# Nonlinear Predictability of Stock Returns? Parametric vs. nonparametric inference in predictive regressions*

Matei Demetrescu† and Benjamin Hillmann‡
Christian-Albrechts-University of Kiel

This version: January 26, 2018

**Abstract**

Most procedures for detecting stock return predictability rely on linear regression models. When assessing the null hypothesis of no predictive power in a possibly nonlinear model, practitioners essentially have two choices. One could resort to a suitable nonparametric test and be prepared to lose power because of leaving the parametric framework. Since the model is linear under the null of no predictability, one could also conduct inference in a linear model, and be prepared to lose power because of the misspecification under the alternative hypothesis. To help decide which approach to use, the paper focuses on size and local power under the additional difficulty that the persistence of the regressors, as quantified by their largest autoregressive roots, is unknown. Regarding nonparametrics, the statistics employed by Juhl (2014, JBES 32, 387-394) and Kasparis et al. (2015, J. Econometrics 185, 468-494), have $\chi^2$ limiting null distributions for both low and high regressor persistence, but are asymptotically dominated in terms of local power by simple linear procedures. We show, theoretically and in simulations, that an overidentified IV testing scheme following Kostakis et al. (2015, Review of Financial Studies 28, 1506-1553) and Breitung and Demetrescu (2015, J. Econometrics 187, 358-375), is particularly well suited for inference in additive predictive models with uncertain predictor persistence. The proposed test is robust to the degree of persistence of the regressors and to time-varying volatility. An analysis of predictability of S&P 500 stock returns finds significant predictability, part of which is nonlinear in nature. For log dividend yields and the long-term rate of return we find a monotonic regression function, while log earnings price ratios exhibit a U-shaped relation. The latter is driven entirely by the 2008 financial crisis, suggesting that, during crises, firm-specific characteristics such as valuation ratios may be inconsistent signals of stock price performance.

**Key words:** Predictive regression, Nonlinear regression function, Unknown persistence, Endogeneity, Time-varying variance, Chi-square distribution

**JEL classification:** C12 (Hypothesis Testing), C22 (Time-Series Models), G17 (Financial Forecasting and Simulation)

---

*We thank Jörg Breitung, Danvee Floro, Michael Funke, Paulo Rodrigues and Julian Schröder for helpful comments, and Ulrich Homm for cooperating on a previous version of the paper as well as many suggestions.

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

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

## 1 Introduction

Predictive regressions for stock returns are an important practical aspect of quantitative finance and financial econometrics; see e.g. Campbell (2008) and Phillips (2015) for recent reviews. Practitioners must however face several methodological and empirical challenges. For instance, the signal-to-noise ratio of the typical (linear) predictive regression involving financial ratios is quite low; see e.g. Figure 1. Moreover, under endogeneity,[1] the properties of the usual OLS-based $t$ statistic in linear regression depend on the degree of persistence of the regressors: Stambaugh (1999) proved the OLS estimator in a predictive regression with endogenous, autocorrelated regressors to exhibit serious bias, and Elliott and Stock (1994) show the distribution of the associated $t$ statistic to depend on the degrees of persistence and endogeneity of the (near-integrated) regressor. Many potential predictors exhibit indeed high persistence without having an exact unit root; and if the regressor is well described as being nearly integrated, limiting distributions depend under endogeneity on a parameter that cannot be consistently estimated. To deal with this issue in the linear case, Amihud and Hurvich (2004); Amihud et al. (2009, 2010) propose bias corrections that apply under stationarity conditions for the predictors, while Campbell and Yogo (2006) propose an improved Bonferroni-based procedure that requires specifying lower and upper bounds for the unknown persistence of the (near-integrated) regressor. Kostakis et al. (2015) adapt the extended IV [IVX] procedure of Magdalinos and Phillips (2009) to the predictive regression framework, and Breitung and Demetrescu (2015) study powerful extensions of the instrumental variable approach.

