in the log excess stock return, which is computed as the log of the monthly return on the S&P 500 index (including dividends) minus the log of the risk-free rate as our dependent variable, based on the predictors in Welch and Goyal, 2008), namely: the log dividend–price ratio (dp); the log dividend yield (dy); the log earnings–price ratio (ep); the log dividend–payout ratio (de); the stock return variance (svar); the book-to-market ratio (bm); the net equity expansion (ntis); the treasury bill rate (tbl); the long-term yield (lty); the long-term return (ltr); the term spread (tms); the default yield spread (dfy); the default return spread (dfr); and inflation (infl); see Welch and Goyal (2008) for full details on how these predictors are generated. We then subsequently revisit the empirical analysis of Campbell and Yogo (2006) who also test for predictability in excess returns using as putative predictors: the dividend–price ratio (dp); the earnings–price ratio (ep); the three-month T-bill rate (r_3), and a measure of the long-short yield spread ($y - r_1$); see Campbell and Yogo (2006) for full data descriptions.[10]

Table 4 presents bootstrap and asymptotic p-values for both one-sided and two-sided IVX predictability tests and the two-sided 2SLS predictability tests of Breitung and Demetrescu (2015) (again using their recommended fractional type-I and sine type-II instruments), in each case computed from predictive regressions with a single predictor based on Welch and Goyal's monthly data (Panel A) and on Campbell and Yogo's data (Panel B). The results in Panel A are obtained from a sample of monthly data from January 1927 to December 2020 ($T = 1128$) and in Panel B for samples of annual data from 1926 to 2002 ($T = 77$), quarterly data from the 4th quarter of 1926 to the 4th quarter of 2002 ($T = 305$) and monthly data from December 1926 to December 2002 ($T = 913$). The asymptotic IVX and 2SLS tests are computed with Eicker–White standard errors to allow for heteroskedasticity in the innovations. The bootstrap 2SLS test is implemented with the fixed regressor wild bootstrap while the bootstrap IVX tests are implemented using the residual wild bootstrap, in each case using conventional standard errors. For each predictor, Table 4 also reports both OLS and IVX estimates of the predictive regression slope parameter, β (denoted $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IVX}$, respectively), together with OLS estimates of the dominant AR root ($\hat{\rho}$) for each predictor and estimates ($\hat{\phi}$) of the endogeneity correlation.[11]

Consider first Panel A. The estimated endogenity correlation is relatively small (between -0.297 and 0.185) for all of the predictors, except dp, ep and bm for which $\hat{\phi}$ is large and negative: -0.980, -0.767 and -0.821, respectively. The

[10] The Welch–Goyal data (updated data up to 2020) were obtained from https://sites.google.com/view/agoyal145/ and the Campbell–Yogo data from https://sites.google.com/site/motohiroyogo/research/asset-pricing.

[11] Here, $\hat{\rho}$ is computed from an error correction parameterisation of an $AR(p)$ model fitted to the predictor, in which p is chosen applying the BIC over $p \in \{0, \ldots, \lfloor 4(T/100)^{0.25} \rfloor\}$, while $\hat{\phi}$ is the OLS estimate of the correlation of the predictive regression residuals and the residuals from the fitted $AR(p)$ model.

Table 4
Empirical results for the Welch and Goyal (2008) and Campbell and Yogo (2006) predictive regressions.

