Residual-augmented IVX predictive regression*

Matei Demetrescu[a†] and Paulo M.M. Rodrigues[b‡]

[a] Christian-Albrechts-University of Kiel

[b] Banco de Portugal and Nova School of Business and Economics, Universidade Nova de Lisboa

October 24, 2017

Abstract

Bias correction in predictive regressions is known to stabilize the empirical size properties of OLS-based predictability tests. This paper shows that bias correction also improves the local 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). Four 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; 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, and non-Gaussian innovations; and fourth, an application of the new procedure to analyze return and rent growth predictability in 19 OECD countries, the US, OECD and Euro area is also provided.

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

---

*__Acknowledgement:__ We thank Danvee Floro, participants of the SOFIE conference in Hong-Kong in 2016, and seminar participants at the University Carlos III in Madrid, for useful comments and suggestions. We are also grateful to Rita Lourenço for providing us with the data used in the empirical section.

†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, and 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 favored 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 the work of 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); Phillips and Lee (2013) and Lee (2016). 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. Given however that the signal-to-noise ratio in predictive regressions is quite low, one should strive to further 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. Yet, 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), Avramov et al. (2013) 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 fourfold: 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; 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, and non-Gaussian innovations, and to contrast them to the properties of available procedures, such as those of Amihud and Hurvich (2004), Campbell and Yogo (2006) and the IVX approach proposed by Kostakis et al. (2015); and iv) to conduct an in-depth analysis of return and rent growth predictability for 19 OECD countries, the US, OECD and Euro area 16 countries composition (EA16).

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 approach. Section 3 presents the large-sample theory under empirically relevant assumptions, including for instance time-varying unconditional variances. Section 4 discusses the finite sample performance of several procedures used to test for predictability. Section 5 presents the empirical analysis of the predictive power of rent-to-price ratios to predict returns and rent growth, 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 the innovations u_t and v_t 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.*, $T^{-1/2} \sum_{t=1}^{[sT]} (u_t, v_t)' \Rightarrow (\sigma_u W_u(s), \sigma_v W_v(s))'$, 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 $T^{-1/2} x_{[sT]} \Rightarrow B_c(s)$, 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 $\gamma \operatorname{E}(\hat{\rho} - \rho)$, 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) [AH] 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 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, where L is the conventional lag operator; the idea is to choose $\varrho := 1 - aT^{-\eta}$, 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 computed as $se\left(\hat{\beta}^{ivx}\right) := \frac{\hat{\sigma}_u \sqrt{\sum_{t=2}^{T} z_{t-1}^2}}{\sum_{t=2}^{T} z_{t-1} x_{t-1}}$. 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) analyze 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 $\beta := bT^{-(1/2+\eta/2)}$, and show that the limiting distribution under such local alternatives is

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

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 neighborhoods where the IVX based test has nontrivial power. This, however, is the price paid for obtaining a pivotal limiting null distribution. While Kostakis et al. (2015) do show that the power loss is moderate, one would of course prefer to further reduce this loss whenever 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 the IVX approach uses for estimation an instrument that is *less persistent* 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 generalizations 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. The key observation is that, 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. Moreover, 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* $\beta := bT^{-(1/2+\eta/2)}$*. 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 in finite samples 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 $\mathrm{E}(\varepsilon_t^2) = (1 - \rho_{uv}^2)\,\sigma_u^2$, 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.*$\square$

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.* $\square$

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 a 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 the correction factor 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 characterize 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 on $(-\infty, 1]$, 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 on $[0, 1]$.*

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, , *i.a.*, 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 Eicker-White standard errors is moreover recommended by Kostakis et al. (2015) to deal with conditional heteroskedasticity – albeit under strict stationarity of the error series v_t. In fact, examining the proofs in the appendix, it can be seen that one may relax the independence assumption to allow for weakly dependent martingale difference sequences at the cost of additional moment restrictions, however we do not pursue this topic here and leave it for future work.

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 $T \to \infty$; 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 characterizes 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. Since persistence needs not always be constant in practice, 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.

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

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 (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 bias the usual standard errors, even asymptotically. Second, while the estimation error $(\hat{v}_t - v_t)$ has no asymptotic effect on 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 are indeed able to use *the same* statistic with the same limiting distribution to cover both cases without having to decide which is which – analogously to the original IVX test of Kostakis et al. (2015).

In specific, we use

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

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

$$\hat{Q}_T = \mathcal{H}_{z\boldsymbol{x}}\mathcal{H}_{\boldsymbol{xx}}^{-1}\mathcal{H}_{\boldsymbol{xx}v}\mathcal{H}_{\boldsymbol{xx}}^{-1}\mathcal{H}_{z\boldsymbol{x}}$$

with $\mathcal{H}_{\boldsymbol{xx}} = \sum_{t=p+1}^{T}\boldsymbol{x}_{t-p}\boldsymbol{x}_{t-p}'$, $\mathcal{H}_{z\boldsymbol{x}} = \sum_{t=p+1}^{T} z_{t-1}\boldsymbol{x}_{t-p}'$, $\mathcal{H}_{\boldsymbol{xx}v} = \sum_{t=p+1}^{T}\boldsymbol{x}_{t-p}\boldsymbol{x}_{t-p}'\hat{\nu}_t^2$, and $\boldsymbol{x}_{t-p} := (x_{t-1}, \ ..., \ x_{t-p})'$. To compute the Eicker-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 Eicker-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* $\frac{\left(\sum_{t=p+1}^{T} z_{t-1}^2 \tilde{\varepsilon}_t^2\right)^{1/2}}{\sum_{t=p+1}^{T} z_{t-1} x_{t-1}}$ *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 the statistics fare better in finite samples 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 $\hat{\Phi}_j$, $j = 1, \ldots, p$, 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,

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

where

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

and

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

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 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 $\rho := 1 - cT^{-1}$, with $c \geq 0$ and fixed. The main objective of this setting is to acknowledge the uncertainty that researchers 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.*,

$$T^{-1/2} x_{[sT]} \Rightarrow \omega \int_0^s e^{-\int_r^s c(t)\mathrm{d}t} \mathrm{d}M(r) := \omega X(s) \tag{14}$$

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*

$$T^{1/2}\left(\tilde{\beta}^{ivx}-\beta\right) \xrightarrow{d} \mathcal{N}\left(0,\sigma_\beta^2\right) \tag{15}$$

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,

$$T^{1/2} se\left(\tilde{\beta}^{ivx}\right) \xrightarrow{p} \sigma_\beta$$

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.

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

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

and

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

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 $Q_T = o_p\left(T^{1+\eta}\right)$ 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 $O_p\left(T^{1+\eta}\right)$ (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.

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

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

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

Setting $X = B_c$, $M = W$, $\omega^2 = 1$, $\sigma_v(s) = \sigma_v$, $\sigma_\varepsilon(s) = \sigma_\varepsilon$ and $c(s) = c$, and using $B_c \mathrm{d}W_v - cB_c\mathrm{d}s$ as shorthand for $\mathrm{d}B_c$, leads to the results for the particular case studied in Section 2.

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

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

with $a_1 \in \{-0.5, 0.5\}$ and $e_t \sim iid\mathcal{N}(0,1)$. We focus on local alternatives of the form $\beta = bT^{-1}$ 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 $\rho := 1 - cT^{-1}$, with $c \in \{0, 10, 20, 30, 40, 50\}$. The correlation causing endogeneity is set to -0.95, which is not an uncommon value in practice; see, *e.g.*, Campbell and Yogo (2006).

The efficient tests of Campbell and Yogo (2006) (denoted as CY) are analyzed, 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 Eicker-White standard errors as recommended by Kostakis et al. (2015). Later on, we shall also consider a version of the original IVX test without Eicker-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 are computed based on standard normal critical values.

4.2 Empirical size and power performance

From Table 1, which presents the results obtained when v_t follows an AR(1) with $a_1 = -0.5$ (negative autocorrelation) in (24) we observe, 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 behavior. 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 for 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.

As a robustness check, we also provide in the appendix results for positive short-run dynamics, *i.e.*, when $a_1 = 0.5$ (see Table B.1). We observe in general some 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.

4.3 Robustness against empirical features of the data

To evaluate the performance of the procedures under other empirically relevant features 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, $\sigma^2_{\varepsilon}(s) = \sigma^2_{\nu}(s) = 1$;
2. an early upward break, $\sigma^2_{\varepsilon}(s) = \sigma^2_{\nu}(s) = 1 + 8\mathbb{I}(s > 0.3)$;
3. a late upward break, $\sigma^2_{\varepsilon}(s) = \sigma^2_{\nu}(s) = 1 + 8\mathbb{I}(s > 0.7)$;
4. an early downward break, $\sigma^2_{\varepsilon}(s) = \sigma^2_{\nu}(s) = 9 - 8\mathbb{I}(s > 0.3)$; and

Table 1: Empirical rejection frequencies against local alternatives (negative short-run AR parameter)

		AHW	CY	t_{ivx}	$\tilde{t}_{ivx}$	AHW	CY	t_{ivx}	$\tilde{t}_{ivx}$
	b			$T = 200$				$T = 500$	
$c = 0$	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 introduced in this paper, all with maximal lag length $p = [4(T/100)^{0.25}]$. The DGP is as in (1) and (2) with $\rho = 1 - cT^{-1}$ and $\beta = bT^{-1}$. For further details see the text.

