Finite-sample size control of IVX-based tests in predictive regressions*

Mehdi Hosseinkouchack
Goethe University Frankfurt†

Matei Demetrescu
Christian-Albrechts-University of Kiel‡

March 14, 2019

Abstract

In regressions with variables of unknown persistence, the use of IVX instruments leads to asymptotically valid inference. The standard normal or chi-squared limiting distributions for the usual t and Wald statistics may however differ markedly from the actual finite-sample distributions which exhibit in particular noncentrality. Convergence to the limiting distributions is shown to occur at a rate depending on the choice of the IVX tuning parameters and can be very slow in practice. A characterization of the leading higher-order terms of the statistic is provided, which motivates finite-sample corrections. Monte Carlo simulations confirm the usefulness of the proposed methods.

Keywords: Near integration, Mild integration, IV estimation, Wald-type test, Bias correction.

JEL classification: C12 (Hypothesis Testing), C22 (Time-Series Models)

1 Introduction

A common inferential task of practical relevance is to decide whether a potential predictor variable does indeed forecast another variable of interest. In the simplest setup, practitioners thus test the null hypothesis of no predictability in the model

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

where the regressor is usually assumed to have an autoregressive structure,

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

with finite initial condition x_1. With financial data, predictors such as dividend yields or earnings-price ratios are often quite persistent, even if still mean-reverting (typically captured by a value of ρ close to unity), and its shocks are contemporaneously correlated with the variable to be predicted

*The authors would like to thank three anonymous referees, as well as Karim Abadir and Uwe Hassler for very useful comments.

† *Corresponding author:* Goethe University Frankfurt, RuW, Postbox 49, Grueneburgplatz 1, 60323 Frankfurt am Main, Germany, email: hosseinkouchack@wiwi.uni-frankfurt.de.

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

(see Phillips, 2015, for a recent review). This biases the OLS estimator of the slope parameter and induces heavy nonnormality of t statistics (Stambaugh, 1999; Elliott and Stock, 1994) such that tests for predictability are too liberal in this framework.

Near to unity asymptotics, obtained by letting $\rho = 1 - c/T$, offer a better approximation of the actual distribution of the OLS t statistic than the standard normal in this situation; cf. Elliott and Stock (1994). The limiting distribution of the OLS estimator and test is explicitly nonnormal under near integration and depends on the mean-reversion parameter c and the correlation between u_t and v_t. Since consistent estimation of c is not possible in such highly persistent cases (Phillips, 1987), the literature suggested several different ways of circumventing the lack of knowledge on ρ. See, among others, Campbell and Yogo (2006); Jansson and Moreira (2006); Maynard and Shimotsu (2009); Camponovo (2015); Phillips (2015); Breitung and Demetrescu (2015).

Building on the work of Phillips and Magdalinos (2007) and Magdalinos and Phillips (2009), so-called extended IV [IVX] estimation and testing for predictive regressions is gaining momentum; see e.g. Gonzalo and Pitarakis (2012); Phillips and Lee (2013); Kostakis et al. (2015); Demetrescu and Rodrigues (2016). For this approach, x_{t-1} is instrumented by the specifically constructed instrument $z_{t-1} = (1 - \varrho L)_+^{-1} \Delta x_{t-1} = \sum_{j=0}^{t-2} \varrho^j \Delta x_{t-1-j}$ with initial condition $z_1 = 0$ and $\varrho = 1 - a/T^\eta$ with $\eta \in (0, 1)$. The idea is to use an instrument, say z_{t-1}, whose persistence is under control (and is typically below that of x_{t-1}). Regularity conditions assumed, the resulting IV estimator follows a mixed Gaussian distribution in the limit, and the limiting null distribution of the corresponding Wald-type statistic is standard chi-squared.

In finite samples, the IVX-based test of no predictability may still be seriously distorted, even if less so than the OLS-based test. This is clearly the case when choosing η too close to unity or a too close to zero: in such constellations, the difference in terms of persistence between the instrument z_t and the regressor x_t may be small and the cure is only marginally better than the disease. E.g. the rule of thumb proposed by Kostakis et al. (2015), which sets $\varrho = 1 - 1/T^{0.95}$, is actually equivalent to a near unit root with mean reversion parameter $\tilde{c}$ between 1 and 2 for a span of sample sizes between $T = 100$ and $T = 10000$. Typically, guidance as to how to choose a and η is simulation-based; while this does offer valuable insights, their scope is ultimately less general than would be desirable. More recently, Camponovo (2015) uses a bootstrap calibration scheme in the spirit of Loh (1987), but resampling schemes are computationally intensive, and bootstrapping near-integrated data generating processes is notoriously difficult. Simple, yet generally valid, rules about the choice of the tuning parameters a and η of the IVX procedures are not (yet) available.

This paper takes the reverse perspective. Rather than providing ranges of possible choices for ϱ, which promise a good approximation of the actual distribution of the IVX statistic via its standard normal limit, we aim to improve the size control of the test built with *user-chosen* ϱ. We therefore examine the behavior of components of the t statistic that vanish in the limit, but still have an effect in finite samples and depend on the parameters a and η. We do so in a setup allowing for time-varying variances and correlations of the errors u_t and v_t. Since the main source of distortions in finite samples appears to be the fact that the finite-sample distribution is not centered at zero (see also Stambaugh, 1999), we focus on correcting for the noncentrality of the t ratio.

One way of doing so is to resort to backward and forward demeaning of the involved variables. In time series analysis, backward (or recursive, or adaptive) demeaning can be traced back to at least the work of So and Shin (1999) where recursive demeaning is shown to reduce bias in estimators of large autoregressive roots. Specifically for (panel) predictive regressions, Westerlund et al. (2017) resorts to forward and backward demeaning to reduce endogeneity bias.[1] While this stabilizes size, we find it has the unpleasant side effect of reducing power in a nontrivial manner. This is argued to be a specific effect of forward demeaning in the context of persistent predictors, and not of IVX. Therefore, we then discuss the use of direct approximations of the higher-order terms affecting the finite-sample behavior of the t statistic. Some depend on the mean-reversion parameter c, which cannot be consistently estimated, so we provide a method of side-stepping this issue. While this is derived under the assumption of homoskedasticity, we find it, in extensive Monte Carlo experiments, to work reasonably well under various patterns of changing error variances.

2 Improved IVX inference

2.1 Preliminaries

Assumption 1 *The observations y_t and x_t, $t = 2, \ldots, T$, are generated according to* (1) *and* (2).

To keep a realistic setup, we allow for error heterogeneity in form of time-varying variances and correlations, as well as short-run dynamics. Concretely, we work under the following assumptions.

Assumption 2 *Let $v_t = \Psi(L)\,\nu_t$ with $\sum_{j\geq 0} j\,|b_j| < \infty$ s.t. $\psi = \Psi(1) \neq 0$, where ν_t is white noise.*

Assumption 3 *Let $(u_t, \nu_t)' = \mathbf{H}\left(\frac{t}{T}\right)(\xi_{1t}, \xi_{2t})'$, with $(\xi_{1t}, \xi_{2t})' \sim iid\,(0, \mathbf{I}_2)$, L_r-bounded for some $r > 4$, where $\mathbf{H}(\cdot) := \{h_{ij}\}_{i,j=1,2}$ is a matrix of piecewise Lipschitz-continuous bounded functions on $(-\infty, 1]$, which is of full rank at all but a finite number of points.*

This would be a typical structure in predictive regressions for stock returns, where the disturbance u_t is not predictable using the past of v_t. We do not assume a particular distribution for the errors but only require finite 4th order moments. Although daily returns may exhibit fat tails, standard predictive regression models are used in conjunction with monthly, quarterly or even annual data, where infinite kurtosis is not an issue. For the same reason, the serial independence assumption we make on the innovations is justifiable.

Let $\mathbf{\Sigma}_t = \mathbf{H}(t/T)\,\mathbf{H}'(t/T) = \begin{pmatrix} \sigma_u^2(t/T) & \sigma_{u\nu}(t/T) \\ \sigma_{u\nu}(t/T) & \sigma_\nu^2(t/T) \end{pmatrix}$ and notice that we have time varying variances, covariances and correlations of the errors, as $\mathrm{Cov}((u_t, \nu_t)') = \mathbf{\Sigma}_t$ is not restricted in any essential way. Importantly, the off-diagonal elements of $\mathbf{\Sigma}_t$ are not imposed to be zero, thereby allowing for the so-called predictive regression endogeneity. Furthermore, the assumption on $\mathbf{H}(s)$ allows for a wide range of variance matrices of the innovations, including e.g. single or multiple (co-) variance shifts, smooth transition (co-) variance shifts or even trending variances. Finally, the absolute summability condition placed on the coefficients of the filter Ψ is standard in the literature involving integrated and near-integrated variables.

[1] In fact, in the panel literature, forward and backward demeaning have a much longer history in dealing with the Nickell bias (Nickell, 1981); see Everaert (2013) for a recent contribution.

Under these assumptions, we have with $\boldsymbol{W}$ a vector of two independent standard Wiener processes

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\begin{pmatrix} u_t \\ \nu_t \end{pmatrix} \Rightarrow \int_0^s \mathbf{H}(r)\mathrm{d}\boldsymbol{W}(r) =: \begin{pmatrix} U_{\mathrm{H}}(s) \\ V_{\mathrm{H}}(s) \end{pmatrix};$$

see Cavaliere et al. (2010). Let $J_{c,\mathrm{H}}(s) = V_{\mathrm{H}}(s) - c\int_0^s e^{-c(s-r)}V_{\mathrm{H}}(r)\mathrm{d}r$ and note that the normalized levels of x_t converge weakly to this heteroskedastic Ornstein-Uhlenbeck type process, i.e.

$$T^{-1/2}x_{[sT]} \Rightarrow \psi J_{c,\mathrm{H}}(s).$$

IVX estimation relies on using the instrument $z_{t-1} = (1-\varrho L)_+^{-1}\Delta x_{t-1} = \sum_{j=0}^{t-2}\Delta x_{t-1-j}$, where $\varrho = 1 - \frac{a}{T^\eta}$ with $a > 0$ and $\eta \in (0,1)$. The textbook IXV t statistic for the null $\beta = 0$ is

$$t_{vx} = \frac{\sum_{t=2}^{T}(z_{t-1}-\bar{z})\, y_t}{\sqrt{\sum_{t=2}^{T}(z_{t-1}-\bar{z})^2\,\hat{u}_t^2}}, \tag{3}$$

with Eicker-White standard errors to account for the heteroskedasticity. The residuals $\hat{u}_t^2$ are computed using the OLS estimator of β, as is common in the predictive IVX regression literature.[2]

What makes the IVX approach interesting for practitioners is that the terms involving c vanish as $T \to \infty$ and pivotal inference on β can be obtained asymptotically. See Kostakis et al. (2015) for details on IVX-based predictive regression under strict stationarity of errors, and Demetrescu and Rodrigues (2016) for a case with time-varying variances with some (nontrivial) restrictions on the correlations. In finite samples, however, the actual distribution is not centered at zero because numerator and denominator correlate, and has somewhat smaller variance, as can be seen in Figure 1. Notice also the slow convergence to the standard normal.

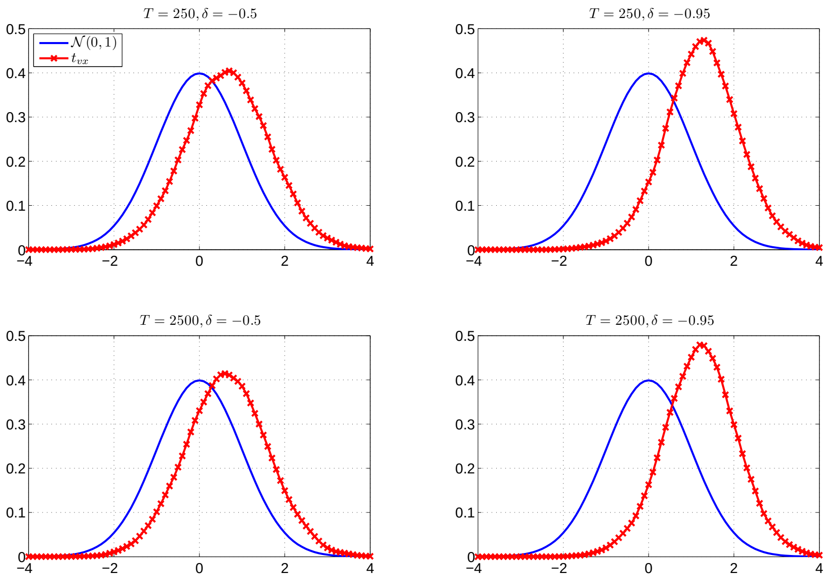

Figure 1: Finite sample null distribution of t_{vx} for $a = 1;\eta = 0.95$ vs. $\mathcal{N}(0,1)$; DGP: (1) with (2), $\rho = 1$, $(u_t, v_t) \sim iid\mathcal{N}(0,((1,\delta);(\delta,1)))$, 25000 replic., different correlations δ and sample sizes.

[2]The IVX estimator has convergence rate $T^{1/2+\eta/2}$ so IVX residuals, although still consistent for $\eta > 0$, are typically less precise; see Kostakis et al. (2015).

2.2 Higher-order terms

We therefore study corrections that make inference even more reliable. To this end, we first characterize the main problematic terms.

Proposition 1 *Under Assumptions 2–3 and $\eta > 1/2$, it holds as $T \to \infty$ that*

$$t_{vx} = Z_T + B_T + C_T + o_p\left(T^{\eta/2 - 1/2}\right)$$

where

$$Z_T \overset{d}{\to} \mathcal{N}(0,1) \quad \text{with} \quad \mathrm{E}(Z_T) = 0,$$

$$B_T \to 0 \quad \text{with} \quad T^{1/2-\eta/2}\,\mathrm{E}(B_T) \to -\frac{1}{\sqrt{2a}} \frac{\int_0^1 \sigma_{u\nu}(s)\,\sigma_u^2(s)\,\sigma_\nu^2(s)\,ds}{\sqrt{\left(\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s\right)^3}},$$

and

$$T^{1/2-\eta/2} C_T \Rightarrow -\sqrt{\frac{2}{a}} \frac{U_{\mathrm{H}}(1)\, J_{c,\mathrm{H}}(1)}{\sqrt{\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s}}.$$

Proof: *See Appendix B.*

The noncentrality is influenced by two main components. The first, B_T, depends on the user-chosen parameters a and η, as well as on what could be interpreted as a particular form of average correlation. In fact, under homoskedasticity ($\mathbf{H} = const.$), the expectation of $T^{1/2-\eta/2}B_T$ is asymptotically equivalent to $\delta/\sqrt{2a}$ with δ the constant correlation of u_t and ν_t. If $\sigma_{u\nu}(s) = 0\ \forall s \in [0,1]$, this expectation is zero and the first term does not affect the centering of the t statistic.

The same holds for the second component: if $\sigma_{u\nu}(s) = 0\ \forall s \in [0,1]$, then U_{H} and $B_{\mathrm{c,H}}$ are independent, therefore the expectation of the limit of the normalized C_T is zero as well. Should there however be contemporaneous correlation, the behavior of C_T—in particular the expectation—does depend on c. Moreover, the dependence is nonlinear, since it is easily shown that

$$\mathrm{E}\left(U_{\mathrm{H}}(1)\, J_{c,\mathrm{H}}(1)\right) = \int_0^1 e^{-c(1-s)} \sigma_{u\nu}(s)\mathrm{d}s.$$

Notice that, as expected, the larger c is, the smaller the expectation would be. This expression simplifies too under homoskedasticity, where $T^{1/2-\eta/2}C_T$ has an asymptotic expectation depending on the (constant) correlation δ, namely $\delta\sqrt{\frac{2}{a}\frac{1-e^{-c}}{c}}$. For $c = 0$, the case with the largest distortions, we see this expectation to be twice as large as that of the normalized B_T, with the relative importance of C_T diminishing as c increases. This component depends however on the mean reversion parameter c which cannot be consistently estimated, unlike the expectation of B_T.

Finally, both components B_T and C_T do vanish, but at rate $T^{1/2-\eta/2}$. This is relatively slow even if not choosing η close to unity. The two are thus relevant for the finite-sample behavior of t_{vx}.

2.3 Corrections

The term C_T from Proposition 1 appears because of the full-sample demeaning of the dependent variable (see the proof of Proposition 1 for details). To deal with this we first discuss recursive demeaning as possible correction. In particular, we use backward recursive demeaning for the regressor and forward demeaning for the dependent variable, in that we write

$$t_{vx}^{rec} = \frac{\sum_{t=2}^{T} (z_{t-1} - \bar{z}_{t-1}) (y_t - \ddot{y}_t)}{\sqrt{\sum_{t=2}^{T} (z_{t-1} - \bar{z}_{t-1})^2 \, \widehat{u}_t^2}},$$

with

$$\ddot{y}_t = \frac{1}{T - t + 1} \sum_{j=t}^{T} y_j \quad \text{and} \quad \bar{z}_t = \frac{1}{t} \sum_{j=1}^{t} z_j.$$

The motivation for such demeaning schemes is that the recursively demeaned regressor and the forward demeaned disturbance are now orthogonal irrespective of the correlation between u_t and ν_t, which is not the case with usual demeaning. Such orthogonal schemes of mean adjustment have been used before in predictive regressions; see e.g. Westerlund et al. (2017) for the panel case.[3]

This effect of recursive adjustment is very much in the spirit of the the proposal of Kostakis et al. (2015), who point out that not demeaning the instrument z_{t-1} (while still demeaning the dependent variable and the predictor itself to account for a nonzero intercept in the predictive regression) reduces the finite-sample correlation between the numerator and the denominator of the IVX t statistic. We shall closer examine the corrections of Kostakis et al. (2015) after analyzing the effect of the orthogonal mean adjustment scheme in:

Proposition 2 *Under Assumptions 2–3 and $\eta > 1/2$, it holds as $T \to \infty$ that*

$$t_{vx}^{rec} = Z_T + B_T + o_p\left(T^{\eta/2 - 1/2}\right) \qquad \text{with } Z_T \text{ and } B_T \text{ from Proposition 1.}$$

Proof: *See the supplement.*

Proposition 2 shows that dependence on c of the leading higher-order terms may in fact be eliminated. Our Monte Carlo study (see Section 3) shows that t_{vx}^{rec} performs quite well in terms of size in spite of the remaining term B_T, so the first correction for noncentrality that we suggest is this orthogonal mean adjustment scheme.