	$t_{zx}^{*,2SLS}$	$t_{zx}^{2SLS,EW}$	$t_{zx}^{*,RWB(-)}$	$t_{zx}^{*,RWB(+)}$	$t_{zx}^{*,RWB}$	$t_{zx}^{EW(-)}$	$t_{zx}^{EW(+)}$	t_{zx}^{EW}	$\hat{\beta}_{OLS}$	$\hat{\beta}_{IVX}$	$\hat{\rho}$	$\hat{\phi}$
PANEL A: Welch and Goyal (2008) **monthly data: January 1927–December 2020**												
dp	0.5125	0.4916	0.5366	0.4634	0.6124	0.7450	0.2550	0.5099	0.002	0.003	0.994	−0.980
dy	0.3395	0.2578	0.8355	0.1645	0.2868	0.9254	**0.0746**	0.1492	0.000	0.000	1.006	−0.060
ep	0.9116	0.9107	0.8271	0.1729	0.2719	0.9142	**0.0858**	0.1716	0.005	0.006	0.989	−0.767
de	0.2721	0.2684	0.2236	0.7764	0.4647	0.3013	0.6987	0.6025	−0.004	−0.004	0.991	−0.073
svar	0.5812	0.5821	0.3511	0.6489	0.7107	0.3856	0.6144	0.7712	−0.159	−0.189	0.577	−0.297
bm	0.5819	0.5805	0.6720	0.3280	0.4530	0.7181	0.2819	0.5939	0.008	0.007	0.987	−0.821
ntis	0.5083	0.5061	**0.0784**	0.9216	0.1609	0.1170	0.8830	0.2339	−0.142	−0.141	0.981	−0.047
tbl	0.3212	0.3135	***0.0340***	0.9660	**0.0874**	***0.0365***	0.9635	**0.0730**	−0.001	−0.001	0.994	−0.053
lty	0.1145	0.1073	***0.0278***	0.9722	**0.0703**	***0.0272***	0.9728	**0.0543**	−0.001	−0.001	0.997	−0.088
ltr	0.4045	0.3992	0.9176	**0.0824**	0.1707	0.9180	**0.0820**	0.1639	0.001	0.001	0.043	0.055
tms	0.8831	0.8855	0.7883	0.2117	0.4144	0.7434	0.2566	0.5132	0.002	0.001	0.962	−0.002
dfy	0.7160	0.7121	0.4588	0.5412	0.9939	0.5019	0.4981	0.9962	0.000	0.000	0.975	−0.265
dfr	0.2367	0.2316	0.8166	0.1834	0.3740	0.8029	0.1971	0.3943	0.002	0.002	−0.102	0.185
infl	0.5870	0.5887	**0.0959**	0.9041	0.1874	**0.0567**	0.9433	0.1113	−0.004	−0.005	0.480	0.033
PANEL B: Campbell and Yogo (2006) **data: 1926–2002**												
Annual data												
dp	0.2610	0.1853	0.9807	***0.0193***	***0.0205***	0.9983	***0.0017***	***0.0035***	0.158	0.166	0.932	−0.721
ep	0.3683	0.3175	0.9780	***0.0220***	***0.0255***	0.9967	***0.0033***	***0.0065***	0.162	0.161	0.855	−0.957
r_3	0.6242	0.6035	0.1178	0.8822	0.2247	**0.0977**	0.9023	0.1954	−0.934	−0.914	0.908	0.091
$y\text{-}r_1$	0.5576	0.5490	0.7721	0.2279	0.4361	0.7824	0.2176	0.4351	1.743	1.570	0.626	−0.248
Quarterly data												
dp	0.6747	0.6470	0.8967	0.1033	0.1296	0.9506	***0.0494***	**0.0987**	0.034	0.035	0.963	−0.942
ep	0.5878	0.5579	0.9558	***0.0442***	**0.0544**	0.9544	***0.0456***	**0.0913**	0.047	0.047	0.958	−0.986
r_3	0.6767	0.6566	0.1267	0.8733	0.2742	0.1206	0.8794	0.2412	−0.233	−0.228	0.965	−0.050
$y\text{-}r_1$	0.6590	0.6517	0.8179	0.1821	0.3378	0.7550	0.2450	0.4900	0.556	0.502	0.800	−0.119
Monthly data												
dp	0.6534	0.6398	0.7974	0.2026	0.2675	0.9092	0.0908	0.1817	0.008	0.008	0.990	−0.954
ep	0.8686	0.8548	0.9467	**0.0533**	**0.0656**	0.9607	***0.0393***	**0.0787**	0.013	0.013	0.989	−0.987
r_3	0.4100	0.4074	**0.0834**	0.9166	0.1775	**0.0833**	0.9167	0.1666	−0.086	−0.086	0.991	−0.058
$y\text{-}r_1$	0.1723	0.1744	0.9348	**0.0652**	0.1304	0.8641	0.1359	0.2719	0.239	0.222	0.938	−0.065

