Residual-augmented IVX predictive regression

Matei Demetrescu*
Christian-Albrechts-University of Kiel

Paulo M. M. Rodrigues†
Banco de Portugal and Universidade Nova de Lisboa

September 30, 2016

Abstract

Bias correction in predictive regressions stabilizes the empirical size properties of OLS-based predictability tests. This paper shows that bias correction also improves the finite sample power of tests, in particular so in the context of the extended instrumental variable (IVX) predictability testing framework introduced by Kostakis et al. (Review of Financial Studies 2015). Concretely, we introduce new IVX-based statistics subject to a bias correction analogous to that proposed by Amihud and Hurvich (Journal of Financial and Quantitative Analysis 2004). Three important contributions are provided: first, we characterize the effects that bias-reduction adjustments have on the asymptotic distributions of the IVX test statistics in a general context allowing for short-run dynamics and heterogeneity; second, we discuss the validity of the procedure when predictors are stationary as well as near-integrated; and third, we conduct an exhaustive Monte Carlo analysis to investigate the small-sample properties of the test procedure and its sensitivity to distinctive features that characterize predictive regressions in practice, such as strong persistence, endogeneity, non-Gaussian innovations and heterogeneity. An application of the new procedure to the Welch and Goyal (Review of Financial Studies 2008) database illustrates its usefulness in practice.

Keywords: Predictability, persistence, persistence change, bias reduction.
JEL classification: C12 (Hypothesis Testing), C22 (Time-Series Models)

*Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany, email: mdeme@stat-econ.uni-kiel.de.

†Banco de Portugal, NOVA School of Business and Economics, Universidade Nova de Lisboa, Portugal, e-mail: pmrodrigues@bportugal.pt.

1 Introduction

Predictive regressions are widely used in economics and finance; see, *e.g.*, Campbell (2008) and Phillips (2015) for surveys. Typically, the variable of interest is regressed on lagged values of a predictor and the existence of predictability assessed through the statistical significance of the resultant estimate of the corresponding slope parameter. However, two important features of predictors need to be taken into consideration in this analysis: i) many predictors are often characterized by highly persistent autoregressive dynamics, and ii) many predictors also exhibit innovations which are strongly correlated to the innovations of the dependent variable. These features raise serious problems of endogeneity which can lead to sizeably biased estimates in finite samples (Stambaugh, 1986 and Mankiw and Shapiro, 1986) and to substantial over-rejections of the null hypothesis of no predictability. The usual asymptotic approximation employing the (standard) normal distribution performs particularly bad when predictors are persistent, even though the largest autoregressive roots of the typical predictor candidate are usually smaller than one – reason for which near-integrated asymptotics has been favoured as an alternative framework for inference (Elliott and Stock, 1994 and Campbell and Yogo, 2006). In the context of near-integrated regressors, the limiting distribution of the slope parameter estimator is not centered at zero, and this bias depends on the mean reversion parameter of the near-integrated regressor. Although near-integrated asymptotics approximates the finite-sample behavior of the t-statistic for no predictability considerably better when predictors are persistent, the exact degree of persistence of a given predictor, and thus the correct critical values for a predictability test, are not known in advance. Moreover, standard estimation or pretests also fail in this context (Cavanagh et al., 1995). Similarly, regression misspecification tests are difficult to conduct; Georgiev et al. (2015) propose for this reason a fixed-regressor wild bootstrap implementation of a residual stationarity test.

These difficulties have led to the proposal of a number of alternative approaches, which differ mainly in the assumptions that characterize the stochastic properties of predictors (*i.e.*, whether these are stationary or near-integrated); see for instance, Campbell and Yogo (2006); Jansson and Moreira (2006); Maynard and Shimotsu (2009); Camponovo (2015); Breitung and Demetrescu (2015) and references therein. The recently proposed extended instrumental variable estimation approach [denoted IVX] motivated by Magdalinos and Phillips (2009) is becoming increasingly popular in predictive regressions, especially because the relevant t-statistic exhibits the same limiting distribution in both, stationary and near-integrated setups and is in this sense invariant to persistence; see, *e.g.*, Kostakis et al. (2015); Gonzalo and Pitarakis (2012); Lee (2016) and Phillips and Lee (2013). The reasoning behind the approach consists in the generation of an instrumental variable whose persistence can be controlled, and this is achieved by suitably filtering the actual predictor.

To some extent, all methods lose some power by having to robustify against unknown persistence; however, as illustrated by Kostakis et al. (2015) the IVX methodology offers a good balance between size control and power loss. Since the noise-to-signal ratio in predictive regressions is quite high, one should still strive to improve this balance. For instance, Demetrescu (2014b) uses a simple variable addition scheme to improve the convergence rates of IVX estimators (and thus the local power of the corresponding t-tests) when the instrument used is relatively close to stationarity. However, for instrument choices closer to near-integration a different approach is required to improve the finite sample power of IVX-based tests without giving up size control.

To this end, we take a closer look at the class of reduced-bias techniques proposed by Amihud and Hurvich (2004) and extended by Amihud et al. (2009, 2010); see, *inter alia*, Bali (2008), Chun (2009), Avramov et al. (2010) and Johannes et al. (2014) for recent empirical applications building on this approach. When compared to other available procedures, the distinctive characteristic of these techniques is that they estimate the predictive slope coefficient and its standard error in a suitably *augmented* predictive regression, so that the bias is reduced to a minimum. While this bias correction was intended to stabilize the size properties of OLS-based predictability tests, we argue that it may also contribute to improve power, in particular so for IVX-based testing.

This paper discusses the large-sample behavior of IVX-statistics subject to bias correction, *i.e.*, the implementation of IVX in an augmented predictive regression context analogous to that of Amihud and Hurvich (2004), considering both stationary and near integrated predictors. Our main objectives are threefold: i) to characterize the effects that our bias-reduction adjustments have on the asymptotic distribution of the IVX-statistics in a general context; ii) to establish the validity of the procedure when predictors are stationary as well as near-integrated; and iii) to provide an exhaustive Monte Carlo analysis to investigate the small-sample properties of the test procedures under distinctive conditions that characterize predictive regressions in practice, such as strong persistence, endogeneity, non-Gaussian innovations and heterogeneity, and to contrast them to the properties of available procedures, such as Amihud and Hurvich (2004), Campbell and Yogo (2006) and the IVX approach proposed by Kostakis et al. (2015). Finally, we revisit the data set used in Welch and Goyal (2008) to illustrate the application of the procedure.

The remainder of the paper is organized as follows. Section 2 briefly describes the characteristic features of predictive regressions and the bias-reduction technique proposed by Amihud and Hurvich (2004), and gives a brief preview of the advantages of the residual-augmented IVX. Section 3 presents the large-sample theory under empirically relevant assumptions, including for instance heterogeneity and time-varying unconditional variances. Section 4 discusses the finite sample performance of several procedures used to test for predictability. Section 5 presents the analysis of the Welch and Goyal data, and section 6 summarizes and concludes. A technical appendix collects the proofs of the main theoretical statements put forward in the paper.

2 Predictive regression framework and tests

2.1 The simplest model

To illustrate the issues with predictive regressions in general and the advantages of our approach in particular, we start by considering the single predictor theoretical model set up analyzed in Stambaugh (1999) and adopted, among many others, by Amihud and Hurvich (2004) and Campbell and Yogo (2006). This setting characterizes the joint dynamics of a stochastic process, $\{y_t\}_{t=2}^{T}$, and its posited predictor, $\{x_t\}_{t=1}^{T-1}$, in a two-equation linear system as,

$$y_t = \beta x_{t-1} + u_t, \quad t = 2, ..., T \tag{1}$$

$$x_t = \rho x_{t-1} + v_t \tag{2}$$

where the innovations $\boldsymbol{\xi}_t := (u_t, v_t)'$ in the two-equation system are typically serially independent Gaussian distributed with mean zero and covariance matrix Σ.

In this setting, predictability is formally analyzed by examining whether the null hypothesis, $\mathrm{H}_0 : \beta = 0$, is statistically rejected through a t-statistic on the OLS estimate $\hat{\beta}$ computed from (1). The usual alternative hypothesis is that $\beta > 0$, focusing on one-sided tests, but two-sided tests $\beta \neq 0$, are also frequently used in the literature. We shall refer to the resultant least-squares statistic as $t_{\hat{\beta}}$ in the sequel. It is a well-documented fact that when the correlation, $\frac{\sigma_{uv}}{\sigma_u \sigma_v}$, between innovations is large and $\rho \simeq 1$, the distribution of $t_{\hat{\beta}}$ largely departs from the typical standard normal limit, posing therefore an interesting challenge on inference; see, *e.g.*, Elliott and Stock (1994) and Stambaugh (1999).

Specifically, under these simple assumptions, weak convergence of the partial sum of $\boldsymbol{\xi}_t$ holds, *i.e.*, $\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} (u_t, v_t)' \Rightarrow (\sigma_u W_u(s), \sigma_v W_v(s))'$, where $(W_u(s), W_v(s))'$ is a vector of dependent standard Wiener processes (see, *e.g.*, Davidson, 1994, Chapter 29). Furthermore, considering that the autoregressive coefficient ρ is local to unity, $\rho := 1 - c/T$, we have, jointly with the above weak convergence, that $\frac{1}{\sqrt{T}} x_{[sT]} \Rightarrow B_c(s)$, where B_c is an Ornstein-Uhlenbeck [OU] process driven by $W_v(s)$, *i.e.*, $B_c(s) := W_v(s) - c \int_0^s e^{-c(s-r)} W_v(r) \mathrm{d}r$. Given these results it follows that the limiting distribution of the OLS based t-test, $t_{\hat{\beta}}$, computed from (1) when the predictor is near-integrated is given by

$$t_{\hat{\beta}} \Rightarrow \sqrt{1 - \frac{\sigma_{uv}^2}{\sigma_u^2 \sigma_v^2}} \mathcal{Z} + \frac{\sigma_{uv}}{\sigma_u \sigma_v} \frac{\int_0^1 B_c(s) \mathrm{d}W_v(s)}{\sqrt{\int_0^1 B_c^2(\mathrm{s}) \mathrm{d}s}}$$

where $\mathcal{Z}$ is a standard normal variate independent of the Wiener process $W_v(r)$ driving $B_c(r)$.

Remark 2.1 *The assumptions of normality and serial independence allow for considerable simplification of the exposition, but shall be relaxed in the following section by allowing for more general forms of serial dependence or heterogeneity.* □

2.2 Residual Augmented Predictive Regressions

Considering (1) - (2) and stationarity of $\{x_t\}$, *i.e.*, the additional assumption that ρ in (2) is fixed and satisfies $|\rho| < 1$, Stambaugh (1986, 1999) shows that the exact OLS bias of $\hat{\beta}$ in (1) is $\gamma \, \mathrm{E}\,(\hat{\rho} - \rho)$, with $\hat{\rho}$ denoting the OLS estimate of ρ and $\gamma := \sigma_{uv}/\sigma_v^2$ is the slope coefficient in a regression of u_t on v_t. Since $\hat{\rho}$ is known to be downward biased in small-samples, and $(u_t, v_t)'$ are typically highly negatively contemporaneously correlated, the autoregressive OLS bias feeds into the small-sample distribution of $\hat{\beta}$ causing over-rejections of the null hypothesis of no predictability, $\mathrm{H}_0 : \beta = 0$.

To correct for this effect, Amihud and Hurvich (2004) propose a simple statistical device that builds upon the OLS estimates obtained from a predictive regression which is augmented with estimates of v_t, the innovations to the predictor in (2). The initial motivation for this type of augmentation is that the null distribution of the t-statistic on $\hat{\beta}$ in the *infeasible* regression

$$y_t = \beta x_{t-1} + \gamma v_t + \varepsilon_t \tag{3}$$

converges asymptotically to a standard normal distribution irrespectively of the stochastic nature of x_t and the degree of contemporaneous correlation of $(u_t, v_t)'$. Although it is tempting to use some proxy of v_t to make this regression feasible, it should be noted that the appealing asymptotic properties of the infeasible test do not automatically extend to the feasible counterpart resulting from the use of the OLS residuals from (2), say $\hat{v}_t$. The reason is that the bias of $\hat{\rho}$ still feeds into the estimation of β via $\hat{v}_t = v_t - (\hat{\rho} - \rho)\, x_{t-1}$ and, as a result, the distribution of the OLS t-statistic for $\beta = 0$ in this regression, is simply a re-scaling of that of $t_{\hat{\beta}}$; see Rodrigues and Rubia (2011); Cai and Wang (2014) and Demetrescu (2014a), for further details.

The distinctive feature of the Amihud and Hurvich (2004) [AH] procedure is that it uses a *bias-adjusted* estimate of v_t to reduce the bias of $\hat{\beta}$. Thus, the resulting feasible regression becomes,

$$y_t = \beta x_{t-1} + \gamma \hat{v}_t^* + \varepsilon_t, \tag{4}$$

where $\hat{v}_t^* := x_t - \hat{\rho}^* x_{t-1}$, with $\hat{\rho}^*$ denoting finite-sample bias-corrected OLS estimates of ρ in (2). The central idea is to obtain a $\hat{\rho}^*$ as close to unbiasedness as possible. The procedure however also requires a correction in the form of specific standard errors which is not easily generalized to higher-order dynamics; see Amihud et al. (2009, 2010).

Remark 2.2 *Augmenting linear regression models with covariates is often motivated in terms of efficiency gains (Faust and Wright, 2011). Arguably, the primary purpose of the residual-augmented regression in* (4) *is to stabilize size, with power gains playing a secondary role. This is partly because the true process of the errors is unobservable and must be replaced by some empirical proxy (which prompts the correction for ensuring size control of the AH procedure). We argue in the following that power gains can indeed be expected in the IVX framework, while* *at the same time controlling for size.* □

2.3 The IVX Test Procedures

2.3.1 The Original IVX Approach

Our interest lies in the evaluation of the impact that the bias correction through augmentation may have on the IVX approach. The IVX procedure, introduced to predictive regressions by Kostakis et al. (2015), centers on the construction of instrumental variables from the potential predictors. This ensures relevance of the instruments while at the same time controlling for persistence. In particular, for the implementation of the procedure, one uses

$$z_t := \sum_{j=0}^{t-2} \varrho^j \Delta x_{t-j} = (1 - \varrho L)_+^{-1} \Delta x_t$$

as instrument for x_t, with L standing for the conventional lag operator; the idea is to choose $\varrho := 1 - a/T^\eta$, with $0 < \eta \leq 1$, and $a \geq 0$ and fixed, such that z_t is by construction only *mildly* integrated when the predictor x_t is (nearly) integrated.

The resulting IVX estimator of β (henceforth $\hat{\beta}^{ivx}$), computed from (1) using z_t as instrument has a slower convergence rate than the conventional OLS estimator, but is mixed Gaussian in the limit irrespective of the degree of endogeneity implied by γ. This estimator is given by,

$$\hat{\beta}^{ivx} := \frac{\sum_{t=2}^{T} z_{t-1} y_t}{\sum_{t=2}^{T} z_{t-1} x_{t-1}} \tag{5}$$

and its standard error is $se\left(\hat{\beta}^{ivx}\right) := \frac{\hat{\sigma}_u \sqrt{\sum_{t=2}^{T} z_{t-1}^2}}{\sum_{t=2}^{T} z_{t-1} x_{t-1}}$; note that Kostakis et al. (2015) suggest the use of OLS residuals $\hat{u}_t$ (whose consistency properties do not depend on the persistence properties of the instrument z_t) for the computation of $\hat{\sigma}_u^2$.

Breitung and Demetrescu (2015) analyse the power function of the IVX-based t-test, computed as $t_{ivx} := \hat{\beta}^{ivx}/se\left(\hat{\beta}^{ivx}\right)$, under local alternatives of the form $\beta := b/T^{1/2+\eta/2}$, and show that the limiting distribution under such local alternatives is

$$t_{ivx} \Rightarrow \mathcal{Z} + b\,\frac{\sigma_v \sqrt{2}}{\sigma_u \sqrt{a}} \left[B_c^2\left(1\right) - \int_0^1 B_c\left(s\right) \mathrm{d}B_c\left(s\right) \right] \tag{6}$$

where $\mathcal{Z}$ is a standard normal variate independent of the OU process $B_c(r)$, a is the noncentrality parameter used in ϱ for the construction of the instrument, and σ_v and σ_u are the standard deviations of v_t and u_t, respectively. Note that the reduced convergence rate of $\hat{\beta}^{ivx}$ has consequences on the type of neighbourhoods where the IVX based test has nontrivial power. This, however, is the trade off for obtaining a pivotal limiting null distribution. While Kostakis et al. (2015) show that the power loss is moderate, one would of course prefer to reduce this loss as much as possible.

2.3.2 The Bias-reduced IVX Approach

Turning our attention to the bias correction approach proposed by Amihud and Hurvich (2004), note that, the residuals $\hat{v}_t^*$ used in the residual-augmented predictive regression in (4) rely on a bias-corrected estimate of ρ in order to reduce the endogeneity of the predictor. Interestingly, since IVX uses a *less persistent* instrument for estimation than the original predictor, it turns out that in order to use the residual augmentation approach in the IVX framework it is not necessary to construct a bias corrected estimator, such as $\hat{\rho}^*$ used by Amihud and Hurvich (2004). This is an important advantage of the IVX procedure since it simplifies the analysis considerably and allows for easy generalisations to higher order dynamics in the predictor as we will show below.

Remark 2.3 *It may be surprising that, although simple augmentation using OLS residuals does not work for the OLS estimation of the predictive regression, it will work for IVX. Essentially, the estimation noise* $(\hat{v}_t - v_t)$ *does not affect the IVX estimator given the lower convergence rate of the latter compared to the OLS estimator. In fact, the improved local power is the same as if the true* v_t *were used in* (4)*: the local power of the test based on the augmented IVX regression is obtained by replacing* σ_u *with* σ_ε *in* (6)*; see the next section for more details. Since* $\sigma_\varepsilon < \sigma_u$ *whenever* $\gamma \neq 0$*, we obtain by construction a larger drift term in the distribution under the local alternative* $\beta := b/T^{1/2+\eta/2}$*. This may not increase the convergence rate, but considering the typically high correlation of the innovations* u_t *and* v_t *(given by* ${}^{\sigma_{uv}}/_{\sigma_u \sigma_v}$*), the ratio* $({}^{\sigma_u}/_{\sigma_\varepsilon})$ *can be considerably larger than unity and power gains in finite samples are to be expected. This is confirmed in the Monte Carlo analysis in Section 4.* □

The implementation of our bias-reduced IVX approach in the simple introductory setup given by (1) and (2), is as follows:

1. Regress x_t on x_{t-1} to obtain the residuals $\hat{v}_t := v_t - (\hat{\rho} - \rho)\, x_{t-1}$, where $\hat{\rho} := \rho + \frac{\sum_{t=2}^T x_{t-1} v_t}{\sum_{t=2}^T x_{t-1}^2}$ is the usual OLS estimator.

2. Regress y_t on $\hat{v}_t$ to obtain $\tilde{y}_t := y_t - \hat{\gamma}\hat{v}_t = \varepsilon_t + \beta x_{t-1} + \gamma v_t - \hat{\gamma}\hat{v}_t$, where $\hat{\gamma} := \frac{\sum_{t=2}^T \hat{v}_t y_t}{\sum_{t=2}^T \hat{v}_t^2}$ is the usual OLS estimator.

3. Regress $\tilde{y}_t$ on x_{t-1} via IVX to obtain $\tilde{\beta}^{ivx}$ and the corresponding t-statistic, $\tilde{t}_{ivx}$; similarly to the original IVX, it helps if the residuals are computed using the OLS estimator, $\hat{\beta}$, of this regression given its consistency and higher convergence rates.