The Monte Carlo study also shows that the local power of the test is low. To see why, examine

$$y_t - \ddot{y}_t = \beta \left(x_{t-1} - \ddot{x}_{t-1}\right) + u_t - \ddot{u}_t$$

with $\ddot{\cdot}_t$ denoting forward demeaned quantities. Under the alternative $\beta \neq 0$, it is the cross-product $(z_{t-1} - \bar{z}_{t-1}) (x_{t-1} - \ddot{x}_{t-1})$ that induces power. But the forward demeaned x_{t-1} may be rewritten as

$$x_{t-1} - \frac{1}{T - t + 1} \sum_{j=t}^{T} x_{j-1} = -\frac{1}{T - t + 1} \sum_{j=t}^{T} \sum_{k=0}^{j-t-1} \Delta x_{j+k};$$

e.g. in the extreme case of x_t being a random walk, $\Delta x_t = \nu_t$ and $(x_{t-1} - \ddot{x}_{t-1})$ can be seen to be uncorrelated with $z_{t-1} - \bar{z}_{t-1}$, so the effect on the test statistic under the alternative is similar to that of a weak instrument. The very correction for the size induces power losses under persistence.

[3] Moreover, recursive adjustment schemes are popular in the related unit root testing literature as a means for reducing noncentrality and improving local power; see e.g. Shin and So (2001).

Turning our attention to the corrections proposed by Kostakis et al. (2015), they result in

$$t_{vx}^{W}=\frac{\sum_{t=2}^{T}z_{t-1}\left(y_{t}-\bar{y}\right)}{\sqrt{\sum_{t=2}^{T}z_{t-1}^{2}\hat{u}_{t}^{2}-T\bar{z}^{2}\hat{\omega}_{u|v}^{2}}}$$

where $\hat{\omega}_{u|v}^{2}=\hat{\omega}_{u}^{2}-\hat{\lambda}_{uv}^{2}\hat{\omega}_{v}^{-2}$ with $\begin{pmatrix}\hat{\omega}_{u}^{2} & \hat{\lambda}_{uv}\\ \hat{\lambda}_{uv} & \hat{\omega}_{v}^{2}\end{pmatrix}$ an estimate of the long-run covariance matrix of $(u_t, v_t)'$. The behavior of the resulting IVX t ratio is discussed in

Proposition 3 *Under Assumptions 2–3 and $\eta > 1/2$, it holds as $T \to \infty$ that*

$$t_{vx}^{W}=Z_{T}+B_{T}+C_{T}+o_{p}\left(T^{\eta/2-1/2}\right)\qquad \textit{with } Z_T,\ B_T \textit{ and } C_T \textit{ from Proposition 1.}$$

Proof: *See Appendix B.*

Examining the proof, it is seen that several of the terms of order $o_p\left(T^{\eta/2-1/2}\right)$ are eliminated by the corrections proposed by Kostakis et al. (2015). This explains our findings in the Monte Carlo section, that the t_{vx}^{W} statistic performs quite reliably, even if it is not free of noncentrality yet.

We therefore move on to propose explicit corrections for B_T and C_T. The term B_T may be dealt with without essential complications, although the involved integrals require nonparametric estimation. Moreover, we find in preliminary simulations (not reported here) that the influence of heteroskedasticity on the behavior of the t statistic is of rather secondary importance, so we shall derive the correction for B_T under the premise of homoskedasticity. Then, only the estimation of the contemporaneous error correlation δ is required. (Our simulations in Section 3 show this to work reasonably well under heteroskedasticity too.) The resulting correction term for B_T is then

$$b_{T}=\delta/\sqrt{2T\left(1-\varrho\right)}$$

where δ may be estimated as the correlation of ν_t and u_t based on $\hat{u}_t$ and $\hat{\nu}_t$ from an AR(p) approximation of x_t with p selected via an information criterion (we resort to the Akaike IC).

The term C_T poses more problems due to its intricate dependence on c and on the time-varying covariance. Even under the simplifying assumption of homoskedasticity we have

$$\mathrm{E}\left(U_{\mathrm{H}}\left(1\right)J_{c}\left(1\right)\right)=\delta\sigma_{u}\sigma_{v}(1-e^{-c})/c.$$

Now, although it does not seem possible to completely remove the effect of C_T under usual demeaning, it is possible to find quantities that have the same expectation, such that the t statistic may be better centered. Recall that $\frac{1}{\psi\sqrt{T}}x_{[sT]}\Rightarrow J_{c,\mathrm{H}}\left(s\right)$, where it is easily seen that $\mathrm{Var}\left(J_{c,\mathrm{I}}\left(s\right)\right)=\frac{1-e^{-2sc}}{2c}$ under homoskedasticity. Therefore, $2\,\mathrm{Var}\left(J_{c,\mathrm{I}}\left(1/2\right)\right)=\frac{1-e^{-c}}{c}$, and it suggests itself to resort to a quantity with this expectation to get a hold of the noncentrality induced by C_T. The natural choice following from this property of the Ornstein-Uhlenbeck process is then $\left(\frac{1}{\psi\sigma_{\nu}\sqrt{T}}x_{[T/2]}\right)^{2}$, leading to

$$c_{T}=\frac{\delta\sqrt{2}}{\sqrt{T\left(1-\varrho\right)}}\cdot\frac{2x_{[T/2]}^{2}}{\omega^{2}T}=2b_{T}\frac{2x_{[T/2]}^{2}}{\omega^{2}T},$$

since $\omega^2 = \psi^2\sigma_\nu^2$ is nothing else than the (stationary) long-run variance of v_t (which may be estimated either based on Δx_t, or—as we proceed in our simulations—on the residuals of a first-order autoregression of x_t).

It should be noted, however, that this delivers a noisy proxy for the mean of C_T: while it will remove the noncentrality due to C_T (at least under homoskedasticity), it will at the same time marginally inflate the variance of the corrected t statistic. The presence of the estimator $\hat{\omega}^2$ in the denominator further inflates the variance: since we employ a nonparametric estimator, its variability in finite samples is large enough to affect the positive effect of the correction. Concretely, it induces outliers in the distribution of the correction and inflates the variance of the corrected statistic. To deal with these issues we add finite-sample modifications which do not affect the asymptotics.

First, we use the fact that the term to be corrected is bounded to $[0,1]$, so we censor $2x_{[T/2]}^2/(\hat{\omega}^2 T)$ on $[0,1]$. While this reduces the variability of the term, it changes its expectation, so we re-normalize it. The limit of $2x_{[T/2]}^2/(\hat{\omega}^2 T)$ is $\chi^2(1)$ under the worst-case scenario of $c=0$, which, upon censoring on $[0,1]$, loses the unity expectation, and has an expectation of $1-\sqrt{2/(\pi e)} \approx 0.5160586$ and a variance of $4\left(\Phi(1)-\Phi(0)\right) - 2/(\pi e) - 2\sqrt{2/(\pi e)} \approx 0.1632968$.

Second, we standardize the t statistic to take the variance of the correction c_T into account.

Finally, should x_t be stationary instead of near-integrated, this bias correction may overcorrect, since, for ρ away from unity, the standard normal asymptotics do relatively well even when ϱ is close to unity; see Kostakis et al. (2015). A practical adjustment of the correction is to restrict ϱ in t_{vx}^* to be smaller than an estimate of ρ. In particular we suggest $\varrho = \min(\varrho, \hat{\rho})$, where $\hat{\rho}$ is the OLS estimator in a first-order autoregression of x_t. Asymptotically, this restriction makes no difference under near-integration, but prevents the bias correction to "overshoot."

We therefore suggest to use, with $\hat{b}_T = \hat{\delta}/\sqrt{2T\left(1-\min\{\varrho;\hat{\rho}\}\right)}$, the statistic

$$\frac{t_{vx} - \hat{b}_T\left(1 + 3.8755\min\left\{1; \frac{2x_{[T/2]}^2}{\omega^2 T}\right\}\right)}{\sqrt{1 + 2\frac{\hat{\delta}}{3}\,0.7831\,\hat{b}_T + 0.6132\,\hat{b}_T^2}},$$

where $\frac{\hat{\delta}}{3}$ is intended to capture finite-sample correlation of $\hat{c}_T$ and t_{vx}, and is tuned to homoskedasticity. Ours simulations below suggest that it works well under heteroskedasticity too.

3 Finite sample

In this section we provide finite sample evidence on the merits of the remedies proposed in this paper. We use a data generating process (DGP) as outlined under equations (1) and (2) with independent innovation process governed by a bi-variate normal distribution with a correlation coefficient of $\delta = -0.95$ (which is typical for predictive regressions with stock returns; see e.g. Phillips, 2015), as well as time-varying volatility. Concretely,

$$\begin{pmatrix} u_t \\ \nu_t \end{pmatrix} \sim iid\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}_t\right), \quad \boldsymbol{\Sigma}_t = \begin{pmatrix} \sigma_u(t/T) & 0 \\ 0 & \sigma_\nu(t/T) \end{pmatrix}\begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}\begin{pmatrix} \sigma_u(t/T) & 0 \\ 0 & \sigma_\nu(t/T) \end{pmatrix};$$

furthermore, $\upsilon_t = \phi \upsilon_{t-1} + \nu_t$ for $\phi = 0.5$. The variance of the innovations is taken to vary in time according to the following scenarios:

1. Homoskedasticity [cst]: $\sigma_u^2(s) = \sigma_\nu^2(s) = 1,\ s \in [0,1]$,
2. Early upward break [e.u.]: $\sigma_u^2(s) = \sigma_\nu^2(s) = 1 \cdot \mathbb{I}\,(t < 0.3\,T) + 4 \cdot \mathbb{I}\,(t \geq 0.3\,T)$,
3. Late upward break [l.u.]: $\sigma_u^2(s) = \sigma_\nu^2(s) = 1 \cdot \mathbb{I}\,(t < 0.7\,T) + 4 \cdot \mathbb{I}\,(t \geq 0.7\,T)$,
4. Early downward break [e.d.]: $\sigma_u^2(s) = \sigma_\nu^2(s) = 4 \cdot \mathbb{I}\,(t < 0.3\,T) + 1 \cdot \mathbb{I}\,(t \geq 0.3\,T)$,
5. Late downward break [l.d.]: $\sigma_u^2(s) = \sigma_\nu^2(s) = 4 \cdot \mathbb{I}\,(t < 0.7\,T) + 1 \cdot \mathbb{I}\,(t \geq 0.7\,T)$.

The size study results are generated using 10,000 replications and considering $c \in \{0, 1, 5, 10, 30, 50\}$ together with $\beta = 0$ for $T = 250$ and 500. To analyze the power of the corrected tests, we consider a sequence of local alternatives characterized by $\beta = \frac{b}{T}\sqrt{1-\delta^2}$, for $b \in \{-26, -24, ..., -2, 0, 2, ..., 26\}$. Note under $b = 0$ the size properties of the test will be recovered. Since the sign of β might be known in practice (as is often the case when the choice of the predictor is motivated by economic theory[4]), we consider local alternatives covering both situations, $\beta < 0$ and $\beta > 0$, alongside with cases where two-sided testing is of interest. All through this section, we fix $a = 1$ and $\eta = 0.95$ following the recommendation of Kostakis et al. (2015).

We compare four versions of the IVX statistic testing the null $\beta = 0$: the original IVX t statistic (t_{vx}), the finite-sample adjusted version of Kostakis et al. (2015) (t_{vx}^W),[5] the IVX t statistic computed with orthogonal mean adjustment (t_{vx}^{rec}), as well as our bias-corrected proposal (t_{vx}^*).

Table 1 shows the finite-sample rejection frequencies at the 5% nominal level for strong negative contemporaneous correlation $\delta = -0.95$.[6] The finite-sample noncentrality of the standard IVX t statistic leads as expected to huge size distortions that only drop to reasonable levels for $c = 10$ if not $c = 30$. The time variation of the variance somewhat influences these distortions, but not by much. Also, they do not drop with increasing T, as predicted by the small rates in Proposition 1. The statistic t_{vx}^W on the other hand shows that the finite-sample corrections introduced in Kostakis et al. (2015) work excellent in the two-sided case. Only for $c = 50$ can one observe a very slight tendency to overreject (with rejection frequencies closer to 6% than to 5% for $T = 250$). However, the t_{vx}^W statistic does not behave too well in each tail taken alone, as it tends to overreject to the right (one sees large rejection frequencies for small c, and even for $c = 50$ we note rejection frequencies above 8%) and to underreject to the left (this is most visible for small c, where the rejection frequencies are below 1%). This also does not significantly improve for larger $T = 500$, and exhibits little variation across the different variance patterns. The statistic with backward and forward recursive demeaning has very good size control (with some exceptions for $c = 0$, where rejection frequencies of 7% may be observed for the test against right-sided alternatives, and some cases of under-rejections: for left-sided testing under downward breaks and $c = 0, 1$ we observe rejection frequencies of 2 or 3%). Finally, the t_{vx}^* statistic has the best size control of all four tests:

[4] For example the effect of dividend yields on excess returns can be expected to be positive, while increasing risk-free rates could have a negative effect effect on excess returns (see e.g. Ang and Bekaert, 2007).

[5] Kostakis et al. (2015) consider a Wald statistic W for which $\left(t_{vx}^W\right)^2 = W$.

[6] The findings are symmetric in the sign of δ; moreover, size behavior improves uniformly for decreasing magnitude of δ so we do not include the exact figures to save space.

while it sometimes underrejects for left-sided testing (in the same situations where the t_{vx}^{rec} statistic was undersized), most rejection frequencies lie between 4 and 6%, with only a handful of cases where the 6% threshold is exceeded, and no rejection frequency above 7%.

Summing up, all three modified statistics may be used in a two-sided testing situation in what concerns size control. For one-sided testing situations, the use of t_{vx}^{W} is not recommended as it overrejects to the right and severely underrejects to the left, which has a dampening effect on power, as we shall see next. The power simulations will also confirm the power-reducing effect of the orthogonal mean adjustment scheme mentioned after Proposition 2.

We present in Figures 2 – 4 plots of local power curves of the four statistics compared for $c = 0, 10, 30$ and all variance patterns and test variants (left-, right-, and two-sided).

For the left-sided tests, it is t_{vx}^{*} that has best power in all cases. Compared to t_{vx} and t_{vx}^{W}, this is because t_{vx}^{*} is centered correctly and therefore not undersized. Here, t_{vx} seems to perform a bit better than t_{vx}^{W}. The test based on backward and forward adjustment has worst power of all. The power gap to the other tests decreases as c increases, but power drops anyway with increasing c.

For right-sided testing, t_{vx} rejects very often, but this is of course due to the extreme liberality compared to the other tests. The test based on t_{vx}^{rec} performs, like before, worst (again, with the power differences decreasing as c increases). To the right, t_{vx}^{W} is typically more powerful than t_{vx}^{*}, but keep in mind that it is also quite oversized, even if not as oversized as the uncorrected t_{vx}.

Finally, examining the two-sided tests, we expectedly observe a combination of the findings for the left- and right-sided tests, with the difference that the t_{vx}^{W} test is now correctly sized and the corresponding test decisions are now reliable. The test based on t_{vx}^{*} is also correctly sized, and the power ranking of the two depends on the sign of β under the alternative. While t_{vx}^{W} is more powerful against right-sided alternatives, but less powerful against left-sided ones, t_{vx}^{*} has a symmetric power curve. Again, the larger c, the closer the power functions of the three corrected tests.

Summing up, we recommend the use of t_{vx}^{*} for one-sided testing. For two-sided testing, one has the choice between t_{vx}^{W} and t_{vx}^{*}, with the symmetry of the power curve being an argument in favor of t_{vx}^{*}, and the higher power against right-sided alternatives (or left-sided, should the correlation δ be positive) being an argument in favor of t_{vx}^{W}.

4 Concluding remarks

A typical approach in the context of predictive regressions where the persistence of the endogenous forecasting variable is unknown, is to turn to IV regressions where a so-called extended instrumental variable with a controlled level of persistence is constructed. The resulting IVX estimator is asymptotically mixed Gaussian and makes for standard asymptotic inference. The small sample sample deviations from the asymptotic limit, however, depend heavily on how the IV estimator is constructed. The literature here relies on simulation based rules to choose the persistence properties of this instrumental variable.