Note: The columns headed $t_{zx}^{*,2SLS}$ and t_{zx}^{2SLS} provide p-values for the bootstrap based 2SLS tests and for the corresponding 2SLS test based on asymptotic critical values, respectively. The columns $t_{zx}^{*,RWB(-)}$, $t_{zx}^{*,RWB(+)}$ and $t_{zx}^{*,RWB}$ correspond to the p-values of the left-sided, right-sided and two-sided residual wild bootstrap based IVX t-tests, respectively. The columns labelled $t_{zx}^{(-)}$, $t_{zx}^{(+)}$ and t_{zx} provide the p-values of the left-sided, right-sided and two-sided IVX tests computed using asymptotic critical values. $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IVX}$ are the OLS and IVX estimates of the predictive regression slope parameter β, $\hat{\rho}$ is the estimate of the largest root from an AR model fitted to the predictor, and $\hat{\phi}$ is the OLS estimate of the correlation of the predictive regression residuals and the residuals from an AR model fitted to the predictor. All bootstrap p-values were computed using 9999 bootstrap replications. Test outcomes significant the 10% (5%) level are highlighted in bold (bold italic).

dominant estimated AR root is also close to unity for most of the predictors: $\hat{\rho} \in [0.962, 1.006]$, with the exception of svar, ltr, dfr and infl for which $\hat{\rho}$ is 0.577, 0.043, −0.102 and 0.480, respectively. Turning to the outcomes of the predictability tests, for both dy and ep we see that the right-sided asymptotic IVX test, $t_{zx}^{EW(+)}$, yields significant evidence of positive predictability at the 10% level, while in both cases the two-sided t_{zx}^{EW} test fails to reject at the 10% level. For both of these series the right-sided RWB $t_{zx}^{*,RWB(+)}$ fails to reject at the 10% level suggesting that these are most likely spurious rejections attributable to the finite-sample oversize of the $t_{zx}^{EW(+)}$ test seen in the simulations in Table 1 for strongly persistent predictors. Among the other putative predictors, the left-sided RWB $t_{zx}^{*,RWB(-)}$ test find evidence of (negative) predictability at the 5% level for tbl and lty and at the 10% level for ntis, in each case statistically stronger evidence than is provided by the corresponding two-sided $t_{zx}^{*,RWB}$ test. Similar conclusions are drawn for the tbl and lty predictors using the asymptotic IVX tests, while no rejections are seen with either the two-sided or one-sided asymptotic IVX tests in the case of ntis. For ltr (infl) both the right-sided RWB $t_{zx}^{*,RWB(+)}$ (left-sided RWB $t_{zx}^{*,RWB(-)}$) test and the right-sided asymptotic $t_{zx}^{EW(+)}$ (left-sided asymptotic $t_{zx}^{EW(-)}$) test find evidence of positive (negative) predictability at the 10% level, not found in the corresponding two-sided tests. To summarise the results in Panel A, we find rather stronger evidence of predictability when using one-sided tests, with 5 (2) of the predictors being found to be statistically significant at the 10% (5%) level using one-sided RWB bootstrap tests, compared to 2 (0) when using the two-sided RWB bootstrap tests. Finally, consistent with the local power results in Fig. 1, the 2SLS tests are insignificant at the 10% level for all of the predictors, regardless of whether asymptotic or bootstrap p-values are used.