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 six 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 Eicker-White heteroskedasticity correction and denote it by $t_{ivx}^{\#}$; t_{ivx} is computed with (the usual) Eicker-White standard errors, and $\tilde{t}_{ivx}$ is computed using the heteroskedasticity-robust standard errors from (12) as before.

Table 2 confirms 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 Eicker-White-type standard errors account for time-varying variances as well.[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.

As second robustness check of the findings Table B.2 in the appendix shows that, 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.

[3]Unreported simulations show that not employing the Eicker-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 2: Empirical rejection frequencies under breaks in variance and persistence, negative short-run AR parameter

		AHW	CY	$t^{\#}_{ivx}$	t_{ivx}	$\tilde{t}_{ivx}$	AHW	CY	$t^{\#}_{ivx}$	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 the Eicker-White correction, and $\tilde{t}_{ivx}$ is the residual-augmented IVX test procedure introduced in this paper, all with maximal lag length $p = [4(T/100)^{0.25}]$. The DGP is as in (1) and (2) with $\rho = 1 - c_t T^{-1}$ and $\beta = bT^{-1}$ and exhibits time-varying variance. For further details see the text.

5 Real estate returns and rent growth predictability

5.1 Background

The real estate market is of considerable economic importance; see, *e.g.*, Englund et al. (2002) and Case et al. (2005). As argued by Helbling and Terrones (2003) and Rapach and Strauss (2006), changes in housing wealth can be more important in their effects on the economy than changes in wealth caused by stock price movements. In effect, some of the most severe systemic financial crises have been associated with boom-bust cycles in real estate markets; see Bordo and Jeanne (2002), Reinhart and Rogoff (2013), and Crowe et al. (2013). Hence, understanding the dynamics of house prices is quite of relevance, not only from an academic perspective.

Price determination in housing markets implies that rents are a fundamental determinant of the housing value and that the rent-to-price ratio (also known as 'cap rate') summarizes market expectations of future housing returns or rent growth (cf. Plazzi et al., 2010, Ghysels et al., 2013 and Engsted and Pedersen, 2015 [EP]). Several recent studies have analyzed the predictive power of the rent-to-price ratio for future housing returns and rent growth in the US housing market (*i.a.*, Gallin, 2008, Plazzi et al., 2010, Cochrane, 2011, and Ghysels et al., 2013), but the European housing market (see, *e.g.*, EP) and the housing market in other countries have received less attention.

The analysis in this section contributes to this literature with further evidence for a large set of OECD countries, the US and the Euro area 16 countries composition (EA16), complementing and consolidating the findings in EP. The framework of analysis follows from the log-linear approximation of one-period gross returns to a housing investment as proposed by Campbell and Shiller (1988), which relates the current log rent-to-price ratio ($r_t - p_t$) to the expected future rate of housing returns and expected future rent growth, such that, the following present value relation is obtained,

$$r_t - p_t = -\frac{\kappa}{1-\rho} + \mathrm{E}_t \sum_{j=0}^{\infty} \rho^j (h_{t+1+j} - \Delta r_{t+1+j}) \tag{25}$$

where h_t, r_t and p_t denote, respectively, the log housing return, the log rent, and the log house price at time t, ρ is the discount factor and κ is a linearization constant; see Campbell et al. (2009), Plazzi et al. (2010), Ghysels et al. (2013) and EP. Note that in (25) the transversality condition (*i.e.*, no-bubble), $\lim_{j\to\infty} \rho^j (r_{t+j} - p_{t+j}) = 0$, is imposed.

Hence, the model in (25) indicates that the rent-to-price ratio is a useful measure of valuation if it can predict future housing returns or rent growth (see, *e.g.*, EP and Gallin, 2008). The intuition behind (25) is that, holding expected housing returns constant, an increase in expected future rents leads to an increase in today's house price which originates a decrease in the rent-to-price ratio. Similarly, holding the expected rent growth constant, an increase in expected future housing returns must imply a lower price today and thereby an increase of the rent-to-price ratio.

The idea of using valuation ratios for prediction is not new. As noted in Ghysels et al. (2013), several ratios have been used in the literature. In particular, for real estate the most commonly used are the rent-to-price ratio (Hamilton and Schwab, 1985, Meese and Wallace, 1997, Geltner and Mei, 1995, Campbell et al., 2009, Himmelberg et al., 2005, Gallin, 2008, Plazzi et al., 2010), the loan-to-value ratio (Lamont and Stein, 1999), and the price-to-income ratio (see, *e.g.*, Malpezzi, 1999).

If expected rent growth and expected housing returns are both stationary, then the rent-to-price ratio should also be stationary, however empirical evidence suggests that this is not necessarily always the case; see Kishor and Morley (2015). Most methods which use ratios as predictors of future returns typically assume stationarity of the ratios, *i.e.*, that the variables used to form the ratios are cointegrated in logs (see, *e.g.*, Campbell et al., 2009, and Plazzi et al., 2010).

For instance, to test for return and rent growth predictability by the rent-to-price ratio in the US housing market, Plazzi et al. (2010) apply a generalized method of moments approach in which they impose the present value restriction. EP on the other hand report the probability of the upper (lower) one-sided alternative, if the estimated predictive coefficient is positive (negative). To conduct inference, they simulate the p-values. If the predictive coefficient is positive (negative) and the null hypothesis is rejected it is concluded that the rent-to-price ratio has positive (negative) predictive power of either housing returns or rent growth depending on the null hypothesis considered. EP's argument for considering this relation through these joint tests is to obtain statistics with better power performance than the usual marginal tests (see also Cochrane, 2008). EP's joint test directly exploits the connection between housing returns, rent growth and the rent-to-price ratio given in (25). In testing the joint hypotheses, EP follow Cochrane (2008), and simulate data under the respective nulls and test the hypotheses of interest using simulated small sample distributions.

In our analysis below we apply the new residual-augmented IVX predictive regression test introduced in Section 2.3.2, and contrast the results obtained with those of the conventional IVX predictive regression test proposed by Kostakis et al. (2015), and the conventional OLS based t-ratio computed with and without Newey-West standard errors (the latter was also used in EP and Ghysels et al., 2013).

5.2 Data

Our analysis focuses on housing returns and rent growth predictability for 19 OECD countries (Australia (AUS), Belgium (BEL), Canada (CAN), Switzerland (CHE), Germany (DEU), Denmark (DNK), Spain (ESP), Finland (FIN), France (FRA) the UK (GBR), Ireland (IRL), Italy (ITA), Korea (KOR), Japan (JPN), the Netherlands (NLD), Norway (NOR), New

Zealand (NZL), Portugal (PRT), Sweden (SWE)) and the US. We also look at overall indexes for the OECD and the EA16. The data consists of seasonally adjusted quarterly real and nominal house prices, and the rent-to-price ratio from 1970Q1 to 2016Q1 for all countries except Australia (begins 1972Q3), Belgium (begins 1976Q2), Korea (begins 1986Q1), Portugal (begins 1988Q1), Spain (begins 1971Q1), Norway (begins 1979Q1), and Sweden (begins 1980Q1). From the real and nominal house prices, inflation in each country is computed, and from the house prices and the rent-to-price ratio real and nominal rent growth as well as housing returns are computed. The latter is computed as $H_{t+1} = \frac{P_{t+1}+R_{t+1}}{P_t}$.

For a clearer understanding of the series under analysis, we computed the averages for the total sample size of each series, as well as for four sub-periods (sub-period I: 1990Q1 - 1999Q4; sub-period II: 2000Q1-2007Q4; sub-period III: 2008Q1-2012Q4; sub-period IV: 2013Q1-2016Q1); see Table 3.