Remark 2.4 *Considering* $\tilde{y}_t$ *as the dependent variable provides a convenient way to think about residual augmented predictive regressions. As discussed in Campbell and Yogo (2006), the unobservable process* $[y_t - E(u_t|v_t)]$ *results from subtracting off the part of the innovation to the predictor variable that is correlated with* y_t*. This provides a less noisy dependent variable in the regression analysis and, therefore, yields power advantages over conventional predictive* *regressions that steam from a relative gain in statistical efficiency. In particular, since* $\mathrm{E}\left(\varepsilon_t^2\right) = \left(1-\rho^2\right)\sigma_u^2$*, the larger the degree of endogenous correlation in the system, the larger the amount of variability in the regressand not related to* x_{t-1} *that can be filtered out – conversely, we can think of the standard predictive regression analysis as a particularly inefficient tool to detect predictability when* ρ *is large. However, since* $[y_t - \mathrm{E}(u_t|v_t)]$ *cannot be directly observed, the feasible representation uses the OLS-based proxy* $\tilde{y}_t$ *in the equation.* □

Remark 2.5 *In practice, one may need to account for non-zero means of* y_t*; this is accomplished by including an intercept in the regression in step 2 and by demeaning the regressor* x_t *in the IVX regression in step 3 (see Kostakis et al., 2015 for the justification of this demeaning procedure in step 3). In the near-integrated case, including an intercept in the autoregression in the first step is typically not needed for the kind of data one has in mind with stock return predictability, where deterministic trends are in general not an empirical issue.* □

Thus, following the three steps above we obtain the bias-corrected IVX estimator, viz.,

$$\tilde{\beta}^{ivx} := \frac{\sum_{t=2}^{T} z_{t-1}\tilde{y}_t}{\sum_{t=2}^{T} z_{t-1}x_{t-1}} = \hat{\beta}^{ivx} - \frac{\hat{\gamma}\sum_{t=2}^{T} z_{t-1}\widehat{v}_t}{\sum_{t=2}^{T} z_{t-1}x_{t-1}} \tag{7}$$

and its corresponding standard error,

$$se\left(\tilde{\beta}^{ivx}\right) := q_T\,\frac{\hat{\sigma}_\varepsilon\sqrt{\sum_{t=2}^{T} z_{t-1}^2}}{\left|\sum_{t=2}^{T} z_{t-1}x_{t-1}\right|} \tag{8}$$

where $\tilde{y}_t := y_t - \widehat{\gamma}\widehat{v}_t$, $\hat{\sigma}_\varepsilon$ is the estimate of the standard deviation of ε_t computed from the residuals $\tilde{\varepsilon}_t := \tilde{y}_t - \hat{\beta}x_{t-1}$ and $\hat{\beta} := \frac{\sum_{t=2}^{T} x_{t-1}\tilde{y}_t}{\sum_{t=2}^{T} x_{t-1}^2}$. Note that the estimator of the standard error in (8) includes the finite sample correction,

$$q_T := 1 + \frac{\left(\hat{\gamma}\hat{\sigma}_v\sum_{t=2}^{T} z_{t-1}x_{t-1}\right)^2}{\hat{\sigma}_\varepsilon^2\sum_{t=2}^{T} z_{t-1}^2\sum_{t=2}^{T} x_{t-1}^2}. \tag{9}$$

A detailed discussion of the importance of q_T will be presented in the following section, but it may be noted that (9) is in principle only required when the predictors used are stationary; see section 3 for details.

Hence, considering (7) and (8) inference can be performed based on the IVX t-statistic,

$$\tilde{t}_{ivx} := \tilde{\beta}^{ivx}/se\left(\tilde{\beta}^{ivx}\right) \tag{10}$$

which turns out to remain standard normal irrespectively of the stationarity or near-integratedness of the regressor.

2.4 Short-run dynamics and heterogeneity

This section looks into the properties of the residual-augmented IVX approach in the empirical relevant cases where predictors may display short-run dynamics and heterogeneity. Hence, in this section we lay out a fairly general setting, which is the framework we will use to characterise the asymptotic properties of the procedures introduced in this paper.

The starting question is how to deal with short-run dynamics in the increments of x_t, since this has implications as to which residuals to use for augmentation in the IVX testing procedure. Here, it is the innovations of v_t (for which a finite-order AR process is a natural choice) that should correlate with u_t rather than v_t itself, like in the case without short-run dynamics. The augmentation approach (described in Section 2.2) relies on decomposing the shocks to the predictive regression as the sum of two orthogonal components; should v_t be one of them, this induces serial correlation in u_t, which is not a plausible feature of the null hypothesis of no predictability. Hence, the general set up considered is formalized in the following assumptions.

Assumption 1 *The data is generated according to* (1) - (2) *with initial condition* x_1 *bounded in probability.*

Assumption 2 *Let*

$$\begin{pmatrix} \varepsilon_t \\ \nu_t \end{pmatrix} := \begin{pmatrix} \sigma_{\varepsilon t}\xi_{\varepsilon t} \\ \sigma_{\nu t}\xi_{\nu t} \end{pmatrix}$$

where $(\xi_{\varepsilon t}, \xi_{\nu t})'$ *is a heterogeneous independent sequence with unity covariance matrix and, for some* $\delta > 0$, *with uniformly bounded moments* $\mathrm{E}\left(\left|\xi_{\varepsilon t}^{4+\delta}\right|\right)$ *and* $\mathrm{E}\left(\left|\xi_{\nu t}^{4+\delta}\right|\right)$. *Furthermore, let* $\sigma_{\varepsilon t} := \sigma_\varepsilon\left({}^{t}\!/_{T}\right)$ *and* $\sigma_{\nu t} := \sigma_\nu\left({}^{t}\!/_{T}\right)$, *where* $\sigma_{\cdot}(\cdot)$ *are piecewise Lipschitz continuous functions, bounded away from zero.*

Assumption 3 *The errors* u_t *and* v_t *are given as*

$$\begin{aligned} v_t &= a_1 v_{t-1} + \ldots + a_{p-1} v_{t-p+1} + \nu_t \\ u_t &= \varepsilon_t + \gamma\nu_t, \qquad\qquad t \in \mathbb{Z}, \end{aligned}$$

where the innovations $(\varepsilon_t, \nu_t)'$ *are contemporaneously orthogonal white noise as indicated in Assumption 2.*

Assumption 4 *The autoregressive parameter* ρ *is either i) fixed when* $|\rho| < 1$, *or ii) time-varying near unity,* $\rho := 1 - {}^{c_t}\!/_{T}$ *with* $c_t := c\left({}^{t}\!/_{T}\right)$ *and* $c(\cdot)$ *is a piecewise Lipschitz function.*

Assumption 2 acknowledges that time series (and in particular financial series) may exhibit permanent volatility changes, which is an important stylized fact of many financial series; see, among others, Guidolin and Timmermann (2006); Teräsvirta and Zhao (2011); Amado and Teräsvirta (2013) and Amado and Teräsvirta (2014). Such forms of nonstationarity typically invalidate the usual standard errors,[1] and we resort to heteroskedasticity robust [HC] standard errors (also known as Eicker-White standard errors) to account for this feature. The use of White standard errors is also recommended by Kostakis et al. (2015) to deal with conditional heteroskedasticity – albeit under strict stationarity of the error series v_t. The AR($p-1$) structure of v_t in Assumption 3 is taken as an approximation to more general data generating processes [DGP]s. In theory, this would require letting $p \to \infty$ at suitable rates as $T \to \infty$; however, dealing with the asymptotics related to the order of augmentation determination is beyond the scope of this paper, but relevant results can be found, for instance, in Chang and Park (2002). Finally, Assumption 4 characterises the persistence properties of the predictor. The flexible near-integrated DGP resulting from Assumption 4 ii) is motivated by the high, yet uncertain persistence of typical predictor series. Moreover, since persistence is not always constant, in particular when close to the unit root region, we allow for time variation in persistence in the near integrated case.

Hence, the implementation of our residual-augmented IVX approach in the general framework described by Assumptions 1 through 4 consists of the following steps:

1. Compute the residuals $\hat{\nu}_t$ from an autoregressive model of order p for the predictor x_t, *viz.*,

$$\hat{\nu}_t = x_t - \sum_{j=1}^{p} \hat{\phi}_j x_{t-j} = \nu_t - \sum_{j=1}^{p} \left(\hat{\phi}_j - \phi_j\right) x_{t-j}, \quad t = p+1, \ldots, T,$$

   with $\hat{\phi}_j$, $j = 1, ..., p$, the OLS autoregressive coefficient estimates. One may use some information criteria in levels to determine the autoregressive order p (we use Akaike's information criteria (AIC) in sections 4 and 5); note that conducting model selection in levels copes with both the stationary and the integrated cases.

2. Regress y_t on $\hat{\nu}_t$ to obtain $\tilde{y}_t$ as regression residuals. From this regression step we also obtain $\hat{\gamma}$, the OLS estimate of γ.

3. Finally, regress $\tilde{y}_t$ on x_{t-1} via IVX and use the provided standard errors (see Equation (12) below) to compute the relevant IVX t-statistic.

From step 3) we thus obtain,

$$\tilde{\beta}^{ivx} := \frac{\sum_{t=p+1}^{T} z_{t-1}\tilde{y}_t}{\sum_{t=p+1}^{T} z_{t-1}x_{t-1}}, \tag{11}$$

which, upon standardization, is used for inference.

Note that under Assumptions 1 to 4, the standard errors need to take into account two specific features of the data. First, time varying variances are likely to bias the usual standard errors asymptotically. Second, while the estimation error $(\hat{v}_t - v_t)$ has no asymptotic effect on

[1] This is especially the case when dealing with (near-) integrated regressors; see, *e.g.*, Cavaliere (2004) and Cavaliere et al. (2010).

the limiting distribution of $\tilde{\beta}^{ivx}$ in the near-integrated context, it does so when x_t is covariance stationary. Yet treating the two cases in a different manner is inconvenient since exact knowledge about which is actually the relevant case is typically not available. Consequently, we derive heteroskedasticity-consistent standard errors for the stationary case and show that these are also valid in the near integrated context. In this way, we use *the same* statistic with the same limiting distribution to cover both cases without having to decide which is which – just like in the original IVX test of Kostakis et al. (2015).

In specific, we use

$$se\left(\tilde{\beta}^{ivx}\right) := \sqrt{\frac{\sum_{t=p+1}^{T} z_{t-1}^2 \tilde{\varepsilon}_t^2 + \hat{\gamma}^2 \hat{Q}_T}{\left(\sum_{t=p+1}^{T} z_{t-1} x_{t-1}\right)^2}} \tag{12}$$

where the finite-sample correction $\hat{Q}_T$ used in (12) is given by

$$\hat{Q}_T = \left(\sum_{t=p+1}^{T} z_{t-1}\boldsymbol{x}_{t-p}'\right)\left(\sum_{t=p+1}^{T} \boldsymbol{x}_{t-p}\boldsymbol{x}_{t-p}'\right)^{-1}\left(\sum_{t=p+1}^{T} \boldsymbol{x}_{t-p}\boldsymbol{x}_{t-p}'\hat{\nu}_t^2\right)\left(\sum_{t=p+1}^{T} \boldsymbol{x}_{t-p}\boldsymbol{x}_{t-p}'\right)^{-1}\left(\sum_{t=p+1}^{T} z_{t-1}\boldsymbol{x}_{t-p}\right)$$

and $\boldsymbol{x}_{t-p} := (x_{t-1}, \ldots, x_{t-p})'$. To compute the White-type standard errors in (12) we make use of the OLS residuals computed from the residual-augmented predictive regression, $\tilde{\varepsilon}_t := \tilde{y}_t - \tilde{\beta}^{ols} x_{t-1}$ where $\tilde{\beta}^{ols} := \frac{\sum_{t=2}^{T} x_{t-1}\tilde{y}_t}{\sum_{t=2}^{T} x_{t-1}^2}$, rather than IVX residuals due to the superconsistency properties of the former in the near-integrated context.

Remark 2.6 *One may resort to alternative HC variance estimators, e.g., with correction for degrees of freedom (HC1). The HC1 version is obtained here by multiplying the estimated variance by* $\frac{T}{T-p-3}$. □

Remark 2.7 *The standard errors in* (12) *are basically the White standard errors that would have been appropriate under stationarity of* x_t*, where the estimation error of* $\hat{\nu}_t$ *does not vanish asymptotically. We show that* $\hat{Q}_T$ *in* (12) *is dominated under near-integration so that the standard error in* (12) *is asymptotically equivalent to the one implied by the near-integrated framework, which turns out to be simply* $\sqrt{\frac{\sum_{t=p+1}^{T} z_{t-1}^2 \tilde{\varepsilon}_t^2}{\left(\sum_{t=p+1}^{T} z_{t-1} x_{t-1}\right)^2}}$ *as can be seen in Section 3.* □

Remark 2.8 *The near-unit root in* x_t *allows us in principle to use the residuals without the need to use the finite sample correction, but in finite samples the statistics fare better if the correction is included (essentially because, in finite samples, any* $|\rho| < 1$ *is "caught between" stationarity and integration).* □

2.5 Extensions to Multiple Predictors

The discussion so far has side-stepped a couple of aspects relevant for empirical work which we address in this section. They are in fact straightforward extensions of the baseline case and we shall omit some of the technical details.

It is often the case that several predictors are simultaneously considered. Thus, the resulting multiple predictive regression is

$$y_t = \boldsymbol{\beta}' \boldsymbol{x}_{t-1} + u_t$$

where $\boldsymbol{x}_{t-1}$ follows a K-dimensional vector autoregressive data generating process of order p, such as,

$$\begin{aligned} \boldsymbol{x}_t &= R\boldsymbol{x}_{t-1} + \boldsymbol{v}_t \\ \boldsymbol{v}_t &= \sum_{j=1}^{p-1} A_j \boldsymbol{v}_{t-j} + \boldsymbol{\nu}_t \end{aligned}$$

which is either stable or (near) integrated as before depending on the properties of the autoregressive coefficient matrix R ($\boldsymbol{v}_t$ is taken to be a stable autoregression in either case). There is endogeneity, possibly in all regressors, expressed as a nonzero coefficient vector in the decomposition

$$u_t := \boldsymbol{\gamma}' \boldsymbol{\nu}_t + \varepsilon_t,$$

and the shocks $\boldsymbol{\nu}_t$ and ε_t are heterogeneous, serially independent obeying a multivariate version of Assumption 3.

The implementation of the IVX approach introduced in this paper in the multiple predictive regression case is as follows.

1. Get the vector of residuals $\hat{\boldsymbol{\nu}}_t$ from a *vector* autoregression of order p,

   $$\hat{\boldsymbol{\nu}}_t := \boldsymbol{x}_t - \sum_{j=1}^{p} \hat{\Phi}_j \boldsymbol{x}_{t-j}, \quad t = p+1, \ldots, T,$$

   with $\hat{\Phi}_j$, $j = 1, ..., p$, the matrix of OLS coefficient estimates. Note that the use of AIC (or some other information criteria) in levels, for determining the order p, is again recommended.

2. Regress y_t on $\hat{\boldsymbol{\nu}}_t$ to obtain the adjusted $\tilde{y}_t$ as,

   $$\tilde{y}_t = y_t - \hat{\boldsymbol{\gamma}}' \hat{\boldsymbol{\nu}}_t$$

   with $\hat{\boldsymbol{\gamma}}$ the OLS estimate of the vector of parameters $\boldsymbol{\gamma}$.

3. Regress $\tilde{y}_t$ on $\boldsymbol{x}_{t-1}$ via IVX with $\boldsymbol{z}_{t-1} := (1 - \varrho L)_+^{-1} \Delta \boldsymbol{x}_{t-1}$ as instruments to obtain $\tilde{\boldsymbol{\beta}}^{ivx}$ and use the standard errors provided in Equation (13) below to conduct inference.

The estimated covariance matrix of $\tilde{\boldsymbol{\beta}}^{ivx}$ in this context is given by the familiar "sandwich" formula,

$$\widehat{\mathrm{Cov}\left(\tilde{\boldsymbol{\beta}}^{ivx}\right)} = B_T^{-1} M_T \left(B_T^{-1}\right)' \tag{13}$$

where

$$B_T = \sum_{t=2}^{T} \boldsymbol{z}_{t-1}\boldsymbol{x}'_{t-1}$$

and

$$\begin{aligned} M_T = \; & \textstyle\sum_{t=2}^{T} \boldsymbol{z}_{t-1}\boldsymbol{z}'_{t-1}\tilde{\varepsilon}_t^2 + \left(\boldsymbol{\gamma}' \otimes \left(\frac{1}{T}\sum_{t=2}^{T} \boldsymbol{z}_{t-1}\boldsymbol{x}'_{t-p,K}\right)\left(\sum_{t=p+1}^{T} \boldsymbol{x}_{t-p,K}\boldsymbol{x}'_{t-p,K}\right)^{-1}\right) \times \\ & \times \left(\textstyle\sum_{t=p+1}^{T} \boldsymbol{\nu}_t\boldsymbol{\nu}'_t \otimes \boldsymbol{x}_{t-p,K}\boldsymbol{x}'_{t-p,K}\right)\left(\boldsymbol{\gamma} \otimes \left(\sum_{t=p+1}^{T} \boldsymbol{x}_{t-p,K}\boldsymbol{x}'_{t-p,K}\right)^{-1}\left(\frac{1}{T}\sum_{t=2}^{T} \boldsymbol{x}_{t-p,K}\boldsymbol{z}'_{t-1}\right)\right) \end{aligned}$$

with $\boldsymbol{x}_{t-p,K}$ corresponding to the vector stacking all p lags of all K regressors, i.e., $\boldsymbol{x}'_{t-p,K} := (x_{t-1,1}, \ldots, x_{t-1,K}, x_{t-2,1}, \ldots, x_{t-2,K}, \ldots, x_{t-p,1}, \ldots, x_{t-p,K})$.

The limiting distribution of $\tilde{\boldsymbol{\beta}}^{ivx}$ is normal in the stationary case and mixed normal in the near-integrated context; the proofs are simple multivariate extensions of the results from the single-regressor case (see the following section) so we do not spell them out. More importantly, individual and joint significance tests have their usual standard normal and χ^2 limiting distributions irrespective of the persistence and heterogeneity of the DGP as long as the robust covariance matrix estimator in (13) is used.

3 Asymptotic results

In this section, we analyze the limiting distributional characteristics of the new reduced-bias IVX tests considering the general framework described in Section 2.4, which also provides us with the results for the simplest case in Section 2.1 as a particular case. We consider two different theoretical frameworks that critically determine the stochastic properties of the predictive variable. On the one hand, we consider stationary predictors, characterized by a fixed coefficient $|\rho| < 1$ in (2), and on the other, we allow for near-integration by considering $\rho := 1 - c/T$, with $c \geq 0$ and fixed. The main objective of this setting is to acknowledge the uncertainty that researches face regarding the stochastic properties of the predictor, *i.e.*, whether it is stationary or near-integrated when $\hat{\rho}$ is close to, but strictly less than unity in finite samples. This setting includes of course the extreme case of a unit-root when the local parameter c equals zero ($c = 0$).

In the following, we maintain the predictive regression framework in (1) but allow for significant departures from Gaussianity and the restrictive AR(1) structure for the regressor. We also allow for heterogeneity in the form of time-varying variances, different shapes of the distributions, and even changes in the persistence of the regressor. Financial variables often exhibit time-varying variances in addition to GARCH effects; Kostakis et al. (2015) discuss the GARCH case considering strict stationarity, whereas we relax the i.i.d. assumption by replacing stationarity with smoothly varying volatility.