Consider next Panel B. As with the predictors in Panel A, the Campbell and Yogo (2006) predictors again appear to be strongly persistent in general. Moreover, although $\hat{\phi}$ is relatively small for both r_3 and $y-r_1$ (in line with the corresponding Welch and Goyal predictors), for dp and ep very strong negative endogeneity correlations are estimated: for monthly data $\hat{\phi}$ for dp and ep is −0.954 and −0.987, respectively (reducing to −0.721 and −0.957, respectively for annual data). With annual data, the null hypothesis of no predictability is rejected for both dp and ep at the 5% level, regardless of whether

one uses right-sided or two-sided tests and for both the asymptotic and RWB implementations of the IVX test. As the data frequency increases the strength of these rejections tends to decline. For both quarterly and monthly data in the case of ep, both the RWB and asymptotic tests now only reject at the 10% level for the two-sided tests and at around the 5% level for right-sided tests. For dp, both the right-sided and two-sided RWB IVX-based tests fail to reject at the 10% level; some rejections are still seen for both quarterly and monthly data with the right-sided asymptotic IVX test, albeit we should treat the results from the latter with a degree of caution given the high persistence and large negative endogeneity correlation observed for dp (cf Table 1). For the monthly data the left-sided RWB and asymptotic IVX tests both indicate rejection of the null hypothesis of no predictability for r_3 at the 10% level, while the right-sided RWB test also indicates rejection of the null for $\mathsf{y} - \mathsf{r}_1$ at the 10% level; in none of these cases does the corresponding two-sided version of the test signal predictability at the 10% level. Finally, as with the results in Panel A, the 2SLS tests are insignificant at the 10% level for all of the predictors considered, regardless of whether asymptotic or bootstrap p-values are used.

6.2. Testing for bubbles in foreign exchange rates

We next re-visit the problem of testing for speculative bubbles in the U.K. pound to U.S. dollar foreign exchange market considered in Pavlidis et al. (2017), using monthly data (downloaded from the Bank of England, www.bankofengland.co.uk) on spot and forward rates for the period from January 1999 to July 2021 ($T = 271$), the start date coinciding with the introduction of the Euro.

Fama (1984) proposes the following regression as a basis for testing for efficiency in foreign exchange markets,

$$s_{t+h} - f_{t,h} = \alpha_h + \beta_h(f_{t,h} - s_t) + u_{t+h}, \tag{26}$$

where $s_{t+h} - f_{t,h}$ is typically referred to as the excess return (or forecast error) (see, for example, Maynard, 2006) and $f_{t,h} - s_t$ is the forward premium, where s_t is (the log of) the spot exchange rate at time t and $f_{t,h}$ is (the log of) the forward rate at time t for maturity at time $t + h$, $h \geq 1$.

In the context of (26), as discussed in Pavlidis et al. (2017), the efficient market hypothesis corresponds to $\beta_h = 0$, while if an exchange rate bubble is present in any time period then $\beta_h > 0$. Pavlidis et al. (2017) therefore apply right-tailed rolling subsample implementations of the IVX tests of Kostakis et al. (2015) to test the null hypothesis $\beta_h = 0$ against the alternative hypothesis that β_h in (26) is positive in at least one subsample of the data. In the context of this testing problem it is important to use only right-tailed tests because, as is well known in this literature, the estimate of β_h can suffer from a severe negative finite sample bias when $\beta_h = 0$; the so-called *forward bias puzzle*. A number of explanations have been posited for this phenomenon including the negative correlation between the risk premium and the forward premium; see Maynard (2003). Consequently, a two-sided test might be inappropriate as a rejection could be due to either a downward bias effecting large negative statistics in some subsamples, or a genuine bubble episode.

Like Pavlidis et al. (2017) we report results for the three periods to maturity available in the dataset, namely, one, three, and six months: $h = \{1, 3, 6\}$. Where $h > 1$ we follow Pavlidis et al. (2017, Appendix A.2, pp. 1221–1223) and estimate the parameters of (26) using a reverse regression (Phillips and Lee, 2013) type approach. Fitting an autoregressive model (with a constant), we found the forward premium, $(f_{t,h} - s_t)$, to be a strongly persistent time series regardless of the maturity period; in particular the dominant autoregressive root was estimated to be $\hat{\rho} = 0.9635$ for $h = 1$, $\hat{\rho} = 0.9821$ for $h = 3$, and $\hat{\rho} = 0.9880$ for $h = 6$. The estimated correlation parameter, $\hat{\phi}$, was found to be relatively small and negative for $h = 1$, but increases in absolute value as h increases: $\hat{\phi} = -0.0861$ for $h = 1$, $\hat{\phi} = -0.3100$ for $h = 3$, and $\hat{\phi} = -0.3686$ for $h = 6$. The estimates of ρ and ϕ were calculated as outlined in footnote 11.