Table 1: Size properties of different tests under short-run dynamics and strong contemporaneous shock correlation

		$T=250$												$T=500$											
		2-sided				left-sided				right-sided				2-sided				left-sided				right-sided			
c	Var	t_{vx}	t_{vx}^{rec}	t_{vx}^{W}	t_{vx}^{*}	t_{vx}	t_{vx}^{rec}	t_{vx}^{W}	t_{vx}^{*}	t_{vx}	t_{vx}^{rec}	t_{vx}^{W}	t_{vx}^{*}	t_{vx}	t_{vx}^{rec}	t_{vx}^{W}	t_{vx}^{*}	t_{vx}	t_{vx}^{rec}	t_{vx}^{W}	t_{vx}^{*}	t_{vx}	t_{vx}^{rec}	t_{vx}^{W}	t_{vx}^{*}
	cst.	20.90	4.90	4.69	4.48	0.11	2.91	0.08	4.08	33.01	6.83	8.82	5.42	20.65	5.09	4.39	4.42	0.09	2.82	0.05	4.50	33.51	7.27	9.00	5.19
	e.u.	16.18	4.97	4.66	4.91	0.20	3.11	0.12	6.28	26.30	6.87	9.38	4.72	14.56	4.80	3.88	4.75	0.16	3.07	0.08	6.16	25.15	6.62	8.35	4.18
0	l.u.	16.01	5.73	4.93	5.87	0.76	3.71	0.50	6.38	25.51	7.64	9.07	5.22	15.49	5.93	4.69	5.49	0.75	3.70	0.48	6.54	24.62	7.43	9.06	4.59
	e.d.	23.55	4.91	4.62	3.52	0.02	2.70	0.02	1.82	37.00	7.14	9.46	5.54	21.79	4.64	4.37	3.28	0.00	2.33	0.01	2.00	35.27	6.99	9.20	5.18
	l.d.	22.38	3.81	4.94	3.35	0.01	1.94	0.00	2.64	35.38	6.15	9.51	5.44	21.45	4.02	4.13	3.22	0.01	1.94	0.00	2.77	34.86	6.34	8.64	4.99
	cst.	15.94	4.48	4.87	6.28	0.24	2.58	0.18	6.42	25.58	6.85	9.22	5.82	15.54	5.01	4.46	6.27	0.23	3.24	0.15	6.59	25.03	7.06	9.22	5.56
	e.u.	12.56	4.73	5.10	5.95	0.33	3.08	0.21	7.04	21.40	6.99	9.55	5.40	12.87	4.94	4.78	6.08	0.32	3.10	0.20	7.44	20.90	7.02	9.19	5.03
1	l.u.	13.14	5.30	5.44	5.73	1.07	3.52	0.70	6.12	21.19	7.29	10.13	5.41	12.42	5.12	4.93	5.47	0.63	3.44	0.46	5.43	19.91	7.28	9.35	5.15
	e.d.	15.76	4.50	4.84	5.50	0.06	2.76	0.04	5.39	25.28	6.70	10.15	5.65	14.54	4.55	4.53	5.24	0.11	2.52	0.01	6.01	23.80	6.78	9.07	5.07
	l.d.	16.47	4.61	4.95	5.44	0.04	2.16	0.03	5.90	26.98	7.20	10.01	5.98	15.24	4.22	4.72	4.96	0.04	2.10	0.04	5.80	24.72	6.30	9.19	5.44
	cst.	9.96	5.10	5.69	6.03	1.15	3.33	0.89	5.47	16.01	6.91	10.12	5.89	9.39	4.70	5.18	5.67	1.24	3.36	1.02	5.41	15.10	6.46	9.65	5.56
	e.u.	8.42	5.29	5.20	6.43	1.04	3.13	0.73	6.47	14.24	6.85	9.54	4.94	8.95	5.16	5.13	5.95	0.91	3.24	0.54	5.67	14.76	7.02	9.95	5.18
5	l.u.	9.18	5.22	5.81	4.00	1.35	3.34	0.94	2.31	15.20	7.10	10.51	5.39	9.06	5.39	5.46	3.75	1.35	3.49	0.91	2.36	14.63	7.05	10.45	5.48
	e.d.	8.65	4.95	5.60	4.65	1.07	3.02	0.77	3.99	14.18	6.61	10.49	5.48	8.07	4.85	4.86	4.66	1.22	3.12	0.89	4.09	13.62	6.36	9.74	4.92
	l.d.	9.19	4.81	5.52	6.82	0.86	2.31	0.54	6.41	14.97	6.80	10.40	5.57	8.14	4.14	4.62	5.92	0.78	2.41	0.57	6.29	14.14	6.31	9.24	4.86
	cst.	7.70	4.84	5.81	4.34	1.68	3.39	1.47	3.15	12.18	6.12	9.80	5.37	7.71	4.92	5.74	4.15	1.78	2.95	1.53	3.23	11.95	6.43	9.60	5.31
	e.u.	7.69	5.17	5.59	4.87	1.61	3.37	1.24	4.01	12.24	6.56	9.57	5.27	7.29	5.28	5.51	4.83	1.70	3.46	1.33	3.94	11.83	6.66	9.18	5.11
10	l.u.	7.88	5.29	6.00	3.65	1.84	3.46	1.24	1.89	12.11	6.82	10.07	5.76	7.44	5.34	5.58	3.26	1.70	3.31	1.14	1.68	11.56	6.96	9.78	5.45
	e.d.	7.50	4.94	5.94	3.47	1.85	3.15	1.53	2.18	11.60	6.70	9.98	5.22	7.04	4.78	5.67	3.29	1.82	3.13	1.56	2.06	11.05	6.62	9.42	5.04
	l.d.	7.85	4.44	6.10	4.86	1.97	2.89	1.68	4.27	11.62	6.11	9.49	5.33	6.31	4.31	4.70	4.20	1.67	2.72	1.33	4.07	10.61	5.87	8.81	4.42
	cst.	6.72	5.17	6.14	4.07	2.82	4.07	2.74	3.13	9.72	6.57	9.18	5.57	6.07	4.81	5.52	3.78	2.90	3.74	2.82	3.11	8.40	6.03	7.84	4.91
	e.u.	6.95	5.31	6.21	4.44	3.47	4.11	3.18	3.78	9.22	6.40	8.63	5.50	6.02	4.86	5.51	3.83	2.83	3.78	2.65	3.26	9.00	6.36	8.29	5.10
30	l.u.	6.68	5.10	5.97	4.29	2.95	3.37	2.52	2.90	9.56	6.61	8.88	5.58	5.99	5.06	5.26	3.64	2.45	3.73	2.06	2.42	8.87	6.41	8.16	5.20
	e.d.	6.28	5.16	5.90	3.65	3.09	4.02	2.86	3.09	8.76	6.38	8.38	5.11	5.77	5.16	5.45	3.43	3.03	3.50	2.85	2.98	8.43	6.30	7.93	4.86
	l.d.	6.00	4.80	5.56	3.97	2.99	3.59	2.76	3.44	8.94	6.16	8.34	5.00	6.03	4.74	5.65	3.84	2.86	3.48	2.73	3.16	8.51	6.22	8.03	4.87
	cst.	6.14	5.06	5.91	4.22	3.26	4.10	3.09	3.37	8.41	5.99	8.11	5.48	5.44	4.82	5.11	3.59	3.12	3.90	3.04	3.16	7.14	5.63	6.84	4.65
	e.u.	6.50	5.26	6.17	4.58	3.33	4.28	3.21	3.61	8.84	6.23	8.40	5.59	5.97	5.18	5.62	4.22	2.96	4.13	2.84	3.15	7.85	6.28	7.65	5.33
50	l.u.	6.60	5.25	6.05	4.70	3.62	4.15	3.31	3.62	8.76	6.13	8.36	5.69	5.43	5.00	5.03	3.70	2.68	3.72	2.37	2.69	7.87	5.93	7.46	5.02
	e.d.	5.97	5.12	5.72	4.17	3.39	3.97	3.28	3.44	7.71	6.58	7.56	5.02	5.78	4.87	5.60	3.88	3.27	3.83	3.18	3.26	8.08	6.21	7.89	5.25
	l.d.	6.35	5.18	6.11	4.71	3.68	4.03	3.63	3.88	7.87	6.30	7.62	5.22	5.65	5.09	5.40	3.75	3.45	4.15	3.35	3.59	7.39	5.87	6.99	4.88

Note: Data generated with (1) and (2) with $v_t = \phi v_t + \nu_t$ for $\phi = 0.5$, where $(u_t, \nu_t) \sim iiN(0, \mathbf{\Sigma}_t)$ and $\mathbf{\Sigma}_t$ exhibits constant correlation $\delta = -0.95$ and time-varying variances. We set $\rho = 1 - {}^{c}/{}_{T}$ for various c and $\varrho = 1 - {}^{1}/{}_{T^{0.95}}$ and use standard normal critical values. See the text for details.

In this paper we provide a structured approach to control the small sample bias of the resulting t statistic for any given instrumental variable and as a result control the size distortions present. First we develop a higher order expansion of the corresponding IVX t statistic and as such provide a theoretical understanding of the small sample deviations of the t statistic from its limit. This in turn suggests ways to center the t statistic at the origin under the null. Combining forward and recursive demeaning does account for most leading terms of the bias at the cost of some loss of power. An explicit correction for the noncentrality achieves similar size control but without the power reduction. These proposals do not assume any parametric restriction on the persistence of the extended instrumental variable, and rather provide, for any given parameterization thereof, a corresponding noncentrality reduction recipe.

Our Monte Carlo study shows that all of these proposals provide substantial remedies to small sample size distortions to the IV t statistic while maintaining relatively good power properties. Further, when the effect of a forecasting variable is negative we suggest using a left-sided t statistic with one of the corrections we provided in this paper, since our Monte Carlo study provides evidence that such a strategy is associated with a better statistical power compared to using a two-sided test. For two-sided alternatives, the Wald test of Kostakis et al. (2015) offers the better balance between size and power; since it is not the best choice for left-sided alternatives, however, an adaptive test procedure or a combination of tests may further improve inference in cases with two-sided alternatives.

Appendix

A Preliminaries

Note that, for large enough T, $\sum_{j=0}^{t-1} \varrho^{kj} = \frac{1-\varrho^{kt}}{1-\varrho^{k}} = \frac{T^\eta}{a}\left(\frac{1-\varrho^{kt}}{1+\varrho+\ldots+\varrho^{k-1}}\right) \leq \frac{1}{ka}T^\eta$ for fixed k and $\varrho = 1-\frac{a}{T^\eta}$ with $\eta \in (0,1)$ and $a > 0$. In fact, for $t/T \to s > 0$, $\varrho^{kt} = \left(1-\frac{a}{T^\eta}\right)^{kt} = \left(\left(1-\frac{a}{T^\eta}\right)^{-\frac{T^\eta}{a}}\right)^{-ka\frac{t}{T}T^{1-\eta}} \to 0$ so $\sum_{j=0}^{t-1}\varrho^{kj} \sim \frac{1}{ka}T^\eta$ in such cases. Let further C denote a generic constant whose value may change from occurrence to occurrence. Before moving on to the proofs, we state three helpful lemmas.

Lemma 1 *Under the assumptions of Proposition 1, it holds that*

$$\sup_{1\leq t\leq T}\left\|T^{-1/2}x_t\right\|_4 \leq C \quad \text{and} \quad \sup_{1\leq t\leq T}\left\|T^{-\eta/2}z_t\right\|_4 \leq C \quad \forall T.$$

Proof: *See Appendix B.*

Lemma 2 *Under the assumptions of Proposition 1, we have under the null $\beta = 0$ that*

$$t_{vx} = \frac{\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\bar{\tilde{z}}\right)\left(u_t-\bar{u}\right)}{\sqrt{\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\bar{\tilde{z}}\right)^2 u_t^2}} + O_p\left(T^{-\eta/2}\right),$$

where $\tilde{z}_{t-1} = (1-\varrho L)_+^{-1}\Delta\tilde{x}_{t-1}$ with $\tilde{x}_t = \sum_{j=0}^{t-2}\rho^j \nu_{t-1-j}$ and $\bar{\tilde{z}}$ the sample average of $\tilde{z}_{t-1}$.

Proof: *See Appendix B.*

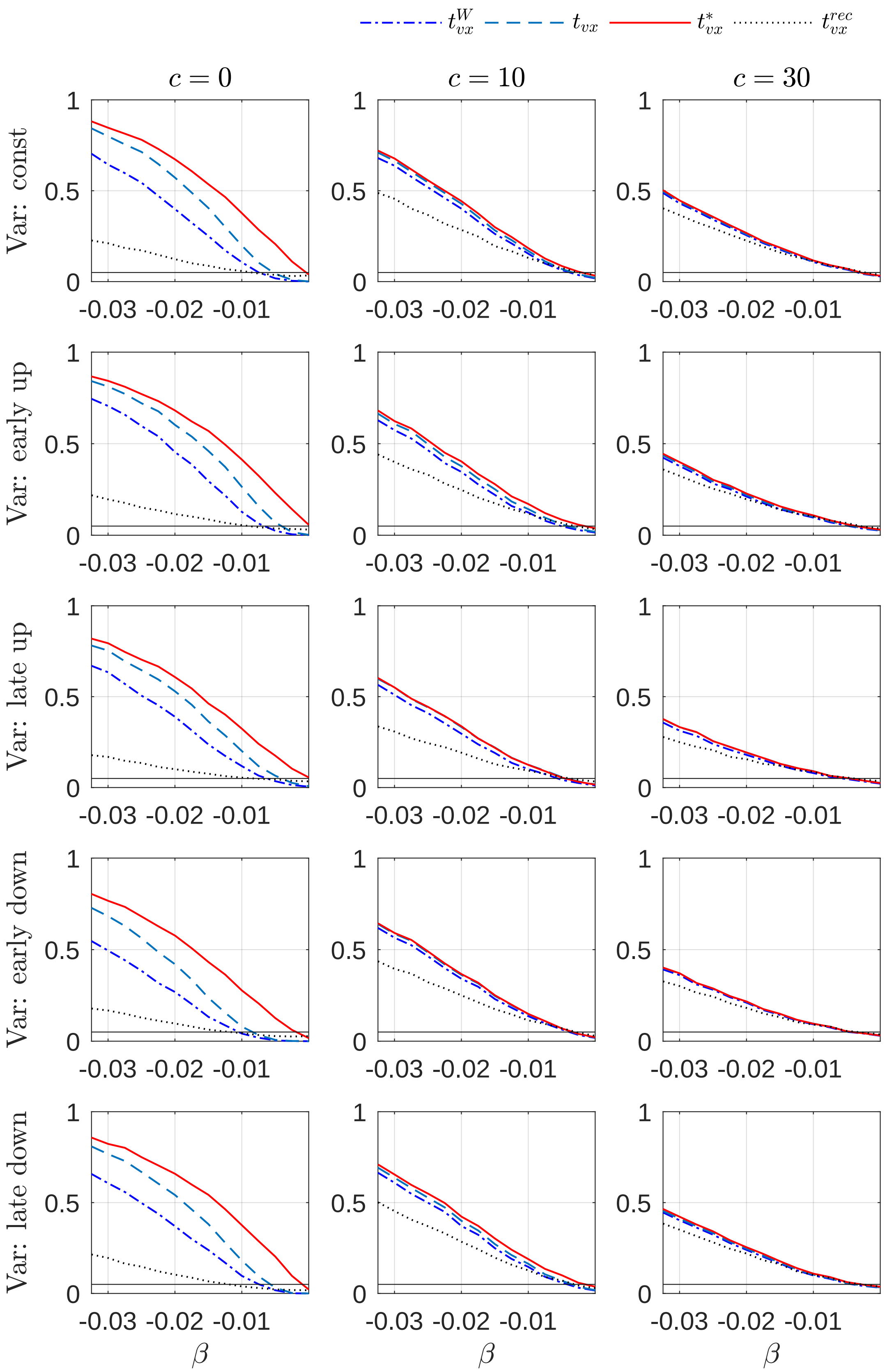

Figure 2: Power properties of different test statistics when $T = 250$ for $H_0 : \beta = 0$ vs $H_1 : \beta < 0$ with $\beta = \frac{b}{T}\sqrt{1-\delta^2}$ for $b \in \{-26, -24, ..., -2, 0\}$. Data generated with (1) and (2) with $v_t = \phi v_t + \nu_t$ for $\phi = 0.5$, where $(u_t, \nu_t) \sim iiN(0, \mathbf{\Sigma}_t)$ and $\mathbf{\Sigma}_t$ exhibits constant correlation $\delta = -0.95$ and time-varying variances. We set $\rho = 1 - {}^{c}/_{T}$ for various c and $\varrho = 1 - {}^{1}/_{T^{0.95}}$ and use standard normal critical values. See the text for details.