Note first that the time-varying properties of the DGP, as stated in Assumptions 1 through 4, imply different behavior in the limit compared to the Gaussian i.i.d. case. In this case, the partial sums of ν_t converge weakly to $M(s) := \int_0^s \sigma_\nu(r)\,\mathrm{d}W_v(r)$, and the partial sums of ε_t to $\int_0^s \sigma_\varepsilon(r)\,\mathrm{d}W_\varepsilon(r)$, with W_ε and W_v independent standard Wiener processes; the "classical" case is only recovered when σ_u and σ_v are constant. Moreover, the suitably normalized regressor can be shown to converge weakly to an Ornstein-Uhlenbeck type process driven by the diffusion $M(s)$, *i.e.*,

$$\frac{1}{\sqrt{T}} x_{[sT]} \Rightarrow \omega \int_0^s e^{-\int_r^s c(t)\mathrm{d}t} \mathrm{d}M(r) := \omega X(s) \tag{14}$$

where $\omega = \left(1 - \sum_{j=1}^{p-1} a_j\right)^{-1}$; see, *e.g.*, Cavaliere (2004) for the case with constant c.

In the case where x_t is stationary, *i.e.*, $|\rho| < 1$ and fixed, the following results can be stated.

Theorem 3.1 *Under Assumptions 1, 2, 3 and 4i), we have, as $T \to \infty$, that*

$$\sqrt{T}\left(\tilde{\beta}^{ivx} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_\beta^2\right) \tag{15}$$

where

$$\sigma_\beta^2 := \frac{\alpha_0 \int_0^1 \sigma_v^2(s)\,\sigma_\varepsilon^2(s)\,\mathrm{d}s + \gamma^2 \boldsymbol{\alpha}_p' \Omega^{-1} \boldsymbol{\alpha}_p \int_0^1 \sigma_v^4(s)\,\mathrm{d}s}{\left[\alpha_0 \int_0^1 \sigma_v^2(s)\,\mathrm{d}s\right]^2} \tag{16}$$

with $\boldsymbol{\alpha}_p := (\alpha_0 \ldots \alpha_{p-1})'$ and $\Omega := \left\{\alpha_{|i-j|}\right\}_{1 \le i,j \le p}$, where $\alpha_h := \sum b_j b_{j+h}$ with b_j the moving average coefficients of x_t, $(1-\rho L)^{-1}\left(1 - a_1 L - \ldots - a_{p-1}L^{p-1}\right) = \sum_{j \ge 0} b_j L^j$. Furthermore,

$$\sqrt{T} se\left(\tilde{\beta}^{ivx}\right) \xrightarrow{p} \sqrt{\sigma_\beta^2}$$

and, under the null hypothesis, $H_0 : \beta = 0$,

$$\tilde{t}_{ivx} \xrightarrow{d} \mathcal{N}(0,1). \tag{17}$$

The limit behavior changes under near-integration as shown in the following Theorem.

rewrite XdX correctly!

Theorem 3.2 *Under Assumptions 1, 2, 3 and 4ii), we have, as $T \to \infty$, that*

$$T^{1/2+\eta/2}\left(\tilde{\beta}^{ivx} - \beta\right) \Rightarrow \mathcal{MN}\left(0, \frac{a \int_0^1 \sigma_\nu^2(s)\,\sigma_\varepsilon^2(s)\,\mathrm{d}s}{2\omega^2\left(X^2(1) - \int_0^1 X(s)\,\mathrm{d}X(s)\right)^2}\right) \tag{18}$$

and

$$se\left(\tilde{\beta}^{ivx}\right) \Rightarrow \sqrt{\frac{a}{2\omega^2}} \frac{\sqrt{\int_0^1 \sigma_\nu^2(s)\,\sigma_\varepsilon^2(s)\,\mathrm{d}s}}{X^2(1) - \int_0^1 X(s)\,\mathrm{d}X(s)} \tag{19}$$

where a and η are fixed, ω^2 plays the role of the long-run variance (and is defined in (14)*),* $X(s) = \int_0^s e^{-\int_r^s c(t)\mathrm{d}t}\sigma_v(r)\,\mathrm{d}W_v(r)$ *and,* $\sigma_\nu^2(s)$ *and* $\sigma_\varepsilon^2(s)$ *are the variances of* v_t *and* ε_t*, respectively. Moreover, under the null hypothesis,* $H_0 : \beta = 0$,

$$\tilde{t}_{ivx} \Rightarrow \mathcal{N}(0,1). \tag{20}$$

The proof of Theorem 3.2 establishes that $Q_T = o_p(T^{1+\eta})$ so that it is dominated in (12) by $\sum_{t=p+1}^T z_{t-1}^2\tilde{\varepsilon}_t^2$ which is of exact order $O_p(T^{1+\eta})$ (see the Appendix for details), and the residuals estimation effect is negligible in the near-integrated case. The near-integrated case is also more interesting for an evaluation of the local power and for comparison with the original IVX.[2] The power function of the residual augmented IVX is provided next.

Theorem 3.3 *Under Assumptions 1, 2, 3 and 4ii), we have for local alternatives* $\beta = b/T^{1/2+\eta/2}$*, as* $T \to \infty$ *that*

$$\tilde{t}_{ivx} \Rightarrow \mathcal{N}\left(b\sqrt{\frac{2\omega^2}{a}}\frac{X^2(1) - \int_0^1 X(s)\,\mathrm{d}X(s)}{\sqrt{\int_0^1 \sigma_\nu^2(s)\,\sigma_\varepsilon^2(s)\,\mathrm{d}s}}\,,\ 1\right). \tag{21}$$

Setting $\omega^2 = 1$, $\sigma_v(s) = \sigma_v$ and $\sigma_\varepsilon(s) = \sigma_\varepsilon$ leads to the results for the particular case studied in Section 2.1.

4 Finite sample performance

4.1 Monte Carlo Setup

This section compares the two versions of the IVX procedure, the original IVX test which we denote as t_{ivx} and the residual augmented version $\tilde{t}_{ivx}$, with extant procedures under several heterogeneous DGPs. As benchmarks we use the tests of Campbell and Yogo (2006) and of Amihud and Hurvich (2004) and Amihud et al. (2010).

Concretely, we generate y_t and x_t as in equations (1) and (2) but allow for an intercept in the predictive regression, *i.e.*,

$$y_t = \mu + \beta x_{t-1} + u_t, \quad t = 2, ..., T \tag{22}$$

$$x_t = \rho x_{t-1} + v_t \tag{23}$$

and

$$v_t = a_1 v_{t-1} + e_t \tag{24}$$

[2]The local power in the stationary case is easily derived and we omit the details.

with $a_1 \in \{-0.5, 0, 0.5\}$ and $e_t \sim \mathcal{N}id(0,1)$. We focus on local alternatives of the form $\beta = {}^{b}\!/_{T}$ for two sample sizes, $T = 200$ and $T = 500$. To study the empirical size of the tests we let $b = 0$, and for the local power evaluation we consider $b \in \{5, 10, 15, 25\}$, and the persistence of the predictor is controlled by $\rho := 1 - {}^{c}\!/_{T}$, with $c \in \{0, 10, 20, 40, 50\}$. The correlation causing endogeneity is set to -0.95, which is not an uncommon value in practice; see, *e.g.*, Lewellen (2004).

The efficient tests of Campbell and Yogo (2006) (denoted as CY) are analysed, and the residual augmented predictive regression based test of Amihud et al. (2010) (denoted as AHW) is computed for a fixed $p = 2$ to keep complexity under control. In comparison, t_{ivx} does not require specifying the lag length, while for $\tilde{t}_{ivx}$ we chooses p via Akaike's information criteria (AIC). Both t_{ivx} and $\tilde{t}_{ivx}$ are computed by demeaning the dependent variable and the regressor, but not the instrument (see Section 2.5 for details). Since all tests are invariant to the intercept μ, we set $\mu = 0$ without loss of generality.

Also, we follow Kostakis et al. (2015) and choose $a = 1$ and $\eta = 0.95$ for the construction of the instruments in both. We employ the proposed standard errors from (12) in the computation of $\tilde{t}_{ivx}$, while, for the classical t_{ivx}, we use White standard errors as recommended by Kostakis et al. (2015). We shall also consider a version of the original IVX test without White standard errors, denoted by $t^{\#}_{ivx}$, to illustrate the impact of neglected time-varying volatility on the performance of this approach.

The rejection frequencies are computed at the nominal 5% level based on 10000 Monte Carlo replications, and all results for the t_{ivx} and $\tilde{t}_{ivx}$ tests in Tables 1 – 4 are computed based on standard normal critical values.

4.2 Empirical size and power performance

Tables 1 and 2 illustrate the empirical size and power properties of the AHW, CY, t_{ivx} and $\tilde{t}_{ivx}$ tests under negative and positive short-run dynamics, *i.e.*, considering (24) with $a_1 = -0.5$ and $a_1 = 0.5$.

From Table 1, which presents the results obtained when v_t follows an AR(1) with $a_1 = -0.5$ (negative autocorrelation) we observe that when $b = 0$ and for the values of c considered that AHW and t_{ivx} are slightly oversized, but that this oversizing decreases as the sample size increases. At the same time, we also observe that $\tilde{t}_{ivx}$ displays slightly conservative behaviour. In this experiment CY presents the largest size distortions as a consequence of the negative short-run dynamics. This feature of the CY test has already been noted in the literature; see, *e.g.,* Jansson and Moreira (2006). Note also that in the unit root case ($c = 0$) there are some significant size distortions also for the t_{ivx} and AHW tests. Regarding the empirical power we observe that the $\tilde{t}_{ivx}$ test displays superior power when $c > 0$, relative to the other procedures.

In the case of positive short-run dynamics, *i.e.*, when $a_1 = 0.5$ (see Table 2) we observe in general size distortions for all tests, with t_{ivx} displaying the most severe distortions when compared to the other procedures, and AHW and $\tilde{t}_{ivx}$ displaying the smallest distortions.

Table 1: Size and power against local alternatives, negative short-run AR parameter

	b	AHW	CY	t_{ivx}	$\tilde{t}_{ivx}$	AHW	CY	t_{ivx}	$\tilde{t}_{ivx}$
		$T = 200$				$T = 500$			
$c = 0$	0	8.9	1.1	10.6	6.30	9.4	2.5	10.4	6.3
	5	17.5	28.3	54.4	37.5	17.3	30.7	53.2	39.0
	10	67.8	94.7	93.5	86.1	65.9	97.4	93.0	87.9
	15	98.2	99.4	98.9	97.3	97.8	99.8	98.7	98.1
	25	100.0	99.95	100.0	99.9	100.0	100.0	100.0	99.9
		$T = 200$				$T = 500$			
$c = 10$	0	6.6	0.0	5.4	5.0	6.8	0.4	4.6	4.6
	5	8.1	0.2	13.8	14.5	7.2	2.8	12.4	14.4
	10	17.1	3.8	33.2	39.6	15.0	14.8	31.0	38.7
	15	37.0	29.2	65.1	78.1	33.2	49.6	61.3	77.4
	25	96.6	94.7	96.8	99.4	95.2	98.8	96.0	99.5
		$T = 200$				$T = 500$			
$c = 20$	0	6.4	0.0	4.1	4.5	6.4	0.0	4.1	4.8
	5	7.1	0.0	10.4	12.3	6.4	0.2	9.4	11.1
	10	13.3	0.0	21.9	26.5	11.3	1.6	20.6	25.4
	15	24.5	0.3	40.5	50.3	19.4	7.9	37.2	47.2
	25	68.8	22.6	84.2	93.9	60.4	54.3	80.1	93.2
		$T = 200$				$T = 500$			
$c = 30$	0	6.0	0.0	4.3	4.9	5.8	0.0	4.0	4.9
	5	6.4	0.0	9.1	10.5	6.0	0.0	8.5	10.3
	10	11.4	0.0	17.7	21.9	9.1	0.0	15.8	20.2
	15	20.1	0.0	32.4	39.3	16.1	0.5	28.4	35.9
	25	54.1	0.3	70.6	81.3	42.4	12.1	63.7	77.1
		$T = 200$				$T = 500$			
$c = 40$	0	6.1	0.1	4.0	4.7	5.5	0.0	4.1	5.0
	5	6.8	0.1	8.9	10.5	5.7	0.0	7.2	9.4
	10	10.5	0.1	16.8	20.0	9.1	0.0	14.3	18.3
	15	18.5	0.1	28.1	34.1	13.5	0.0	24.3	30.2
	25	45.1	0.1	60.8	71.4	34.9	0.8	52.5	65.2
		$T = 200$				$T = 500$			
$c = 50$	0	5.9	0.1	3.6	4.4	5.5	0.0	3.7	5.0
	5	6.5	0.1	7.8	9.7	6.2	0.0	7.1	9.5
	10	10.4	0.1	15.3	19.4	8.1	0.0	12.5	16.5
	15	16.6	0.1	26.4	32.1	12.1	0.0	20.5	26.3
	25	41.6	0.1	55.5	64.9	30.2	0.0	45.1	56.3

Notes: AHW denotes the (2-sided) Amihud, Hurwich and Wang test with lag length $p = 2$; CY denotes the Campbell and Yogo test, t_{ivx} is IVX test computed following Kostakis et al. (2015) and $\tilde{t}_{ivx}$ the residual-augmented IVX test procedure, all with maximal lag length $p = [4(T/100)^{0.25}]$. The DGP is as in (1) and (2) with $\rho = 1 - {}^{c}/_{T}$ and $\beta = {}^{b}/_{T}$. For further details see the text.

Table 2: Size and power against local alternatives, positive short-run AR parameter

		AHW	CY	t_{ivx}	$\tilde{t}_{ivx}$	AHW	CY	t_{ivx}	$\tilde{t}_{ivx}$
	b		$T = 200$				$T = 500$		
	0	6.5	4.6	11.1	6.6	6.3	4.1	10.6	6.3
	5	94.7	100.0	98.4	96.1	95.7	100.0	98.5	97.6
$c = 0$	10	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0
	15	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	25	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
			$T = 200$				$T = 500$		
	0	6.3	4.1	8.7	5.7	6.5	3.7	8.6	6.2
	5	26.5	64.4	79.0	72.9	27.3	66.0	79.9	74.9
$c = 10$	10	99.5	100.0	100.0	99.7	99.6	100.0	100.0	99.9
	15	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	25	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
			$T = 200$				$T = 500$		
	0	5.7	3.1	7.2	5.6	5.9	3.1	7.5	5.9
	5	16.4	28.6	48.7	43.9	16.4	31.6	49.2	44.5
$c = 20$	10	70.2	94.4	98.8	97.7	74.9	96.7	99.3	98.7
	15	100.0	100.0	100.0	100.0	96.0	100.0	100.0	100.0
	25	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
			$T = 200$				$T = 500$		
	0	6.0	2.2	7.2	5.9	5.8	2.5	7.1	5.6
	5	13.3	16.1	35.6	32.3	13.2	18.7	37.2	34.1
$c = 30$	10	47.6	63.2	86.8	85.4	50.5	72.9	89.8	89.2
	15	94.1	98.2	100.0	99.9	97.0	99.6	100.0	100.0
	25	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
			$T = 200$				$T = 500$		
	0	5.5	1.6	6.7	5.5	5.2	1.8	6.5	5.5
	5	10.2	10.4	28.4	26.5	11.0	12.2	29.7	27.4
$c = 40$	10	35.7	40.2	71.9	70.2	38.5	50.3	76.3	75.0
	15	79.5	82.8	98.4	98.3	84.4	91.8	99.2	99.2
	25	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0
			$T = 200$				$T = 500$		
	0	6.1	1.3	6.6	5.7	5.3	1.4	6.7	5.7
	5	9.7	7.2	24.7	22.9	9.5	8.7	25.9	24.5
$c = 50$	10	28.1	26.8	61.0	59.0	30.4	33.8	64.9	63.3
	15	64.3	62.3	93.0	92.7	71.2	75.7	95.9	95.7
	25	99.9	99.1	100.0	100.0	100.0	100.0	100.0	100.0

Note: See Table 1.

4.3 Robustness against empirical features of the data

To evaluate the performance of the procedures under other empirically relevant features, in Tables 3 and 4 we report results for the empirical size under DGPs with time-varying volatility and time-varying persistence. In specific, we consider five common variance patterns, namely:

1. constant, $\sigma_\varepsilon^2(s) = \sigma_\nu^2(s) = 1$;
2. an early upward break, $\sigma_\varepsilon^2(s) = \sigma_\nu^2(s) = 1 + 8\mathbb{I}(s > 0.3)$;
3. a late upward break, $\sigma_\varepsilon^2(s) = \sigma_\nu^2(s) = 1 + 8\mathbb{I}(s > 0.7)$;
4. an early downward break, $\sigma_\varepsilon^2(s) = \sigma_\nu^2(s) = 9 - 8\mathbb{I}(s > 0.3)$; and
5. a late downward break, $\sigma_\varepsilon^2(s) = \sigma_\nu^2(s) = 9 - 8\mathbb{I}(s > 0.7)$,

where $\mathbb{I}(\cdot)$ is an indicator function; and to allow for time-varying persistence, we also consider 6 patterns for the localization parameter c:

1. constant close to integration, $c(s) = 5$;
2. small break towards stationarity, $c(s) = 5 + 5\mathbb{I}(s > 0.5)$;
3. large break towards stationarity, $c(s) = 5 + 20\mathbb{I}(s > 0.5)$;
4. constant close to stationarity, $c(s) = 25$;
5. small break towards integration, $c(s) = 25 - 5\mathbb{I}(s > 0.5)$;
6. large break towards integration, $c(s) = 25 - 20\mathbb{I}(s > 0.5)$.

To gauge the necessity of a correction for time-varying variances, we now compute, in addition, the IVX test without White heteroskedasticity correction and denote it by $t_{ivx}^{\#}$; t_{ivx} is computed with (the usual) White standard errors, and $\tilde{t}_{ivx}$ is computed using the heteroskedasticity-robust standard errors from (12) as before.

Tables 3 and 4 confirm the conclusions obtained under the homogenous DGPs. The test based on t_{ivx} exhibits practically the same behavior under the variance patterns employed here, but can be oversized for constant small c (here, it is the closeness to the unit root that matters and not the breaks in c). On the other hand, the size control of $\tilde{t}_{ivx}$ is overall quite good, for all persistence patterns, and the White-type standard errors account for time-varying variances as well.[3]

[3] Unreported simulations show that not employing the White-type standard errors for the $\tilde{t}_{ivx}$ test under time-varying variances leads to size distortions similar to those of the $t_{ivx}^{\#}$ test.