In terms of nominal housing returns, most countries obtained the highest average return rates in sub-period II (2000Q1 - 2007Q4). In specific, in this sub-period the highest returns where observed in Spain (3.1 %), Ireland (2.6%), the United Kingdom (2.5%) and France (2.5%). Overall, during this period 16 countries had nominal average quarterly return rates greater than 1.5%, although two countries displayed negative quarterly housing return rates (Germany (DEU) -0.09% and Japan (JPN) -1.07%). In sub-period I and sub-period IV, most countries present more moderate returns. In sub-periods I and IV only 3 (Ireland (IRL), the Netherlands (NLD) and Portugal (PRT)) and 5 (Australia (AUS), Ireland (IRL), Newzealand (NZL), Sweden (SWE) and the UK (GBR)) of the 20 countries considered displayed returns larger than 1.5% and in both periods 3 countries displayed negative returns (Finland (FIN), Japan (JPN) and Switzerland (CHE) in sub-period I and Finland (FIN), France (FRA) and Italy (ITA) in sub-period IV). Sub-period III (2008Q1 - 2012Q4) registered overall the worst performance (9 of the 20 countries considered as well as the OECD and the EA16 index displayed negative returns over this period). Ireland (IRE) (-3.4%), Spain (ESP) (-1.95%), Denmark (DNK) (-0.9%), Portugal (PRT) (-0.87%) and the Netherlands (NLD) (-0.85%) displayed the lowest return rates. On the other hand, Norway (NOR) (1.2%), Switzerland (CHE)(1%), Canada (CAN) (0.88%) and Germany (DEU) (0.85%) registered the highest returns over this sub-period. The evolution of real returns is qualitatively very similar.

Table 3: Average housing returns, rent growth and rent-to-price ratio.

	AUS	BEL	CAN	DNK	FIN	FRA	DEU	IRL	ITA	JPN	KOR	NLD	NZL	NOR	PRT	ESP	SWE	CHE	GBR	USA	OECD	EA16
Nominal Returns																						
1990-2000	0.875	1.248	0.163	1.244	-0.020	0.261	0.529	2.325	0.750	-0.232	0.093	2.266	1.090	1.337	1.570	1.183	0.632	-0.702	0.832	0.793	0.611	0.793
2000-2007	2.452	1.864	2.014	2.073	1.233	2.524	-0.093	2.594	1.928	-1.069	1.515	1.527	2.433	2.012	0.514	3.122	2.206	0.568	2.537	1.501	1.333	1.570
2008-2012	0.623	0.563	0.882	-0.906	0.612	0.042	0.846	-3.418	-0.467	-0.229	0.652	-0.852	0.175	1.206	-0.867	-1.953	0.744	1.042	-0.160	-0.720	-0.161	-0.189
2012-2016	1.930	0.272	1.216	1.244	-0.003	-0.312	1.272	2.318	-0.865	0.524	0.419	0.267	2.215	0.997	0.745	0.151	2.443	0.577	1.862	1.474	1.091	0.291
Total	**2.028**	**1.396**	**1.613**	**1.452**	**1.462**	**1.487**	**0.699**	**2.022**	**1.904**	**0.666**	**0.841**	**1.307**	**2.202**	**1.745**	**0.904**	**2.230**	**1.695**	**0.824**	**2.209**	**1.228**	**1.357**	**1.431**
Real Returns																						
1990-2000	0.314	0.744	-0.330	0.770	-0.618	-0.057	0.041	1.662	-0.326	-0.415	-1.626	1.694	0.691	0.727	0.203	0.138	-0.210	-1.165	0.092	0.235	0.026	0.120
2000-2007	1.746	1.256	1.594	1.617	0.830	2.026	-0.454	1.776	1.285	-0.877	0.776	0.912	1.958	1.577	-0.341	2.273	1.864	0.333	2.158	0.927	0.860	1.003
2008-2012	-0.017	0.118	0.557	-1.431	0.010	-0.189	0.529	-3.156	-0.925	0.074	-0.093	-1.153	-0.339	0.704	-1.123	-2.416	0.324	1.070	-0.767	-1.135	-0.566	-0.545
2012-2016	1.513	0.091	0.851	1.107	-0.326	-0.336	1.103	1.945	-0.909	0.392	0.240	0.080	2.035	0.397	0.603	0.317	2.200	0.702	1.551	1.250	0.835	0.213
Total	**0.740**	**0.502**	**0.649**	**0.401**	**0.289**	**0.467**	**0.073**	**0.646**	**0.257**	**0.119**	**-0.224**	**0.502**	**0.790**	**0.632**	**-0.075**	**0.588**	**0.498**	**0.242**	**0.892**	**0.352**	**0.379**	**0.330**
Nominal Rents																						
1990-2000	0.568	0.754	0.465	0.687	0.194	0.750	1.132	-0.051	1.433	0.374	1.032	1.048	0.810	0.688	2.011	1.601	1.611	0.785	1.529	0.843	0.907	1.143
2000-2007	0.784	0.532	0.338	0.621	0.714	0.593	0.271	2.924	0.600	-0.056	0.564	0.636	0.215	0.746	0.657	1.037	0.479	0.388	0.701	0.788	0.600	0.628
2008-2012	1.279	0.382	0.343	0.694	0.338	0.404	0.292	-1.202	0.568	-0.094	0.704	0.540	0.495	0.660	0.519	0.425	0.606	0.357	0.616	0.319	0.410	0.395
2012-2016	0.504	0.343	0.311	0.453	0.707	0.219	0.327	0.206	0.055	-0.082	0.628	0.908	0.545	0.664	0.511	-0.124	0.395	0.194	0.621	0.710	0.490	0.230
Total	**1.397**	**0.896**	**0.714**	**1.188**	**1.129**	**1.219**	**0.782**	**1.279**	**1.578**	**0.659**	**0.832**	**1.124**	**1.543**	**1.017**	**1.273**	**1.550**	**1.262**	**0.774**	**1.773**	**1.157**	**1.148**	**1.195**
Real Rents																						
1990-2000	0.007	0.250	-0.028	0.214	-0.404	0.432	0.644	-0.714	0.358	0.191	-0.687	0.477	0.411	0.079	0.644	0.555	0.770	0.321	0.789	0.284	0.321	0.470
2000-2007	0.078	-0.076	-0.082	0.165	0.311	0.095	-0.090	2.105	-0.043	0.136	-0.175	0.021	-0.261	0.310	-0.197	0.188	0.137	0.153	0.321	0.213	0.128	0.061
2008-2012	0.639	-0.063	0.019	0.169	-0.264	0.173	-0.024	-0.939	0.110	0.209	-0.041	0.240	-0.018	0.157	0.262	-0.038	0.185	0.384	0.009	-0.096	0.005	0.040
2012-2016	0.086	0.162	-0.053	0.316	0.385	0.195	0.158	-0.167	0.010	-0.215	0.449	0.721	0.365	0.063	0.369	0.041	0.152	0.319	0.309	0.485	0.234	0.151
Total	**0.122**	**0.166**	**-0.250**	**0.136**	**-0.044**	**0.199**	**0.156**	**-0.097**	**-0.069**	**0.112**	**-0.234**	**0.318**	**0.132**	**0.092**	**0.294**	**-0.093**	**0.362**	**0.193**	**0.455**	**0.282**	**0.170**	**0.094**
Rent_to_price																						
1990-2000	1.706	1.512	1.344	1.701	1.414	1.587	0.799	2.888	1.173	0.648	1.015	1.813	1.898	1.731	1.020	1.632	1.502	0.899	1.795	1.365	1.214	1.213
2000-2007	1.125	1.110	1.101	1.099	1.107	1.191	0.966	1.195	1.118	0.918	1.027	1.028	1.248	1.041	0.986	1.236	1.074	1.021	1.154	1.119	1.074	1.080
2008-2012	0.971	0.863	0.771	1.039	0.954	0.943	0.986	1.726	1.031	1.066	0.837	1.067	0.942	0.808	1.209	1.099	0.796	0.939	1.003	1.239	1.063	1.008
2012-2016	0.924	0.856	0.683	1.089	0.985	1.001	0.843	1.893	1.238	1.023	0.888	1.361	0.818	0.722	1.395	1.450	0.693	0.843	0.959	1.207	1.048	1.067
Total	**1.587**	**1.344**	**1.468**	**1.381**	**1.401**	**1.362**	**0.789**	**2.923**	**1.267**	**0.806**	**0.959**	**1.547**	**1.652**	**1.341**	**1.076**	**2.081**	**1.118**	**0.873**	**1.452**	**1.249**	**1.166**	**1.186**
Rent_to_price (real)																						
1990-2000	1.708	1.716	1.505	1.884	1.583	1.697	0.905	2.608	1.282	0.705	1.207	1.979	2.168	1.914	1.114	1.748	1.693	0.961	2.008	1.536	1.350	1.327
2000-2007	1.126	1.260	1.233	1.217	1.240	1.273	1.094	1.079	1.222	0.999	1.223	1.123	1.426	1.151	1.076	1.324	1.210	1.092	1.291	1.259	1.195	1.182
2008-2012	0.972	0.979	0.863	1.150	1.069	1.009	1.117	1.558	1.127	1.160	0.996	1.165	1.076	0.894	1.320	1.177	0.897	1.004	1.123	1.395	1.183	1.104
2012-2016	0.925	0.971	0.765	1.206	1.103	1.071	0.955	1.710	1.353	1.113	1.057	1.486	0.935	0.799	1.523	1.553	0.781	0.901	1.073	1.358	1.165	1.168
Total	**1.589**	**1.525**	**1.644**	**1.530**	**1.569**	**1.456**	**0.893**	**2.640**	**1.384**	**0.877**	**1.141**	**1.689**	**1.887**	**1.483**	**1.175**	**2.229**	**1.260**	**0.934**	**1.624**	**1.406**	**1.296**	**1.298**