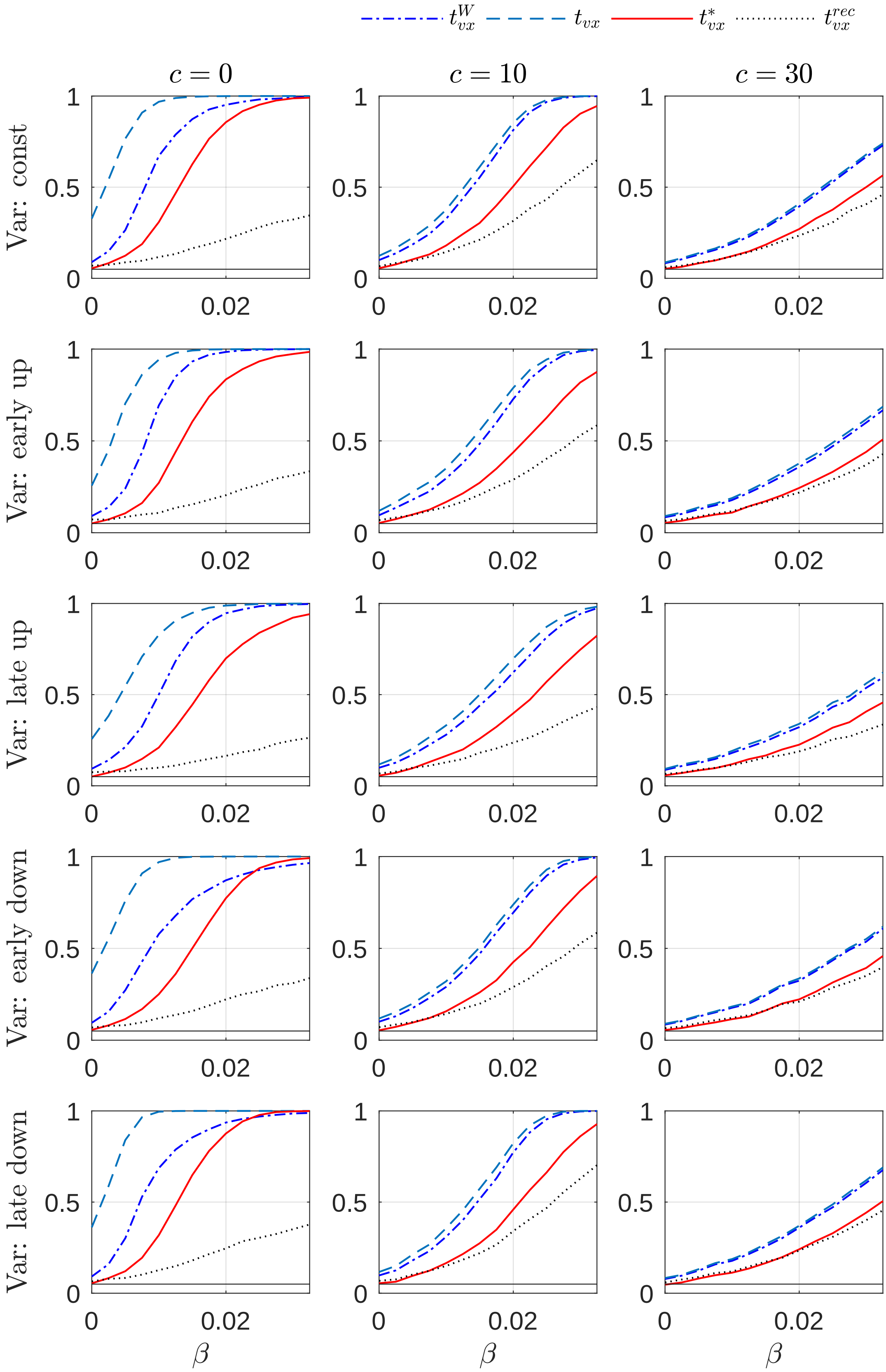

Figure 3: Power properties of different test statistics when $T = 250$ for $H_0 : \beta = 0$ vs $H_1 : \beta > 0$ with $\beta = \frac{b}{T}\sqrt{1-\delta^2}$ for $b \in \{0, 2, 4, ..., 26\}$. Data generated with (1) and (2) with $v_t = \phi v_t + \nu_t$ for $\phi = 0.5$, where $(u_t, \nu_t) \sim iiN(0, \mathbf{\Sigma}_t)$ and $\mathbf{\Sigma}_t$ exhibits constant correlation $\delta = -0.95$ and time-varying variances. We set $\rho = 1 - {}^{c}/_{T}$ for various c and $\varrho = 1 - {}^{1}/_{T^{0.95}}$ and use standard normal critical values. See the text for details.

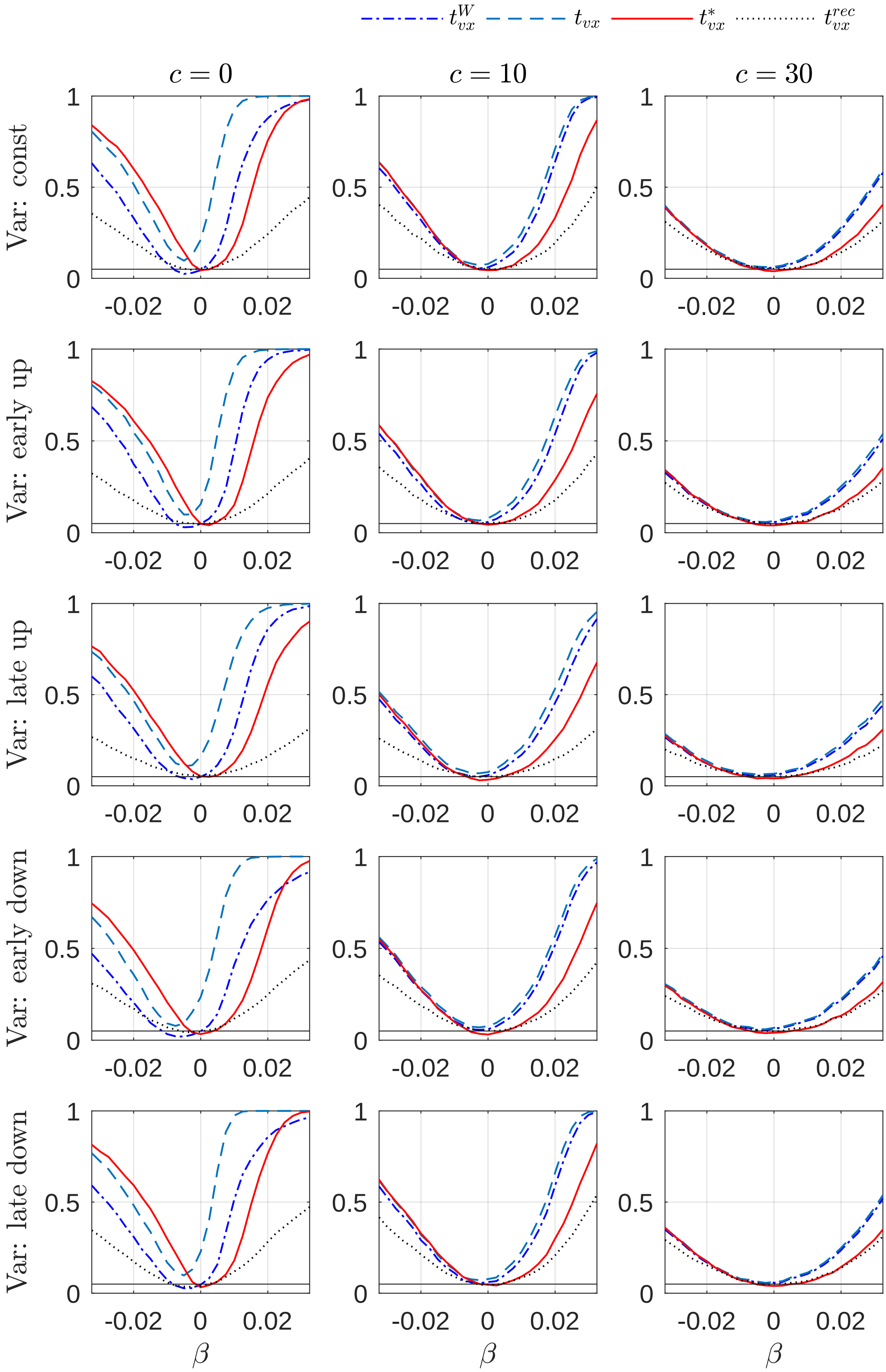

Figure 4: Power properties of different test statistics when $T = 250$ for $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$ with $\beta = \frac{b}{T}\sqrt{1-\delta^2}$ for $b \in \{-26, -24, ..., -2, 0, 2, , ..., 26\}$. Data generated with (1) and (2) with $v_t = \phi v_t + \nu_t$ for $\phi = 0.5$, where $(u_t, \nu_t) \sim iiN(0, \mathbf{\Sigma}_t)$ and $\mathbf{\Sigma}_t$ exhibits constant correlation $\delta = -0.95$ and time-varying variances. We set $\rho = 1 - {}^{c}/_{T}$ for various c and $\varrho = 1 - {}^{1}/_{T^{0.95}}$ and use standard normal critical values. See the text for details.

Lemma 3 *Under the assumptions of Proposition 1, we have*

$$T^{1/2-\eta/2}\,\mathrm{E}\left(B_T\right) \to -\frac{1}{\sqrt{2a}}\frac{\int_0^1 \sigma_{u\nu}\left(s\right)\sigma_u^2\left(s\right)\sigma_\nu^2\left(s\right)ds}{\sqrt{\left(\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s\right)^3}}.$$

Proof: *See the supplement.*

B Proofs

Proof of Lemma 1

Use the Phillips-Solo decomposition of v_t to write $v_t := \psi\nu_t + \Delta\tilde{v}_t$ where $\tilde{v}_t$ is a linear process in ν_t with absolutely summable coefficients (and as such uniformly L_r-bounded itself), leading to

$$x_t = \psi\sum_{j=0}^{t-1}\rho^j\nu_{t-j} + \tilde{v}_t - \rho^{t-1}\tilde{v}_1 + (\rho-1)\sum_{j=1}^{t-1}\rho^{j-1}\tilde{v}_{t-j}.$$

Hence, Minkowski's norm inequality gives

$$\left\|x_t\right\|_r = \psi\left\|\sum_{j=0}^{t-1}\rho^j\nu_{t-j}\right\|_r + \left\|\tilde{v}_t\right\|_r + \rho^{t-1}\left\|\tilde{v}_1\right\|_r + |\rho-1|\sum_{j=1}^{t-1}\rho^{j-1}\left\|\tilde{v}_{t-j}\right\|_r$$

such that $\left\|x_t\right\|_r \le \psi\left\|\sum_{j=0}^{t-1}\rho^j\nu_{t-j}\right\|_r + C$. For $r=4$ we have with the iid property of ν_t

$$\mathrm{E}\left(x_t^4\right) = \sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\sum_{k=0}^{t-1}\sum_{l=0}^{t-1}\rho^i\rho^j\rho^k\rho^l\,\mathrm{E}\left(\nu_{t-i}\nu_{t-j}\nu_{t-k}\nu_{t-l}\right) \le CT^2$$

and $T^{-1/2}x_t$ is thus uniformly L_4-bounded as required. A similar reasoning leads to the second part of the result.

Proof of Lemma 2

With $\tilde{x}_{t-1} = \sum_{j=0}^{t-2}\rho^j\nu_{t-1-j} = (1-\rho L)_+^{-1}\nu_{t-1}$ (i.e. the DGP without short run dynamics), use the Phillips-Solo decomposition of v_t (see the proof of Lemma 1), to obtain that

$$x_{t-1} = \sum_{j=0}^{t-2}\rho^j v_{t-1-j} = \psi\sum_{j=0}^{t-2}\rho^j\nu_{t-1-j} + \sum_{j=0}^{t-2}\rho^j\Delta\tilde{v}_{t-1-j}$$

such that $\Delta x_{t-1} = \psi\Delta\tilde{x}_{t-1} + \Delta q_{t-1}$ where

$$q_{t-1} = \sum_{j=0}^{t-2}\rho^j\Delta\tilde{v}_{t-1-j} = \tilde{v}_{t-1} - \rho^{t-3}\tilde{v}_1 + (\rho-1)\sum_{j=0}^{t-4}\rho^{j-1}\tilde{v}_{t-2-j}.$$

Note that q_t is uniformly L_4-bounded. Now, Demetrescu et al. (2018, Lemma A.1) show under slightly stricter assumptions than ours that $\sum_{t=2}^{T} z_{t-1} = O_p\left(T^{1/2+\eta}\right)$ such that $\bar{z} = O_p\left(T^{\eta-1/2}\right)$ and hence

$$\sum_{t=2}^{T}\left(\tilde{z}_{t-1} - \bar{z}\right)u_t \;=\; \sum_{t=2}^{T} z_{t-1}\left(u_t - \bar{u}\right) = \sum_{t=2}^{T} z_{t-1}u_t - \bar{u}\sum_{t=2}^{T} z_{t-1} = \Theta_p\left(T^{1/2+\eta/2}\right).$$

Moreover, $\sqrt{\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)^2\hat{u}_t^2}$ has the same order: using the OLS residuals $\hat{u}_t = u_t - x_{t-1}\left(\hat{\beta}-\beta\right)$, we obtain

$$\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)^2\hat{u}_t^2 = \sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)^2 u_t^2 - 2\left(\hat{\beta}-\beta\right)\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)^2 x_{t-1}u_t + \left(\hat{\beta}-\beta\right)^2\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)^2 x_{t-1}^2,$$

and, using Lemma 1, together with the superconsistency of the OLS estimator $\hat{\beta}$,[7] it follows that

$$\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)^2\hat{u}_t^2 = \sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)^2 u_t^2 + O_p\left(T^{1/2+\eta}\right).$$

Then, with

$$\begin{aligned}\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)^2 u_t^2 &= \frac{1}{T^{1+\eta}}\sum_{t=2}^{T} z_{t-1}^2 u_t^2 - \frac{2\bar{z}}{T^{1+\eta}}\sum_{t=2}^{T} z_{t-1}u_t^2 + \frac{\bar{z}^2}{T^{1+\eta}}\sum_{t=2}^{T} u_t^2 \\ &= \frac{1}{T^{1+\eta}}\sum_{t=2}^{T} z_{t-1}^2 u_t^2 + o_p(1)\end{aligned}$$

since $\sum_{t=2}^{T} z_{t-1}u_t^2 = \sum_{t=2}^{T} z_{t-1}\left(u_t^2-\sigma_{u,t}^2\right) + \sum_{t=2}^{T}\sigma_{u,t}^2 z_{t-1} = O_p\left(\max\left\{T^{1/2+\eta/2}; T^{1/2+\eta}\right\}\right)$.

At the same time, it is easily shown using the exact same arguments as in the proof of Lemma A.4 item 2 of Demetrescu and Rodrigues (2016) that

$$\frac{1}{T^{1+\eta}}\sum_{t=2}^{T} z_{t-1}^2 u_t^2 \xrightarrow{p} \frac{\psi^2}{2a}\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s > 0, \tag{4}$$

so it follows that

$$\frac{\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)u_t}{\sqrt{\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)^2\hat{u}_t^2}} = \frac{\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)u_t}{\sqrt{\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)^2 u_t^2}} + O_p\left(T^{-1/2}\right). \tag{5}$$

Note now that $z_{t-1} = \psi\tilde{z}_{t-1} + \sum_{j=0}^{t-3}\varrho^j\Delta q_{t-1-j} = \psi\tilde{z}_{t-1} + r_{t-1}$ where the sum $\sum_{j=0}^{t-3}\varrho^j\Delta q_{t-1-j}$ may be re-arranged to give

$$r_{t-1} = q_{t-1} - \varrho^{t-3}q_1 + (\varrho-1)\sum_{j=0}^{t-4}\varrho^{j-1}q_{t-1-j};$$

since q_t is uniformly L_4-bounded, it is easily seen that r_t is itself uniformly L_4-bounded. We now show that

$$\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)^2 u_t^2 = \frac{\psi^2}{T^{1+\eta}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\bar{\tilde{z}}\right)u_t^2 + O_p\left(T^{-\eta/2}\right) \tag{6}$$

and

$$\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\left(z_{t-1}-\bar{z}\right)u_t = \frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T} z_{t-1}\left(u_t-\bar{u}\right) = \frac{\psi}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}\left(u_t-\bar{u}\right) + O_p\left(T^{-\eta/2}\right). \tag{7}$$

[7]Showing that $T\left(\hat{\beta}-\beta\right) \Rightarrow \left(\psi\int_0^1 J_{c,\mathrm{H}}^2(s)\mathrm{d}s\right)^{-1}\int_0^1 J_{c,\mathrm{H}}(s)\mathrm{d}U_{\mathrm{H}}(s)$ is a standard exercise and we omit the details.

For (6), the difference between the terms on the l.h.s. and r.h.s. is given by

$$\frac{2\psi}{T^{1+\eta}}\sum_{t=2}^{T}\tilde{z}_{t-1}r_{t-1}u_t^2+\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}r_{t-1}^2u_t^2=A_{1T}+A_{2T}.$$

Then,

$$\mathrm{E}\left(|A_{1T}|\right)\leq\frac{C}{T^{1+\eta}}\sum_{t=2}^{T}\sqrt{\mathrm{E}\left(\left|\tilde{z}_{t-1}^2\right|\right)\mathrm{E}\left(\left|r_{t-1}^2\right|\right)}\,\mathrm{E}\left(\left|u_t^2\right|\right)=O\left(T^{-\eta/2}\right)$$

and, using L_r-boundedness again (and again), $A_{2T}=O_p\left(T^{-\eta}\right)$.

For (7), the vanishing term is given by

$$\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}r_{t-1}\left(u_t-\bar{u}\right)=\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}r_{t-1}u_t-\bar{u}\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}r_{t-1}=O_p\left(T^{-\eta/2}\right)$$

since the elements of the first sum have the md property and are uniformly L_4-bounded, while for the second summand we have $\left\|\sum_{t=2}^{T}r_{t-1}\right\|_4=O\left(T\right)$ but $\bar{u}=O_p\left(T^{-1/2}\right)$ as required.