Still, predictability of stock returns is elusive in practice, in spite of the well-developed theoretical foundations. As emphasized by Lettau and Van Nieuwerburgh (2008), predictability is plagued by inconsistencies between in-sample and out-of-sample performance of predictive regressions. To explain this, there is a growing trend in the literature positing the existence of a nonlinear dimension in the breakdown of the linear present value model; see e.g. McMillan (2001), McMillan (2003) or Kanas (2005); in fact, the idea that a nonlinear model may improve forecasting performance for stock returns can be traced back at least to Chung and Zhou (1996). This is by no means far-fetched: the relation between stock returns and dividend yields, say, grows exponentially over time (Ang and Bekaert, 2007), such that the basic linear model may indeed be a less appropriate representation of the true relation between financial ratios and asset returns.[2]

Indeed, examining Figure 2 giving local linear regression curve estimates, one may notice that the slope appears to vary over the range of the predictors, in particular it is smaller in magnitude around the center of the distribution of the regressors, where most observations lie. A linear fit would be closer to the slope around the center, and would therefore forecast badly if the out-of-sample predictor comes from the tails. And we may notice in Figure 1 that especially the log(D/P) series is decisively below its long-run average from the late 90s onwards, which is largely the period where out-of-sample predictability is not found; see e.g. Welch and Goyal (2008).

It is therefore quite reasonable to allow for nonlinear regression functions when testing the null hypothesis of no predictability. But technical developments on testing predictive ability have mostly

[1] In this context, one speaks of endogenous regressors when the shocks to the series to be predicted are contemporaneously correlated with the innovations to the regressors.

[2] Slope parameter instability is another possible explanation for such phenomena; see, among others, Viceira (1997), Paye and Timmermann (2006), Ang and Bekaert (2007), and Henkel et al. (2011); this may also be interpreted as a nonlinear model, with the slope coefficient depending on time or other variables.

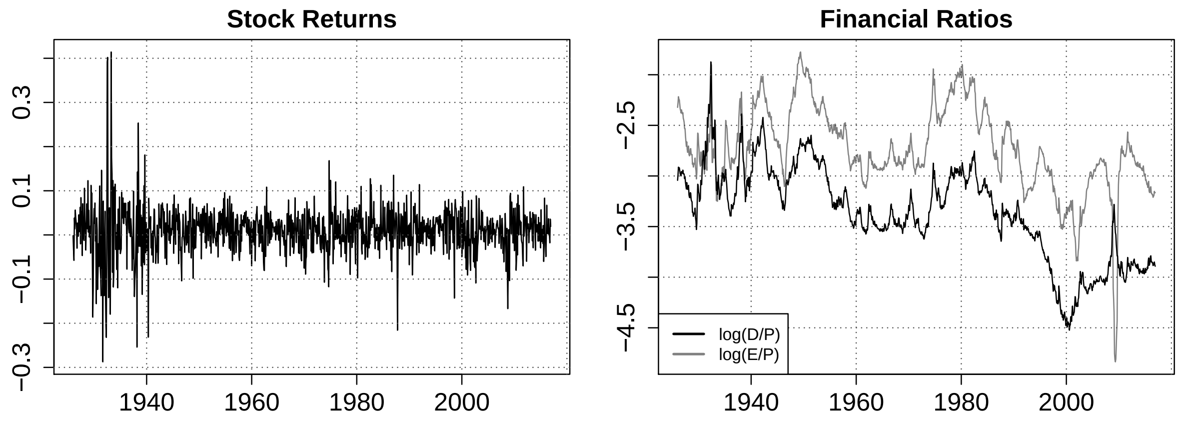


Figure 1: S&P 500 returns, log dividend yield and log earnings/price ratio – monthly observations, December 1926 to December 2016; see Section 5 for details