Regarding the nominal rent growth, the number of countries with negative average nominal rent growth is relatively small (only Ireland (IRL) in sub-period I, Japan (JPN) in sub-period II, Ireland (IRL) and Japan (JPN) in sub-period III, and Japan (JPN) and Spain (ESP) in sub-period IV, display average negative rent growth). Taking the OECD average rent growth as benchmark we observe that the average rent growth in EA16 in sub-periods I and II was higher than in the OECD, but lower in sub-periods III and IV. Overall, in sub-periods I, II, III and IV, 8, 10, 12 and 9 of the 20 countries considered, respectively, displayed larger average rent growth than the OECD. The highest values in sub-period I are observed for Portugal (PRT) (2%), Sweden (SWE)(1.6%), Spain (ESP)(1.6%) and the United Kingdom (GBR)(1.5%); in sub-period II, for Ireland (IRE)(2.9%), Spain (ESP)(1%), the US (0.8%), and Australia (AUS)(0.8%); in sub-period III for Australia (AUS)(1.3%), Korea (KOR)(0.7%), Denmark (DNK)(0.7%) and Norway (NOR)(0.7%); and finally in sub-period IV for the Netherlands (NLD)(0.9%), the US (0.7%), Finland (FIN)(0.7%) and Norway (NOR)(0.7%). The evolution of real rents is qualitatively similar.

Hence, given the heterogeneous evolution of rents and house price dynamics across countries in the next section we analyze the predictive power of the rent-to-price ratio to predict these series.

5.3 Testing for predictability

5.3.1 Complete sample analysis

Table 4 presents the findings for nominal housing returns, nominal rent growth and inflation predictability by the rent-to-price ratio. The present value relation in (25) suggests that the rent-to-price ratio should predict returns with a positive sign and rent growth with a negative sign. However, regardless of the test used, the results suggest that this is not always the case (this was also noted by EP).

Regarding nominal housing return predictability the marginal t-tests, t_{OLS} and t_{NW}, suggest that the rent-to-price ratio is a significant predictor in 17 of the 22 countries and economic areas under analysis, whereas the residual-augmented IVX and IVX only find predictability in 12 and 16 of the 22 countries and economic areas considered, respectively. For instance, t_{NW} finds predictability for Italy (ITA), the Netherlands (NLD), OECD and EA16, which is not confirmed by both the residual-augmented IVX and the IVX. On the other hand, the residual-augmented IVX and IVX find predictability for nominal housing returns in Japan (JPN), and Switzerland (CHE) which was not detected by t_{NW}.

Table 4: Predictability of nominal returns, nominal rent growth and inflation. Sample: 1970 - 2016.

	returns				rent growth				Inflation			
	$\tilde{t}_{IVX}$	t_{IVX}	t_{OLS}	t_{NW}	$\tilde{t}_{IVX}$	t_{IVX}	t_{OLS}	t_{NW}	$\tilde{t}_{IVX}$	t_{IVX}	t_{OLS}	t_{NW}
AUS	3.9277***	3.2302***	5.6628***	4.4544***	5.3643***	4.2232***	7.5065***	6.5025***	1.0505	0.6146	1.0609	0.8071
BEL	4.0463 ***	3.4893***	4.2513***	4.3315***	10.464***	6.1540***	5.6817***	6.9821***	1.5256	1.8714*	1.6533*	1.0249
CAN	2.0477**	1.7801*	6.1063***	5.2854***	1.4555	0.9278	7.1563***	4.5841***	0.6043	0.2527	4.2003***	2.8877***
DNK	1.4534	1.9480*	3.2563***	2.2019**	2.8002***	2.1076**	4.0075***	3.5159***	1.2366	0.8328	1.6395	1.1044
FIN	1.867*	2.1114**	4.6749***	4.5125***	1.4283	0.9602	2.9298***	2.4397**	1.3421	1.0936	3.1299***	2.2653**
FRA	1.3778	1.6908*	5.0813***	3.5692***	0.8868	0.6342	3.9175***	3.6501***	-0.5826	-0.7248	1.1151	0.8097
DEU	-3.9911***	-3.0917***	-2.7055**	-1.9690**	-6.8974***	-4.9888***	-7.4356***	-10.4520***	-6.1334***	-5.4224***	-7.2386***	-5.5412***
IRL	2.0808**	2.5885***	7.9874***	6.9395***	-0.2296	-0.5224	0.1716	0.1397	-0.7814	-1.0963	0.8580	0.5982
ITA	-0.6061	-0.6073	3.2715***	1.9693**	-2.2646**	-3.1020***	-1.6117	-1.7397*	-3.8721***	-4.1515***	-7.0302***	-7.2114***
JPN	-3.0279***	-2.8805***	-0.8188	-0.8236	-9.569 ***	-7.6329***	-5.3873***	-6.1726***	-8.0175***	-8.2329***	-6.1746***	-5.7231***
KOR	-1.2629	-1.6044	-1.0539	-0.7484	-4.6463***	-6.1721***	-6.7180***	-3.7339***	1.5949	1.7069*	1.2997	0.9205
NLD	1.5339	2.3149**	4.0122***	4.4249***	1.7738*	1.0922	1.8160**	2.2335**	-2.3900**	-2.4186**	-1.7329*	-1.3064
NZL	3.8034 ***	3.0548***	5.3499***	4.5357***	4.4515***	3.0720***	5.5373***	7.2056***	3.0944***	2.2087**	3.5074***	2.8420***
NOR	2.5964***	1.8511*	3.1292***	2.9667***	5.2489***	2.8476***	4.8848***	3.5437***	1.0594	0.9428	1.8842*	1.7282*
PRT	-1.6618*	-1.7741*	-3.6626***	-2.4250**	-4.9672***	-2.9261***	-6.1860***	-3.6955***	-4.9640***	-5.9560***	-11.4406***	-9.7824***
ESP	2.5417**	2.7177***	10.1658***	7.4108***	2.2766**	2.2639**	11.9506***	10.0029***	-1.4792	-1.8243*	-0.8312	-0.6426
SWE	-0.2771	-0.3886	0.2132	0.1881	1.9098*	1.5422	-0.5011	-0.7642	2.3352**	2.1294**	0.0077	-0.1064
CHE	-2.7634***	-2.1911**	-1.4933	-1.2536	-8.8303***	-9.6615***	-10.0321***	-7.4992***	-9.1415***	-8.7407***	-9.4543***	-7.1974***
GBR	1.566	1.8527*	4.1801***	3.7540***	1.4713	0.7984	2.8170***	3.4401***	0.2001	-0.1449	0.4287	0.2704
USA	-1.3772	-0.7907	-0.0779	-0.0426	-1.4280	-2.8236***	-2.7984***	-1.9670**	-3.5993***	-3.4562***	-3.8353***	-4.1383***
OECD	-0.5785	-0.1029	5.4989***	4.0903***	-1.2246	-1.5299	3.4218***	3.9040***	-2.2683**	-2.8253***	-1.7533*	-1.6592*
EA16	-0.1993	0.1106	5.7127***	4.0081***	-1.4208	-1.7441*	3.0531***	3.0407***	-2.5242**	-3.1315***	-2.1571**	-1.8423*

Note:***, ** and * denote significance at the 1%, 5% and 10% significance level, respectively.

Moreover, considering the results of the residual-augmented IVX which is based on a bias corrected estimator we observe that 12 (55%) of the parameter estimates for the 22 returns series considered present, as suggested by the present value relation in (25) a positive parameter estimate ($\hat{\beta}_h > 0$), of which however only eight are statistically significant. From the results of the residual-augmented IVX we also observe that of the 12 countries and regions for which the rent-to-price ratio has statistically significant predictive power, 4 display a negative $\hat{\beta}_h$. From Table 4 it can further be observed that there seems to be stronger nominal rent growth predictability than nominal housing return predictability which was also noted in EP. Note that based on t_{NW} the only cases for which the rent-to-price ratio is not a statistically significant predictor are Ireland (IRL) and Sweden (SWE). The residual IVX detects predictability in nominal rent growth for 14 of the series of which six display significant negative parameter estimates ($\hat{\beta}_r < 0$).