Proof of Proposition 1

Begin by applying Lemma 2, and write

$$\frac{\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}\left(u_t-\bar{u}\right)}{\sqrt{\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\bar{\tilde{z}}\right)^2u_t^2}}=\frac{\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}u_t}{\sqrt{\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\bar{\tilde{z}}\right)^2u_t^2}}-\bar{u}\frac{\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}}{\sqrt{\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\bar{\tilde{z}}\right)^2u_t^2}}.$$

For the second term, we have (using the arguments in the proof of Lemma A.1 of Demetrescu et al., 2018) that

$$\frac{1}{T^{1/2+\eta}}\sum_{t=2}^{T}\tilde{z}_{t-1}=\frac{1}{T^{1/2}a}\tilde{x}_{T-1}+o_p(1)\Rightarrow\frac{1}{a}J_{c,\mathrm{H}}\left(1\right),$$

such that

$$T^{1/2-\eta/2}\frac{\bar{u}}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}\Rightarrow\frac{1}{a}U_{\mathrm{H}}\left(1\right)J_{c,\mathrm{H}}\left(1\right)$$

and the limit of C_T follows since $\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\bar{\tilde{z}}\right)^2u_t^2=\frac{1}{2a}\int_0^1\sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s+o_p\left(1\right)$; cf. Eq. (4).

For the first term, use a first-order Taylor series expansion with rest term in differential form for the function $\frac{a}{\sqrt{x}}$ about $x_0=\frac{1}{2a}\int_0^1\sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s$, where $x=\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\bar{\tilde{z}}\right)^2u_t^2$ and $a=\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}u_t$, such that

$$\frac{\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}u_t}{\sqrt{\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\bar{\tilde{z}}\right)^2u_t^2}}=Z_T+B_T+R_T$$

where

$$Z_T=\frac{\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}u_t}{\sqrt{\frac{1}{2a}\int_0^1\sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s}},\qquad(8)$$

$$B_T = -\frac{1}{2}\frac{1}{\sqrt{\left(\frac{1}{2a}\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s\right)^3}}\left(\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}u_t\right)\left(\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\tilde{z}_{t-1}^2u_t^2 - \frac{1}{2a}\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s\right) \tag{9}$$

and, with ξ_T between $\frac{1}{2a}\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s$ and $\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\bar{\tilde{z}}\right)^2 u_t^2$,

$$\begin{aligned} R_T &= \frac{3}{4}\frac{1}{\sqrt{\xi_T^5}}\left(\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}\left(u_t-\bar{u}\right)\right)\left(\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\bar{\tilde{z}}\right)^2u_t^2 - \frac{1}{2a}\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s\right)^2 \\ &\quad -\frac{1}{2}\frac{1}{\sqrt{\left(\frac{1}{2a}\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s\right)^3}}\left(\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\tilde{z}_{t-1}u_t\right)\left(\frac{-2}{T^{1+\eta}}\sum_{t=2}^{T}\tilde{z}_{t-1}\bar{\tilde{z}}u_t^2 + \frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\bar{\tilde{z}}^2u_t^2\right). \end{aligned}$$

Given that $\xi_T = O_p(1)$, $\frac{\bar{\tilde{z}}^2}{T^{1+\eta}}\sum_{t=2}^{T}u_t^2 = O_p\left(T^{\eta-1}\right) = \frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\tilde{z}_{t-1}\bar{\tilde{z}}u_t^2$, we have that $R_T = o_p\left(T^{\eta/2-1/2}\right)$. Using lemma (3) we have that $\mathrm{E}\left(B_T\right)$ is $O\left(T^{\eta/2-1/2}\right)$ but not $o\left(T^{\eta/2-1/2}\right)$, and R_T is therefore dominated by B_T and Z_T.

Then, Z_T has the required limiting standard normal (see Lemmata A.3 item (ii) and A.4 item 2 in Demetrescu and Rodrigues, 2016), and has zero mean for all T; the result follows with Lemma 3.

Proof of Proposition 3

Write under the null

$$t_{vx}^W = \frac{\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}z_{t-1}\left(u_t-\bar{u}\right)}{\sqrt{\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}z_{t-1}^2\hat{u}_t^2}} + \frac{1}{2}\frac{\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}z_{t-1}\left(u_t-\bar{u}\right)}{\sqrt{\xi_T}}\left(T^{-\eta}\bar{z}^2\hat{\omega}_{u|v}^2\right)$$

where $\xi_T = \frac{1}{T^{1+\eta}}\sum_{t=2}^{T}z_{t-1}^2\hat{u}_t^2 - \gamma_T T\bar{z}^2\hat{\omega}_{u|v}^2$ for $\gamma_T \in [0,1]$ is obviously $O_p(1)$. For the first summand, we obtain after some straightforward algebra the same expansion as used for establishing Proposition 1, with one important difference. Namely, the term analog to R_T in the proof of Proposition 1 will not have any components involving $\bar{z}$ from the denominator.

For the second summand is shown to vanish fast enough, such that the results follows. Recall that $\bar{z} = O_p\left(T^{\eta-1/2}\right)$, such that, not surprisingly, $\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}z_{t-1}\left(u_t-\bar{u}\right) = O_p(1)$ such that, as required, $T^{-\eta}\bar{z}^2\hat{\omega}_{u|v}^2 = O_p\left(T^{\eta-1}\right)$.

References

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

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

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

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

Cavaliere, G., 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.

Demetrescu, M., I. Georgiev, P. M. M. Rodrigues, and A. M. R. Taylor (2018). Testing for episodic predictability in stock returns. Working paper.

Demetrescu, M. and P. M. M. Rodrigues (2016). Residual-augmented IVX predictive regression. Working paper 5/2016, Banco de Portugal.

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.

Everaert, G. (2013). Orthogonal to backward mean transformation for dynamic panel data models. *The Econometrics Journal 16*(2), 179–221.

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

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

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

Loh, W.-Y. (1987). Calibrating confidence coefficients. *Journal of the American Statistical Association 82*(397), 155–162.

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.

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

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*(6), 1417–1426.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika 74*(3), 535–547.

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 T. Magdalinos (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics 136*(1), 115–130.

Shin, D. W. and B. S. So (2001). Recursive mean adjustment for unit root tests. *Journal of Time Series Analysis 22*(5), 595–612.

So, B. S. and D. W. Shin (1999). Recursive mean adjustment in time-series inferences. *Statistics & Probability Letters 43*(1), 65–73.

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

Westerlund, J., H. Karabiyik, and P. Narayan (2017). Testing for predictability in panels with general predictors. *Journal of Applied Econometrics 32*(3), 554–574.

Supplement to

Finite-sample size control of IVX-based tests in predictive regressions

Proof of Lemma 3

For computing the limit of $\mathrm{E}\left(B_T\right)$, note that

$$\begin{aligned}\Delta\tilde{x}_{t-1} &= (\rho-1)\,\tilde{x}_{t-2}+\nu_{t-1}\\ &= (\rho-1)\sum_{j=1}^{t-2}\rho^{j-1}\nu_{t-j-1}+\nu_{t-1}\end{aligned}$$

and thus

$$\begin{aligned}\tilde{z}_{t-1} &= (1-\varrho L)_+^{-1}\,\Delta\tilde{x}_{t-1}\\ &= (\rho-1)\,(1-\varrho L)_+^{-1}\sum_{j=1}^{t-2}\rho^{j-1}\nu_{t-j-1}+(1-\varrho L)_+^{-1}\,\nu_{t-1}\\ &= (\rho-1)\,(1-\varrho L)_+^{-1}\sum_{j=1}^{t-2}\rho^{j-1}\nu_{t-j-1}+\sum_{j=0}^{t-2}\varrho^{j}\nu_{t-j-1}\\ &= \sum_{j=1}^{t-1}\frac{\varrho^{t-1-j}\,(1-\varrho)-\rho^{t-1-j}\,(1-\rho)}{\rho-\varrho}\nu_j\equiv\sum_{j=1}^{t-1}c_{j,t-1}\nu_j\end{aligned}$$

by defining $c_{t-1,t-1}=1$. We may then focus on $\mathrm{E}\left(\sum_{t=2}^{T}\tilde{z}_{t-1}u_t\cdot\sum_{t=2}^{T}\tilde{z}_{t-1}^2u_t^2\right)=S_0$,

$$\begin{aligned}S_0 &= \mathrm{E}\left(\sum_{t=2}^{T}\tilde{z}_{t-1}^3u_t^3\right)+\mathrm{E}\left(\sum_{t=2}^{T}\sum_{s=2}^{t-1}\tilde{z}_{t-1}u_t\tilde{z}_{s-1}^2u_s^2\right)+\mathrm{E}\left(\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\tilde{z}_{t-1}u_t\tilde{z}_{s-1}^2u_s^2\right)\\ &= \mathrm{E}\left(\sum_{t=2}^{T}\tilde{z}_{t-1}^3u_t^3\right)+\mathrm{E}\left(\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\tilde{z}_{t-1}u_t\tilde{z}_{s-1}^2u_s^2\right),\end{aligned}$$

since $\mathrm{E}\left(\sum_{t=2}^{T}\tilde{z}_{t-1}u_t\right)=0$.

Let $S_0=S_{0,1}+S_{0,2}$, with $S_{0,1}=\mathrm{E}\left(\sum_{t=2}^{T}\tilde{z}_{t-1}^3u_t^3\right)$ and $S_{0,2}=\mathrm{E}\left(\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\tilde{z}_{t-1}u_t\tilde{z}_{s-1}^2u_s^2\right)$. Let's work out $S_{0,2}$ first. Recall that

$$\tilde{z}_{t-1}=\sum_{j=1}^{t-1}c_{j,t-1}\upsilon_j\text{ with }c_{j,t-1}=\frac{\varrho^{t-1-j}\,(1-\varrho)-\rho^{t-1-j}\,(1-\rho)}{\rho-\varrho}.$$

Using the independence of shocks we obtain

$$\begin{aligned}\mathrm{E}\left(\tilde{z}_{t-1}u_t\tilde{z}_{s-1}^2\right) &= \mathrm{E}\left(\left(\sum_{j=1}^{t-1}c_{j,t-1}\upsilon_j\right)u_t\left(\sum_{j=1}^{t-1}c_{j,s-1}\upsilon_j+\sum_{j=t}^{s-1}c_{j,s-1}\upsilon_j\right)^2\right)\\ &= 2\,\mathrm{E}\left(\left(\sum_{j=1}^{t-1}c_{j,t-1}\upsilon_j\right)u_t\left(\sum_{j=1}^{t-1}c_{j,s-1}\upsilon_j\right)\left(\sum_{j=t}^{s-1}c_{j,s-1}\upsilon_j\right)\right)\\ &= 2\sigma_{u\upsilon}^2\left(\frac{t}{T}\right)c_{t,s-1}\sum_{j=1}^{t-1}c_{j,t-1}c_{j,s-1}\sigma_\upsilon^2\left(\frac{j}{T}\right)\end{aligned}$$

which implies

$$\mathrm{E}\left(\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\tilde{z}_{t-1}u_t\tilde{z}_{s-1}^2u_s^2\right)=2\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\sigma_{uv}\left(\frac{t}{T}\right)c_{t,s-1}c_{j,t-1}c_{j,s-1}\sigma_v^2\left(\frac{j}{T}\right).$$

Now since $c_{j,t-1}=\frac{\varrho^{t-1-j}(1-\varrho)-\rho^{t-1-j}(1-\rho)}{\rho-\varrho}=\frac{1-\varrho}{\rho-\varrho}\left(\varrho^{t-1-j}-\frac{1-\rho}{1-\varrho}\rho^{t-1-j}\right)$ we have

$$\begin{aligned}c_{t,s-1}c_{j,t-1}c_{j,s-1} &= \varrho^{2s-3-2j}-\frac{1-\rho}{1-\varrho}\varrho^{s-2-j}\rho^{s-1-j}-\frac{1-\rho}{1-\varrho}\varrho^{2s-2-t-j}\rho^{t-1-j} \\ &\quad +\left(\frac{1-\rho}{1-\varrho}\right)^2\varrho^{s-1-t}\rho^{t+s-2-2j}-\frac{1-\rho}{1-\varrho}\varrho^{t+s-2-2j}\rho^{s-1-t} \\ &\quad +\left(\frac{1-\rho}{1-\varrho}\right)^2\rho^{2s-2-t-j}\varrho^{t-1-j} \\ &\quad +\left(\frac{1-\rho}{1-\varrho}\right)^2\rho^{s-2-j}\varrho^{s-1-j}-\left(\frac{1-\rho}{1-\varrho}\right)^3\rho^{2s-2j-3} \\ &= \sum_{k=1}^{8}\alpha_k,\end{aligned} \tag{10}$$

where we have droppped the term $\left(\frac{1-\varrho}{\rho-\varrho}\right)^3=1+O\left(T^{\eta-1}\right)$. Further, those terms that have a factor of $\left(\frac{1-\rho}{1-\varrho}\right)^2$ and the last term with $\left(\frac{1-\rho}{1-\varrho}\right)^3$ are clearly dominated by the other terms and hence vanish. In the following we will, therefore, look only at the terms that are associated with α_1, α_2, α_3 and α_5.

We start by focusing on the first term stemming from $c_{t,s-1}c_{j,t-1}c_{j,s-1}$, namely $\varrho^{2s-2j-3}$.

$$\begin{aligned}S_{0,2}^{(\alpha_1)} &= \sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\sigma_{uv}\left(\frac{t}{T}\right)\varrho^{s-1-t}\varrho^{t-1-j}\varrho^{s-1-j}\sigma_v^2\left(\frac{j}{T}\right) \\ &= \sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\varrho^{2s-2j-3}\sigma_v^2\left(\frac{j}{T}\right).\end{aligned}$$

Note here that $1<j<t<s\leq T$ hence

$$\begin{aligned}\left|\sigma_u^2\left(\frac{s}{T}\right)-\sigma_u^2\left(\frac{t}{T}\right)\right| &< C\frac{|s-t|}{T}<C\frac{|s-j|}{T}, \\ \left|\sigma_v^2\left(\frac{j}{T}\right)-\sigma_v^2\left(\frac{t}{T}\right)\right| &< C\frac{|t-j|}{T}<C\frac{|s-j|}{T},\end{aligned}$$

and

$$\left|\sigma_u^2\left(\frac{s}{T}\right)-\sigma_u^2\left(\frac{t}{T}\right)\right|\left|\sigma_v^2\left(\frac{j}{T}\right)-\sigma_v^2\left(\frac{t}{T}\right)\right|<C\frac{(s-j)^2}{T^2},$$

therefore

$$\begin{aligned}&\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\varrho^{2s-2j-3}\sigma_v^2\left(\frac{j}{T}\right) \\ =\ &\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\left(\sigma_u^2\left(\frac{s}{T}\right)-\sigma_u^2\left(\frac{t}{T}\right)+\sigma_u^2\left(\frac{t}{T}\right)\right)\varrho^{2s-2j-3}\left(\sigma_v^2\left(\frac{j}{T}\right)-\sigma_v^2\left(\frac{t}{T}\right)+\sigma_v^2\left(\frac{t}{T}\right)\right) \\ =\ &\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\left(\sigma_u^2\left(\frac{s}{T}\right)-\sigma_u^2\left(\frac{t}{T}\right)\right)\left(\sigma_v^2\left(\frac{j}{T}\right)-\sigma_v^2\left(\frac{t}{T}\right)\right)\varrho^{2s-2j-3} \\ &+\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\left(\sigma_u^2\left(\frac{s}{T}\right)-\sigma_u^2\left(\frac{t}{T}\right)\right)\sigma_v^2\left(\frac{t}{T}\right)\varrho^{2s-2j-3}\end{aligned}$$

$$+\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{t}{T}\right)\varrho^{2s-2j-3}\left(\sigma_v^2\left(\frac{j}{T}\right)-\sigma_v^2\left(\frac{t}{T}\right)\right)$$
$$+\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{t}{T}\right)\varrho^{2s-2j-3}\left(\sigma_v^2\left(\frac{t}{T}\right)\right),$$

which implies that

$$\begin{aligned}
&\left|\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\varrho^{2s-2j-3}\sigma_v^2\left(\frac{j}{T}\right)-\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2j-3}\right|\\
\leq\ &\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\left|\sigma_u^2\left(\frac{s}{T}\right)-\sigma_u^2\left(\frac{t}{T}\right)\right|\left|\sigma_v^2\left(\frac{j}{T}\right)-\sigma_v^2\left(\frac{t}{T}\right)\right|\varrho^{2s-2j-3}\\
&+\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\left|\sigma_u^2\left(\frac{s}{T}\right)-\sigma_u^2\left(\frac{t}{T}\right)\right|\sigma_v^2\left(\frac{t}{T}\right)\varrho^{2s-2j-3}\\
&+\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{t}{T}\right)\left|\sigma_v^2\left(\frac{j}{T}\right)-\sigma_v^2\left(\frac{t}{T}\right)\right|\varrho^{2s-2j-3}\\
\leq\ &\frac{C}{T^2}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-j)^2\varrho^{2s-2j-3}+\frac{C}{T}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-t)\varrho^{2s-2j-3}+\frac{C}{T}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2j-3}(j-t).
\end{aligned}$$