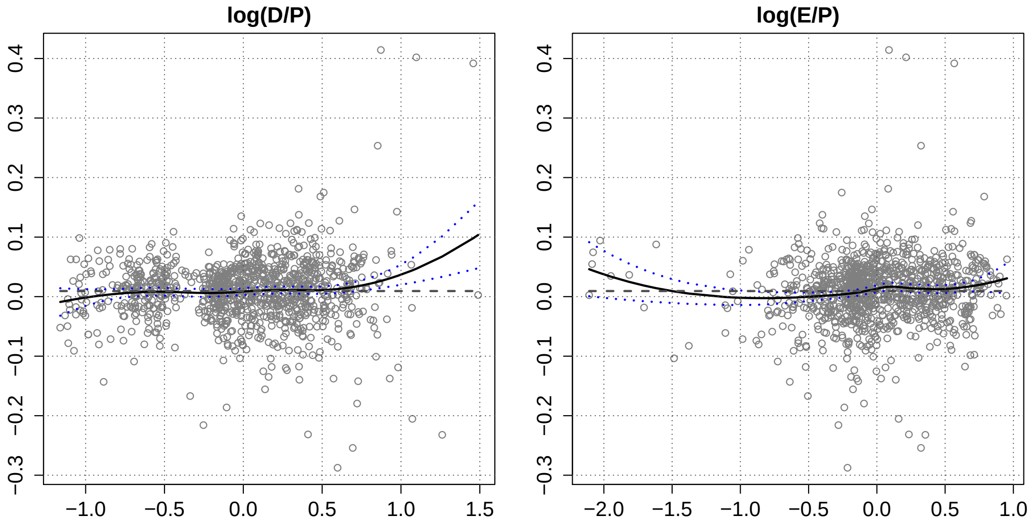


Figure 2: Local linear regression curves of stock returns against lagged financial ratios; see Section 5 for details on the computation

been confined to linear models sofar. Only more recently did Juhl (2014) propose a U statistic to test the predictive ability of a near-integrated or stationary regressor; building on the work of Wang and Phillips (2012) and Fan and Li (1999), he shows that the limiting distribution of the U statistic is the same for either stationary or near-integrated regressors.[3] Kasparis et al. (2015) study the Nadaraya-Watson estimator and related test statistics under several types of persistence; their test statistics can immediately be used for inference in nonlinear predictive regressions, with the same advantage that the asymptotic distributions are the same irrespective of the actual persistence of the regressor. Practitioners have in fact continusouly pitted parametric vs. nonparametric methods even earlier, and there is evidence that linear models cannot fully capture predictability because of nonlinear predictable components; see e.g. McMillan (2003) or Chen and Hong (2010).

In this paper we argue however that, contrary to the intuition that nonparametric methods can capture nonlinear dynamics better, mispecified linear models actually have better chances of un-

[3]In the predictive regression setup, the U statistics of Wang and Phillips (2012) and Juhl (2014) are essentially the same. Wang and Phillips (2012), however, discuss a more general testing problem, aiming to detect misspecification in nonlinear cointegrating regressions with near-integrated regressors; similarly, Fan and Li (1999) discuss misspecification tests in the stationary case.

covering significant predictive relations. To make the point, we focus on an additive nonlinear predictive regression model,

$$y_t = \beta_0 + \beta_1 f(x_{t-1}) + u_t, \quad t = 2, \ldots T. \tag{1}$$

This nests of course the linear case. Such models have been recently studied by Chang and Park (2010) and Shi and Phillips (2012) for integrated regressors. In line with the literature, the regressor $x_{t-1}$ has an autoregressive structure,

$$x_t - \mu = \rho(x_{t-1} - \mu) + e_t, \quad t = 2, \ldots, T. \tag{2}$$

with $e_t$ a short-memory linear process given as $e_t = \sum_{j \geq 0} b_j v_{t-j}$ with suitable summability conditions on the Wold coefficients $b_j$ and $v_t$ is zero-mean uncorrelated such that $\mu$ is the mean of the process irrespective of its persistence. We take $\rho$ to be either fixed and belonging to the stationarity region (cf. Amihud et al., 2009), or near to unity, $\rho = 1 - {}^{c}\!/\!_{T}$ (cf. Campbell and Yogo, 2006), thus allowing for flexibility in modelling the persistence of the posited predictor. Moreover, $x_{t-1}$ is allowed to be endogenous as captured by non-zero contemporaneous correlation of the zero-mean serially uncorrelated sequence $(u_t, v_t)'$; such endogeneity is a typical feature of predictive regressions for stock returns, say. The situation relevant in practice is when $\rho$ and $f$ are unknown, and the paper studies inference on $\beta_1$ under such circumstances.