Overall, the number of significant statistics is greater for the t_{NW} (and t_{OLS}) statistics than for the IVX and residual-augmented IVX procedures when nominal housing returns and rent growth are considered. We note that, in contrast to the t_{NW}, the residual-augmented IVX based test does not find significant results for Denmark (DNK), France (FRA), Italy (ITA), the Netherlands (NLD), the UK (GBR), OECD and EA16 when housing returns are considered and for Canada (CAN), Finland (FIN), France (FRA), the UK (GBR), the US, OECD and EA16 when the rent growth is analyzed. However, the residual-augmented IVX finds significant results for JPN and CHE for housing returns, and for Sweden (SWE) for rent growth, whereas t_{NW} (and t_{OLS}) does not.

Table 5 suggests that the evidence of predictability when real housing returns and real rent growth are considered is considerably weaker than when nominal data is used (Table 4). When real housing returns are used, the residual-augmented IVX test only finds evidence of predictability for Belgium (BEL) and Korea (KOR) (whereas when nominal housing returns where used evidence was found for 12 countries), and when real rent growth is considered for Belgium (BEL), Canada (CAN), Finland (FIN), Korea (KOR), the Netherlands (NLD), Spain (ESP) and the UK (GBR) (whereas for nominal rent growth 14 countries displayed evidence of predictability). According to EP, the present value relation in (25) shows that if the rent-to-price ratio does not predict future housing returns, then it must predict future rent growth. In other words, since the rent-to-price ratio varies over time then either expected housing returns or expected rent growth, or both, must also vary over time, *i.e.*, the null hypothesis cannot consist of having unpredictable housing returns and unpredictable rent growth.

Following EP, we define the log inflation from t to $t+1$ as π_{t+1}, and rewrite (25) as,

$$r_t - p_t = E_t \sum_{j=0}^{\infty} \rho^j \left[(h_{t+1+j} - \pi_{t+1+j}) - (\Delta r_{t+1+j} - \pi_{t+1+j}) \right] - \frac{\kappa}{1-\rho}. \tag{26}$$

Table 5: Predictability of real returns and real rent growth. Sample: 1970 - 2016.

	returns				rent growth			
	$\tilde{t}_{IVX}$	t_{IVX}	t_{OLS}	t_{NW}	$\tilde{t}_{IVX}$	t_{IVX}	t_{OLS}	t_{NW}
AUS	0.7911	0.9513	1.5821	1.2546	-1.561	-1.4470	-2.6721***	-2.2998**
BEL	1.7735*	2.0288	2.8593	2.9338***	2.8056***	2.7458***	2.6576***	2.2002**
CAN	0.5747	1.2510	3.2879***	2.9627***	-1.6606*	-1.8218	-6.9834***	-5.3039***
DNK	0.6780	1.6221	2.3002**	1.7023*	0.9597	0.8664	0.4741	0.6414
FIN	-0.2343	0.2671	1.0822	1.0481	-2.0634**	-2.2739**	-3.1675***	-2.9397***
FRA	1.1579	1.7585	3.4674***	2.6664***	0.5434	0.1995	-0.1390	-0.1500
DEU	1.2653	1.2358	3.2327***	2.5844***	0.4760	-0.0979	-0.7236	0.7376
IRL	0.9648	1.6866*	4.7997***	4.7689***	-0.5763	-1.2168	-2.2951**	-1.8346**
ITA	0.0833	0.6962	3.0703***	1.8902*	-0.1787	-0.7660	-3.4858***	-2.9067***
JPN	0.1097	0.7145	1.8381*	1.4868	0.6352	0.9974	0.4911	0.3814
KOR	-1.8803*	-1.2095	-0.4742	-0.3485	-4.2956***	-4.9650***	-4.8700	-3.4732***
NLD	1.2947	2.5492**	3.9431***	4.6140***	2.3185**	1.2580	1.2099	1.6394
NZL	1.2225	1.0919	1.7428*	1.5582	0.5131	0.2618	0.4133	0.5771
NOR	0.6207	0.9665	1.6055	1.5288	-0.2195	-0.3641	-0.5465	-0.8595
PRT	0.0222	0.2859	0.4312	0.0811	-0.0424	-0.7013	-1.8154*	-1.2596
ESP	0.7196	1.4404	5.2743***	4.2027***	-2.4588**	-2.3168**	-5.7807***	-5.0957***
SWE	-1.4855	-1.0309	0.6229	0.5793	0.8705	0.4172	-0.2738	-0.5994
CHE	-1.2272	0.6277	1.3494	1.0314	-1.1354	-1.4908	-1.7135	-1.3937
GBR	1.2671	1.8526*	2.9444***	2.9649***	1.9019*	1.0496	1.4450	2.0741**
USA	-0.3000	1.2267	1.9475*	1.2960	0.7087	-0.4773	-0.4653	-0.3911
OECD	0.6953	1.5722	4.2769***	3.1460***	0.8471	0.6648	0.4995	0.6320
EA16	0.7073	1.3910	4.0927***	3.3997***	0.4153	-0.0883	-2.1022**	-1.4812

Note: ***, ** and * denote significance at the 1%, 5% and 10% significance level, respectively.

The representation in (26) suggests that if the rent-to-price ratio predicts nominal and real housing returns or rent growth differently, then it must be due to the rent-to-price ratio having predictive power for future inflation. For example, if nominal housing returns are predictable but inflation is not, then the result must be that real housing returns are predictable. From Table 5 we note that the evidence of predictability found with the nominal returns for Germany (DEU), Japan (JPN), the Netherlands (NZL), Portugal (PRT) and Switzerland (CHE) is not observed when real returns are used, but significant results are observed when analyzing predictability of inflation for these countries. As also indicated in EP from the present value relation in (25) it is clear that a negative return coefficient using nominal data can only turn positive if the inflation coefficient is negative and numerically larger than the return coefficient. Note that the rent-to-price ratio predicts inflation with a negative sign for Germany (DEU), Japan (JPN), Portugal (PRT) and Switzerland (CHE), which displayed negative return coefficients when nominal data was used but are not significant for real housing returns.

Table 6: Predictability of nominal returns, nominal rent growth and inflation. Sub-period 1970 - 2007.

	Returns				Rents				Inflation			
	$\tilde{t}_{IVX}$	t_{IVX}	t_{OLS}	t_{NW}	$\tilde{t}_{IVX}$	t_{IVX}	t_{OLS}	t_{NW}	$\tilde{t}_{IVX}$	t_{IVX}	t_{OLS}	t_{NW}
AUS	2.4384**	1.9566*	3.2466***	2.8141***	4.5257***	4.0785***	7.0468***	6.7236***	0.7427	0.5966	1.2713	1.0822
BEL	0.4928	-0.0622	0.3357	0.2555	5.1867***	3.5230***	2.1108**	2.2614**	1.9685**	1.5817	1.6306	1.4154
CAN	1.1114	1.1531	4.9224***	4.1091***	-0.6238	-0.7294	3.4496***	2.1917**	2.2029**	2.2614**	10.4784***	7.3114***
DNK	0.0038	0.1914	0.6554	0.3172	1.1621	1.0955	2.1902**	1.8392*	0.4278	0.4168	1.5175	1.1581
FIN	0.8772	1.2433	2.6934***	2.2328**	0.8537	0.3356	1.7369*	1.4696	3.9933***	3.5757***	5.3996***	4.4630***
FRA	-2.9666***	-2.6505***	-1.8165*	-1.5912	-1.9308*	-2.0666**	-0.6955	-0.7270	5.6488***	3.3326***	5.9388***	6.7835***
DEU	-8.8727***	-5.5330***	-5.9832***	-4.8704***	-9.5014***	-5.0172***	-5.9061***	-10.6130***	-4.8153***	-3.3362***	-3.6887***	-2.8935***
IRL	3.0601***	2.9215***	6.3611***	7.5837***	-0.5824	-0.6287	-0.8457	-0.6866	3.4425***	4.7432***	8.6092***	5.2424***
ITA	-0.4019	-0.3026	2.1538**	1.4524	-2.1832**	-2.9655***	-2.5542**	-2.7908***	-0.1329	-0.2336	4.3495***	3.3610***
JPN	-2.9458***	-2.5845***	0.5232	0.4175	-4.8269***	-4.0811***	-0.6669	-0.3870	-2.0087**	-1.6024	2.2316**	1.8960*
KOR	-1.9792**	-2.3536**	-1.7973*	-1.2205	-6.1556***	-7.3703***	-8.9731***	-5.6375***	-1.1315	-1.1318	-2.0313**	-1.5855
NLD	0.8091	1.0725	1.8862*	1.8949*	0.8671	0.6778	0.9075	1.0555	4.3727***	5.2745***	5.9511***	4.1685***
NZL	0.6743	0.5759	1.9251*	1.8672*	1.9062*	1.4107	3.5214***	6.1470***	2.2909**	1.6307	3.8395***	5.3995***
NOR	0.1984	0.4168	0.9589	0.8196	2.5583**	1.4842	2.8576***	2.2234**	-0.0499	0.0871	0.0603	0.3521
PRT	-2.6493***	-2.6229***	-5.5354***	-4.1292***	-4.1219***	-3.3788***	-7.7343***	-5.3630***	4.9534***	3.3149***	2.6359***	2.7662***
ESP	2.8871***	2.4682**	6.4131***	5.0254***	2.9774***	2.1097**	7.5457	5.4946***	3.6399***	4.0415***	9.3903***	6.5672***
SWE	-1.7019*	-1.4978	-0.6891	-0.8601	-0.3252	-0.4229	-3.8652***	-3.4293***	0.7631	0.5142	-2.0401**	-1.9807**
CHE	-2.7980***	-2.4074**	-1.5221	-1.4053	-9.0914***	-9.4067***	-9.1105***	-7.3020***	-4.6785***	-5.1967***	-4.5029***	-2.9179***
GBR	-0.6064	-0.1397	1.0196	0.9397	-0.2935	-0.7744	0.2079	0.0858	-0.4012	-0.2234	1.3095	0.8659
USA	-2.0437**	-2.6004***	-2.0517**	-1.4489	-2.0177**	-3.1518***	-3.3645***	-2.1559**	1.3462	1.2713	2.3056**	2.0630**
OECD	-1.5313	-1.1592	2.7791***	2.3366**	-1.9731**	-2.4928**	0.4871	0.2770	0.9914	1.0775	7.0549***	5.5041***
EA16	-0.8979	-0.6702	2.3133**	1.7552*	-2.2271**	-2.6926***	-0.2803	-0.4003	2.6773***	2.5643**	9.6049***	7.9521***