Table 3: Size under breaks in variance and persistence, negative short-run AR parameter

		AHW	CY	$t^{\#}_{ivx}$		$\tilde{t}_{ivx}$	AHW	CY	$t^{\#}_{ivx}$		$\tilde{t}_{ivx}$
c	Var			$T = 200$					$T = 500$		
	const	7.6	0.1	9	9.6	5.5	7.4	1.2	10.4	10.7	5.9
	early up	11.5	0.1	13.2	9.8	6.4	11.2	1.6	13.5	9.9	6.6
const small	late up	24.1	0.6	17.9	9.6	5.8	25.2	3.9	19.3	10	6.1
	early down	21.5	0.4	15.1	8.8	5.5	22.1	3.0	16.4	9.4	5.9
	late down	10.7	0.4	11.3	9.3	5.6	11.1	2.3	12.3	9.6	6.3
				$T = 200$					$T = 500$		
	const	7.0	0.0	8.3	8.8	5.9	7.3	0.7	9.6	9.9	6.3
	early up	11.7	0.0	12.1	9.3	5.9	11.5	1.5	12.9	9.6	6.6
up small	late up	23.2	0.1	16.4	9.3	5.6	24.1	2.3	17.4	9.4	5.8
	early down	22.2	0.2	14.9	8.2	6	22.2	2.8	17	9.3	6.9
	late down	10.9	0.1	11	8.6	6.3	11.3	1.8	12.1	9.2	6.6
				$T = 200$					$T = 500$		
	const	6.6	0.0	7.2	7.9	5.4	6.8	0.3	8.9	9	6.2
	early up	10.9	0.0	10.3	8.1	5.5	11.5	0.3	11.5	8.7	5.9
up large	late up	21.5	0.0	13.6	8.5	4.8	21.5	0.3	14.2	8.8	5.2
	early down	22.3	0.2	14.7	7.8	6.9	22.8	2.7	17.1	8	6.9
	late down	11.6	0.0	10.5	7.5	6.3	11.3	1.1	11.4	8.4	6.9
				$T = 200$					$T = 500$		
	const	6.2	0.0	5.6	6.1	5.3	5.6	0.0	6.7	6.7	5.5
	early up	10.6	0.0	10.2	7.8	6.1	10.4	0.0	11.3	8.1	6.2
const large	late up	24.4	0.1	15.6	7.8	6.3	24.5	0.1	16.8	7.9	6.7
	early down	24.0	0.0	11	5.5	5.5	23.2	0.0	13.4	6.4	6.4
	late down	11.1	0.0	7.8	5.8	5.6	11.0	0.0	8.4	5.9	5.4
				$T = 200$					$T = 500$		
	const	6.1	0.0	5.9	6.2	5.5	6.1	0.0	7.1	7.4	5.6
	early up	10.9	0.0	10.4	8	6	11.1	0.1	11.1	7.9	6.1
down small	late up	23.6	0.1	16.4	8.2	6.9	23.9	0.2	16.9	8.2	6.6
	early down	23.4	0.0	10.7	5.4	5.6	23.2	0.1	12.5	6.2	5.8
	late down	10.6	0.0	7.5	5.9	5.4	10.8	0.0	9.3	6.6	5.8
				$T = 200$					$T = 500$		
	const	7.0	0.0	7.2	7.6	5.1	7.4	0.2	9.1	9.3	5.9
	early up	11.2	0.1	12.4	9.4	6.2	11.4	1.3	13.6	9.4	6.6
down large	late up	25.0	0.4	19.9	9.1	7.1	25.4	4.3	21.4	9	7.3
	early down	21.3	0.0	10	6	4.3	21.3	0.2	11.8	6.8	4.5
	late down	10.2	0.0	8.9	7.5	4.6	10.3	0.3	9.5	8.2	4.6

Notes: AHW denotes the (2-sided) Amihud, Hurwich and Wang test with lag length $p = 2$; CY denotes the Campbell and Yogo test, $t^{\#}_{ivx}$ is IVX test computed following Kostakis et al. (2015) but without White correction, and $\tilde{t}_{ivx}$ is the residual-augmented IVX test procedure, all with maximal lag length $p = [4(T/100)^{0.25}]$. The DGP is as in (1) and (2) with $\rho = 1 - c_t/T$ and $\beta = b/T$ and exhibits time-varying variance. For further details see the text.

Table 4: Size under breaks in variance and persistence, positive short-run AR parameter

		AHW	CY	$t^{\#}_{ivx}$		$\tilde{t}_{ivx}$	AHW	CY	$t^{\#}_{ivx}$		$\tilde{t}_{ivx}$
c	Var			$T = 200$					$T = 500$		
	const	6.6	4.5	10.3	10.7	6.1	6.3	4.2	10.5	10.7	6
	early up	10.0	6.8	14.6	11.2	6.8	10.0	6.6	13.9	10	6.5
const small	late up	22.6	10.4	19.5	11.5	6.7	23.4	9.6	20.4	10.9	6.4
	early down	19.9	9.0	17.1	10.7	6	19.9	8.4	18	10.5	6.2
	late down	9.8	6.7	12.5	10.6	6.2	9.5	6.1	12.8	10.3	6.4
				$T = 200$					$T = 500$		
	const	5.8	4.3	10.2	10.9	5.9	6.4	4.2	10.3	10.5	6.1
	early up	10.2	6.4	14	11.1	6.3	10.1	6.1	13.5	10.1	6.8
up small	late up	21.7	8.7	17.3	10.7	6.4	21.9	8.4	18.5	10.6	6.1
	early down	19.9	9.3	17.7	10.7	7	19.7	9.4	18.3	10.1	7
	late down	9.7	7.1	13.1	10.3	6.6	9.7	6.6	13.5	10.3	7.2
				$T = 200$					$T = 500$		
	const	5.9	4.1	9.3	9.7	5.8	5.9	3.7	9.8	10.2	6.4
	early up	9.9	5.7	12.3	10.1	5.9	10.4	5.2	11.7	9.1	6
up large	late up	20.5	6.1	14.1	9.9	5.5	20.7	5.9	14.6	9.7	5.5
	early down	20.9	9.7	18.6	10.2	7.6	20.6	10.4	19.5	9.6	7.7
	late down	10.1	7.1	12.3	9.4	6.9	10.4	6.7	12.6	8.8	6.6
				$T = 200$					$T = 500$		
	const	5.8	2.6	8.4	9	6.3	5.6	2.8	8	8.2	6.1
	early up	10.9	5.3	10.8	8.4	5.9	10.5	5.4	11.6	8	6.2
const large	late up	22.7	8.0	17.2	8.8	7	24.1	9.1	18.8	9.4	7.4
	early down	23.1	4.5	14.1	7.8	6.6	22.3	5.6	15.4	7.7	6.3
	late down	10.7	3.8	10.2	7.9	6.4	10.1	4.1	10.3	7.5	5.7
				$T = 200$					$T = 500$		
	const	5.9	2.9	8.5	8.9	5.9	6.0	3.0	8.4	8.5	5.9
	early up	10.5	5.5	11.8	8.8	5.9	10.7	5.6	11.8	8.6	6.3
down small	late up	23.3	8.8	18.5	9.5	7.3	24.2	9.7	19.2	9.1	7.2
	early down	22.2	4.7	15.1	8.7	7	21.8	5.6	15.9	8.4	6.8
	late down	10.1	3.9	10.6	8.5	6.9	10.3	4.3	10.5	7.8	5.8
				$T = 200$					$T = 500$		
	const	6.3	3.9	9.5	10	5.8	6.2	3.6	9.4	9.5	5.5
	early up	10.3	7.1	13.5	9.9	6.8	11.0	6.7	14.4	9.7	6.9
down large	late up	25.0	12.8	21.7	10.4	7.9	24.8	12.7	22.9	9.6	7.4
	early down	20.6	4.4	12.6	9	5.3	19.8	4.5	13.7	9	5.3
	late down	9.7	4.7	10.6	10.3	5.6	9.6	4.2	10.5	9.4	4.9

Note: See Table 3.

IVX without robust standard errors can be seriously oversized, which, again, was expected; the worst effect is observed for late upward breaks in the variance. AHW exhibits a similar pattern, to an even larger extent. We note that breaks in the persistence parameter c tend to rather have a dampening effect, if any. CY is severely undersized, in line with the previous experiments for negative short-run correlation. For positive short-run correlation, CY now controls size fairly well except for late upward and early downward breaks in the variance; the other three tests do not appear to be sensitive to the sign of the short-run serial correlation of the predictor. The effects are practically the same for both sample sizes, indicating that the size distortions are not finite-sample in nature.

5 Excess return predictability

The objective of this empirical part is to re-examine the predictive power of several variables used in Welch and Goyal (2008), updated with information up to December 2013.[4] using the approaches discussed in the previous sections. We look at the claims by Welch and Goyal (2008) that "evidence suggests that most models are unstable or even spurious" and that "models are no longer significant even in-sample."

5.1 Background

According to the findings of Welch and Goyal (2008), most predictive models have performed poorly in sample over the last 30 years. As they argue for many models any earlier apparent statistical significance was often based exclusively on years up to and especially on the years of the Oil shock 1973-1975 (Welch and Goyal, 2008, p. 1456).

Ang and Bekaert (2007), considering a sample from 1935 to 2001, report results for several subsamples and for the full sample. Since interest rate data is hard to interpret before the 1951 Treasury Accord, Ang and Bekaert (2007) (as well as Lewellen, 2004) consider 1952 as their starting date. Furthermore, Ang and Bekaert (2007) also indicate that the majority of studies establish strong evidence of predictability when data before or up to the early 1990s is used. For instance, Lettau and Ludvigsson (2001) and Goyal and Welch (2003) point out that the predictive power of the dividend yield weakens with the addition of the 1990s decade.

Several researchers suggest that the disapearance of stock return predictability is due to parameter instability or structural breaks and identify the disapearance around 1991 (see, *e.g.,* Pesaran and Timmermann, 2002; and Lettau and Nieuwerburgh, 2008). A related hypothesis is that predictability was arbitraged away once discovered, in a scenario similar to the attenuation of the January effect. Welch and Goyal (2008) argue that predictability has not been significant in- or out-of-sample over the past 30 years. Still others take a more drastic view and argue that it was never actually there (*e.g.,* Bossaerts and Hillion, 1999 and Goyal and Welch, 2003).

[4]We thank A. Goyal for making this data available on his Web site.

Henkel et al. (2011) reveal that predictability is a phenomenon whose strength is distinctively time-varying. The dividend yield and commonly used term structure variables are effective predictors almost exclusively during recessions. According to these authors, the robust prominence of busines cycles in these results suggests a potentially substantial tie to the literature on the dynamics of expected returns. Campbell and Cochrane (1999), Menzly et al. (2004) and Bekaert et al. (2009) show that risk premiums are countercyclical and that the time series behaviour of risk premium is higher during recessions.

Since a time-varying predictive relation is the byproduct of the interacting dynamics of expected returns and of the predictors, the complex behaviour of the predictors themselves must be considered when testing for predictability. The underlying fundamentals are the potential micro-level objectives of firms and central banks whose activities jointly determine aggregate predictor variables. The business cycle is an important driver of these micromotives and this lead Henkel et al. (2011) to re-examine predictability using a regime-switching framework capable of matching the time-varying dynamics of predictors to the dynamics of expected returns. It is found that predictors are less persistent and more volatile during recessions. Several features of their analysis stand out: the random walk model of stock prices prevailed in the 1970s based on CRSP data from the 1960s era expansion; predictability emerged in research of the late 1970s and mid-1980s, following several recessions; and predictability was subsequently doubted following the long booms of the 1980s and 1990s.

Hence, in line with Ang and Bekaert (2007) and given the availability of data, we revisit the impact of the addition of the 1990s first, followed by the analysis of the effects of adding the period from January 2000 to September 2007 and finally the remaining sample period (October 2007 to December 2013). Moreover, in order to remove the possible impact of the Oil shock (1973-1975) we repeat the analysis starting in 1976.

Given the available empirical evidence of change in strength of predictability of some variables over time, in what follows we split the sample into eigth periods. These changes appear to be accompanied by changes in the persistence of the considered regressors.[5] In particular, we consider the eight time periods: i) Jan 1952 - Dec 1989; ii) Jan 1952 - Dec 1999; iii) Jan 1952 - Sep 2007; iv) Jan 1952 - Dec 2013; v) Jan 1976 - Dec 1989; vi) Jan 1976 - Dec 1999; vii) Jan 1976 - Sep 2007; and viii) Jan 1976 - Dec 2013.

5.2 Data

The dependent variable is the equity premium (or excess return), *i.e.*, the total rate of return on the stock market minus the prevailing short-term interest rate. Stock returns are the continuously compounded returns on the S&P 500 index, including dividends, and the risk-free rate is the Treasury-bill rate.

The independent variables used are: i) the 12-month moving sums of dividends (D12) paid

[5]See the results in Appendix B for more details.

on the S&P 500 index; ii) the dividend price-ratio (d/p) computed as the difference between the log of dividends and the log of prices; iii) the dividend yield (d/y) computed as the difference between the log of dividends and the log of lagged prices; iv) the 12-month moving sums of earnings on the S&P 500 index (E12); v) the earnings price-ratio (e/p) computed as the difference between the log of earnings and the log of prices; vi) the dividend payout-ratio (d/e) computed as the difference between the log of dividends and the log of earnings; vii) the stock variance (svar) computed as the sum of squared daily returns on the S&P 500; viii) the cross-sectional beta premium (csp) which measures the relative valuations of high- and low-beta stocks; ix) the book-to-market ratio (b/m) computed as the ratio of book value to market value for the Dow Jones industrial average. To include corporate issuing activity we also use x) the net equity expansion (ntis) computed as the ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks; and xi) the percent equity issuing (eqis), which is the ratio of equity issuing activity as a fraction of total issuing activity.

A further set of predictors used is: the treasury bills (tbl) rates; the long term government bond yield (lty); the term spread (tms) which is the difference between the long term yield on government bonds and the treasury-bill; the default yield spread (dfy) which is the difference between BAA and AAA-rated corporate bond yields. The default return spread (dfr) is the difference between long-term corporate bond and long-term government bond returns; inflation (infl) which corresponds to the consumer price index (all urban consumers); and long-term government bond returns (ltr). For details on the construction of these variables and for a greater description see Welch and Goyal (2008).

5.3 Findings

Tables 5 and 6 report the predictability test results computed from t_{ivx}, $\tilde{t}_{ivx}$ and the OLS based tests procedures over four subperiods of analysis starting in January 1952. From Table 5 it is interesting to observe that the OLS based test procedure finds most evidence of predictability in the subsample from January 1952 to December 1989, and as we add information the number of significant predictors decreases. Note that in the subsample from January 1952 to December 1989, based on this procedure, nine variables (d/p, d/y, d/e, tbl, tms, ntis, infl, ltr, svar) seemed to be significant; whereas in the following subperiods (January 1952 to December 1999; to September 2007, and to December 2013) the number of significant variables reduced to six (tbl, lty, tms, ntis, infl, ltr), to two (infl, ltr) and increases again to six (tbl, lty, tms, infl, ltr, svar), respectively. However, if we look at the results obtained with the two IVX approaches, the number of significant predictors is smaller. The original IVX approach for the four periods under analysis (January 1952 to December 1989; January 1952 to December 1999; January 1952 to September 2007 and January 1952 to December 2013) finds 5, 5, 2 and 4 significant predictors, respectively; whereas the residual augmented IVX approach proposed in this paper finds 5, 6, 2 and 5, respectively.

Performing the same analysis, but starting now in January 1975 instead of January 1952, the OLS based approach finds 1, 3, 1 and 1 significant predictors in the four subsamples under analysis (January 1975 to December 1989; January 1975 to December 1999; January 1975 to September 2007 and January 1976 to December 2013), respectively. Thus, based on this statistic the period between January 1976 to December 1999 is the one which presents more evidence of predictability. Using the IVX based approaches, the number of signifcicant predictors is 1, 5, 1 and 3, for the original IVX and 3, 4, 2 and 0 for the residual augmented IVX approach, for the four subperiods under analysis, respectively. Hence, both IVX based approaches also identify the period between 1976 and 1999 as the period with strongest evidence of predictability.

The results in Table 6 agree to a certain extent with the conclusions put forward by Welch and Goyal (2008) that apparent statistical significance was often based exclusively on years up to and especially on the years of the Oil Shock of 1973-1975.

6 Conclusions

This paper introduced a new IVX test statistic computed from a residual augmented predictive regression as considered in Amihud and Hurvich (2004) and reexamined the empirical evidence on predictability of stock returns of Welch and Goyal (2008) using these new robust methods.

To this end we resorted to IVX estimation and testing, and proposed a residual-augmented variant that allows practitioners to distinguish more reliably between the null of no predictability and the alternative. The method is asymptotically correct under near-integration as well as under stationarity of the regressor, has improved local power under high regressor persistence, and allows, *e.g.*, for heterogeneity of the data in the form of time-varying variances.

The results derived here on bias correction can be generalized for other types of instrumental variable estimation than just IVX. The IV framework of Breitung and Demetrescu (2015), who distinguish between type-I instruments that are less persistent than the initial regressor (the IVX instrument is actually of type I; see Breitung and Demetrescu, 2015), and type-II instruments that are (stochastically) trending, yet exogenous, allows for a quick discussion: a careful examination of the arguments presented here shows that they are easily extended for type-I instruments, but type-II instruments behave like the OLS estimator where residual-augmentation is not improving on the test procedure even asymptotically.

The provided Monte Carlo evidence shows that the asymptotic improvements are a good indicative of the finite-sample performance, also in the presence of time-varying volatility or time varying persistence. Finally, the empirical analysis showed that the bias-adjusted IVX procedure detected predictability more often than the original IVX procedure.

Table 5: Testing for Predictability (starting date January 1952)

	Jan 1952 - Dec 1989			Jan 1952 - Dec 1999			Jan 1952 - Sep 2007			Jan 1952 - Dec 2013		
	$\tilde{t}_{ivx}$	t_{ivx}	OLS	$\tilde{t}_{ivx}$	t_{ivx}	OLS	$\tilde{t}_{ivx}$	t_{ivx}	OLS	$\tilde{t}_{ivx}$	t_{ivx}	OLS
D12	-0.648	-0.038	-0.939	-0.233	1.423	0.591	-0.175	1.098	0.090	-0.205	1.074	0.054
E12	-0.843	-0.289	-1.254	-0.116	1.287	0.408	-0.135	1.019	-0.017	-0.171	0.924	-0.138
d/p	-0.674	0.253	2.103**	-0.540	-1.153	0.143	-0.455	-0.386	1.093	-0.664	-0.414	1.067
d/y	0.877	0.477	2.286**	0.078	-1.069	0.207	0.677	-0.284	1.153	0.744	-0.243	1.223
e/p	-1.049	-0.332	0.588	-1.243	-1.220	-0.188	-0.461	-0.296	0.921	-0.307	-0.507	0.548
d/e	1.557	1.343	2.109**	0.707	-0.069	0.680	0.450	-0.323	0.474	0.356	0.093	0.660
b/m	-0.472	0.027	0.479	-0.892	-1.451	-0.802	-0.530	-0.894	-0.091	-0.551	-0.847	-0.015
tbl	-2.368	-1.820	-2.819***	-2.666***	-1.651*	-2.925***	-2.275	-1.567	-2.683	-2.201**	-1.726*	-2.596***
lty	-1.454	-0.818	-1.633	-1.817*	-0.587	-1.668*	-1.492	-0.590	-1.638	-1.552	-0.766	-1.752*
tms	2.660***	2.824***	2.867***	2.466**	2.768***	2.726***	2.173	2.406	2.291	2.019**	2.399**	2.180**
dfy	1.292	1.653	1.339	0.742	1.411	1.072	0.760	1.351	1.012	0.082	0.868	0.461
dfr	0.940	0.934	0.492	1.174	1.451	0.881	0.093	0.535	-0.175	0.021	0.814	0.090
ntis	-2.069**	-2.477**	-2.181**	-1.723*	-2.509**	-2.128**	-1.636	-2.353	-1.895	-0.349	-1.201	-0.541
infl	-2.093**	-1.721*	-3.124***	-2.673***	-1.793*	-3.566***	-2.588***	-1.873*	-3.558***	-2.091**	-1.363	-2.772***
ltr	2.620***	2.820***	2.785***	2.611***	2.770***	2.714***	2.462**	2.559**	2.485**	2.120**	2.450**	2.378**
svar	-1.950*	-1.958*	-2.012**	-1.264	-1.403	-1.507	-1.206	-1.454	-1.618	-1.648*	-2.964***	-3.139***

Table 6: Testing for Predictability (starting date January 1975)

	Jan 1975 - Dec 1989			Jan 1975 - Dec 1999			Jan 1975 - Sep 2007			Jan 1975 - Dec 2013		
	$\tilde{t}_{ivx}$	t_{ivx}	OLS	$\tilde{t}_{ivx}$	t_{ivx}	OLS	$\tilde{t}_{ivx}$	t_{ivx}	OLS	$\tilde{t}_{ivx}$	t_{ivx}	OLS
D12	-0.191	0.311	0.156	-0.062	1.854*	1.358	0.229	1.437	0.578	0.038	1.271	0.340
E12	-0.315	0.055	-0.170	0.618	1.996**	1.450	0.431	1.489	0.605	0.154	1.128	0.183
d/p	-0.302	0.668	0.939	-0.393	-1.475	-0.774	-0.260	-0.416	0.581	-0.467	-0.417	0.621
d/y	0.599	0.632	1.077	-0.288	-1.535	-0.792	0.456	-0.495	0.589	0.534	-0.407	0.739
e/p	-0.455	0.087	0.332	-0.987	-1.301	-0.616	-0.021	-0.261	0.754	-0.101	-0.482	0.415
d/e	0.591	0.852	0.795	-0.696	-0.031	-0.275	-0.512	-0.349	-0.309	-0.069	0.194	0.204
b/m	0.103	0.040	0.300	-0.415	-1.609	-0.955	-0.189	-1.088	-0.145	-0.181	-1.022	-0.036
tbl	-1.892*	-1.451	-1.506	-2.273**	-2.002**	-1.973**	-1.520	-1.628	-1.273	-1.242	-1.712*	-1.230
lty	-1.431	-0.863	-0.935	-2.084**	-1.746*	-1.778*	-1.277	-1.375	-0.966	-1.126	-1.508	-0.992
tms	1.402	1.432	1.423	0.889	1.181	1.141	0.858	1.014	0.983	0.854	1.144	1.007
dfy	1.239	0.859	1.453	0.813	-0.574	0.388	0.501	-0.590	0.536	-0.123	-0.916	-0.035
dfr	0.866	0.066	0.259	1.639	0.083	0.645	-0.094	-0.873	-0.733	-0.210	-0.421	-0.345
ntis	-2.579***	-2.257**	-2.196**	-1.997**	-1.368	-1.627	-1.729*	-1.416	-1.543	0.021	-0.076	-0.057
infl	-1.532	-1.415	-1.394	-1.879*	-2.513**	-2.201**	-1.956*	-2.443**	-2.074**	-1.035	-1.749*	-1.234
ltr	1.415	1.285	1.428	1.536	1.027	1.389	1.382	0.922	1.279	1.217	0.949	1.300
svar	-2.174**	-1.628	-1.584	-1.105	-1.477	-1.237	-1.112	-1.560	-1.353	-1.636	-3.092***	-2.951***

References

Amado, C. and T. Teräsvirta (2013). Modelling volatility by variance decomposition. *Journal of Econometrics 175*(2), 142–153.