Now note that

$$\begin{aligned}
\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-j)^2\varrho^{2s-2j} &= -\frac{2\varrho^4\left(4+7\varrho^2+\varrho^4\right)\left(1-\varrho^{2T}\right)}{\left(1-\varrho^2\right)^5}+\frac{2\varrho^4\left(2+\varrho^2\right)+\varrho^{2+2T}\left(1+12\varrho^2+5\varrho^4\right)}{\left(1-\varrho^2\right)^4}T\\
&\quad+\frac{2\varrho^{2+2T}\left(1+2\varrho^2\right)}{\left(1-\varrho^2\right)^3}T^2+\frac{\varrho^{2+2T}}{\left(1-\varrho^2\right)^2}T^3\\
&= O\left(T^{4\eta+1}\right),
\end{aligned}$$

and in the same way we can show that

$$\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-t)\varrho^{2s-2j-3}=O\left(T^{3\eta+1}\right)=\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2j-3}(j-t).$$

Therefore

$$\begin{aligned}
&\left|\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\varrho^{2s-2j-3}\sigma_v^2\left(\frac{j}{T}\right)-\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2j-3}\right|\\
\leq\ &\frac{C}{T^2}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-j)^2\varrho^{2s-2j-3}+\frac{C}{T}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-t)\varrho^{2s-2j-3}+\frac{C}{T}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2j-3}(j-t)\\
=\ &O\left(T^{4\eta-1}\right)+O\left(T^{3\eta}\right)=O\left(T^{3\eta}\right),\text{ since }3\eta>4\eta-1.
\end{aligned}$$

Therefore

$$\begin{aligned}
&T^{1/2-\eta/2}\frac{1}{T^{1/2+\eta/2}}\frac{1}{T^{1+\eta}}\times\\
&\left|\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\varrho^{2s-2j-3}\sigma_v^2\left(\frac{j}{T}\right)-\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2j-3}\right|\\
=\ &T^{1/2-\eta/2}O\left(T^{-\frac{3}{2}\eta-\frac{3}{2}}\right)O\left(T^{3\eta}\right)=O\left(T^{\eta-1}\right).
\end{aligned}$$

Next observe that

$$\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2j}=\frac{\varrho^{4}-\varrho^{2t+2}+\varrho^{2T+2}-\varrho^{2T-2t+4}}{\left(1-\varrho^{2}\right)^{2}}.$$

Hence we have

$$\begin{aligned}
&\frac{1}{T^{1/2+\eta/2}}\frac{1}{T^{1+\eta}}\sum_{t=2}^{T-1}\sigma_{uv}^{2}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2j}\\
=\;&\frac{1}{T^{3/2+3\eta/2}}\frac{\varrho^{4}}{\left(1-\varrho\right)^{2}\left(1+\varrho\right)^{2}}\sum_{t=2}^{T-1}\sigma_{uv}^{2}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)\\
&-\frac{1}{T^{3/2+3\eta/2}}\frac{1}{\left(1-\varrho^{2}\right)^{2}}\sum_{t=2}^{T-1}\sigma_{uv}^{2}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)\varrho^{2t+2}\\
&+\frac{1}{T^{3/2+3\eta/2}}\frac{\varrho^{2T+2}}{\left(1-\varrho^{2}\right)^{2}}\sum_{t=2}^{T-1}\sigma_{uv}^{2}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)\\
&-\frac{1}{T^{3/2+3\eta/2}}\frac{1}{\left(1-\varrho^{2}\right)^{2}}\sum_{t=2}^{T-1}\sigma_{uv}^{2}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)\varrho^{2T-2t+4}\\
=\;&\frac{1}{T^{3/2+3\eta/2}}\frac{\left(1-aT^{-\eta}\right)^{4}}{\left(aT^{-\eta}\right)^{2}\left(2+aT^{-\eta}\right)^{2}}\sum_{t=2}^{T-1}\sigma_{uv}^{2}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)+R_{1}^{(\alpha_1)}+R_{2}^{(\alpha_1)}+R_{3}^{(\alpha_1)}.
\end{aligned}$$

For $R_1^{(\alpha_1)}$ note that

$$\begin{aligned}
\left|R_{1}^{(\alpha_1)}\right| &= \frac{1}{T^{3/2+3\eta/2}}\frac{1}{\left(1-\varrho^{2}\right)^{2}}\sum_{t=2}^{T-1}\sigma_{uv}^{2}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)\varrho^{2t+2}\\
&\leq \frac{C}{T^{3/2-\eta/2}}\sum_{t=2}^{T-1}\varrho^{2t}=O\left(T^{-3/2+3\eta/2}\right).
\end{aligned}$$

$R_2^{(\alpha_1)}=\frac{1}{T^{3/2+3\eta/2}}\frac{\varrho^{2T+2}}{(1-\varrho^2)^2}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)$ is clearly dominated by $R_1^{(\alpha_1)}$ and for $R_3^{(\alpha_1)}$ we have

$$\begin{aligned}
\left|R_{3}^{(\alpha_1)}\right| &= \frac{1}{T^{3/2+3\eta/2}}\frac{1}{\left(1-\varrho^{2}\right)^{2}}\sum_{t=2}^{T-1}\sigma_{uv}^{2}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)\varrho^{2T-2t+4}\\
&\leq \frac{C}{T^{3/2-\eta/2}}\sum_{t=2}^{T-1}\varrho^{2T-2t}=O\left(T^{-3/2+3\eta/2}\right).
\end{aligned}$$

Further, note that $T^{1/2-\eta/2}\left|R_1^{(\alpha_1)}+R_2^{(\alpha_1)}+R_3^{(\alpha_1)}\right|=O\left(T^{-1+\eta}\right)$, hence

$$\begin{aligned}
&T^{1/2-\eta/2}\frac{1}{T^{1/2+\eta/2}}\frac{1}{T^{1+\eta}}\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2j-3}\\
=\;&T^{1/2-\eta/2}\frac{1}{T^{1/2+\eta/2}}\frac{1}{T^{1+\eta}}\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)\frac{\varrho^{4}-\varrho^{2t+2}+\varrho^{2T+2}-\varrho^{2T-2t+4}}{\left(1-\varrho^{2}\right)^{2}}\\
=\;&T^{1/2-\eta/2}\frac{1}{T^{1/2+3\eta/2}}\frac{\varrho}{\left(1-\varrho^{2}\right)^{2}}\frac{1}{T}\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)+O\left(T^{\eta-1}\right)\\
=\;&\frac{1}{a^{2}}\frac{\left(1-aT^{-\eta}\right)}{\left(2-aT^{-\eta}\right)^{2}}\frac{1}{T}\sum_{t=2}^{T-1}\sigma_{uv}\left(\frac{t}{T}\right)\sigma_{u}^{2}\left(\frac{t}{T}\right)\sigma_{v}^{2}\left(\frac{t}{T}\right)+O\left(T^{\eta-1}\right)\\
\to\;&\frac{1}{4a^{2}}\int_{0}^{1}\sigma_{uv}\left(x\right)\sigma_{u}^{2}\left(x\right)\sigma_{v}^{2}\left(x\right)dx.
\end{aligned}$$

We now turn to the second term stemming from $c_{t,s-1}c_{j,t-1}c_{j,s-1}$ and given as α_2 below (10), leading to the analysis of

$$S_{0,2}^{(\alpha_2)} = -\frac{1-\rho}{1-\varrho}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\sigma_{uv}\left(\frac{t}{T}\right)\varrho^{s-2-j}\rho^{s-1-j}\sigma_v^2\left(\frac{j}{T}\right).$$

Following the same lines of arguments as for $S_{0,2}^{(\alpha_1)}$ and noting that

$$\begin{aligned}\frac{C}{T}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-t)\,\varrho^{s-2-j}\rho^{s-1-j} &= O\left(T^{3\eta}\right)\\ \frac{C}{T}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(j-t)\,\varrho^{s-2-j}\rho^{s-1-j} &= O\left(T^{3\eta}\right)\end{aligned}$$

and as

$$\frac{C}{T^2}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-j)^2\,\varrho^{s-2-j}\rho^{s-1-j} = O\left(T^{5\eta-2}\right)$$

we have that

$$T^{1/2-\eta/2}\frac{1}{T^{1/2+\eta/2}}\frac{1}{T^{1+\eta}}\cdot\frac{1}{\varrho^2\rho}\times$$
$$\left|\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)(\varrho\rho)^{s-j}\sigma_v^2\left(\frac{j}{T}\right)-\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(\varrho\rho)^{s-j}\right|$$

is of order $T^{1/2-\eta/2}O\left(T^{-\frac{3}{2}\eta-\frac{3}{2}}\right)O\left(T^{4\eta-1}\right) = O\left(T^{\eta-1}\right)$.

Next observe that

$$\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{s-2-j}\rho^{s-1-j} = \frac{\rho-\rho^t\varrho^{t-1}+\varrho^{T-1}\rho^T-\varrho^{T-t}\rho^{T-t+1}}{(1-\rho\varrho)^2}.$$

Therefore

$$\begin{aligned}&\frac{1}{T^{1/2+\eta/2}}\frac{1}{T^{1+\eta}}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{s-2-j}\rho^{s-1-j}\\
=\;&\frac{1}{T^{3/2+3\eta/2}}\frac{\rho}{(1-\rho\varrho)^2}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\\
&-\frac{1}{T^{3/2+3\eta/2}}\frac{1}{(1-\rho\varrho)^2}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\rho^t\varrho^{t-1}\\
&+\frac{1}{T^{3/2+3\eta/2}}\frac{\varrho^{T-1}\rho^T}{(1-\rho\varrho)^2}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\\
&-\frac{1}{T^{3/2+3\eta/2}}\frac{1}{(1-\rho\varrho)^2}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\varrho^{T-t}\rho^{T-t+1}\\
=\;&\frac{1}{T^{3/2+3\eta/2}}\frac{\rho}{(1-\rho\varrho)^2}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)+R_1^{(\alpha_2)}+R_2^{(\alpha_2)}+R_3^{(\alpha_2)},\end{aligned}$$

for which again one can elementarily show that $R_1^{(\alpha_2)}$, $R_2^{(\alpha_2)}$ and $R_3^{(\alpha_2)}$ are $o(1)$ and since $\frac{\rho}{(1-\rho\varrho)^2} = O\left(T^{2\eta}\right)$, we have

$$\frac{T^{1/2-\eta/2}}{T^{3/2+3\eta/2}}\frac{\rho}{(1-\rho\varrho)^2}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\to\frac{1}{a^2}\int_0^1\sigma_{uv}(x)\,\sigma_u^2(x)\,\sigma_v^2(x)\,dx.$$

Since $\frac{1-\rho}{1-\varrho} = O\left(T^{\eta-1}\right)$, we obtain $T^{1/2-\eta/2} S_{0,2}^{(\alpha_2)} = O\left(T^{\eta-1}\right)$.

We now turn to the third term stemming from $c_{t,s-1}c_{j,t-1}c_{j,s-1}$ and given as α_3 below (10), $\alpha_3 = \frac{1-\rho}{1-\varrho}\varrho^{2s-2-t-j}\rho^{t-1-j}$, i.e. we analyze

$$S_{0,2}^{(\alpha_3)} = -\frac{1-\rho}{1-\varrho}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\sigma_{uv}\left(\frac{t}{T}\right)\varrho^{2s-2-t-j}\rho^{t-1-j}\sigma_v^2\left(\frac{j}{T}\right).$$

Here note that

$$\begin{aligned}
\frac{C}{T}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-t)\,\varrho^{2s-2-t-j}\rho^{t-1-j} &= O\left(T^{3\eta}\right)\\
\frac{C}{T}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(j-t)\,\varrho^{2s-2-t-j}\rho^{t-1-j} &= O\left(T^{3\eta}\right)
\end{aligned}$$

and as

$$\frac{C}{T^2}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-j)^2\,\varrho^{2s-2-t-j}\rho^{t-1-j} = O\left(T^{5\eta-2}\right)$$

we have

$$\begin{aligned}
& T^{1/2-\eta/2}\frac{1}{T^{1/2+\eta/2}}\frac{1}{T^{1+\eta}}\times\\
& \left|\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\varrho^{2s-2-t-j}\rho^{t-1-j}\sigma_v^2\left(\frac{j}{T}\right)-\right.\\
& \left.\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2-t-j}\rho^{t-1-j}\right|\\
= \;& T^{1/2-\eta/2}O\left(T^{-\frac{3}{2}\eta-\frac{3}{2}}\right)O\left(T^{3\eta}\right) = O\left(T^{\eta-1}\right).
\end{aligned}$$

Next observe that

$$\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2-t-j}\rho^{t-1-j} = \frac{\left(1-\varrho^{2T-2t}\right)\left(\varrho\rho-\varrho^t\rho^t\right)}{\left(1-\varrho^2\right)\rho\left(1-\varrho\rho\right)}.$$

Therefore

$$\begin{aligned}
& \frac{1}{T^{1/2+\eta/2}}\frac{1}{T^{1+\eta}}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{2s-2-t-j}\rho^{t-1-j}\\
= \;& \frac{1}{T^{3/2+3\eta/2}}\frac{\varrho}{\left(1-\varrho^2\right)\rho\left(1-\varrho\rho\right)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\\
& -\frac{1}{T^{3/2+3\eta/2}}\frac{1}{\left(1-\varrho^2\right)\rho\left(1-\varrho\rho\right)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\rho^t\varrho^t\\
& +\frac{1}{T^{3/2+3\eta/2}}\frac{1}{\left(1-\varrho^2\right)\rho\left(1-\varrho\rho\right)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\varrho^{2T-2t}\rho^t\varrho^t\\
& -\frac{1}{T^{3/2+3\eta/2}}\frac{\varrho}{\left(1-\varrho^2\right)\left(1-\varrho\rho\right)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\varrho^{2T-2t}\\
= \;& \frac{1}{T^{3/2+3\eta/2}}\frac{\varrho}{\left(1-\varrho^2\right)\left(1-\varrho\rho\right)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)+R_1^{(\alpha_3)}+R_2^{(\alpha_3)}+R_3^{(\alpha_3)},
\end{aligned}$$

for which again one can elementarily show that $R_1^{(\alpha_3)}$, $R_2^{(\alpha_3)}$ and $R_3^{(\alpha_3)}$ are $o\left(1\right)$ and since $\frac{\varrho}{(1-\varrho^2)\rho(1-\varrho\rho)} =$

$O\left(T^{2\eta}\right)$, we have

$$\frac{T^{1/2-\eta/2}}{T^{3/2+3\eta/2}}\frac{\varrho}{(1-\varrho^2)(1-\rho^2)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\to\frac{1}{a^2}\int_0^1\sigma_{uv}(x)\,\sigma_u^2(x)\,\sigma_v^2(x)\,dx.$$

Now as $\frac{1-\rho}{1-\varrho}=O\left(T^{\eta-1}\right)$ we obtain $T^{1/2-\eta/2}S_{0,2}^{(\alpha_3)}=O\left(T^{\eta-1}\right)$.

Finally we turn to the α_5 given under (10): $\frac{1-\rho}{1-\varrho}\varrho^{t+s-2-2j}\rho^{s-1-t}$

$$S_{0,2}^{(\alpha_5)}=-\frac{1-\rho}{1-\varrho}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\sigma_{uv}\left(\frac{t}{T}\right)\varrho^{t+s-2-2j}\rho^{s-1-t}\sigma_v^2\left(\frac{j}{T}\right).$$

Here note that

$$\frac{C}{T}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-t)\,\varrho^{t+s-2-2j}\rho^{s-1-t}=O\left(T^{3\eta}\right)=\frac{C}{T}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(j-t)\,\varrho^{t+s-2-2j}\rho^{s-1-t},$$

and as $\frac{C}{T^2}\sum_{t=2}^{T-1}\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}(s-j)^2\varrho^{t+s-2-2j}\rho^{s-1-t}=O\left(T^{5\eta-2}\right)$ we have we have

$$\begin{aligned}
& T^{1/2-\eta/2}\frac{1}{T^{1/2+\eta/2}}\frac{1}{T^{1+\eta}}\times\\
& \left|\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\sigma_u^2\left(\frac{s}{T}\right)\varrho^{t+s-2-2j}\rho^{s-1-t}\sigma_v^2\left(\frac{j}{T}\right)\right.\\
& \left.-\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{t+s-2-2j}\rho^{s-1-t}\right|\\
= & \; T^{1/2-\eta/2}O\left(T^{-\frac{3}{2}\eta-\frac{3}{2}}\right)O\left(T^{3\eta}\right)=O\left(T^{\eta-1}\right).
\end{aligned}$$

Next observe that

$$\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{t+s-2-2j}\rho^{s-1-t}=\frac{\left(\varrho-\varrho^{2t-1}\right)\left(1-\varrho^{T-t}\rho^{T-t}\right)}{(1-\varrho^2)(1-\varrho\rho)}.$$

Therefore

$$\begin{aligned}
& \frac{1}{T^{1/2+\eta/2}}\frac{1}{T^{1+\eta}}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\sum_{s=t+1}^{T}\sum_{j=1}^{t-1}\varrho^{t+s-2-2j}\rho^{s-1-t}\\
= & \frac{1}{T^{3/2+3\eta/2}}\frac{\varrho}{(1-\varrho^2)(1-\varrho\rho)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\\
& -\frac{1}{T^{3/2+3\eta/2}}\frac{1}{(1-\varrho^2)(1-\varrho\rho)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\varrho^{2t-1}\\
& +\frac{1}{T^{3/2+3\eta/2}}\frac{1}{(1-\varrho^2)(1-\varrho\rho)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\varrho^{T+t-1}\rho^{T-t}\\
& -\frac{1}{T^{3/2+3\eta/2}}\frac{\varrho}{(1-\varrho^2)(1-\varrho\rho)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\varrho^{T-t}\rho^{T-t}\\
= & \frac{1}{T^{3/2+3\eta/2}}\frac{\varrho}{(1-\varrho^2)(1-\varrho\rho)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)+R_1^{(\alpha_5)}+R_2^{(\alpha_5)}+R_3^{(\alpha_5)},
\end{aligned}$$

for which again one can elementarily show that $R_1^{(\alpha_5)}$, $R_2^{(\alpha_5)}$ and $R_3^{(\alpha_5)}$ are $o(1)$ and since $\frac{\varrho}{(1-\varrho^2)(1-\varrho\rho)}=$

$O\left(T^{2\eta}\right)$, we have

$$\frac{T^{1/2-\eta/2}}{T^{3/2+3\eta/2}}\frac{\varrho}{\left(1-\varrho^2\right)\left(1-\rho^2\right)}\sum_{t=2}^{T-1}\sigma_{uv}^2\left(\frac{t}{T}\right)\sigma_u^2\left(\frac{t}{T}\right)\sigma_v^2\left(\frac{t}{T}\right)\to\frac{1}{a^2}\int_0^1\sigma_{uv}\left(x\right)\sigma_u^2\left(x\right)\sigma_v^2\left(x\right)dx.$$

Now as $\frac{1-\rho}{1-\varrho}=O\left(T^{\eta-1}\right)$ we obtain $T^{1/2-\eta/2}S_{0,2}^{(\alpha_5)}=O\left(T^{\eta-1}\right)$.