5.3.2 Sub-period analysis (1970 - 2007)

Considering the strong impact of the financial crisis of 2008 on the real estate market, we repeat the previous analysis considering only the period before the crisis (1970 - 2007).

The first observation we can make from the results in Table 6 is that in comparison to the full sample, the t_{NW} statistic detects less evidence of nominal housing return and rent-growth predictability than when the complete sample was considered, *i.e.*, only 11 statistics display significant results for the former and 14 for the latter, which are in line with the total number of significant statistics put forward by the residual-augmented IVX test.

Although the number of significant statistics based on the residual augmented IVX and t_{NW} is very similar, the residual-augmented IVX in contrast to t_{NW} does not find significant results for Canada (CAN), Finland (FIN), the Netherlands (NLD), New Zealand (NZL), OECD and EA16 when housing returns are considered and for Canada (CAN), Finland (FIN) and Sweden (SWE) when rent growth is analyzed. On the other hand, the residual-augmented IVX finds significant results for France (FRA), Japan (JPN), Korea (KOR), Sweden (SWE), Switzerland (CHE) and the US for housing returns, and for France (FRA), Japan (JPN), OECD and EA16 for rent growth whereas t_{NW} (and t_{OLS}) does not.

Table 7: Predictability of real returns and real rent growth. Sub-period 1970 - 2007.

	Returns $\tilde{t}_{IVX}$	t_{IVX}	t_{OLS}	t_{NW}	Rents $\tilde{t}_{IVX}$	t_{IVX}	t_{OLS}	t_{NW}
AUS	-0.1648	0.0646	-0.1155	-0.0353	-0.3454	-0.3201	-0.8419	-0.6171
BEL	-0.3413	-0.4619	0.3288	0.2109	2.2845**	2.2364**	1.9961**	1.5627
CAN	0.4576	1.1383	2.9726***	2.7215***	-1.1039	-1.0246	-6.0812***	-4.3530***
DNK	-0.3698	0.3620	0.4770	0.1916	1.4994	1.5787	1.4596	1.7050*
FIN	-0.9877	-0.3764	-0.1881	-0.2140	-2.4519**	-2.4456**	-3.0393***	-3.1143***
FRA	-1.8994*	-1.3011	-1.5917	-1.4276	0.4515	0.0118	-0.5969	-1.2831
DEU	0.3333	0.0803	0.0319	0.1076	0.4121	0.3801	-0.1261	-0.0127
IRL	1.3691	1.8366*	3.3144***	4.3291***	-0.8742	-1.3416	-2.9628***	-2.3736**
ITA	0.2411	0.9340	2.3574**	1.6846*	-0.0927	-0.7123	-3.0159***	-2.8121***
JPN	-1.0191	-0.3752	1.5750	1.2277	0.8227	1.4813	1.2419	1.2566
KOR	-1.8152*	-1.3650	-0.3335	-0.2104	-3.0210***	-3.7473***	-3.5332***	-2.8331***
NLD	0.7700	1.4804	2.2131**	2.3571**	1.9049*	1.3150	1.4055	1.9273*
NZL	-0.5105	-0.3574	-0.5248	-0.9812	0.2294	0.0015	0.0912	0.0005
NOR	-0.9665	-0.0959	-0.0662	-0.3237	-0.3346	-0.4867	-0.8177	-1.2210
PRT	-1.0108	-0.6515	-1.4601	-1.1480	-0.7704	-1.4571	-3.8569***	-2.5731**
ESP	0.6779	1.1517	2.4872**	2.0923**	-2.5764**	-2.5808***	-6.4541***	-6.1274***
SWE	-1.6640*	-1.0824	1.4656	1.0425	0.0000	-0.3560	-1.4641	-1.6336
CHE	-1.0098	0.2809	0.8878	0.7650	-1.2002	-1.4405	-2.0173**	-1.7267*
GBR	0.0970	0.8205	1.1417	1.2194	1.4868	0.5646	0.4226	0.5342
USA	-0.4992	0.2975	0.8867	0.2145	0.1602	-0.4547	-0.6641	-0.7987
OECD	0.4735	1.2329	2.8696***	2.2966**	0.8879	0.6737	-0.4660	-0.8132
EA16	0.4565	1.0240	1.8370*	1.5263	0.4911	-0.0043	-2.4985**	-2.2401**

A considerable decrease in the number of significant statistics is also observed when real housing returns and real rent growth is considered (see Table 7). Note that the residual-augmented IVX only detects predictability for France (FRA), Korea (KOR) and Sweden (SWE), whereas t_{NW} finds predictability for Canada (CAN), Ireland (IRL), the Netherlands (NLD), Spain (ESP) and OECD. For rent-growth there is a small increase in the number of significant statistics, the residual-augmented IVX finds predictability for Belgium (BEL), Finland (FIN), Korea (KOR), and Spain (ESP), whereas t_{NW} finds evidence for Canada (CAN), Finland (FIN), Ireland (IRL), Italy (ITA), Korea (KOR), Portugal (PRT), Spain (ESP), Switzerland (CHE) and EA16.

5.4 Discussion

From the results of the predictability tests performed in the previous section we conclude, using the four competing estimators as well as a sample split, that five countries actually emerge for which the rent-to-price ratio is relentless in predicting housing returns and rent growth in nominal terms. In specific one can summarize the results as follows:

a) The findings reveal that the rent-to-price ratio (in nominal terms) is a dominant and stable predictor for both future housing returns and rent growth for Australia, Germany, Portugal, Spain and Japan. For these countries, the predictive ability of the rent-to-price ratio consistently predicts regardless of the estimator employed and, except for Japan, is robust to the sub-sample split considered.

b) Of these five countries where the rent-to-price ratio is found to be stable in terms of predictive power and the sign of the coefficient, three countries (Germany, Japan and Portugal) are shown to have a negative predictive relationship with future housing returns. That is, using robust inference methods, the pattern that stands out is that rent-to-price ratios negatively predict housing returns: a divergence from the dynamic Gordon Growth model, as stated by EP (2015).

c) Interestingly, rent growth in Belgium turns out to be the most predictable variable of all. All four competing estimators predict future rent in both nominal and real terms and this result holds in the sub-sample analysis. Rent growth is also consistently predictable in Spain and Korea, but the predictive coefficient of rent growth in Spain changes signs when shifting the analysis from nominal to real terms.

6 Concluding remarks

This paper introduced a new IVX-based statistic computed from a residual augmented predictive regression motivated by Amihud and Hurvich (2004), and reexamined the empirical evidence on returns and rent growth predictability using these new robust methods.

The residual-augmented IVX variant 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 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 analysis of OECD housing price data showed that the bias-adjusted IVX procedure detected predictability more often than the original IVX procedure, but less often than non-robust procedures. Overall, this analysis reveals, among other things, that the rent-to-price ratio (in nominal terms) is a useful predictor for both future housing returns and rent growth for Australia, Germany, Portugal, Spain and Japan. We leave to further work the check of whether adding other putative predictors (such as, among others, disposable income, mortgage rates, unemployment, investment in housing and short-term interest rates) strengthens the evidence on predictability.