Amado, C. and T. Teräsvirta (2014). Modelling changes in the unconditional variance of long stock return series. *Journal of Empirical Finance 25*(1), 15–35.

Amihud, Y. and C. M. Hurvich (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis 39*(4), 813–841.

Amihud, Y., C. M. Hurvich, and Y. Wang (2009). Multiple-predictor regressions: Hypothesis testing. *Review of Financial Studies 22*(1), 413–434.

Amihud, Y., C. M. Hurvich, and Y. Wang (2010). Predictive regression with order-p autoregressive predictors. *Journal of Empirical Finance 17*(3), 513–525.

Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica, 61, 821-856.*.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica, 62, 1383-1414*.

Ang, A. and G. Bekaert (2007). Return predictability: Is it there? *Review of Financial Studies 20*(3), 651–707.

Avramov, D., L. Barras, and R. Kosowski (2010). Hedge fund predictability under the magnifying glass: The economic value of forecasting individual fund returns. Working paper.

Bali, T. G. (2008). The intertemporal relation between expected returns and risk. *Journal of Financial Economics 87*(1), 101–131.

Bekaert, G., E. Engstrom, and Y. Xing (2009). Uncertainty, and asset prices. *Journal of Financial Economics 91*(1), 59–82.

Bossaerts, P. and P. Hillion (1999). Implementing statistical criteria to select return forecasting models: What do we learn? *Review of Financial Studies 12*(2), 405–428.

Breitung, J. and M. Demetrescu (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics 187*(1), 358–375.

Busetti, F. and A. M. R. Taylor (2004). Tests of stationarity against a change in persistence. *Journal of Econometrics, 123, 33-66*.

Cai, Z. and Y. Wang (2014). Corrigendum to “Testing predictive regression models with nonstationary regressors” [J. Econometrics 178 (2014) 4–14]. *Journal of Econometrics 181*(2), 194.

Campbell, J. Y. (2008). Viewpoint: Estimating the equity premium. *Canadian Journal of Economics/Revue canadienne d'économique 41*(1), 1–21.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy 107*(1), 205–251.

Campbell, J. Y. and M. Yogo (2006). Efficient tests of stock return predictability. *Journal of Financial Economics 81*(1), 27–60.

Camponovo, L. (2015). Differencing transformations and inference in predictive regression models. *Econometric Theory 31*(6), 1331–1358.

Cavaliere, G. (2004). Unit root tests under time-varying variances. *Econometric Reviews 23*(3), 259–292.

Cavaliere, G., A. Rahbek, and A. M. R. Taylor (2010). Testing for Co-Integration in Vector Autoregressions with Non-Stationary Volatility. *Journal of Econometrics 158*(1), 7–24.

Cavanagh, C. L., G. Elliott, and J. H. Stock (1995). Inference in models with nearly integrated regressors. *Econometric Theory 11*(5), 1131–1147.

Chang, Y. and J. Y. Park (2002). On the asymptotics of ADF tests for unit roots. *Econometric Reviews 21*(4), 431–447.

Chun, S. (2009). Are international equity market returns predictable? *Boston University working paper*.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford university press.

Davies, R. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64, 247-54.*.

Demetrescu, M. (2014a). Corrigendum to “Testing predictive regression models with nonstationary regressors” [J. Econometrics 178 (2014) 4–14]. Working paper, University of Bonn.

Demetrescu, M. (2014b). Enhancing the local power of IVX-based tests in predictive regressions. *Economics Letters 124*(2), 269–273.

Elliott, G. and J. H. Stock (1994). Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory 10*(3-4), 672–700.

Faust, J. and J. H. Wright (2011). Efficient prediction of excess returns. *Review of Economics and Statistics 93*(2), 647–659.

Georgiev, I., D. I. Harvey, S. J. Leybourne, and A. M. R. Taylor (2015). A bootstrap stationarity test for predictive regression invalidity. Essex Finance Centre Working Papers 16666, University of Essex, Essex Business School.

Gonzalo, J. and J.-Y. Pitarakis (2012). Regime-specific predictability in predictive regressions. *Journal of Business & Economic Statistics 30*(2), 229–241.

Goyal, A. and I. Welch (2003). Predicting the equity premium with dividend ratios. *Management Science 49*(5), 639–654.

Guidolin, M. and A. Timmermann (2006). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics 21*(1), 1–22.

Hansen, B. (1991). Tests for parameter instability in regressions with i(1) processes. *Journal of Business and Economic Statistics 10, 321-335..*

Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory 8*(4), 489–500.

Harvey, D. L., S. J. Leybourne, and A. M. R. Taylor (2006). Modified tests for a change in persistence. *Journal of Econometrics, 134, 441-469..*

Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of Financial Economics 99*(3), 560–580.

Jansson, M. and M. J. Moreira (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica 74*(3), 681–714.

Johannes, M., A. Korteweg, and N. Polson (2014). Sequential learning, predictability, and optimal portfolio returns. *The Journal of Finance 69*(2), 611–644.

Kim, J. (2000). Detection of change in persistence of a linear time series. *Journal of Econometrics, 95, 97-116..*

Kim, J. Y., J. Belaire-Franch, and R. Amador (2002). Corringendum to detection of change in persistence of a linear time series. *Journal of Econometrics, 109, 389-392..*

Kostakis, A., T. Magdalinos, and M. P. Stamatogiannis (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies 28*(5), 1506–1553.

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

Lettau, M. and S. Ludvigsson (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance 56*(3), 815–850.

Lettau, M. and S. V. Nieuwerburgh (2008). Reconciling the return predictability evidence. *The Review of Financial Studies 21*(4), 1607–1652.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics 74*(2), 209–235.

Magdalinos, T. and P. C. B. Phillips (2009). Limit Theory for Cointegrated Systems with Moderately Integrated and Moderately Explosive Regressors. *Econometric Theory 25*(2), 482–526.

Mankiw, G. N. and M. D. Shapiro (1986). Do we reject too often?: Small sample properties of tests of rational expectations models. *Economics Letters 20*(2), 139–145.

Maynard, A. and K. Shimotsu (2009). Covariance-based orthogonality tests for regressors with unknown persistence. *Econometric Theory 25*(1), 63–116.

Menzly, L., T. Santos, and P. Veronesi (2004). Understanding predictability. *Journal of Political Economy 112*(1), 1–47.

Park, J. Y. (1990). Testing for unit roots and cointegration by variable addition in t. fomby and f. rhodes (eds) advances in econometrics: Cointegration, spurious regressions and unit roots. *London Jai Press, 107-134*.

Pesaran, M. H. and A. Timmermann (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance 9*(5), 495–510.

Phillips, P. C. B. (2015). Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics 13*(3), 521–555.

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

Phillips, P. C. B. and K. L. Xu (2006). Inference in Autoregression Under Heteroskedasticity. *Journal of Time Series Analysis 27*(2), 289–308.

Rodrigues, P. M. M. and A. Rubia (2011). A class of robust tests in augmented predictive regressions. Working Paper 26/2011, Banco de Portugal.

Stambaugh, R. F. (1986). Bias in regressions with lagged stochastic regressors. *CRSP Working Paper No 156, University of Chicago*.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375–421.

Teräsvirta, T. and Z. Zhao (2011). Stylized facts of return series, robust estimates and three popular models of volatility. *Applied Financial Economics 21*(1-2), 67–94.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

A Technical Appendix

A.1 Preliminary Results

Throughout the proofs, we consider that $\sum_{j=0}^{t-1} \varrho^{kj} = \frac{1-\varrho^{kt}}{1-\varrho^{k}} = \frac{T^{\eta}}{a}\left(\frac{1-\varrho^{kt}}{1+\varrho+\ldots+\varrho^{k-1}}\right) \leq \frac{1}{ka}T^{\eta}$ for large enough T and fixed k, where $\varrho := 1 - \frac{a}{T^{\eta}}$ with $\eta \in (0,1)$ and $a > 0$ and fixed. Furthermore, let C denote a generic constant whose value may change from occurrence to occurrence.

Lemma A.1 *Under the assumptions of Theorem 3.1, as $T \to \infty$, it follows that*

1. $\frac{1}{T}\sum_{t=p+1}^{T} x_{t-1}\boldsymbol{x}_{t-p} \xrightarrow{p} \boldsymbol{\alpha}_p' \int_0^1 \sigma_v^2 \mathrm{d}s$, *where* $\boldsymbol{\alpha}_p := (\alpha_0, \ldots, \alpha_{p-1})$ *and* $\boldsymbol{x}_{t-p} := (x_{t-1},\ \ldots,\ x_{t-p})'$ *and α_h is as defined in Theorem 3.1;*
2. $\frac{1}{T}\sum_{t=p+1}^{T} \boldsymbol{x}_{t-p}\boldsymbol{x}_{t-p}' \xrightarrow{p} \boldsymbol{\Omega} \int_0^1 \sigma_v^2 \mathrm{d}s$, *where $\boldsymbol{\Omega}$ is a $p \times p$ matrix with generic element $a_{ij} = \alpha_{|i-j|}$;*
3. $\frac{1}{T}\sum_{t=p+1}^{T} \boldsymbol{x}_{t-p}\boldsymbol{x}_{t-p}'\nu_t^2 \xrightarrow{p} \boldsymbol{\Omega} \int_0^1 \sigma_v^4(s)\,\mathrm{d}s$;
4. $\frac{1}{T}\sum_{t=p+1}^{T} z_{t-1}^2\varepsilon_t^2 \xrightarrow{p} \alpha_0 \int_0^1 \sigma_v^2(s)\,\sigma_\varepsilon^2(s)\,\mathrm{d}s$.

Proof of Lemma A.1

Phillips and Xu (2006) show in their Lemma 1 that $\frac{1}{T}\sum_{t=h+1}^{T} x_t x_{t-h} \xrightarrow{p} \alpha_h \int_0^1 \sigma_v^2 \mathrm{d}s$, $h = 0, 1, \ldots, p-1$; this suffices to establish the results in the first two items. The result in item 3 also follows directly from Lemma 1 of Phillips and Xu (2006), and the proof can be adapted in a straightforward manner to establish the result in item 4. ■

Lemma A.2 *Under the assumptions of Theorem 3.2, as $T \to \infty$, it follows that*

$$\frac{\sum_{t=2}^{T} \tilde{z}_{t-1}\varepsilon_t}{\sqrt{\sum_{t=2}^{T} \tilde{z}_{t-1}^2\varepsilon_t^2}} \xrightarrow{d} \mathcal{N}(0,1)$$

where $\tilde{z}_t = \sum_{j=0}^{t-1} \varrho^j \nu_{t-j}$.

Proof of Lemma A.2

Consider $s_T^2 := \frac{1}{T^{1+\eta}} \sum_{t=2}^{T} \sum_{j=0}^{t-2} \varrho^{2j} \sigma_{\nu,t-1-j}^2 \sigma_{\varepsilon,t}^2$ and note that s_T^2 is bounded and bounded away from zero, since

$$\frac{\min_{1\leq t\leq T} \sigma_{\nu,t}^2 \min_{1\leq t\leq T} \sigma_{\varepsilon,t}^2}{T^{1+\eta}} \sum_{t=2}^{T} \sum_{j=0}^{t-2} \varrho^{2j} \leq s_T^2 \leq \frac{\max_{1\leq t\leq T} \sigma_{\nu,t}^2 \max_{1\leq t\leq T} \sigma_{\varepsilon,t}^2}{T^{1+\eta}} \sum_{t=2}^{T} \sum_{j=0}^{t-2} \varrho^{2j}$$

where $\sum_{t=2}^{T} \sum_{j=0}^{t-2} \varrho^{2j} \sim CT^{1+\eta}$.

Since,

$$\frac{\sum_{t=2}^{T} \tilde{z}_{t-1}\varepsilon_t}{\sqrt{\sum_{t=2}^{T} \tilde{z}_{t-1}^2 \varepsilon_t^2}} = \frac{1}{T^{1/2+\eta/2}} \sum_{t=2}^{T} \frac{\tilde{z}_{t-1}\varepsilon_t}{s_T} \sqrt{\frac{\sum_{t=2}^{T} \sum_{j=0}^{t-2} \varrho^{2j} \sigma_{\nu,t-1-j}^2 \sigma_{\varepsilon,t}^2}{\sum_{t=2}^{T} \tilde{z}_{t-1}^2 \varepsilon_t^2}}, \tag{25}$$

we show next that $\frac{1}{T^{1/2+\eta/2}} \sum_{t=2}^{T} \frac{\tilde{z}_{t-1}\varepsilon_t}{s_T}$ follows a limiting standard normal distribution by resorting to a central limit theorem for martingale difference [md] arrays (Davidson, 1994, Theorem 24.3). However, to apply it, we need to show that, i) $\max_t \frac{1}{T^{1/2+\eta/2}} \left| \frac{\tilde{z}_{t-1}\varepsilon_t}{s_T} \right| \xrightarrow{p} 0$ and ii) $\frac{1}{T^{1+\eta}} \sum_{t=2}^{T} \frac{\tilde{z}_{t-1}^2 \varepsilon_t^2}{s_T^2} \xrightarrow{p} 1$.

Given that the result in ii) also implies

$$\sqrt{\frac{\sum_{t=2}^{T} \sum_{j=0}^{t-2} \varrho^{2j} \sigma_{\nu,t-1-j}^2 \sigma_{\varepsilon,t}^2}{\sum_{t=2}^{T} \tilde{z}_{t-1}^2 \varepsilon_t^2}} \xrightarrow{p} 1, \tag{26}$$

hence the result in (25) would follow.

To verify i), note that uniform boundedness of moments of order $2+\delta^*$ for some $\delta^* > 0$ of $T^{-\eta/2}\tilde{z}_{t-1}\varepsilon_t$ suffices to establish this condition. An application of Hölder's inequality shows that uniformly bounded 4th order moments of $T^{-\eta/2}\tilde{z}_{t-1}$ and uniform $L_{4+\delta^*}$-boundedness of ε_t suffices, since δ^* may be chosen arbitrarily close to zero, so we check the uniform boundedness of

$$\mathrm{E}\left(\frac{\tilde{z}_{t-1}^4}{T^{2\eta}}\right) = \frac{1}{T^{2\eta}} \sum_{j=0}^{t-2} \sum_{k=0}^{t-2} \sum_{l=0}^{t-2} \sum_{m=0}^{t-2} \varrho^j \varrho^k \varrho^l \varrho^m \, \mathrm{E}\left(\nu_{t-j}\nu_{t-k}\nu_{t-l}\nu_{t-m}\right). \tag{27}$$

Due to the serial independence of ν_t, the expectation $\mathrm{E}\left(\nu_{t-j}\nu_{t-k}\nu_{t-l}\nu_{t-m}\right)$ is nonzero only if the indices are pairwise equal, thus we can simplify (27) as,

$$\mathrm{E}\left(\frac{\tilde{z}_{t-1}^4}{T^{2\eta}}\right) = \frac{1}{T^{2\eta}} \sum_{j=0}^{t-2} \sum_{k=0}^{t-2} \varrho^{2j} \varrho^{2k} \, \mathrm{E}\left(\nu_{t-j}^2 \nu_{t-k}^2\right).$$

Since ν_t is uniformly L_4-bounded, the expectations on the r.h.s. are uniformly bounded for any

t, k and j, therefore,

$$0 \leq E\left(\frac{\tilde{z}_{t-1}^4}{T^{2\eta}}\right) \leq C\frac{1}{T^{2\eta}}\sum_{j=0}^{t-2}\sum_{k=0}^{t-2}\varrho^{2j}\varrho^{2k} = C\frac{1}{T^{2\eta}}\left(\sum_{j=0}^{t-2}\varrho^{2j}\right)^2 \leq C\frac{1}{T^{2\eta}}\left(\sum_{j=0}^{T-2}\varrho^{2j}\right)^2 \leq C$$

which suffices for the required uniform L_4-boundedness.