Putting all the terms together we obtain

$$T^{1/2-\eta/2}E\left(S_{0,2}\right)\to\frac{1}{4a^2}\int_0^1\sigma_{uv}\left(x\right)\sigma_u^2\left(x\right)\sigma_v^2\left(x\right)dx.$$

We may now turn to $S_{0,1}$. Assuming Lipschitz continuity for the third moments we have

$$\begin{aligned}S_{0,1} &= \sum_{t=2}^{T}\sum_{j=1}^{t-1}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}c_{j,t-1}c_{k,t-1}c_{l,t-1}\,\mathrm{E}\left(v_jv_kv_l\right)\mathrm{E}\left(u_t^3\right)\\ &= \sum_{t=2}^{T}\sum_{j=1}^{t-1}c_{j,t-1}^3\,\mathrm{E}\left(v_j^3\right)\mathrm{E}\left(u_t^3\right).\end{aligned}$$

Now again as for the analysis of $S_{0,2}$ we have to look at all the terms stemming from the expansion of $c_{j,t-1}^3$. We start by considering the effect of the first term, namely ϱ^{3t-3j}.

$$\begin{aligned}&\sum_{t=2}^{T}\sum_{j=1}^{t-1}\varrho^{3t-3j}\,\mathrm{E}\left(v_j^3\right)\mathrm{E}\left(u_t^3\right)\\ =&\sum_{t=2}^{T}\sigma_u^3\left(\frac{t}{T}\right)\sum_{j=1}^{t-1}\varrho^{3t-3j}\sigma_v^3\left(\frac{j}{T}\right)\\ =&\sum_{t=2}^{T}\sigma_u^3\left(\frac{t}{T}\right)\sum_{j=1}^{t-1}\varrho^{3t-3j}\left(\sigma_v^3\left(\frac{j}{T}\right)-\sigma_v^3\left(\frac{t}{T}\right)+\sigma_v^3\left(\frac{t}{T}\right)\right)\\ =&\sum_{t=2}^{T}\sigma_u^3\left(\frac{t}{T}\right)\sum_{j=1}^{t-1}\varrho^{3t-3j}\left(\sigma_v^3\left(\frac{j}{T}\right)-\sigma_v^3\left(\frac{t}{T}\right)\right)+\sum_{t=2}^{T}\sigma_u^3\left(\frac{t}{T}\right)\sum_{j=1}^{t-1}\varrho^{3t-3j}\sigma_v^3\left(\frac{t}{T}\right)\end{aligned}$$

which implies that

$$\begin{aligned}&\left|\sum_{t=2}^{T}\sum_{j=1}^{t-1}\varrho^{3t-3j}\,\mathrm{E}\left(v_j^3\right)\mathrm{E}\left(u_t^3\right)-\sum_{t=2}^{T}\sigma_u^3\left(\frac{t}{T}\right)\sigma_v^3\left(\frac{t}{T}\right)\sum_{j=1}^{t-1}\varrho^{3t-3j}\right|\\ \leq&\sum_{t=2}^{T}\sigma_u^3\left(\frac{t}{T}\right)\sum_{j=1}^{t-1}\varrho^{3t-3j}\left|\sigma_v^3\left(\frac{j}{T}\right)-\sigma_v^3\left(\frac{t}{T}\right)\right|\\ \leq&\frac{C}{T}\sum_{t=2}^{T}\sum_{j=1}^{t-1}\varrho^{3t-3j}\left(t-j\right)=O\left(T^{3\eta}\right)\end{aligned}$$

On the other hand $\sum_{j=1}^{t-1}\varrho^{3t-3j}=\frac{\varrho^3-\varrho^{3t}}{1-\varrho^3}$, hence

$$\begin{aligned}&T^{1/2-\eta/2}\times\frac{1}{T^{1/2+\eta/2}}\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\sum_{j=1}^{t-1}\varrho^{3t-3j}\,\mathrm{E}\left(v_j^3\right)\mathrm{E}\left(u_t^3\right)\\ =&T^{-2\eta}\frac{1}{T}\sum_{t=2}^{T}\sigma_u^3\left(\frac{t}{T}\right)\sigma_v^3\left(\frac{t}{T}\right)\sum_{j=1}^{t-1}\varrho^{3t-3j}+O\left(T^{\eta-1}\right)\\ =&T^{-2\eta}\frac{1}{T}\sum_{t=2}^{T}\sigma_u^3\left(\frac{t}{T}\right)\sigma_v^3\left(\frac{t}{T}\right)\frac{\varrho^3-\varrho^{3t}}{1-\varrho^3}+O\left(T^{\eta-1}\right)\end{aligned}$$

$$= \frac{\varrho^3}{1-\varrho^3} T^{-2\eta} \frac{1}{T} \sum_{t=2}^{T} \sigma_u^3 \left(\frac{t}{T}\right) \sigma_v^3 \left(\frac{t}{T}\right) - \frac{1}{1-\varrho^3} T^{-2\eta} \frac{1}{T} \sum_{t=2}^{T} \sigma_u^3 \left(\frac{t}{T}\right) \sigma_v^3 \left(\frac{t}{T}\right) \varrho^{3t} + O\left(T^{\eta-1}\right)$$

First note that $\frac{\varrho^3}{1-\varrho^3} T^{-2\eta} \frac{1}{T} \sum_{t=2}^{T} \sigma_u^3 \left(\frac{t}{T}\right) \sigma_v^3 \left(\frac{t}{T}\right) = O\left(T^{-\eta}\right)$. Further

$$\left| \sum_{t=2}^{T} \sigma_u^3 \left(\frac{t}{T}\right) \sigma_v^3 \left(\frac{t}{T}\right) \varrho^{3t} \right| \leq C \left| \sum_{t=2}^{T} \varrho^{3t} \right| = O\left(T^{\eta}\right).$$

Therefore $T^{1/2-\eta/2} \times \frac{1}{T^{1/2+\eta/2}} \frac{1}{T^{1+\eta}} \sum_{t=2}^{T} \sum_{j=1}^{t-1} \varrho^{3t-3j} \mathrm{E}\left(v_j^3\right) \mathrm{E}\left(u_t^3\right) = o(1)$. The other terms stemming from expanding $c_{j,t-1}^3$ can be shown, in a similar way, to vanish.

Proof of Proposition 2

We prove first that, the null $\beta = 0$, the following holds:

$$t_{vx}^{rec} = \frac{\sum_{t=2}^{T} \left(\tilde{z}_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} \tilde{z}_j\right) \left(u_t - \frac{\sum_{j=t}^{T} u_j}{T-t+1}\right)}{\sqrt{\sum_{t=2}^{T} \left(\tilde{z}_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} \tilde{z}_j\right)^2 u_t^2}} + O_p\left(T^{-\eta/2}\right),$$

where $\tilde{z}_{t-1} = (1-\varrho L)_+^{-1} \Delta \tilde{x}_{t-1}$ with $\tilde{x}_t = \sum_{j=0}^{t-2} \rho^j \nu_{t-1-j}$.
Using arguments like those given in the proof of Lemma 2, we conclude that

$$\frac{\sum_{t=2}^{T} \left(z_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} z_j\right) \left(u_t - \frac{\sum_{j=t}^{T} u_j}{T-t+1}\right)}{\sqrt{\sum_{t=2}^{T} \left(z_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} z_j\right)^2 \widehat{u}_t^2}} = \frac{\sum_{t=2}^{T} \left(z_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} z_j\right) \left(u_t - \frac{\sum_{j=t}^{T} u_j}{T-t+1}\right)}{\sqrt{\sum_{t=2}^{T} \left(z_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} z_j\right)^2 u_t^2}} + O_p\left(T^{-1/2}\right).$$

Using $z_{t-1} = \psi \tilde{z}_{t-1} + r_{t-1}$ from Lemma (2) we have that

$$\begin{aligned}
&\frac{1}{T^{1+\eta}} \sum_{t=2}^{T} \left(z_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} z_j\right)^2 u_t^2 = \\
&= \frac{1}{T^{1+\eta}} \sum_{t=2}^{T} \left(\psi \tilde{z}_{t-1} + r_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} (\psi \tilde{z}_j + r_j)\right)^2 u_t^2 \\
&= \frac{\psi^2}{T^{1+\eta}} \sum_{t=2}^{T} \left(\tilde{z}_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} \tilde{z}_j\right)^2 u_t^2 + \frac{1}{T^{1+\eta}} \sum_{t=2}^{T} \left(r_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} r_j\right)^2 u_t^2 \\
&\quad + \frac{2\psi}{T^{1+\eta}} \sum_{t=2}^{T} \left(\tilde{z}_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} \tilde{z}_j\right) \left(r_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} r_j\right) u_t^2. \\
&= \frac{\psi^2}{T^{1+\eta}} \sum_{t=2}^{T} \left(\tilde{z}_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} \tilde{z}_j\right)^2 u_t^2 + A_{1T} + A_{2T}.
\end{aligned}$$

Examine A_{1T} first,

$$A_{1T} = \frac{1}{T^{1+\eta}} \sum_{t=2}^{T} r_{t-1}^2 u_t^2 + \frac{1}{T^{1+\eta}} \sum_{t=2}^{T} \left(\frac{1}{t-1} \sum_{j=1}^{t-1} r_j\right)^2 u_t^2 - \frac{2}{T^{1+\eta}} \sum_{t=2}^{T} \frac{r_{t-1}}{t-1} \sum_{j=1}^{t-1} r_j u_t^2.$$

Thanks to the independence of r_{t-1} and u_t and since r_t is uniformly L_4−bounded, we have

$$\mathrm{E}\left(\left| \frac{1}{T^{1+\eta}} \sum_{t=2}^{T} r_{t-1}^2 u_t^2 \right|\right) \leq \frac{C}{T^{1+\eta}} \sum_{t=2}^{T} \sqrt{\mathrm{E}\left(r_{t-1}^4\right)} \mathrm{E}\left(u_t^2\right) = O\left(T^{-\eta}\right).$$

For the second term of A_{1T}, note that

$$\begin{aligned} \mathrm{E}\left(\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\frac{1}{t-1}\sum_{j=1}^{t-1}r_j\right)^2 u_t^2\right) &= \frac{\sigma_u^2}{T^{1+\eta}}\sum_{t=2}^{T}\frac{1}{(t-1)^2}\mathrm{E}\left(\left(\sum_{j=1}^{t-1}r_j\right)^2\right) \\ &\leq \frac{C}{T^{\eta}}, \end{aligned}$$

while, for the last term, we have

$$\begin{aligned} \frac{2}{T^{1+\eta}}\mathrm{E}\left(\left|\sum_{t=2}^{T}\frac{r_{t-1}}{t-1}\sum_{j=1}^{t-1}r_j u_t^2\right|\right) &\leq \frac{2\sigma_u^2}{T^{1+\eta}}\sum_{t=2}^{T}\frac{1}{t-1}\sqrt{\mathrm{E}\left(r_{t-1}^2\right)}\sqrt{\mathrm{E}\left(\left(\sum_{j=1}^{t-1}r_j\right)^2\right)} \\ &= O\left(T^{-\eta}\right). \end{aligned}$$

Moving on to A_{2T}, write

$$\begin{aligned} \mathrm{E}(|A_{2T}|) &\leq \frac{2\psi}{T^{1+\eta}}\mathrm{E}\left(\left|\sum_{t=2}^{T}\tilde{z}_{t-1}r_{t-1}u_t^2\right|\right) + \frac{2\psi}{T^{1+\eta}}\mathrm{E}\left(\left|\sum_{t=2}^{T}\frac{\tilde{z}_{t-1}u_t^2}{t-1}\sum_{j=1}^{t-1}r_j\right|\right) \\ &\quad + \frac{2\psi}{T^{1+\eta}}\mathrm{E}\left(\left|\sum_{t=2}^{T}\frac{r_{t-1}u_t^2}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right|\right) + \frac{2\psi}{T^{1+\eta}}\mathrm{E}\left(\left|\sum_{t=2}^{T}\frac{1}{(t-1)^2}\sum_{j=1}^{t-1}\tilde{z}_j\sum_{j=1}^{t-1}r_j u_t^2\right|\right) \\ &\leq \frac{2\psi}{T^{1+\eta}}\sum_{t=2}^{T}\sqrt{\mathrm{E}\left(\tilde{z}_{t-1}^2\right)\mathrm{E}\left(r_{t-1}^2\right)}\,\mathrm{E}(u_t^2) + \frac{2\psi}{T^{1+\eta}}\sum_{t=2}^{T}\sqrt{\mathrm{E}\left(\tilde{z}_{t-1}^2\right)\mathrm{E}(u_t^4)}\sqrt{\frac{\mathrm{E}\left(\left(\sum_{j=1}^{t-1}r_j\right)^2\right)}{(t-1)^2}} \\ &\quad + \frac{2\psi}{T^{1+\eta}}\sum_{t=2}^{T}\sqrt{\mathrm{E}(r_{t-1}^2)\,\mathrm{E}(u_t^4)}\sqrt{\frac{\mathrm{E}\left(\left(\sum_{j=1}^{t-1}\tilde{z}_j\right)^2\right)}{(t-1)^2}} \\ &\quad + \frac{2\psi}{T^{1+\eta}}\sum_{t=2}^{T}\sqrt{\frac{\mathrm{E}\left(\left(\sum_{j=1}^{t-1}\tilde{z}_j\right)^2\right)}{(t-1)^2}\frac{\mathrm{E}\left(\left(\sum_{j=1}^{t-1}r_j\right)^2\right)}{(t-1)^2}\mathrm{E}(u_t^4)} \\ &= O\left(T^{-\eta/2}\right) \end{aligned}$$

where we have everywhere used the L_4 boundedness and also $\mathrm{Var}\left(\sum_{j=1}^{t-1}\tilde{z}_j\right) \leq CtT^{2\eta}$ which can easily be established using the expansion for $\tilde{z}_{t-1}$ given under the proof of Proposition 1. Therefore

$$\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(z_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}z_j\right)^2 u_t^2 = \frac{\psi^2}{T^{1+\eta}}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2 u_t^2 + O_p\left(T^{-\eta/2}\right). \qquad (11)$$

Furthermore

$$\begin{aligned} &\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\left(z_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}z_j\right)\left(u_t-\frac{\sum_{j=t}^{T}u_j}{T-t+1}\right) \\ &= \frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\left(\psi\tilde{z}_{t-1}+r_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\left(\psi\tilde{z}_j+r_j\right)\right)\left(u_t-\frac{\sum_{j=t}^{T}u_j}{T-t+1}\right) \\ &= \frac{1}{T^{1/2+\eta/2}}\psi\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)\left(u_t-\frac{\sum_{j=t}^{T}u_j}{T-t+1}\right) \end{aligned}$$

$$+\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\left(r_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}r_j\right)\left(u_t-\frac{\sum_{j=t}^{T}u_j}{T-t+1}\right).$$

We now re-arrange the sum terms to exploit the serial independence of u_t, leading after some algebra exploiting the L_r boundedness of r_t to

$$\mathrm{Var}\left(\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\left(r_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}r_j\right)\left(u_t-\frac{\sum_{j=t}^{T}u_j}{T-t+1}\right)\right)=O\left(T^{-\eta}\right)$$

and therefore

$$\begin{aligned}&\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\left(z_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}z_j\right)\left(u_t-\frac{\sum_{j=t}^{T}u_j}{T-t+1}\right)\\&=\frac{1}{T^{1/2+\eta/2}}\psi\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)\left(u_t-\frac{\sum_{j=t}^{T}u_j}{T-t+1}\right)+O_p\left(T^{-\eta/2}\right).\end{aligned}\tag{12}$$

The desired representation follows from equations (11) and (12).