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.

Avramov, D., L. Barras, and R. Kosowski (2013). Hedge fund return predictability under the magnifying glass. *Journal of Financial and Quantitative Analysis 48*(4), 1057–1083.

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

Bordo, M. D. and O. Jeanne (2002). Boom-busts in asset prices, economic instability, and monetary policy. Working paper, National Bureau of Economic Research, Cambridge, MA, NBER (8966).

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

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, D., D. Morris, J. Gallin, and R. Martin (2009). What moves housing markets: a variance decomposition of the rent-price ratio. *Journal of Urban Economics 66*(2), 90–102.

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 R. J. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies 1*(3), 195–228.

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.

Case, K. E., J. M. Quigley, and R. J. Shiller (2005). Comparing wealth effects: the stock market versus the housing market. *The B.E. Journal of Macroeconomics 5*(1), article 1.

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.

Cochrane, J. H. (2008). The dog that did not bark. A defence of return predictability. *Review of Financial Studies 21*(4), 1533–1575.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance 66*(4), 1047–1108.

Crowe, C., G. Dell'Ariccia, D. Igan, and P. Rabanal (2013). How to deal with real estate booms: Lessons from country experiences. *Journal of Financial Stability 9*(3), 300–319.

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

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.

Englund, P., M. Hwang, and J. M. Quigley (2002). Hedging housing risk. *Journal of Real Estate Finance and Economics 24*(1), 167–200.

Engsted, T. and T. Q. Pedersen (2015). Predicting returns and rent growth in the housing market using the rent-price ratio: Evidence from the OECD countries. *Journal of International Money and Finance 53*, 257–275.

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

Gallin, J. (2008). The long-run relationship between house prices and rents. *Real Estate Economics 36*(4), 635–658.

Geltner, D. and J. Mei (1995). The present value model with time-varying discount rates: Implications for commercial property valuation and investment decisions. *Journal of Real Estate Finance and Economics 11*(2), 119–135.

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.

Ghysels, E., A. Plazzi, W. Torous, and R. Valkanov (2013). Forecasting real estate prices. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Vol.2*, pp. 509–580.

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

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.

Hamilton, B. W. and R. M. Schwab (1985). Expected appreciation in urban housing markets. *Journal of Urban Economics 18*(1), 103–118.

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

Helbling, T. and M. Terrones (2003). Real and financial effects of bursting asset price bubbles. In *IMF World Economic Outlook, Chapter II, April 2003*.

Himmelberg, C., C. Mayer, and T. Sinai (2005). Assessing high house prices: Bubbles, fundamentals, and misperceptions. *Journal of Economic Perspectives 19*, 67–92.

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.

Kishor, N. and J. Morley (2015). What factors drive the price-rent ratio for the housing market? a modified present-value analysis. *Journal of Economic Dynamics $ Control 58*, 235–249.

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

Lamont, O. and J. Stein (1999). Leverage and house-price dynamics in U.S. cities. *RAND Journal of Economics 30*(3), 498–514.

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

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.

Malpezzi, S. (1999). A simple error correction model of house prices. *Journal of Housing Economics 8*(1), 27–62.

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.

Meese, R. A. and N. E. Wallace (1997). The construction of residential housing price indices: A comparison of repeat-sales, hedonic-regression, and hybrid approaches. *The Journal of Real Estate Finance and Economics 14*(1), 51–73.

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.

Plazzi, A., W. Torous, and R. Valkanov (2010). Expected returns and expected growth in rents of commercial real estate. *Review of Financial Studies 23*(9), 3469–3519.

Rapach, D. and J. K. Strauss (2006). The long-run relationship between consumption and housing wealth in the eight district states. *Federal Reserve Bank of St Louis, Regional Economic Development 2*(2), 140–147.

Reinhart, C. M. and K. S. Rogoff (2013). Banking crises: An equal opportunity menace. *Journal of Banking and Finance 37*(11), 4557–4573.

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.

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. $\frac{1}{T}\sum_{t=p+1}^{T} x_{t-1}\boldsymbol{x}_{t-p} \xrightarrow{p} \boldsymbol{\alpha}_p' \int_0^1 \sigma_v^2(s)\,\mathrm{d}s$, *where* $\boldsymbol{\alpha}_p := (\alpha_0, \ldots, \alpha_{p-1})$, $\boldsymbol{x}_{t-p} := (x_{t-1}, \ \ldots, \ x_{t-p})'$ *and α_h is as defined in Theorem 3.1;*
2. $\frac{1}{T}\sum_{t=p+1}^{T} \boldsymbol{x}_{t-p}\boldsymbol{x}_{t-p}' \xrightarrow{p} \boldsymbol{\Omega} \int_0^1 \sigma_v^2(s)\,\mathrm{d}s$, *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 $T^{-1}\sum_{t=h+1}^{T} x_t x_{t-h} \xrightarrow{p} \alpha_h \int_0^1 \sigma_v^2(s)\,\mathrm{d}s$, $h = 0, 1, \ldots, p-1$; 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*

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

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,

$$\frac{\sum_{t=2}^{T} \tilde{z}_{t-1}\varepsilon_t}{\left(\sum_{t=2}^{T} \tilde{z}_{t-1}^2\varepsilon_t^2\right)^{\frac{1}{2}}} = \frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\frac{\tilde{z}_{t-1}\varepsilon_t}{s_T}\left(\frac{\sum_{t=2}^{T}\sum_{j=0}^{t-2}\varrho^{2j}\sigma_{\nu,t-1-j}^2\sigma_{\varepsilon,t}^2}{\sum_{t=2}^{T}\tilde{z}_{t-1}^2\varepsilon_t^2}\right)^{\frac{1}{2}}, \tag{A.1}$$

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

$$\left(\frac{\sum_{t=2}^{T} \sum_{j=0}^{t-2} \varrho^{2j} \sigma_{\nu,t-1-j}^2 \sigma_{\varepsilon,t}^2}{\sum_{t=2}^{T} \tilde{z}_{t-1}^2 \varepsilon_t^2} \right)^{\frac{1}{2}} \xrightarrow{p} 1, \tag{A.2}$$

hence the result in (A.1) will 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

$$\mathrm{E}\left(\frac{\tilde{z}_{t-1}^4}{T^{2\eta}} \right) = \frac{1}{T^{2\eta}} \sum_{j=0}^{t-2} \sum_{k=0}^{t-2} \sum_{l=0}^{t-2} \sum_{m=0}^{t-2} \varrho^j \varrho^k \varrho^l \varrho^m \, \mathrm{E}\left(\nu_{t-j}\nu_{t-k}\nu_{t-l}\nu_{t-m} \right). \tag{A.3}$$

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 (A.3) 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,

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

which suffices for the required uniform L_4-boundedness.

To check condition ii), it suffices to show that

$$\frac{1}{T^{1+\eta}} \sum_{t=2}^{T} \tilde{z}_{t-1}^2 \varepsilon_t^2 - s_T^2 \xrightarrow{p} 0 \tag{A.4}$$

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 (A.4), 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,

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

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,

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

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

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

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

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

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

For B_{T2} we proceed similarly,

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

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 $t=s$, $t=2,\dots,T$, we obtain

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

and for $s=t-1$, $t=3,\dots,T$, 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 $s=t+1$, $t=2,\dots,T-1$ (or equivalently $t=s-1$, $s=3,\dots,T$), 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 $r=2,\dots,T-2$, 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) $\frac{\sum_{t=2}^{T} z_{t-1}\varepsilon_t}{\left(\sum_{t=2}^{T} z_{t-1}^2\varepsilon_t^2\right)^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)$; *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 Eicker-White standard errors in the heteroskedastic near-integrated case, $W.s.e := \frac{\left(\sum_{t=2}^{T} z_{t-1}^2\hat{\varepsilon}_t^2\right)^{1/2}}{\sum_{t=2}^{T} z_{t-1}^2}$ 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, Eicker-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

$$\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\bar{z}_{t-1}^2\varepsilon_t^2 = \frac{\omega^2}{T^{1+\eta}}\sum_{t=2}^{T}\tilde{z}_{t-1}^2\varepsilon_t^2 + o_p(1) \tag{A.5}$$

and

$$\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\bar{z}_{t-1}\varepsilon_t = \frac{\omega}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}\varepsilon_t + o_p(1). \tag{A.6}$$