To check condition ii), it suffices to show that

$$\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\tilde{z}_{t-1}^2\varepsilon_t^2 - s_T^2 \xrightarrow{p} 0 \tag{28}$$

because s_T^2 is bounded and bounded away from zero (we learn from Lemma A.4 below that $s_T^2 \to \frac{1}{2a}\int_0^1 \sigma_\nu^2(s)\,\sigma_u^2(s)\,\mathrm{d}s$, but the exact limit does not matter here). To prove (28), write

$$\begin{aligned}\sum_{t=2}^{T}\tilde{z}_{t-1}^2\varepsilon_t^2 &= \sum_{t=2}^{T}\sum_{j=0}^{t-2}\sum_{k=0}^{t-2}\varrho^j\varrho^k\nu_{t-1-j}\nu_{t-1-k}\left(\varepsilon_t^2 - \sigma_{\varepsilon,t}^2\right) + \sum_{t=2}^{T}\sum_{j=0}^{t-2}\sum_{k=0}^{t-2}\varrho^j\varrho^k\nu_{t-1-j}\nu_{t-1-k}\sigma_{\varepsilon,t}^2 \\ &=: A_T + B_T.\end{aligned}$$

Note that $\sum_{j=0}^{t-2}\sum_{k=0}^{t-2}\varrho^j\varrho^k\nu_{t-1-j}\nu_{t-1-k}\left(\varepsilon_t^2 - \sigma_{\varepsilon,t}^2\right)$ builds an md array and as such, is uncorrelated in t. Hence, showing $\frac{1}{T^{1+\eta}}A_T$ to vanish is not difficult, given that from the uncorrelatedness of the summands we can write that,

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{T^{1+\eta}}A_T\right) &= \frac{1}{T^{2+2\eta}}\sum_{t=2}^{T}\operatorname{Var}\left(\sum_{j=0}^{t-2}\sum_{k=0}^{t-2}\varrho^j\varrho^k\nu_{t-1-j}\nu_{t-1-k}\left(\varepsilon_t^2 - \sigma_{\varepsilon,t}^2\right)\right) \\ &= \frac{1}{T^{2+2\eta}}\sum_{t=2}^{T}\mathrm{E}\left(\left(\sum_{j=0}^{t-2}\sum_{k=0}^{t-2}\varrho^j\varrho^k\nu_{t-1-j}\nu_{t-1-k}\right)^2\right)\mathrm{E}\left(\left(\varepsilon_t^2 - \sigma_{\varepsilon,t}^2\right)^2\right).\end{aligned}$$

Now, ε_t is uniformly L_4-bounded and

$$\mathrm{E}\left(\left(\sum_{j=0}^{t-2}\sum_{k=0}^{t-2}\varrho^j\varrho^k\nu_{t-1-j}\nu_{t-1-k}\right)^2\right) = \sum_{j=0}^{t-2}\sum_{k=0}^{t-2}\sum_{l=0}^{t-2}\sum_{m=0}^{t-2}\varrho^j\varrho^k\varrho^l\varrho^m\,\mathrm{E}\left(\nu_{t-1-j}\nu_{t-1-k}\nu_{t-1-l}\nu_{t-1-m}\right)$$

where the expectation on the r.h.s. is, as before, uniformly bounded and nonzero only if the indices are pairwise equal. Hence,

$$0 \leq \mathrm{E}\left(\left(\sum_{j=0}^{t-2}\sum_{k=0}^{t-2}\varrho^j\varrho^k\nu_{t-1-j}\nu_{t-1-k}\right)^2\right) \leq C\sum_{j=0}^{t-2}\sum_{k=0}^{t-2}\varrho^{2j}\varrho^{2k} \leq CT^{2\eta}$$

leading to $\operatorname{Var}\left(\frac{1}{T^{1+\eta}}A_T\right) \to 0$ and thus $A_T = o_p\left(T^{1+\eta}\right)$.

Regarding B_T, note that,

$$\begin{aligned} B_T &= T^{1+\eta}s_T^2 + \sum_{t=2}^{T}\sum_{j=0}^{t-2}\varrho^{2j}\left(\nu_{t-1-j}^2 - \sigma_{\nu,t-1-j}^2\right)\sigma_{\varepsilon,t}^2 + \sum_{t=2}^{T}\sum_{\substack{j=0 \\ j\neq k}}^{t-2}\sum_{k=0}^{t-2}\varrho^{j}\varrho^{k}\nu_{t-1-j}\nu_{t-1-k}\sigma_{\varepsilon,t}^2 \\ &= T^{1+\eta}s_T^2 + B_{T1} + B_{T2}. \end{aligned}$$

For B_{T1} we have from the serial independence and L_4-boundedness of ν_t that

$$\begin{aligned} \mathrm{E}\left(\left(\sum_{j=0}^{t-2}\varrho^{2j}\left(\nu_{t-1-j}^2 - \sigma_{\nu,t-1-j}^2\right)\sigma_{\varepsilon,t}^2\right)^2\right) &= \sigma_{\varepsilon,t}^4\sum_{j=0}^{t-2}\varrho^{4j}\,\mathrm{E}\left(\left(\nu_{t-1-j}^2 - \sigma_{\nu,t-1-j}^2\right)^2\right) \\ &\leq CT^{\eta} \end{aligned}$$

and thus $\mathrm{E}\left(\left|\sum_{j=0}^{t-2}\varrho^{2j}\left(\nu_{t-1-j}^2 - \sigma_{\nu,t-1-j}^2\right)\sigma_{\varepsilon,t}^2\right|\right) \leq CT^{\eta/2}$. Hence,

$$\mathrm{E}\left(\left|\frac{1}{T^{1+\eta}}B_{T1}\right|\right) \leq \frac{C}{T^{1+\eta}}\sum_{t=2}^{T}T^{\eta/2} \to 0$$

and Markov's inequality indicates that $B_{T1} = o_p\left(T^{1+\eta}\right)$.

For B_{T2} we proceed similarly,

$$\begin{aligned} &\mathrm{E}\left(\left(\sum_{t=2}^{T}\sum_{\substack{j=0 \\ j\neq k}}^{t-2}\sum_{k=0}^{t-2}\varrho^{j}\varrho^{k}\nu_{t-1-j}\nu_{t-1-k}\sigma_{\varepsilon,t}^2\right)^2\right) \\ &\quad = \sum_{t=2}^{T}\sum_{s=2}^{T}\sum_{\substack{j=0 \\ j\neq k}}^{t-2}\sum_{k=0}^{t-2}\sum_{\substack{l=0 \\ l\neq m}}^{s-2}\sum_{m=0}^{s-2}\varrho^{j}\varrho^{k}\varrho^{l}\varrho^{m}\sigma_{\varepsilon,t}^2\sigma_{\varepsilon,s}^2\,\mathrm{E}\left(\nu_{t-1-j}\nu_{t-1-k}\nu_{s-1-l}\nu_{s-1-m}\right), \end{aligned}$$

where the expectations on the r.h.s. are nonzero if $t-j=s-l$ and $t-k=s-m$ or if $t-j=s-m$ and $t-k=s-l$ (with $t-j=t-k$ and $s-l=s-m$ being excluded by the requirement that $j\neq k$ and $l\neq m$). Note that, for any t,s,j,k,l,m with $j\neq k$ and $l\neq m$,

$$\sigma_{\varepsilon,t}^2\sigma_{\varepsilon,s}^2\,\mathrm{E}\left(\nu_{t-1-j}\nu_{t-1-k}\nu_{s-1-l}\nu_{s-1-m}\right) \leq \left(\max_t \sigma_{\varepsilon,t}^2\right)^2\left(\max_t \sigma_{\nu,t}^2\right)^2 \leq C.$$

Let us now focus on the terms for which $t-s=j-l=k-m$. Thus, for $t=s$, $t=2,\ldots,T$, we obtain

$$\sum_{\substack{j=0 \\ j\neq k, l\neq m, t-s=j-l=k-m}}^{t-2}\sum_{k=0}^{t-2}\sum_{l=0}^{s-2}\sum_{m=0}^{s-2}\varrho^{j}\varrho^{k}\varrho^{l}\varrho^{m} = \sum_{\substack{j=0 \\ j\neq k}}^{t-2}\sum_{k=0}^{t-2}\varrho^{2j}\varrho^{2k} \leq \left(\sum_{j=0}^{t-2}\varrho^{2j}\right)^2;$$

and for $s = t-1$, $t = 3, \ldots, T$, we have analogously that,

$$\sum_{\substack{j=0 \\ j \neq k, l \neq m, t-s=j-l=k-m}}^{t-2} \sum_{k=0}^{t-2} \sum_{l=0}^{s-2} \sum_{m=0}^{s-2} \varrho^j \varrho^k \varrho^l \varrho^m \leq \varrho^2 \left(\sum_{j=0}^{t-3} \varrho^{2j} \right)^2$$

while, for $s = t+1$, $t = 2, \ldots, T-1$ (or equivalently $t = s-1$, $s = 3, \ldots, T$), it follows that,

$$\sum_{\substack{j=0 \\ j \neq k, l \neq m, t-s=j-l=k-m}}^{t-2} \sum_{k=0}^{t-2} \sum_{l=0}^{s-2} \sum_{m=0}^{s-2} \varrho^j \varrho^k \varrho^l \varrho^m \leq \varrho^2 \left(\sum_{l=0}^{s-3} \varrho^{2l} \right)^2 .$$

Repeating the discussion for $s = t \pm r$ for $r = 2, \ldots, T-2$, we have

$$\sum_{\substack{j=0 \\ j \neq k, l \neq m, t-s=j-l=k-m}}^{t-2} \sum_{k=0}^{t-2} \sum_{l=0}^{s-2} \sum_{m=0}^{s-2} \varrho^j \varrho^k \varrho^l \varrho^m \leq 2\varrho^{2r} \left(\sum_{j=0}^{t-r-2} \varrho^{2j} \right)^2 ,$$

leading to

$$\sum_{t=2}^{T} \sum_{s=2}^{T} \sum_{\substack{j=0 \\ j \neq k, l \neq m, t-s=j-l=k-m}}^{t-2} \sum_{k=0}^{t-2} \sum_{l=0}^{s-2} \sum_{m=0}^{s-2} \varrho^j \varrho^k \varrho^l \varrho^m \leq \sum_{t=2}^{T} \left(\sum_{j=0}^{t-2} \varrho^{2j} \right)^2 + 2 \sum_{r=1}^{T-2} \varrho^{2r} \sum_{t=2+r}^{T} \left(\sum_{j=0}^{t-r-2} \varrho^{2j} \right)^2 .$$

The same holds when imposing $t - s = j - m = k - l$, such that, with $\sum_{j=0}^{t-r-2} \varrho^{2j} \leq \sum_{j=0}^{T-1} \varrho^{2j}$ and $\sum_{t=2+r}^{T} C \leq CT$, thus, we ultimately have

$$\mathrm{E} \left(\left(\sum_{t=2}^{T} \sum_{j=0}^{t-2} \sum_{\substack{k=0 \\ j \neq k}}^{t-2} \varrho^j \varrho^k \nu_{t-1-j} \nu_{t-1-k} \sigma_{\varepsilon,t}^2 \right)^2 \right) \leq CT^{1+3\eta}$$

and consequently $B_{T2} = o_p\left(T^{1+\eta}\right)$ when $\eta < 1$, as required to complete the proof. ■

Lemma A.3 *Under the assumptions of Theorem 3.2, it follows, as* $T \to \infty$, *that i)* $\frac{\sum_{t=2}^{T} z_{t-1}\varepsilon_t}{\sqrt{\sum_{t=2}^{T} z_{t-1}^2 \varepsilon_t^2}} \xrightarrow{d} \mathcal{N}(0,1)$; *and ii)* $\frac{\sum_{t=2}^{T} z_{t-1}u_t}{\sqrt{\sum_{t=2}^{T} z_{t-1}^2 u_t^2}} \xrightarrow{d} \mathcal{N}(0,1)$.

Lemma A.3 suggests the use of White standard errors in the heteroskedastic near-integrated case, $W.s.e := \frac{\sqrt{\sum_{t=2}^{T} z_{t-1}^2 \hat{\varepsilon}_t^2}}{\sum_{t=2}^{T} z_{t-1}^2}$ with $\hat{\varepsilon}_t$ the OLS residuals guaranteeing $\sup_{2 \leq t \leq T} |\hat{\varepsilon}_t - \varepsilon_t| \xrightarrow{p} 0$ both in cases with and without intercept, and also better finite-sample behavior; see Kostakis et al. (2015). For the stable case, White standard errors are "mandatory" under time heteroskedasticity (Phillips and Xu, 2006).

Proof of Lemma A.3

We first resort to the Phillips-Solo decomposition of v_t and write $v_t = \omega\nu_t + \Delta\tilde{v}_t$ where $\tilde{v}_t$ is a linear process in ν_t with exponentially decaying coefficients. Let also $\bar{z}_t := (1-\varrho L)_+^{-1} v_t$. Thus, denoting $\tilde{z}_t = \sum_{j=0}^{t-1}\varrho^j\nu_{t-j}$ like in Lemma A.2, it follows that,

$$\begin{aligned}\bar{z}_t &= \omega\sum_{j=0}^{t-1}\varrho^j\nu_{t-j} + \left(\tilde{v}_t + (\varrho-1)\sum_{j=1}^{t-1}\varrho^{j-1}\tilde{v}_{t-j} - \varrho^{t-1}\tilde{v}_1\right)\\ &= \omega\tilde{z}_t + d_t,\end{aligned}$$

and it can then easily be shown that $\mathrm{Var}\left(\sum_{j=1}^{t-1}\varrho^{j-1}\tilde{v}_{t-j}\right) \leq CT^{\eta}$ such that d_t is uniformly L_2-bounded given that $\varrho - 1 = -aT^{-\eta}$. Similarly, $T^{-\eta/2}\tilde{z}_t$ is uniformly L_2-bounded itself. We now show that

$$\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\bar{z}_{t-1}^2\varepsilon_t^2 = \frac{\omega^2}{T^{1+\eta}}\sum_{t=2}^{T}\tilde{z}_{t-1}^2\varepsilon_t^2 + o_p(1) \tag{29}$$

and

$$\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\bar{z}_{t-1}\varepsilon_t = \frac{\omega}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}\varepsilon_t + o_p(1)\,. \tag{30}$$

Let us consider first (29). Note that,

$$\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\bar{z}_{t-1}^2\varepsilon_t^2 = \frac{\omega^2}{T^{1+\eta}}\sum_{t=2}^{T}\tilde{z}_{t-1}^2\varepsilon_t^2 + \frac{2\omega}{T^{1+\eta}}\sum_{t=2}^{T}\tilde{z}_{t-1}d_{t-1}\varepsilon_t^2 + \frac{1}{T^{1+\eta}}\sum_{t=2}^{T}d_{t-1}^2\varepsilon_t^2.$$

Since,

$$\mathrm{E}\left(\left|d_{t-1}^2\varepsilon_t^2\right|\right) = \mathrm{E}\left(d_{t-1}^2\right)\mathrm{E}\left(\varepsilon_t^2\right)$$

and

$$\mathrm{E}\left(\left|\tilde{z}_{t-1}d_{t-1}\varepsilon_t^2\right|\right) \leq \sqrt{\mathrm{E}\left(\tilde{z}_{t-1}^2\right)\mathrm{E}\left(d_{t-1}^2\right)}\,\mathrm{E}\left(\varepsilon_t^2\right)$$

due to the independence of ε_t and d_{t-1} and of ε_t and z_{t-1}. With $\mathrm{E}\left(d_{t-1}^2\right)$, $\mathrm{E}\left(\varepsilon_t^2\right)$ and $T^{-\eta}\,\mathrm{E}\left(\tilde{z}_{t-1}^2\right)$ being uniformly bounded, (29) then follows. To establish (30), write

$$\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\bar{z}_{t-1}\varepsilon_t = \frac{\omega}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}\varepsilon_t + \frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}d_{t-1}\varepsilon_t$$

and note that $d_{t-1}\varepsilon_t$ has the md property. Hence, $\sum_{t=2}^{T}d_{t-1}\varepsilon_t = O_p\left(T^{1/2}\right)$ due to the uniform L_2-boundedness and independence of ε_t and d_{t-1}. Thus, from (29) and (30) we obtain that

$$\frac{\sum_{t=2}^{T}\bar{z}_{t-1}\varepsilon_t}{\sqrt{\sum_{t=2}^{T}\bar{z}_{t-1}^2\varepsilon_t^2}} - \frac{\sum_{t=2}^{T}\tilde{z}_{t-1}\varepsilon_t}{\sqrt{\sum_{t=2}^{T}\tilde{z}_{t-1}^2\varepsilon_t^2}} \xrightarrow{p} 0. \tag{31}$$

In a second step we use the same reasoning to show that

$$\frac{\sum_{t=2}^{T}\bar{z}_{t-1}\varepsilon_t}{\sqrt{\sum_{t=2}^{T}\bar{z}_{t-1}^2\varepsilon_t^2}}-\frac{\sum_{t=2}^{T}z_{t-1}\varepsilon_t}{\sqrt{\sum_{t=2}^{T}z_{t-1}^2\varepsilon_t^2}}\xrightarrow{p}0. \tag{32}$$

Write to this end $z_t := \bar{z}_t + r_t$ where $r_t := -(1-\varrho L)_+^{-1}\frac{c_t}{T}x_{t-1}$ with

$$\mathrm{Var}\left(\frac{1}{\sqrt{T}}x_t\right)=\frac{1}{T}\sum_{j=1}^{t}\sum_{k=1}^{t}\left(1-\frac{c_{t-j}}{T}\right)^j\left(1-\frac{c_{t-k}}{T}\right)^k\mathrm{E}\left(v_{t-j}v_{t-k}\right)\leq\frac{1}{T}\sum_{j=1}^{t}\sum_{k=1}^{t}\left|\mathrm{E}\left(v_{t-j}v_{t-k}\right)\right|.$$

Given the uniform L_2-boundedness of the innovations ν_t and the exponential decay of the Wold coefficients of v_t, $\left|\mathrm{E}\left(v_{t-j}v_{t-k}\right)\right|\leq Ce^{|j-k|}$ $\forall t$ and $\frac{1}{\sqrt{T}}x_t$ is easily shown to be uniformly L_2-bounded.

The key in establishing (32) is to note that r_{t-1} is independent of ε_t and uniformly L_2-bounded, and that $T^{-\eta}\,\mathrm{E}\left(z_{t-1}^2\right)$ is uniformly bounded too whenever $T^{-\eta}\,\mathrm{E}\left(\bar{z}_{t-1}^2\right)$ and $\mathrm{E}\left(r_t^2\right)$ are. The arguments employed to show (31) thus apply for z_t and $\bar{z}_t$ as well, and (32) holds.

Summing up, $\frac{\sum_{t=2}^{T}z_{t-1}\varepsilon_t}{\sqrt{\sum_{t=2}^{T}z_{t-1}^2\varepsilon_t^2}}$ and $\frac{\sum_{t=2}^{T}\tilde{z}_{t-1}\varepsilon_t}{\sqrt{\sum_{t=2}^{T}\tilde{z}_{t-1}^2\varepsilon_t^2}}$ are asymptotically equivalent and the result follows from Lemma A.2.

The proof of the result in ii) follows along the same lines and we omit the details.■

Lemma A.4 *Under the assumptions of Theorem 3.2, it holds, as $T\to\infty$, that*

1. $\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}z_{t-1}^2\varepsilon_t^2\xrightarrow{p}\frac{\omega^2}{2a}\int_0^1\sigma_\nu^2(s)\,\sigma_\varepsilon^2(s)\,\mathrm{d}s$;
2. $\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}z_{t-1}^2u_t^2\xrightarrow{p}\frac{\omega^2}{2a}\int_0^1\sigma_\nu^2(s)\,\sigma_u^2(s)\,\mathrm{d}s$ *where* $\sigma_u^2(s)=\sigma_\varepsilon^2(s)+\gamma^2\sigma_\nu^2(s)$;
3. $\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}z_{t-1}x_{t-1}\Rightarrow\frac{\omega^2}{a}\left(X^2(1)-\int_0^1X(s)\,\mathrm{d}X(s)\right)$

where $X(r)$ is an Ornstein-Uhlenbeck process as defined in (14).