Write now

$$\begin{aligned}\frac{\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)\left(u_t-\frac{\sum_{j=t}^{T}u_j}{T-t+1}\right)}{\sqrt{\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2u_t^2}}&=\frac{\sum_{t=2}^{T}\tilde{z}_{t-1}u_t}{\sqrt{D_T}}-\frac{\sum_{t=2}^{T}\frac{1}{T-t+1}\tilde{z}_{t-1}\sum_{j=t}^{T}u_j}{\sqrt{D_T}}\\&\quad-\frac{\sum_{t=2}^{T}\frac{1}{t-1}u_t\sum_{j=1}^{t-1}\tilde{z}_j}{\sqrt{D_T}}+\frac{\sum_{t=2}^{T}\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\frac{1}{T-t+1}\sum_{j=t}^{T}u_j}{\sqrt{D_T}}\\&=\frac{\sum_{t=2}^{T}\tilde{z}_{t-1}u_t}{\sqrt{D_T}}+\frac{C_{1,T}}{\sqrt{D_T}}+\frac{C_{2,T}}{\sqrt{D_T}}+\frac{C_{3,T}}{\sqrt{D_T}}.\end{aligned}$$

First let's look at the variance of $C_{1,T}=\sum_{t=2}^{T}\frac{1}{T-t+1}\tilde{z}_{t-1}\sum_{j=t}^{T}u_j$:

$$\begin{aligned}\mathrm{Var}\left(C_{1,T}\right)&=\mathrm{Var}\left(\sum_{j=2}^{T}u_j\sum_{t=2}^{j}\frac{\tilde{z}_{t-1}}{T-t+1}\right)\\&\le\max_{s\in[0,1]}\sigma_u^2(s)\sum_{j=2}^{T}\mathrm{Var}\left(\sum_{t=2}^{j}\frac{\tilde{z}_{t-1}}{T-t+1}\right)\\&=\max_{s\in[0,1]}\sigma_u^2(s)\sum_{j=2}^{T}\mathrm{Var}\left(\sum_{t=2}^{j}\sum_{k=1}^{t-1}\frac{c_{k,t-1}}{T-t+1}\nu_k\right)\end{aligned}$$

with $c_{k,t}$ from the proof of Proposition 1. Hence

$$\begin{aligned}\mathrm{Var}\left(C_{1,T}\right)&\le\max_{s\in[0,1]}\sigma_u^2(s)\sum_{j=2}^{T}\mathrm{Var}\left(\sum_{k=1}^{j-1}\nu_k\sum_{t=k+1}^{j}\frac{c_{k,t-1}}{T-t+1}\right)\\&=\max_{s\in[0,1]}\sigma_u^2(s)\max_{s\in[0,1]}\sigma_\nu^2(s)\sum_{j=2}^{T}\sum_{k=1}^{j-1}\left(\sum_{t=k+1}^{j}\frac{c_{k,t-1}}{T-t+1}\right)^2\\&=O\left(\left(\frac{1-\varrho}{\rho-\varrho}\right)^2\sum_{j=2}^{T}\sum_{k=1}^{j-1}\left(\sum_{t=k+1}^{j}\frac{\varrho^{t-1-k}}{T-t+1}\right)^2\right).\end{aligned}$$

With

$$\begin{aligned}
\sum_{j=2}^{T}\sum_{k=1}^{j-1}\left(\sum_{t=k+1}^{j}\frac{\varrho^{t-1-k}}{T-t+1}\right)^{2} &\leq \sum_{j=2}^{T}\sum_{k=1}^{j-1}\left(\sum_{t=1}^{T}\frac{\varrho^{t-1-k}}{T-t+1}\right)^{2}\\
&= \sum_{j=2}^{T}\sum_{k=1}^{j-1}\varrho^{-k}\left(\sum_{t=1}^{T}\frac{\varrho^{t-1}}{T-t+1}\right)^{2}\\
&= \sum_{j=2}^{T}\sum_{k=1}^{j-1}\varrho^{-k}\left(\varrho^{T}\sum_{t=1}^{T-1}\frac{\varrho^{-t}}{t}+\frac{1}{T}\right)^{2}\\
&= \sum_{j=2}^{T}\sum_{k=1}^{j-1}\varrho^{-k}\left(\varrho^{T}\frac{1}{T}\sum_{t=1}^{T-1}\frac{\left(\varrho^{T}\right)^{-t/T}}{t/T}+\frac{1}{T}\right)^{2}\\
&= \sum_{j=2}^{T}\sum_{k=1}^{j-1}\varrho^{-k}\left(\varrho^{T}O\left(\int_{1/T}^{1-1/T}\frac{\left(\varrho^{-T}\right)^{x}}{x}\right)+\frac{1}{T}\right)^{2},
\end{aligned}$$

and, with $\mathrm{Ei}\left(x\right)=\int_{x}^{\infty}\frac{\exp(-z)}{z}\mathrm{d}z$,

$$\begin{aligned}
\int_{1/T}^{1-1/T}\frac{\left(\varrho^{-T}\right)^{x}}{x}\mathrm{d}x &= \mathrm{Ei}\left(\left(1-\frac{1}{T}\right)\log\left(\varrho^{-T}\right)\right)-\mathrm{Ei}\left(\frac{1}{T}\log\left(\varrho^{-T}\right)\right)\\
&= O\left(\log T\right),
\end{aligned}$$

leading to

$$\begin{aligned}
&\sum_{j=2}^{T}\sum_{k=1}^{j-1}\varrho^{-k}\left(\varrho^{T}O\left(\int_{1/T}^{1-1/T}\frac{\left(\varrho^{-T}\right)^{x}}{x}\right)+\frac{1}{T}\right)^{2}\\
&= O\left(\log^{2}T\right)\sum_{j=2}^{T}\sum_{k=1}^{j-1}\varrho^{2T-k}\\
&= O\left(\log^{2}T\right)\sum_{j=2}^{T}\frac{\varrho^{2T}-\varrho^{2T-j+1}}{\varrho-1}\\
&= O\left(\log^{2}T\right)\left(\frac{\left(T-1\right)\varrho^{2T}}{\varrho-1}+\frac{\varrho}{\varrho-1}\frac{\varrho^{2T-1}-\varrho^{T}}{\varrho-1}\right)\\
&= O\left(Te^{-aT^{1-\eta}}\log^{2}T\right),
\end{aligned}$$

which in turn implies that

$$T^{-1/2-\eta}C_{1,T}=o\left(1\right).$$

Now we turn to the variance of $C_{2,T}=\sum_{t=2}^{T}\frac{1}{t-1}u_{t}\sum_{j=1}^{t-1}\tilde{z}_{j}$:

$$\begin{aligned}
\mathrm{Var}\left(C_{2,T}\right) &= \sum_{t=2}^{T}\mathrm{Var}\left(\sum_{j=1}^{t-1}\frac{1}{t-1}u_{t}\tilde{z}_{j}\right)\\
&\leq \max_{s\in[0,1]}\sigma_{u}^{2}(s)\sum_{t=2}^{T}\frac{1}{\left(t-1\right)^{2}}\mathrm{E}\left[\left(\sum_{j=1}^{t-1}\sum_{k=1}^{j}c_{k,j}\nu_{k}\right)^{2}\right]\\
&\leq \max_{s\in[0,1]}\sigma_{u}^{2}(s)\max_{s\in[0,1]}\sigma_{\nu}^{2}(s)\sum_{t=2}^{T}\frac{1}{\left(t-1\right)^{2}}\sum_{k=1}^{t-1}\left(\sum_{j=k}^{t-1}c_{k,j}\right)^{2}
\end{aligned}$$

like above. Since $\sum_{j=k}^{t-1} c_{k,j} = \frac{\rho^{t-k}-\varrho^{t-k}}{\rho-\varrho}$, we obtain

$$\begin{aligned}
\sum_{k=1}^{t-1}\left(\sum_{j=k}^{t-1} c_{k,j}\right)^2 &= \sum_{k=1}^{t-1}\left(\frac{\rho^{t-k}-\varrho^{t-k}}{\rho-\varrho}\right)^2 \\
&= \frac{1}{(\rho-\varrho)^2}\sum_{k=1}^{t-1}\left(\rho^{2t-2k}+\varrho^{2t-2k}-2(\rho\varrho)^{t-k}\right) \\
&= \frac{1}{(\rho-\varrho)^2}\left(\frac{\rho^2-\rho^{2t}}{1-\rho^2}+\frac{\varrho^2-\varrho^{2t}}{1-\varrho^2}-2\frac{(\varrho\rho)^2-(\rho\varrho)^{2t}}{1-(\rho\varrho)^2}\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
\sum_{t=2}^{T}\frac{1}{(t-1)^2}\frac{\rho^{2-2t}}{(\rho-\varrho)^2} &= \frac{\rho^2}{(\rho-\varrho)^2}\frac{1}{T^2}\sum_{t=1}^{T-1}\frac{\left(\rho^{-2T}\right)^{t/T}}{(t/T)^2} \\
&= O\left(T^{2\eta-1}\int_{1/T}^{1-1/T}\frac{\left(\rho^{-2T}\right)^x}{x^2}dx\right) \\
&= O\left(T^{2\eta}\right),
\end{aligned}$$

where

$$\begin{aligned}
\int_{1/T}^{1-1/T}\frac{\left(\rho^{-2T}\right)^x}{x^2}dx &= \log\left(\rho^{-2T}\right)\operatorname{Ei}\left(x\log\left(\rho^{-2T}\right)\right)-\frac{\rho^{-2Tx}}{x}\Big|_{1/T}^{1-1/T} \\
&= \log\left(\rho^{-2T}\right)\operatorname{Ei}\left(\left(1-\frac{1}{T}\right)\log\left(\rho^{-2T}\right)\right)-\frac{T\rho^{-2(T-1)}}{T-1} \\
&\quad -\log\left(\rho^{-2T}\right)\operatorname{Ei}\left(\frac{1}{T}\log\left(\rho^{-2T}\right)\right)-T\rho^{-2} \\
&= O(T).
\end{aligned}$$

Also we have that

$$\begin{aligned}
\int_{1/T}^{1-1/T}\frac{\left(\varrho^{-2T}\right)^x}{x^2}dx &= \log\left(\varrho^{-2T}\right)\operatorname{Ei}\left(\left(1-\frac{1}{T}\right)\log\left(\varrho^{-2T}\right)\right)-\frac{T\varrho^{-2(T-1)}}{T-1} \\
&\quad -\log\left(\varrho^{-2T}\right)\operatorname{Ei}\left(\frac{1}{T}\log\left(\varrho^{-2T}\right)\right)-T\varrho^{-2} \\
&= O(T)
\end{aligned}$$

The other term is of the same order, implying that

$$T^{-\eta-1/2}C_{2,T} = O_p\left(T^{-1/2}\right).$$

Finally, we look at the variance of $C_{3,T} = \sum_{t=2}^{T}\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\frac{1}{T-t+1}\sum_{j=t}^{T}u_j$:

$$\operatorname{Var}\left(C_{3,T}\right) \leq \max_{s\in[0,1]}\sigma_u^2(s)\max_{s\in[0,1]}\sigma_\nu^2(s)\sum_{t=2}^{T}\frac{1}{(t-1)^2}\operatorname{E}\left[\left(\sum_{j=1}^{t-1}\tilde{z}_j\right)^2\right],$$

which can be treated as $C_{2,T}$, leading to

$$T^{-\eta-1/2}C_{3,T} = O_p\left(T^{-1/2}\right).$$

For D_T, appearing in the denominator we have

$$\begin{aligned}
T^{-\eta-1}D_T &= T^{-\eta-1}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2\left(u_t^2-\sigma_u^2(t/T)\right) \\
&\quad +T^{-\eta-1}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2\sigma_u^2(t/T) \\
&= D_{1,T}+D_{2,T}
\end{aligned}$$

Examine

$$\begin{aligned}
D_{2,T} &= T^{-\eta-1}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2\sigma_u^2\left(\frac{t}{T}\right) \\
&= \frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\tilde{z}_{t-1}^2\sigma_u^2\left(\frac{t}{T}\right)+\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2\sigma_u^2\left(\frac{t}{T}\right)-\frac{2}{T^{1+\eta}}\sum_{t=2}^{T}\frac{\tilde{z}_{t-1}}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\sigma_u^2\left(\frac{t}{T}\right).
\end{aligned}$$

It is not difficult to show (using for example arguments analogous to the derivations for $C_{1,T}$) that, for all $2\leq t\leq T$ and a suitable constant C,

$$\mathrm{Var}\left(\sum_{j=1}^{t-1}\tilde{z}_j\right)\leq CtT^{2\eta}.$$

Therefore,

$$\mathrm{E}\left(\left|T^{-\eta-1}\sum_{t=2}^{T}\left(\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2\sigma_u^2\left(\frac{t}{T}\right)\right|\right)\leq\frac{C}{T^{\eta+1}}\sum_{t=2}^{T}\mathrm{Var}\left(\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)\leq\frac{C}{T^{1-\eta}}\sum_{t=2}^{T}\frac{1}{t-1}$$

and, thanks to Markov's inequality,

$$T^{-\eta-1}\sum_{t=2}^{T}\left(\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2\sigma_u^2\left(\frac{t}{T}\right)=O_p\left(T^{\eta-1}\log T\right).$$

Moreover, recall that $T^{-\eta/2}\tilde{z}_{t-1}$ is uniformly L_4-bounded, hence, with the Cauchy-Schwarz inequality, we have that

$$\mathrm{E}\left(\left|\frac{\tilde{z}_{t-1}}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right|\right)\leq\frac{CT^{\eta/2}}{t-1}\sqrt{\mathrm{E}\left(\left(\sum_{j=1}^{t-1}\tilde{z}_j\right)^2\right)}\leq\frac{CT^{3\eta/2}}{\sqrt{t}}$$

and

$$\mathrm{E}\left(\left|T^{-\eta-1}\sum_{t=2}^{T}\frac{\tilde{z}_{t-1}}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\sigma_u^2\left(\frac{t}{T}\right)\right|\right)\leq\frac{CT^{3\eta/2}}{T^{\eta+1}}\sum_{t=2}^{T}\frac{1}{\sqrt{t}}=CT^{\eta/2-1/2}.$$

Furthermore, it is not difficult to show that $D_{1,T}\overset{p}{\to}0$ thanks to the md property of the summands. We thus have that

$$T^{-\eta-1}D_T\overset{p}{\to}\frac{1}{2a}\int_0^1\sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s$$

and consequently

$$\frac{C_{1,T}}{\sqrt{D_T}}+\frac{C_{2,T}}{\sqrt{D_T}}+\frac{C_{3,T}}{\sqrt{D_T}}=O_p\left(T^{-1/2}\right).$$

We are now left to consider the asymptotic behavior of $\frac{\sum_{t=2}^{T}\tilde{z}_{t-1}u_t}{\sqrt{D_T}}$. To do so we follow the proof of

Proposition 1 and write

$$\frac{T^{-1/2-\eta/2}\sum_{t=2}^{T}\tilde{z}_{t-1}u_t}{\sqrt{T^{-1-\eta}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2 u_t^2}} = Z_T + B_T + R_T$$

where Z_T and B_T are the same as equations (8) and (9) which have been treated in the proof of Proposition 1, and

$$\begin{aligned} R_T &= \frac{3}{4}\frac{T^{-1/2-\eta/2}\sum_{t=2}^{T}\tilde{z}_{t-1}u_t}{\sqrt{\xi_T^5}}\left(T^{-1-\eta}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2 u_t^2 - \frac{1}{2a}\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s\right)^2 \\ &\quad - \frac{T^{-1/2-\eta/2}\sum_{t=2}^{T}\tilde{z}_{t-1}u_t}{2\sqrt{\left(\frac{1}{2a}\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s\right)^3}}\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\left(\left(\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2 u_t^2 - \frac{2u_t^2\tilde{z}_{t-1}}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right) \end{aligned}$$

with ξ_T between $\frac{1}{2a}\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s$ and $T^{-1-\eta}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2 u_t^2$, i.e. $\xi_T = O_p(1)$. The term B_T vanishes due to the component $\left(T^{-1-\eta}\sum_{t=2}^{T}\left(\tilde{z}_{t-1}-\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2 u_t^2 - \frac{1}{2a}\int_0^1 \sigma_u^2(s)\sigma_\nu^2(s)\mathrm{d}s\right)$, appearing squared in the first summand of R_T which is therefore, in turn, dominated by B_T. To analyze the second summand, examine in turn

$$\mathrm{E}\left(\left|\left(\frac{1}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right)^2 u_t^2\right|\right) \le \frac{CT^{2\eta}}{t-1}$$

thanks to independence of $\tilde{z}_j$ and u_t for $j \le t-1$, and, like above,

$$\mathrm{E}\left(\left|\frac{2u_t^2\tilde{z}_{t-1}}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right|\right) \le \frac{C}{t-1}\mathrm{E}\left(\left|\tilde{z}_{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right|\right) = \frac{C}{t-1}\sum_{j=1}^{t-1}\mathrm{E}\left(\left|\tilde{z}_{t-1}\tilde{z}_j\right|\right).$$

Some algebra indicates that

$$\mathrm{E}\left(\left|\tilde{z}_{t-1}\tilde{z}_j\right|\right) \le CT^{\eta}\varrho^{t-j-1}$$

such that, summing up,

$$\mathrm{E}\left(\left|\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\frac{2u_t^2\tilde{z}_{t-1}}{t-1}\sum_{j=1}^{t-1}\tilde{z}_j\right|\right) \le \frac{CT^{\eta}}{T^{1+\eta}}\sum_{t=2}^{T}\frac{1}{t-1}\sum_{j=1}^{t-1}\varrho^{t-j-1} = \frac{C}{T}\sum_{t=2}^{T}\frac{1}{t-1}\frac{1-\varrho^t}{1-\varrho} = O\left(T^{\eta-1}\log T\right)$$

as required for $R_T = o_p\left(T^{\eta/2-1/2}\right)$. The result follows from the proof of Proposition 1.