Table 5 reports bootstrap (both RWB and FRWB) p-values for the maximum rolling, and forward and backward recursive subsample IVX statistics from Section 3.2, in each case implemented as upper-tailed (right-sided) tests, using $B = 9999$ bootstrap replications. Results are reported for four values of the tuning parameters $\Delta\tau$ (the *window fraction* used for the sequence of rolling statistics) and τ_L and $(1 - \tau_U)$ (the *warm-in* parameters for the forward and backward recursive sequences, respectively), namely 1/6, 1/4, 1/3, and 1/2.

The results in Table 5 show that for $h = 1$ none of the subsample IVX tests provide evidence, at any conventional significance level, of exuberant behaviour in the foreign exchange rate. However, for the longer maturities considered, $h = 3$ and $h = 6$, statistically significant results are found in the case of the rolling tests, suggesting the presence of a potential bubble. Specifically, for both $\Delta\tau = 1/6$ and $1/2$ the rolling tests signal the presence of exuberant behaviour in the foreign exchange rate, although they do not for $\Delta\tau = \{1/4, 1/3\}$. The strongest rejections are observed for $h = 6$. In this case, the RWB based rolling tests find evidence of exuberant behaviour for all four window widths. The FRWB based rolling tests also reject the null that $\beta_6 = 0$ for all window widths, except $\Delta\tau = 1/4$. The forward and backward recursive tests do not reject the null that $\beta_h = 0$ for any of the warm-in parameters and maturities considered, suggesting that the start and end points of the bubble episode are likely bounded away from the end points of the full sample.

Pavlidis et al. (2017) find no evidence of speculative bubbles for any of the maturity periods considered for data covering the period January 1979 to December 2013, and so, although we consider a different sample period, it would seem instructive to investigate where in the sample the rejections that we find above occur. To that end, Fig. 2 plots the sequence of rolling IVX subsample statistics computed for a window fraction $\Delta\tau = 1/2$ and maturity period $h = 6$. Plotted on the graph are the 5% and 10% RWB critical values for the maximum of the rolling test statistics together with the upper-tail 5% and 10% pointwise RWB critical values. This plot indicates that the rejection of the null hypothesis

Table 5
Testing for bubbles in exchange rates between 1999 and 2021. P-values for the rolling, and forward and backward recursive statistics.

$\Delta\tau,\ \tau_L,\ (1-\tau_U)$	$\mathcal{T}_U^{R,RWB}$	$\mathcal{T}_U^{R,FRWB}$	$\mathcal{T}_U^{F,RWB}$	$\mathcal{T}_U^{F,FRWB}$	$\mathcal{T}_U^{B,RWB}$	$\mathcal{T}_U^{B,FRWB}$
$\boldsymbol{h=1}$						
1/6	0.1770	0.4929	0.8851	0.5917	0.8817	0.6292
1/4	0.2321	0.6094	0.8085	0.4214	0.8124	0.6035
1/3	0.2913	0.6451	0.7530	0.4059	0.7837	0.5769
1/2	0.1547	0.3251	0.4454	0.3517	0.6795	0.5592
$\boldsymbol{h=3}$						
1/6	***0.0498***	**0.0847**	0.8152	0.8719	0.3858	0.3733
1/4	0.2443	0.3973	0.7681	0.7932	0.3512	0.3551
1/3	0.1126	0.2198	0.7364	0.7328	0.3275	0.3348
1/2	**0.0588**	***0.0347***	0.5129	0.6676	0.2866	0.3230
$\boldsymbol{h=6}$						
1/6	***0.0169***	***0.0078***	0.8915	0.8628	0.2125	0.1745
1/4	**0.0950**	0.1390	0.8484	0.7848	0.1799	0.1497
1/3	***0.0347***	**0.0587**	0.8129	0.7575	0.1752	0.1427
1/2	***0.0121***	***0.0010***	0.6205	0.6768	0.1631	0.1419