Proof of Lemma A.4

1. To obtain the limit of $\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}z_{t-1}^2\varepsilon_t^2$, we use from the proof of Lemma A.3 (see (26)) the fact that

$$\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}z_{t-1}^2\varepsilon_t^2=\omega^2\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\sum_{j=0}^{t-2}\varrho^{2j}\sigma_{\nu,t-1-j}^2\sigma_{\varepsilon,t}^2+o_p(1).$$

The Lipschitz property implies that $\left|\sigma_{\nu,t-1-j}^2-\sigma_{\nu,t}^2\right|\leq C\frac{j}{T}$ such that

$$0\leq\frac{1}{T^{1+\eta}}\left|\sum_{t=2}^{T}\sum_{j=0}^{t-2}\varrho^{2j}\sigma_{\nu,t-1-j}^2\sigma_{\varepsilon,t}^2-\sum_{t=2}^{T}\sigma_{\nu,t}^2\sigma_{\varepsilon,t}^2\sum_{j=0}^{t-2}\varrho^{2j}\right|\leq C\frac{1}{T^{2+\eta}}\sum_{t=2}^{T}\sum_{j=0}^{t-2}j\varrho^{2j}.$$

On the r.h.s. we have immediately, as T$\rightarrow\infty$, that

$$\frac{1}{T^{2+\eta}}\sum_{t=2}^{T}\sum_{j=0}^{t-2}j\varrho^{2j}\rightarrow 0$$

given that $\sum_{j=0}^{t-2}j\varrho^{2j}=\frac{t\varrho^{2(t-3)}(\varrho-1)-\left(\varrho^{2(t-2)}-1\right)}{(\varrho^2-1)^2}$, where $\left|\frac{t\varrho^{2(t-3)}(\varrho-1)}{(\varrho^2-1)^2}\right|\leq CT^{1+\eta}\varrho^{2(t-3)}$ and $\left|\frac{\varrho^{2(t-2)}-1}{(\varrho^2-1)^2}\right|\leq CT^{2\eta}$. We also observe that,

$$\begin{aligned}\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\sigma_{\nu,t}^2\sigma_{\varepsilon,t}^2\sum_{j=0}^{t-2}\varrho^{2j} &= \frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\sigma_{\nu,t}^2\sigma_{\varepsilon,t}^2\frac{T^\eta}{a}\left(\frac{1-\varrho^{2(t-1)}}{1+\varrho}\right)\\ &= \frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\sigma_{\nu,t}^2\sigma_{\varepsilon,t}^2\frac{T^\eta}{a\left(1+\varrho\right)}-\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\sigma_{\nu,t}^2\sigma_{\varepsilon,t}^2\frac{T^\eta}{a}\left(\frac{\varrho^{2(t-1)}}{1+\varrho}\right).\end{aligned}$$

The first summand on the r.h.s. is easily seen to converge to $\frac{1}{2a}\int_0^1\sigma_\nu^2\left(s\right)\sigma_\varepsilon^2\left(s\right)\mathrm{d}s$, while, for the second, we have

$$\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\sigma_{\nu,t}^2\sigma_{\varepsilon,t}^2\frac{T^\eta}{a}\left(\frac{\varrho^{2(t-1)}}{1+\varrho}\right)\leq\frac{C}{aT}\sum_{t=2}^{T}\varrho^{2(t-1)}=O\left(T^{\eta-1}\right)=o\left(1\right)$$

as required to complete the proof.

2. The proof of 2 is analogous to the proof of 1 and is therefore omitted.

3. Let $S_t:=\sum_{j=2}^{t}z_t$. We first follow Breitung and Demetrescu (2015, Proof of Corollary 1.2) and show that

$$\frac{1}{T^{1/2+\eta}}S_t=\frac{1}{a\sqrt{T}}x_t+o_p\left(1\right)$$

where the $o_p\left(1\right)$ term is uniform. The arguments are essentially the same as there; the only difference is having to show that $\mathrm{E}\left(\left|x_t-x_{t-j}\right|\right)\leq C\sqrt{j}$ for all t and j, which is obvious in their i.i.d. setup, but marginally more difficult here. To this end, recall that $\Delta x_t:=v_t-\frac{c_{t-1}}{T}x_{t-1}$and use Liapunov's and Minkowski's inequalities to conclude that,

$$\begin{aligned}\mathrm{E}\left(\left|x_t-x_{t-j}\right|\right) &\leq \sqrt{\mathrm{E}\left(\left(x_t-x_{t-j}\right)^2\right)}=\sqrt{\mathrm{E}\left(\left(\sum_{k=0}^{j-1}v_{t-j}-\frac{1}{T}\sum_{k=0}^{j-1}c_{t-k-1}x_{t-k-1}\right)^2\right)}\\ &\leq \sqrt{\mathrm{E}\left(\left(\sum_{k=0}^{j-1}v_{t-j}\right)^2\right)}+\frac{1}{\sqrt{T}}\sum_{k=0}^{j-1}\left|c_{t-k-1}\right|\sqrt{\mathrm{E}\left(\left(\frac{x_{t-k-1}}{\sqrt{T}}\right)^2\right)};\end{aligned}$$

and therefore using the uniform boundedness of the variance of $\frac{x_{t-k-1}}{\sqrt{T}}$, it follows indeed that $\mathrm{E}\left(\left|x_t-x_{t-j}\right|\right)\leq C\sqrt{j}$ as required.

We then follow Breitung and Demetrescu (2015, Proof of Theorem 2) and obtain via partial summation that,

$$\begin{aligned}\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T} z_{t-1}x_{t-1} &= \frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}\left(S_{t-1}-S_{t-2}\right)x_{t-1} \\ &= \frac{1}{T^{1+\eta}}\left(S_{T-1}x_{T-1}-S_{p-1}x_{p}\right)-\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T} S_{t-2}\Delta x_{t-1}.\end{aligned}$$

Now, since $S_{p-1}x_p = O_p(1)$ it is negligible in the limit; furthermore note that,

$$\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T} S_{t-2}\Delta x_{t-1} = \frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T} S_{t-2}v_{t-1} - \frac{1}{T^{2+\eta}}\sum_{t=p+1}^{T} c_{t-2}S_{t-2}x_{t-2}.$$

For the first summand on the r.h.s., we have using the Phillips-Solo device for the AR process v_{t-1} that,

$$\begin{aligned}\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T} S_{t-2}v_{t-1} &= \frac{\omega}{T^{1+\eta}}\sum_{t=p+1}^{T} S_{t-2}\nu_{t-1} + \frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T} S_{t-2}\Delta\tilde{v}_{t-1} \\ &=: A_T + B_T,\end{aligned}$$

where $\tilde{v}_t$ is a linear process with exponentially decaying coefficients.

Since ν_{t-1} is independent of S_{t-2} and the conditions of Hansen (1992) are fulfilled, we have that,

$$A_T \Rightarrow \frac{\omega^2}{a}\int_0^1 X(s)\,\mathrm{d}M(s).$$

Using the partial summation formula on B_T, it follows that,

$$B_T = \frac{1}{T^{1+\eta}}\left(\tilde{v}_{T-1}S_{T-2}-\tilde{v}_{p-1}S_{p-1}\right) - \frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}\tilde{v}_{t-2}\Delta S_{t-2}.$$

Since $\sup_{1\le t\le T}|S_t| = T^{\eta}\sup_{1\le t\le T}|x_t| + o_p\left(T^{1/2+\eta}\right) = O_p\left(T^{1/2+\eta}\right)$ and $\tilde{v}_{p-1}S_{p-1} = O_p(1)$, it follows that the first summand on the r.h.s. of the above equation is negligible; for the second, we have

$$\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}\tilde{v}_{t-2}\Delta S_{t-2} = \frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}\tilde{v}_{t-2}z_{t-2}.$$

Clearly, $\tilde{v}_{t-2}$ is uniformly L_2-bounded, and it is easily shown that $T^{-\eta/2}z_t$ is uniformly L_2-bounded as well. Then, the Cauchy-Schwarz inequality indicates that $\mathrm{E}\left(|\tilde{v}_{t-2}z_{t-2}|\right) < CT^{\eta/2}$ such that

$$\mathrm{E}\left(\left|\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}\tilde{v}_{t-2}\Delta S_{t-2}\right|\right) \le CT^{-\eta/2}$$

and $\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}\tilde{v}_{t-2}\Delta S_{t-2}$ vanishes in probability.

Hence

$$\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T} z_{t-1}x_{t-1} = \frac{1}{a}\frac{x_{T-1}^{2}}{T} - \frac{1}{a}\left(\frac{a\omega}{T^{1+\eta}}\sum_{t=p+1}^{T} S_{t-2}\nu_{t-1} - \frac{1}{T^{2}}\sum_{t=p+1}^{T} c_{t-2}x_{t-2}^{2}\right) + o_p\left(1\right).$$

Using the weak convergence of S_t and x_t we obtain

$$\begin{aligned}\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T} z_{t-1}x_{t-1} \;&\Rightarrow\; \frac{\omega^2}{a}X^2\left(1\right) - \frac{\omega^2}{a}\left(\int_0^1 X\left(s\right)\mathrm{d}M\left(s\right) - \int_0^1 c\left(s\right)X^2\left(s\right)\mathrm{d}s\right)\\ &\equiv\; \frac{\omega^2}{a}\left(X^2\left(1\right) - \int_0^1 X\left(s\right)\mathrm{d}X\left(s\right)\right).\end{aligned}$$

Note that, interestingly, $\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T} S_{t-2}v_{t-1}$ converges to an Itô-type integral without bias term, unlike $\frac{1}{T^{1}}\sum_{t=p+1}^{T} x_{t-2}v_{t-1}$ under serial correlation. This is because S_t and x_t require different normalizations, which is essentially the expression of the same mechanism ensuring mixed Gaussianity of the unadjusted IVX estimator. ■

Proof of Theorem 3.1

Consider

$$\tilde{\beta}^{ivx} := \frac{\sum_{t=p+1}^{T} z_{t-1}\tilde{y}_t}{\sum_{t=p+1}^{T} z_{t-1}x_{t-1}}. \tag{33}$$

Since $\tilde{y}_t := y_t - \hat{\gamma}\hat{\nu}_t = \beta x_{t-1} + \gamma\nu_t - \hat{\gamma}\hat{\nu}_t + \varepsilon_t$ it follows that we can express $\tilde{\beta}^{ivx}$ as,

$$\tilde{\beta}^{ivx} := \frac{\sum_{t=p+1}^{T} z_{t-1}\tilde{y}_t}{\sum_{t=p+1}^{T} z_{t-1}x_{t-1}} = \beta + \frac{\sum_{t=p+1}^{T} z_{t-1}(\gamma\nu_t - \hat{\gamma}\hat{\nu}_t + \varepsilon_t)}{\sum_{t=p+1}^{T} z_{t-1}x_{t-1}}. \tag{34}$$

Write for the stable autoregression case

$$\hat{\nu}_t := \nu_t - \left(\hat{\boldsymbol{a}} - \boldsymbol{a}\right)'\boldsymbol{x}_{t-p}$$

with $\boldsymbol{x}_{t-p}$ stacking the p lags of x_t and $\boldsymbol{a}$ the corresponding coefficients (of $(1-\rho L)\,A\left(L\right)$), *i.e.* the pure autoregressive representation of x_t.

Then, analyze

$$\begin{aligned} z_{t-1} \;&=\; \sum_{j=0}^{t-3}\varrho^{j}\Delta x_{t-1-j}\\ &=\; x_{t-1} - \varrho^{t-3}x_1 + (\varrho-1)\sum_{j=0}^{t-4}\varrho^{j}x_{t-2-j}.\end{aligned}$$

We have that

$$(\varrho - 1) \sum_{j=0}^{t-4} \varrho^j x_{t-2-j} = -\frac{a}{T^\eta} \sum_{j=0}^{t-4} \varrho^j x_{t-2-j} = -\frac{a}{T^\eta} d_{t-2}$$

where d_{t-2} is here, with x_t a stable autoregression, a mildly integrated process which is known to be $O_p\left(T^{\eta/2}\right)$. Furthermore, $\varrho^{t-3} \to 0$ when t goes to infinity at suitable rates; in the derivations below, the effect will be quantified precisely whenever needed, but it is important to keep in mind that $z_{t-1} \approx x_{t-1}$ which is a stable autoregression.

We thus have for the numerator of $\tilde{\beta}^{ivx} - \beta$ in (34) that,

$$\sum_{t=p+1}^{T} z_{t-1} \left(\varepsilon_t + \gamma \nu_t - \hat{\gamma} \hat{\nu}_t\right) = \sum_{t=p+1}^{T} z_{t-1} \varepsilon_t - \gamma \sum_{t=p+1}^{T} z_{t-1} \left(\hat{\nu}_t - \nu_t\right) - (\hat{\gamma} - \gamma) \sum_{t=p+1}^{T} z_{t-1} \hat{\nu}_t. \quad (35)$$

The first two summands in (35) deliver a normal distribution. This is because

$$\begin{aligned} \frac{1}{T^{1/2}} \sum_{t=p+1}^{T} z_{t-1} \varepsilon_t &= \frac{1}{T^{1/2}} \sum_{t=p+1}^{T} x_{t-1} \varepsilon_t - \frac{a}{T^{1/2+\eta}} \sum_{t=p+1}^{T} d_{t-2} \varepsilon_t + \frac{x_1}{T^{1/2}} \sum_{t=p+1}^{T} \varrho^{t-3} \varepsilon_t \\ &= \frac{1}{T^{1/2}} \sum_{t=p+1}^{T} x_{t-1} \varepsilon_t + o_p(1) \end{aligned}$$

with $\sum_{t=p+1}^{T} d_{t-2} \varepsilon_t = O_p\left(T^{1/2+\eta/2}\right)$ given the results in the proofs of Lemmas A.2 and A.3, and $\sum_{t=p+1}^{T} \varrho^{t-3} \varepsilon_t = O_p\left(T^{\eta/2}\right)$ given that $\text{Var}\left(\sum_{t=p+1}^{T} \varrho^{t-3} \varepsilon_t\right) = O_p\left(\sum_{t=p+1}^{T} \varrho^{2t}\right) = O_p\left(T^{\eta}\right)$. Furthermore,

$$\frac{1}{T^{1/2}} \sum_{t=p+1}^{T} z_{t-1} \left(\hat{\nu}_t - \nu_t\right) = -\left(\frac{1}{T} \sum_{t=p+1}^{T} z_{t-1} \boldsymbol{x}'_{t-p}\right) \sqrt{T} \left(\hat{\boldsymbol{a}} - \boldsymbol{a}\right),$$

where the OLS autoregressive estimators,

$$\sqrt{T} \left(\hat{\boldsymbol{a}} - \boldsymbol{a}\right) = \left(\frac{1}{T} \sum_{t=p+1}^{T} \boldsymbol{x}_{t-p} \boldsymbol{x}'_{t-p}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \boldsymbol{x}_{t-p} \nu_t,$$

following standard arguments can be shown to have a limiting multivariate normal distribution. We now show that $\frac{1}{T} \sum_{t=2}^{T} z_{t-1} \boldsymbol{x}_{t-p}$ does not converge to a vector of zeros, such that the limiting distribution of $\frac{1}{T^{1/2}} \sum_{t=p+1}^{T} z_{t-1} \left(\hat{\nu}_t - \nu_t\right)$ is driven by $\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \boldsymbol{x}_{t-p} \nu_t$. Given that

$$\frac{1}{T} \sum_{t=p+1}^{T} z_{t-1} \boldsymbol{x}_{t-p} = \frac{1}{T} \sum_{t=p+1}^{T} x_{t-1} \boldsymbol{x}_{t-p} - \frac{1}{T} \sum_{t=p+1}^{T} \varrho^{t-3} x_1 \boldsymbol{x}_{t-p} - \frac{a}{T^{1+\eta}} \sum_{t=p+1}^{T} d_{t-2} \boldsymbol{x}_{t-p},$$

the first summand on the r.h.s. gives the desired limit (see Lemma A.1). The second is easily seen to vanish since $\mathrm{E}\left(x_1 x_t\right)$ vanishes at exponential rate (in t). For the third, we show that $\sum_{t=p+1}^{T} d_{t-2}\boldsymbol{x}_{t-p} = O_p\left(T\right)$ as follows. By resorting to the Phillips-Solo device, it is tedious, yet straightforward to show that

$$\frac{1}{T}\sum_{t=p+1}^{T} d_{t-2}\boldsymbol{x}_{t-p} = O_p\left(\frac{1}{T}\sum_{t=p+1}^{T}\tilde{d}_{t-2}\nu_{t-p}\right) \qquad \text{where} \qquad \tilde{d}_{t-2} := \sum_{j=0}^{t-3}\varrho^j\nu_{t-2-j}.$$

Then,

$$\frac{1}{T}\sum_{t=p+1}^{T}\tilde{d}_{t-2}\nu_{t-p} = \frac{1}{T}\sum_{t=p+1}^{T}\tilde{d}_{t-p-1}\nu_{t-p} + O_p\left(1\right),$$

and the proofs of Lemmas A.2 and A.3 provide the arguments leading to $\frac{1}{T}\sum_{t=p+2}^{T}\tilde{d}_{t-p-1}\nu_{t-p} = O_p\left(\frac{T^{1/2+\eta/2}}{T}\right) = O_p\left(1\right)$ as required.

The third summand in (35) is

$$\begin{aligned}\frac{\hat{\gamma}-\gamma}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}\hat{\nu}_t &= \left(\hat{\gamma}-\gamma\right)\left(\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}\nu_t + \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}\left(\hat{\nu}_t-\nu_t\right)\right)\\ &= o_p\left(1\right)\end{aligned}$$

since $\hat{\gamma}$ is easily shown to be consistent for γ, $\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}\nu_t = O_p\left(1\right)$ like in the case of $\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}\varepsilon_t$, and $\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}\left(\hat{\nu}_t-\nu_t\right) = O_p\left(1\right)$ as above. Hence,

$$\begin{aligned}&\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}\left(\varepsilon_t+\gamma\nu_t-\hat{\gamma}\hat{\nu}_t\right)\\ &\quad= \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}\varepsilon_t + \gamma\left(\frac{1}{T}\sum_{t=p+1}^{T} z_{t-1}\boldsymbol{x}_{t-p}'\right)\left(\frac{1}{T}\sum_{t=p+1}^{T}\boldsymbol{x}_{t-p}\boldsymbol{x}_{t-p}'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\boldsymbol{x}_{t-p}\nu_t + o_p\left(1\right).\end{aligned}$$

Furthermore, it is shown along the lines of the discussion of $T^{-1}\sum_{p+1}^{T} z_{t-1}\boldsymbol{x}_{t-p}$ that

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}\varepsilon_t = \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} x_{t-1}\varepsilon_t + o_p\left(1\right).$$

for both $\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}\varepsilon_t$ and $\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\boldsymbol{x}_{t-p}\nu_t$, Theorem 24.3 in Davidson (1994) is easily checked to apply (see Lemma A.1 for the convergence of the sample covariance matrices); since $\boldsymbol{x}_{t-p}\nu_t$ and $z_{t-1}\varepsilon_t$ are orthogonal thanks to the uncorrelatedness of ν_t and ε_t, the term $\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}\left(\varepsilon_t+\gamma\nu_t-\hat{\gamma}\hat{\nu}_t\right)$ is asymptotically normal with mean zero and asymptotic variance

$$\alpha_0\int_0^1\sigma_v^2\left(s\right)\sigma_\varepsilon^2\left(s\right)\mathrm{d}s + \gamma^2\left(\alpha_0\ldots\alpha_{p-1}\right)\Omega^{-1}\left(\alpha_0\ldots\alpha_{p-1}\right)'\int_0^1\sigma_v^4\left(s\right)\mathrm{d}s.$$

Checking that

$$\frac{1}{T}\sum_{t=p+1}^{T} z_{t-1}^2\hat{\varepsilon}_t^2 + \frac{1}{T}\hat{\gamma}^2\hat{Q}_T$$

estimates the above asymptotic variance consistently is straightforward and we omit the details.
■

Proof of Theorem 3.2

Standard OLS algebra shows that the residuals $\hat{\nu}_t$ are numerically the same as in the autoregressive representation of x_t if resorting to the error-correction representation, which is more convenient with near-integration. We may thus write

$$\hat{\nu}_t := \nu_t - \left(\hat{\phi} - \phi\right) x_{t-1} - (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})' \Delta\boldsymbol{x}_{t-p+1}$$

with $\Delta\boldsymbol{x}_{t-p+1}$ stacking the first $p-1$ lags of Δx_t and $\phi := \frac{1}{\omega}(\rho - 1)$ (the vector $\boldsymbol{\alpha}$ depends on all autoregressive coefficients of x_t, but its exact value is irrelevant here).