Let us consider first (A.5). 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
$$\mathrm{E}\left(\left|\tilde{z}_{t-1}d_{t-1}\varepsilon_t^2\right|\right) \leq \left(\mathrm{E}\left(\tilde{z}_{t-1}^2\right)\mathrm{E}\left(d_{t-1}^2\right)\right)^{1/2}\mathrm{E}\left(\varepsilon_t^2\right)$$
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, (A.5) then follows. To establish (A.6), 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 (A.5) and (A.6) we obtain that

$$\frac{\sum_{t=2}^{T}\bar{z}_{t-1}\varepsilon_t}{\left(\sum_{t=2}^{T}\bar{z}_{t-1}^2\varepsilon_t^2\right)^{1/2}} - \frac{\sum_{t=2}^{T}\tilde{z}_{t-1}\varepsilon_t}{\left(\sum_{t=2}^{T}\tilde{z}_{t-1}^2\varepsilon_t^2\right)^{1/2}} \xrightarrow{p} 0. \tag{A.7}$$

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

$$\frac{\sum_{t=2}^{T}\bar{z}_{t-1}\varepsilon_t}{\left(\sum_{t=2}^{T}\bar{z}_{t-1}^2\varepsilon_t^2\right)^{1/2}} - \frac{\sum_{t=2}^{T}z_{t-1}\varepsilon_t}{\left(\sum_{t=2}^{T}z_{t-1}^2\varepsilon_t^2\right)^{1/2}} \xrightarrow{p} 0. \tag{A.8}$$

Write to this end $z_t := \bar{z}_t + r_t$ where $r_t := -\left(1-\varrho L\right)_+^{-1}\frac{c_t}{T}x_{t-1}$ 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 (A.8) 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 (A.7) thus apply for z_t and $\bar{z}_t$ as well, and (A.8) holds.

Summing up, $\frac{\sum_{t=2}^{T}z_{t-1}\varepsilon_t}{\left(\sum_{t=2}^{T}z_{t-1}^2\varepsilon_t^2\right)^{1/2}}$ and $\frac{\sum_{t=2}^{T}\tilde{z}_{t-1}\varepsilon_t}{\left(\sum_{t=2}^{T}\tilde{z}_{t-1}^2\varepsilon_t^2\right)^{1/2}}$ 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. $\frac{1}{T^{1+\eta}}\sum_{t=p+1}^{T}z_{t-1}x_{t-1} \Rightarrow \frac{\omega^2}{a}\left(X^2(1) - \int_0^1 X(s)\,\mathrm{d}X(s)\right)$

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 (A.2)) 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 $\left|\sigma^2_{\nu,t-1-j} - \sigma^2_{\nu,t}\right| \leq C\frac{j}{T}$ such that

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

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,

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

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

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

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

$$\frac{1}{T^{1/2+\eta}} S_t = \frac{1}{a\sqrt{T}} x_t + R_{t,T}$$

where $\sqrt{\mathrm{E}\left(|R_{t,T}|^2\right)} \rightarrow 0$ as $T \rightarrow \infty$ uniformly in $t = 1, \ldots, T$. The arguments are essentially the same as there; the only difference is having to show that $\sqrt{\mathrm{E}\left(|x_t - x_{t-j}|^2\right)} \leq C\sqrt{j}$ for all t and j, which is obvious in their strictly stationary 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 Minkowski's inequality to conclude that,

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

and therefore using the uniform boundedness of the variance of $\frac{x_{t-k-1}}{\sqrt{T}}$, it follows indeed that

$\sqrt{\mathrm{E}\left(|x_t - x_{t-j}|^2\right)} \leq C\sqrt{j}$ 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.

Then,

$$A_T = \frac{\omega}{aT} \sum_{t=p+1}^{T} x_{t-2} \nu_{t-1} + \frac{\omega}{\sqrt{T}} \sum_{t=p+1}^{T} R_{t-2,T} \nu_{t-1}.$$

It is furthermore seen from the expression of $R_{t,T}$ (Breitung and Demetrescu, 2015, Proof of Theorem 2) that $R_{t,T}$ is independent of u_{t+j}, v_{t+j} $\forall 1 \leq j \leq T - t$ whenever $(u_t, v_t)'$ is serially independent, such that $R_{t-2,T} \nu_{t-1}$ are the elements of a martingale difference array with uniformly vanishing variance, so $\mathrm{Var}\left(\sum_{t=p+1}^{T} R_{t-2,T} \nu_{t-1}\right) = o_p(T)$ as required for the summand involving $R_{t,T}$ to vanish. Since ν_{t-1} is independent of x_{t-2} and the conditions of Hansen (1992) are fulfilled, $T^{-1} \sum_{t=p+1}^{T} x_{t-2} \nu_{t-1}$ converges weakly, and we obtain

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

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 $\sup_{1 \leq t \leq T} |S_t| = T^{\eta} \sup_{1 \leq t \leq T} |x_t| + o_p\left(T^{1/2+\eta}\right) = O_p\left(T^{1/2+\eta}\right)$ 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 $\mathrm{E}\left(\left|\tilde{v}_{t-2}z_{t-2}\right|\right) < CT^{\eta/2}$ such that

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

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

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

as required. 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

$$\tilde{\beta}^{ivx} := \frac{\sum_{t=p+1}^{T} z_{t-1}\tilde{y}_t}{\sum_{t=p+1}^{T} z_{t-1}x_{t-1}}. \tag{A.9}$$

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,

$$\tilde{\beta}^{ivx} := \frac{\sum_{t=p+1}^{T} z_{t-1}\tilde{y}_t}{\sum_{t=p+1}^{T} z_{t-1}x_{t-1}} = \beta + \frac{\sum_{t=p+1}^{T} z_{t-1}(\gamma\nu_t - \hat{\gamma}\hat{\nu}_t + \varepsilon_t)}{\sum_{t=p+1}^{T} z_{t-1}x_{t-1}}. \tag{A.10}$$

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 $\left(1-\rho L\right)A\left(L\right)$), *i.e.* the pure autoregressive representation of x_t.

Then, analyze

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

We have that

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

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 (A.10) that,

$$\sum_{t=p+1}^{T} z_{t-1}\left(\varepsilon_t + \gamma\nu_t - \hat{\gamma}\hat{\nu}_t\right) = \sum_{t=p+1}^{T} z_{t-1}\varepsilon_t - \gamma \sum_{t=p+1}^{T} z_{t-1}\left(\hat{\nu}_t - \nu_t\right) - \left(\hat{\gamma} - \gamma\right) \sum_{t=p+1}^{T} z_{t-1}\hat{\nu}_t. \tag{A.11}$$

The first two summands in (A.11) 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 $\mathrm{Var}\left(\sum_{t=p+1}^{T} \varrho^{t-3}\varepsilon_t\right) = O_p\left(\sum_{t=p+1}^{T} \varrho^{2t}\right) = O_p\left(T^{\eta}\right)$. 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 $\sum_{t=p+1}^{T} d_{t-2}\boldsymbol{x}_{t-p} = O_p(T)$ 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,

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

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

The third summand in (A.11) is

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

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

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

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

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

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, it follows that the term $\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} z_{t-1}(\varepsilon_t+\gamma\nu_t-\hat{\gamma}\hat{\nu}_t)$ is asymptotically normal with mean zero and asymptotic variance

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

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 (A.11), *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 $\omega^2\int_0^1\sigma_v^2(s)\,\sigma_\varepsilon^2(s)\,\mathrm{d}s$, and we may re-write

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

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

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

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 $\mathrm{E}\left(\frac{1}{T^2}\sum_{t=p+1}^{T} c_t z_{t-1}x_{t-2}\right)=O\left(1\right)$. Moreover, $\frac{1}{T}\sum_{t=p+1}^{T} z_{t-1}v_{t-1}$ is itself $O_p\left(1\right)$, 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,

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

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 Additional Tables

Table B.1: 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		$T = 200$				$T = 500$		
$c = 0$	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
	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
			$T = 200$				$T = 500$		
$c = 10$	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
	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
			$T = 200$				$T = 500$		
$c = 20$	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
	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
			$T = 200$				$T = 500$		
$c = 30$	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
	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
			$T = 200$				$T = 500$		
$c = 40$	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
	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
			$T = 200$				$T = 500$		
$c = 50$	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
	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.

Table B.2: Size under breaks in variance and persistence, positive short-run AR parameter

		AHW	CY	$t_{ivx}^{\#}$	t_{ivx}	$\tilde{t}_{ivx}$	AHW	CY	$t_{ivx}^{\#}$	t_{ivx}	$\tilde{t}_{ivx}$
c	Var			$T = 200$					$T = 500$		
const small	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
	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
				$T = 200$					$T = 500$		
up small	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
	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
				$T = 200$					$T = 500$		
up large	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
	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 large	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
	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
				$T = 200$					$T = 500$		
down small	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
	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
				$T = 200$					$T = 500$		
down large	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
	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 2.