Notes: The columns headed $\mathcal{T}_U^{R,k}$, $k = RWB, FRWB$, provide p-values for the residual (RWB) and fixed regressor ($FRWB$) wild bootstrap based rolling (R) upper tail tests. The columns $\mathcal{T}_U^{F,k}$, and $\mathcal{T}_U^{B,k}$, $k = RWB, FRWB$, correspond to forward (F) and backward (B) recursive residual (RWB) and fixed regressor ($FRWB$) wild bootstrap based upper tail tests, respectively. All bootstrap p-values were computed using 9999 bootstrap replications. Test outcomes significant at the 10% (5%) level are highlighted in bold (bold italic).

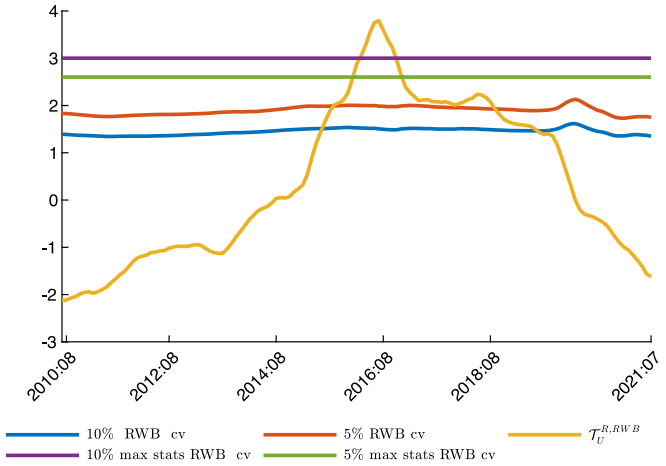


Fig. 2. Plot of the sequence of upper-tailed rolling statistics for testing the null hypothesis of no bubble in the U.K. pound–U.S. dollar foreign exchange market for a six month maturity ($h = 6$).

$H_0 : \beta_6 = 0$ by the maximum of the rolling tests occurs between January 2016 and November 2016 (after the end of the sample period considered in Pavlidis et al., 2017) when a 10% significance level is considered and between February 2016 and September 2016 for a 5% significance level. The sequence of rolling statistics displays a steady and sustained increase in magnitude from mid-2013 onwards, with the statistics exceeding the pointwise 10% (5%) significance level between June 2015 and May 2019 (September 2015 and August 2018). These findings are, on the face of it, consistent with a bubble episode in the U.K. pound–U.S. dollar exchange rate which collapsed at or around the time of the Brexit vote in summer 2016.

7. Conclusions

We have extended the IVX-based predictability tests of Kostakis et al. (2015) in three directions. First, we have shown that, provided either a suitable bootstrap implementation is employed or Eicker–White standard errors are used, these tests still deliver asymptotically valid inference, regardless of the degree of persistence or endogeneity of the predictor, under considerably weaker assumptions on the innovations, including quite general forms of conditional and unconditional heteroskedasticity, than required by Kostakis et al. (2015) in their analysis. Second, we have developed asymptotically valid residual and fixed regressor wild bootstrap implementations of the IVX tests. Simulation evidence has been provided which demonstrates that tests based around a residual wild bootstrap resampling scheme perform

particularly well in finite samples. Third, we have shown how sub-sample implementations of the IVX approach can be used to develop asymptotically valid one-sided and two-sided tests for temporary windows of predictability.

We finish with three suggestions for further research. First, we have focused on the case of a single predictive regressor. As we have noted, the methods we have discussed readily extend to the case of multiple regressors, provided, as assumed in Kostakis et al. (2015), they all belong to the same persistence class. However, based on the results in this paper, we conjecture that our bootstrap IVX tests should also retain asymptotic validity in the scenario where some of the regressors are weakly persistent and others strongly persistent, and where the strongly persistent regressors could cointegrate. The practitioner would not need to know which of the regressors were weakly persistent and which were strongly persistent, or the form of any cointegrating relations present. A formal proof of this conjecture is likely very involved but constitutes an important next step in this research agenda, with technical material in this paper providing important groundwork.