We have the same representation as in (35), *i.e.*,

$$\sum_{t=p+1}^{T} z_{t-1}\left(\varepsilon_t + \gamma\nu_t - \hat{\gamma}\hat{\nu}_t\right) \; = \; \sum_{t=p+1}^{T} z_{t-1}\varepsilon_t - \gamma\sum_{t=p+1}^{T} z_{t-1}\left(\hat{\nu}_t - \nu_t\right) - \left(\hat{\gamma} - \gamma\right)\sum_{t=p+1}^{T} z_{t-1}\hat{\nu}_t,$$

yet z_t is now a mildly integrated variable. Still, Lemmas A.3 and A.4 show that

$$\frac{1}{T^{1/2+\eta/2}}\sum_{t=p+1}^{T} z_{t-1}\varepsilon_t$$

is asymptotically normal with variance $\omega^2\int_0^1 \sigma_v^2(s)\,\sigma_\varepsilon^2(s)\,\mathrm{d}s$, whereas the remaining term can be re-written as

$$\frac{1}{T^{1/2+\eta/2}}\sum_{t=p+1}^{T} z_{t-1}\left(\hat{\nu}_t - \nu_t\right) = -\frac{1}{T^{1/2+\eta/2}}\sum_{t=p+1}^{T} z_{t-1}x_{t-1}\left(\hat{\phi} - \phi\right) - \frac{1}{T^{1/2+\eta/2}}\sum_{t=p+1}^{T} z_{t-1}\Delta\boldsymbol{x}'_{t-p+1}\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\right).$$

In the limit, this vanishes because $\left(\hat{\phi} - \phi\right)$ is $O_p\left(T^{-1}\right)$ and $\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\right) = O_p\left(T^{-1/2}\right)$ as standard analysis of near-unit root autoregressions shows, while, at the same time,

$$\sum_{t=p+1}^{T} z_{t-1}x_{t-1} = O_p\left(T^{1+\eta}\right)$$

(see Lemma A.4.3) and we only need to show that

$$\sum_{t=p+1}^{T} z_{t-1} \Delta \boldsymbol{x}'_{t-p+1} = O_p(T).$$

This is known to be the case when z_{t-1} is a near-integrated or stationary variable; we discuss here the case where z_t is an IVX instrument. Examining $\sum_{t=p+2}^{T} z_{t-1} \Delta x_{t-1}$ as a representative for the whole vector,

$$\frac{1}{T} \sum_{t=p+1}^{T} z_{t-1} \Delta x_{t-1} = \frac{1}{T} \sum_{t=p+1}^{T} z_{t-1} v_{t-1} + \frac{1}{T^2} \sum_{t=p+1}^{T} c_t z_{t-1} x_{t-2},$$

it is easily shown that both $\frac{z_t}{\sqrt{T}}$ and $\frac{x_t}{\sqrt{T}}$ are uniformly L_2-bounded, hence $\mathrm{E}\left(\frac{1}{T^2} \sum_{t=p+1}^{T} c_t z_{t-1} x_{t-2}\right) = O(1)$. Moreover, $\frac{1}{T} \sum_{t=p+1}^{T} z_{t-1} v_{t-1}$ is itself $O_p(1)$, which can be shown along the lines of the discussion for $\frac{1}{T} \sum q_{t-2} \boldsymbol{x}_{t-p}$ in the proof of Theorem 3.1. ■

Proof of Theorem 3.3

Since the residual effect of ε_t and ν_t is easily checked to be negligible, the correction Q_T is negligible under the local alternative as well and we have for the residual-augmented IVX t-statistic that,

$$\begin{aligned} \tilde{t}_{\beta_1}^{ivx} &= \frac{\sum_{t=p+1}^{T} z_{t-1} (\varepsilon_t + \beta_1 x_{t-1})}{\sqrt{\sum_{t=p+1}^{T} z_{t-1}^2 \varepsilon_t^2}} + o_p(1) \\ &= \frac{\sum_{t=p+1}^{T} z_{t-1} \varepsilon_t}{\sqrt{\sum_{t=p+1}^{T} z_{t-1}^2 \varepsilon_t^2}} + b \frac{\frac{1}{T^{1+\eta}} \sum_{t=p+1}^{T} z_{t-1} x_{t-1}}{\sqrt{\frac{1}{T^{1+\eta}} \sum_{t=p+1}^{T} z_{t-1}^2 \varepsilon_t^2}} + o_p(1). \end{aligned}$$

The first summand on the r.h.s. converges to a standard normal distribution, $\mathcal{Z}$; note that $\mathcal{Z}$ would indeed be independent of the limit process of the regressor x_t since $z_{t-1}\varepsilon_t$ and ν_t are orthogonal. Thus, the result follows with Lemma A.4, items 1 and 3. ■

B Tests for Persistence Change

In this section, for completeness, we provide a brief overview of the persistence change tests of Harvey et al. (2006), which where used to evaluate whether the series under analysis had undergone some persistence change over time.

B.1 The generic persistence change model

We follow Harvey et al. (2006) and Busetti and Taylor (2004) and consider the following data generation process (DGP),

$$
\begin{aligned}
x_t &= d_t'\beta + r_t \\
r_t &= \rho_t r_{t-1} + v_t
\end{aligned}
$$

where $r_0 = 0$, d_t is a set of deterministic variables, such as a constant or, if necessary, a constant and time trend, v_t is taken to satisfy Assumption 3 (together with 2), and ρ_t obeys Assumption 4 in the most general case. For compatibility with the existing literature on testing for changes in persistence we shall assume the variance functions in Assumption 2 to be constant throughout.

Four relevant hypothesis can be considered:

1. $H_1 : x_t$ is $I(1)$ (i.e. nonstationary) throughout the sample period. Harvey et al. (2006) set $\rho_t = 1 - c/T$, $c \geq 0$, so as to allow for unit root and near unit root behaviour.

2. $H_{01} : x_t$ is $I(0)$ changing to $I(1)$ (in other words, stationary changing to nonstationary) at time $[\tau^* T]$; that is $\rho_t = \rho$, $\rho < 1$ for $t \leq [\tau^* T]$ and $\rho_t = 1 - c/T$ for $t > [\tau^* T]$. The change point proportion, τ^*, is assumed to be an unknown point in $\Lambda = [\tau_l, \tau_u]$, an interval in (0,1) which is symmetric around 0.5;

3. $H_{10} : x_t$ is $I(1)$ changing to $I(0)$ (i.e. nonstationary changing to stationary) at time $[\tau^* T]$;

4. $H_0 : x_t$ is $I(0)$ (stationary) throughout the sample period.

B.2 The ratio-based persistence change tests

In the context of no breaks, Kim (2000), Kim et al. (2002) and Busetti and Taylor (2004) introduced tests for the constant $I(0)$ DGP (H_0) against the $I(0) - I(1)$ change (H_{01}) which are based on the ratio statistic,

$$
K_{[\tau T]} = \frac{(T - [\tau T])^{-2} \sum_{t=[\tau T]+1}^{T} \left(\sum_{i=[\tau T]+1}^{t} \tilde{v}_{i\tau} \right)^2}{[\tau T]^{-2} \sum_{t=1}^{[\tau T]} \left(\sum_{i=1}^{t} \hat{v}_{i\tau} \right)^2}
$$

where $\hat{v}_{i\tau}$ is the residual from the OLS regression of x_t on d_t for $t = 1, ..., [\tau T]$ and $\tilde{v}_{i\tau}$ is the OLS residual from the regression of x_t on d_t for $t = [\tau T] + 1, ..., T$.

Since the true change point, τ^*, is assumed unknown Kim (2000), Kim et al. (2002) and Busetti and Taylor (2004) consider three statistics based on the sequence of statistics $\{K_{[\tau T]},$

$\tau \in \Lambda\}$, where $\Lambda = [\tau_l, \tau_u]$ is a compact subset of [0,1], *i.e.*,

$$MS = T_*^{-1} \sum_{s=[\tau_l]}^{[\tau_u]} K_{[sT]}; \tag{36}$$

$$ME = \ln \left\{ T_*^{-1} \sum_{s=[\tau_l]}^{[\tau_u]} \exp\left[\frac{1}{2} K_{[sT]}\right] \right\}; \tag{37}$$

$$MX = \max_{s \in \{[\tau_l],...,[\tau_u]\}} K_{[sT]} \tag{38}$$

where $T_* = [\tau_u] - [\tau_l] + 1$, and τ_l and τ_u correspond to the (arbitrary) lower and upper values assumed for τ^*. Limit results and critical values for the statistics in (36) - (38) can be found in Harvey et al. (2006).

Remark B.1 *The procedure in (36) corresponds to the mean score approach of Hansen (1991), (37) is the mean exponential approach of Andrews and Ploberger (1994) and finally (38) is the maximum Chow approach of Davies (1977); see also Andrews (1993).* □

In order to test H_0 against the I(1) - I(0) (H_{10}) hypothesis, Busetti and Taylor (2004) suggest the sequence of reciprocals of K_t, $t = [\tau_l T], ..., [\tau_u T]$. They define MS^R, ME^R and MX^R as the respective analogues of MS, ME and MX, with $K_{[\tau T]}$ replaced by $K_{[\tau T]}^{-1}$ throughout. Furthermore, to test against an unknown direction of change (that is either a change from I(0) to I(1) or vice versa), they also propose $MS^M = \max\left[MS, MS^R\right]$, $ME^M = \max\left[ME, ME^R\right]$, and $MX^M = \max\left[MX, MX^R\right]$. Thus, tests which reject for large values of MS, ME, and MX can be used to detect H_{01}, tests which reject for large values of MS^R, ME^R and MX^R can be used to detect H_{10}, and MS^M, ME^M, and MX^M can be used to detect either H_{01} or H_{10}.

Harvey et al. (2006) also introduce a set of modified test statistics such that the cdfs of the statistics under the null (H_0) and alternative (H_1) coincide asymptotically at an asymptotic critical value associated with a given significance level.

The first modified tests proposed where $MS_m = \exp(-b_1 J_{1T}) MS$, $ME_m = \exp(-b_2 J_{1T}) ME$ and $MX_m = \exp(-b_3 J_{1T}) MX$, where b_k, $k = 1, 2, 3$ are fixed constants and the modification also makes use of the unit root test proposed by Park (1990), defined as $J_{1,T}$ which consists of T^{-1} times the Wald statistic for testing the joint hypothesis $\gamma_{k+1} = ... = \gamma_9 = 0$ in the regression,

$$x_t = z_t'\beta + \sum_{i=k+1}^{9} \gamma_i t^i + error, \qquad t = 1, ..., T.$$

Note that under H_0, $J_{1,T}$ is $O_p(T^{-1})$ so that $\exp(-b_k J_{1T}) \to 1$, $k = 1, 2, 3$, and therefore MS_m, ME_m and MX_m are simply equivalent to the MS, ME and MX statistics.

The choice of b_k, $k = 1, 2, 3$ ensures that, for a significance level, $100a\%$, the corresponding asymptotic upper-tail critical value of MS_m, ME_m and MX_m under either H_0 or H_1 is identical to the corresponding upper-tail critical values of MS, ME and MX under H_0. These statistics have the same limiting distribution under H_0.

A further variante of modified procedures proposed by Harvey et al. (2006) is obtained by replacing $J_{1,T}$ with $J_{\min} = \min_{\tau\in\Lambda} J_{1,[\tau T]}$, where $J_{1,[\tau T]}$ is T^{-1} times the Wald statistic for testing the joint hypothesis $\gamma_{k+1} = ... = \gamma_9 = 0$ in the regression,

$$x_t = z_t'\beta + \sum_{i=k+1}^{9} \gamma_i t^i + error, \qquad t = 1, ..., [\tau T].$$

Note that also in his case, under H$_0$, $J_{\min}$ is $O_p(T^{-1})$ so that $\exp(-b_k^* J_{\min}) \to 1$, $k = 1, 2, 3$. Therefore, $MS_{m\ \min} = \exp(-b_1^* J_{\min})MS$, $ME_{m\ \min} = \exp(-b_2^* J_{\min})ME$ and $MX_{m\ \min} = \exp(-b_3^* J_{\min})MX$.

The reciprocal versions of these test, MS_m^R, ME_m^R, MX_m^R and $MS_{m\ \min}^R$, $ME_{m\ \min}^R$, $MX_{m\ \min}^R$, are constructed in a similar way, i.e., $MS_m^R = \exp(-b_1 J_{1T})MS^R$, $ME_m^R = \exp(-b_2 J_{1T})ME^R$ and $MX_m^R = \exp(-b_3 J_{1T})MX^R$; as well as $MS_{m\ \min}^R = \exp(-b_1^* J_{\min}^R)MS^R$, $ME_{m\ \min}^R = \exp(-b_2^* J_{\min}^R)ME^R$ and $MX_{m\ \min}^R = \exp(-b_3^* J_{\min}^R)MX^R$, where $J_{\min}^R = \min_{\tau\in\Lambda} J_{[\tau T],T}$ and $J_{[\tau T],T}$ is T^{-1} times the Wald statistic for testing the joint hypothesis $\gamma_{k+1} = ... = \gamma_9 = 0$ in the regression,

$$x_t = z_t'\beta + \sum_{i=k+1}^{9} \gamma_i t^i + error, \qquad t = [\tau T] + 1, ..., T.$$

Finally, the modified tests against an unknown direction of change are simply given as, $MS_m^M = \exp(-b_1 J_{1T})MS^M$, $ME_m^M = \exp(-b_2 J_{1T})ME^M$, and $MX_m^M = \exp(-b_3 J_{1T})MX^M$; as well as $MS_{m\ \min}^M = \exp(-b_1^* \min[J_{\min}, J_{\min}^R])MS^M$, $ME_{m\ \min}^M = \exp(-b_2^* \min[J_{\min}, J_{\min}^R])ME^M$ and $MX_{m\ \min}^M = \exp(-b_3^* \min[J_{\min}, J_{\min}^R])MX^M$.

B.3 Test outcomes

Table 7 gives the test outcomes for the null of constant persistence of the predictors considered in Section 5. We decided upon visual inspection whether a constant or a constant with linear trend is to be modeled as deterministic component d_t. Except for E12, there is serious evidence of time-varying persistence of the examined series.

Table 7: Persistence Change Test Results

	D12$^\tau$	E12$^\tau$	dp$^\tau$	dy$^\tau$	ep$^\tau$	de$^\mu$	b/m$^\tau$	tbl$^\mu$	lty$^\mu$	tms$^\tau$	dfy$^\tau$	dfr$^\tau$	ntis$^\mu$	infl$^\mu$	ltr$^\mu$	svar$^\mu$
MS	9.119*	2.491	3.813*	3.768*	6.713*	5.863*	7.436*	6.098*	4.313	3.179*	7.003*	3.606*	10.967*	5.137*	2.720	41.441*
ME	36.044*	1.595	16.517*	16.212*	98.137*	8.523*	24.970*	17.216*	6.814*	5.126*	25.232*	9.818*	64.374*	12.256*	3.035	158.885*
MX	80.592*	7.037	44.904*	44.289*	208.545*	22.654*	62.046*	41.968*	19.183*	17.032*	59.253*	28.275*	139.791*	32.410*	14.903	329.544*
MS^R	2.635	0.563	2.764	2.874*	2.033	0.791	39.887*	2.095	4.984*	0.781	1.699	1.513	0.515	21.232*	5.099*	0.282
ME^R	2.180*	0.299	9.612*	11.079*	1.476	0.431	200.415*	2.171	9.380*	0.455	1.731	0.885	0.303	38.791*	8.701*	0.156
MX^R	7.520	1.440	29.262*	32.893*	6.062	1.613	413.101*	11.815	27.018*	2.885	7.135	4.809	2.168	86.771*	25.546*	1.037
MS^M	9.119*	2.491	3.813*	3.768*	6.713*	5.863	39.887*	6.098*	4.984	3.179	7.003*	3.606*	10.967*	21.232*	5.099	41.441*
ME^M	36.044*	1.595	16.517*	16.212*	98.137*	8.523*	200.415*	17.216*	9.380*	5.126*	25.232*	9.818*	64.374*	38.791*	8.701*	158.885*
MX^M	80.592*	7.037	44.904*	44.289*	208.545*	22.654	413.101*	41.968*	27.018*	17.032*	59.253*	28.275*	139.791*	86.771*	25.546*	329.544*
MS$_m$	4.063*	2.275	0.399	0.378	3.097*	5.376*	0.308	2.793	0.267	3.003*	4.290*	3.322*	8.239*	4.535	2.705	40.467*
ME$_m$	6.614*	1.318	0.145	0.130	19.376*	7.393*	0.031	4.791	0.072	4.549*	9.026*	8.265*	40.297*	9.992*	3.007	152.813*
MX$_m$	22.079*	6.084	1.210	1.112	60.420*	20.224*	0.377	15.125	0.506	15.549*	27.026*	24.790*	96.194*	27.536*	14.792	319.456*
MS$_m^R$	1.203	0.515	0.310	0.309	0.960	0.726	1.816	0.968	0.318	0.739	1.056	1.397	0.388	18.767*	5.071*	0.276
ME$_m^R$	0.435	0.249	0.107	0.113	0.316	0.374	0.348	0.603	0.098	0.406	0.651	0.751	0.189	31.612*	8.620*	0.150
MX$_m^R$	2.217	1.256	0.968	1.018	1.884	1.441	3.355	4.269	0.719	2.648	3.402	4.247	1.493	73.753*	25.357*	1.005
MS$_m^M$	3.730*	2.253	0.315	0.296	2.854	5.308	1.177	2.491	0.205	2.985	4.073*	3.293	7.901*	18.403*	5.066	40.326*
ME$_m^M$	5.627*	1.294	0.093	0.082	16.599*	7.267	0.133	4.105	0.057	4.498*	8.183*	8.130*	38.079*	30.854*	8.611*	152.096*
MX$_m^M$	19.458*	5.998	0.851	0.776	53.540*	19.850	1.527	12.787	0.391	15.411*	25.033*	24.474*	90.458*	71.772*	25.326*	317.826*
MS$_{m,\min}$	2.641	1.917	2.150	2.051	4.787*	4.341	4.879*	4.631*	0.429	2.957*	5.316*	3.150*	9.967*	4.695*	2.713	41.021*
ME$_{m,\min}$	3.967*	1.000	5.953*	5.491*	53.749*	5.465*	11.790*	11.462*	0.225	4.507*	15.445*	7.716*	55.894*	10.729*	3.023	156.510*
MX$_{m,\min}$	14.103*	4.869	20.055*	18.832*	129.624*	15.597	34.300*	29.823*	1.092	15.385*	40.212*	23.375*	124.152*	28.983*	14.851	325.401*
MS$_{m,\min}^R$	1.423	0.531	1.623	1.737	1.531	0.754	29.706*	1.183	0.985	0.748	1.419	1.396	0.430	21.142*	5.082*	0.279
ME$_{m,\min}^R$	0.737	0.270	3.763*	4.565*	0.896	0.401	119.304*	0.935	0.862	0.421	1.260	0.768	0.232	38.550*	8.657*	0.153
MX$_{m,\min}^R$	3.219	1.330	14.050*	16.439*	4.101	1.519	275.314*	5.821	3.632	2.717	5.565	4.305	1.732	86.316*	25.436*	1.020
MS$_{m,\min}^M$	3.895*	2.299	1.827	1.880	4.537*	5.475	26.554*	4.109	0.514	2.993	5.459*	3.227	9.562*	21.106*	5.079	40.840*
ME$_{m,\min}^M$	8.220*	1.387	4.601*	4.843*	49.690*	7.736*	98.843*	9.855*	0.378	4.616*	16.369*	8.096*	53.036*	38.467*	8.652*	155.636*
MX$_{m,\min}^M$	24.490*	6.290	16.033*	16.729*	120.513*	20.824	233.713*	25.845	1.660	15.654*	41.810*	24.204*	118.130*	86.140*	25.420*	323.680*

Notes: * denotes significance at the 5% significance level. Critical values used taken from Harvey et al., 2006, p.451) for $T = \infty$. The superscripts $^\tau$ and $^\mu$ used on the variables analysed, indicates that the test results are obtained considering either detrended or demaned data, respectively.