Second, unlike Kostakis et al. (2015), we have not discussed the case of mildly integrated regressors. While we hold it to be plausible that our results may be extended to cover the mildly integrated case, the corresponding derivations may be quite lengthy and we leave them for further work. Among other things, one would have to consider several distinct cases depending on whether the IVX filter depends on a coefficient which is closer to or further away from unity than the largest autoregressive root of the predictor, together with the interplay of the true regressor's persistence with the mixed fourth moments of the innovations series, which, as shown by Proposition 3, is quite different under weak and strong persistence.

Third, the finite sample efficacy of the residual wild bootstrap IVX tests proposed in this paper will depend, in part, on the finite sample properties of the autoregressive parameter estimates obtained in Step 2 of Algorithm 1. The OLS estimates we have employed are known to suffer from non-negligible finite sample biases. It might be useful to explore a refinement of Algorithm 1 based on the bootstrap-after-bootstrap approach of Kilian (1998) to investigate if this can further improve on the finite sample properties of our proposed bootstrap tests.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2022.02.007.

References

Amihud, Y., Hurvich, C.M., 2004. Predictive regressions: A reduced-bias estimation method. J. Financ. Quant. Anal. 39, 813–841.

Andersen, T.G., Varneskov, R.T., 2021a. Consistent inference for predictive regressions in persistent economic systems. J. Econometrics 224, 215–244.

Andersen, T.G., Varneskov, R.T., 2021b. Testing for parameter instability and structural change in persistent predictive regressions. J. Econometrics forthcoming.

Andersen, T.G., Varneskov, R.T., 2021c. Consistent local spectrum (LCM) inference for predictive return regressions, NBER Working Paper 28569, downloadable from https://www.nber.org/papers/w28569.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. J. Finance 59, 1481–1509.

Bauer, D., Maynard, A., 2012. Persistence-robust surplus-lag granger causality testing. J. Econometrics 169, 293–300.

Breitung, J., Demetrescu, M., 2015. Instrumental variable and variable addition based inference in predictive regressions. J. Econometrics 187, 358–375.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Rev. Financ. Stud. 21, 1509–1531.

Campbell, J.Y., Yogo, M., 2006. Efficient tests of stock return predictability. J. Financ. Econ. 81, 27–60.

Carnero, M.A., Peña, D., Ruiz, E., 2004. Persistence and kurtosis in GARCH and stochastic volatility models. J. Financ. Econom. 2, 319–342.

Cavanagh, C.L., Elliott, G., Stock, J.H., 1995. Inference in models with nearly integrated regressors. Econom. Theory 11, 1131–1147.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press, Oxford.

Demetrescu, M., Georgiev, I., Rodrigues, P.M.M., Taylor, A.M.R., 2022. Testing for episodic predictability in stock returns. J. Econometrics 227, 85–113.

Demetrescu, M., Hillmann, B., 2022. Nonlinear predictability of stock returns? Parametric vs. nonparametric inference in predictive regressions. J. Bus. Econom. Statist. 40, 382–397.

Demetrescu, M., Hosseinkouchack, M., 2021. Finite-sample size control of IVX-based tests in predictive regressions. Econom. Theory 37, 769–793.

Elliott, G., Müller, U.K., Watson, M.W., 2015. Nearly optimal tests when a nuisance parameter is present under the null hypothesis. Econometrica 83, 771–811.

Fama, E.F., 1984. Forward and spot exchange rates. J. Monetary Econ. 14 (3), 319–338.

Fan, R., Lee, J.H., 2019. Predictive quantile regressions under persistence and conditional heteroskedasticity. J. Econometrics 213, 261–280.

Georgiev, I., Harvey, D.I., Leybourne, S.J., Taylor, A.M.R., 2018. Testing for parameter instability in predictive regression models. J. Econometrics 204, 101–118.

Georgiev, I., Harvey, D.I., Leybourne, S.J., Taylor, A.M.R., 2019. A bootstrap stationarity test for predictive regression invalidity. J. Bus. Econom. Statist. 37, 528–541.

Gordon, M., 1962. The Investment, Financing, and Valuation of the Corporation. Homewood, IL, Irwin.

Hansen, B.E., 2000. Sample splitting and threshold estimation. Econometrica 68, 575–603.

Harvey, D.I., Leybourne, S.J., Taylor, A.M.R., 2021. Simple tests for stock return predictability with good size and power properties. J. Econometrics 224, 198–214.

Jansson, M., Moreira, M.J., 2006. Optimal inference in regression models with nearly integrated regressors. Econometrica 74, 681–714.

Johannes, M., Korteweg, A., Polson, N., 2014. Sequential learning, predictability, and optimal portfolio returns. J. Finance 69, 611–644.

Kilian, L., 1998. Small-sample confidence intervals for impulse response functions. Rev. Econ. Stat. 80, 218–230.

Kostakis, A., Magdalinos, T., Stamatogiannis, M.P., 2015. Robust econometric inference for stock return predictability. Rev. Financ. Stud. 28, 1506–1553.

Lee, J.H., 2016. Predictive quantile regression with persistent covariates: IVX-QR approach. J. Econometrics 192, 105–118.

Magdalinos, T., 2020. Least squares and IVX limit theory in systems of predictive regressions with GARCH innovations. Econom. Theory forthcoming.

Maynard, A., 2003. Testing for forward-rate unbiasedness: On regression in levels and in returns. Rev. Econ. Stat. 85 (2), 313–327.

Maynard, A., 2006. The forward premium anomaly: Statistical artefact or economic puzzle? New evidence from robust tests. Can. J. Econ. / Revue Canadienne D'Economique 39 (4), 1244–1281.

Maynard, A., Phillips, P.C.B., 2001. Rethinking an old empirical puzzle: econometric evidence on the forward discount anomaly. J. Appl. Econometrics 16, 671–708.
Maynard, A., Shimotsu, K., 2009. Covariance-based orthogonality tests for regressors with unknown persistence. Econom. Theory 25, 63–116.
Merlevède, F., Peligrad, M., Utev, S., 2006. Recent advances in invariance principles for stationary sequences. Probab. Surv. 3, 1–36.
Nelson, C.R., Kim, M.J., 1993. Predictable stock returns: The role of small sample bias. J. Finance 48, 641–661.
Pastor, L., Stambaugh, R.F., 2009. Predictive systems: Living with imperfect predictors. J. Finance 64, 1583–1628.
Pavlidis, E.G., Paya, I., Peel, D.A., 2017. Testing for speculative bubbles using spot and forward prices. Internat. Econom. Rev. 58, 1191–1226.
Pettenuzzo, D., Timmermann, A., Valkanov, R., 2014. Forecasting stock returns under economic con- straints. J. Financ. Econ. 114, 517–553.
Phillips, P.C.B., Lee, J.H., 2013. Predictive regression under various degrees of persistence and robust long-horizon regression. J. Econometrics 177, 250–264.
Phillips, P.C.B., Magdalinos, T., 2008. Limit theory for explosively cointegrated systems. Econom. Theory 24, 865–887.
Phillips, P.C.B., Magdalinos, T., 2009. Econometric inference in the vicinity of unity. CoFie Working Paper 7, Singapore Management University.
Phillips, P.C.B., Shi, S.-P., Yu, J., 2015. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the SP500. Internat. Econom. Rev. 56, 1043–1078.
Smeekes, S., Westerlund, J., 2019. Robust block bootstrap panel predictability tests. Econometric Rev. 38, 1089–1107.
Stambaugh, R.F., 1999. Predictive regressions. J. Financ. Econ. 54, 375–421.
Van der Weide, R., 2002. Go-GARCH: A multivariate generalized orthogonal GARCH model. J. Appl. Econometrics 17 (5), 549–564.
Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21, 1455–1508.
Xu, K.-L., Guo, J., 2021. A new test for multiple predictive regression. Working paper, downloadable from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3991366.