The main interest lies in achieving correctly-sized inference without significant losses of power, without having to specify the functional form of $f$ (up to regularity conditions) and without having to decide whether the regressor is stationary or (near) integrated. In particular, it is the local power that is relevant (cf. Phillips and Lee, 2013) since the prototypical predictive regression (stock returns on dividend yields) exhibits small coefficients and a low signal-to noise ratio (cf. Figure 1 again).

Bluntly put, the practitioner's dilemma is as follows. Should one run a linear regression of $y_t$ on $x_{t-1}$ instead of estimating the nonlinear relation (1), power may be lost when testing $\beta_1 = 0$ due to having misspecified $f$ to be linear under the alternative. If, on the other hand, one uses a nonparametric regression method, power is lost again due to lower convergence rates, say, than for a parametric (even misspecified) setup; moreover, the problem worsens in a nonparametric multiple regression setup, where convergence rates may further be reduced by the curse of dimensionality. One has the choice between a rock and a hard place: should inference rely on a misspecified (under the alternative) linear model, or on typically less powerful nonparametric techniques?

To answer the question of which approach to use, we exploit the fact that modelling $f$ appropriately is not always a strictly necessary step in answering the "yes/no" question on predictive ability, even in a possible nonlinear framework; modelling $f$ could well be done *after* deciding whether there is predictive power at all. This is because of two reasons. First, the null hypothesis of no predictability is, trivially, linear: $y_t = \beta_0 + 0 \cdot x_{t-1} + u_t$. Second, *local* alternatives in (1) imply *local* departures from linearity, so power losses due to misspecification may not be as serious as in a textbook situation. We consider power against sequences of local alternatives of the form $\beta_1 = {}^{b}\!/\!_{T^\nu}$ for suitable $\nu > 0$ depending on the properties of the regression function under the alternative (the case $b = 0$ recovers the null hypothesis and thus size). Since $f$ is unknown, we make minimal identifying assumptions, and consider in a first step alternatives where the regression function $f$ is monotonic and asymptotically homogenous of some degree $\alpha > 0$ in the sense of Park and Phillips (1999). For the typical financial predictive regression with predictors such as the log dividend price ratio (for which theory predicts an upwards sloping relation), monotonicity is a reasonable requirement. Among others, this allows $f$ to be a (signed) power function, and includes piecewise linearity as a particular case. A (misspecified) linear model can be interpreted as a linear approximation of the unknown function $f$, and our approach may alternatively be formulated in terms of the Taylor expansion approach of Luukkonen et al. (1988). This perspective then allows us to deal with potential violations of monotonicity by simply using a higher-order expansion. Therefore, we are able to relax the monotonicity requirement in a second step.

In more detail, our contributions are as follows. Section 2 provides the model framework. To keep this paper self-contained, the section also gives a preliminary analysis showing that, not surprisingly, the local power of OLS-based tests building on knowing the true functional form depends on the homogeneity degree $\alpha$ of the regression function $f$ for nearly integrated regressors with nontrivial power given against alternatives of the type $\beta_1 = bT^{-\frac{\alpha+1}{2}}$. Then, for nearly integrated regressors, we show the local power of the test assuming a linear relation to be nontrivial in the same $T^{-\frac{\alpha+1}{2}}$ neighbourhoods of the null as that of the unfeasible OLS; this is quite surprising given the misspecification to linearity of the predictive model. Moreover, the test statistics considered by Juhl (2014) and Kasparis et al. (2015) for testing predictive ability of unknown functional form have reduced power in the sense that it has power equal to size in neighbourhoods of the null of type $T^{-\frac{\alpha+1}{2}}$. The same ranking holds, at a different scale, in the stationary case, where parametric procedures typically exhibit $\sqrt{T}$ convergence rates compared to lower rates in the nonparametric approach.

Since linear OLS-based testing is size-distorted in the misspecified linear regression case, we shall resort in Section 3 to robust IV estimation and testing as advocated by Breitung and Demetrescu (2015) to provide a test with asymptotic size control and, as we shall see, power in the "optimal" neighbourhoods of the null.[4] Breitung and Demetrescu propose a two-stage least squares [2SLS] procedure using two instrument variables: one instrument (so-called of type I) is designed to work when $x_t$ is stationary, whereas the other (so-called of type II) works under regressor near integration.[5] While they discuss several type-I instruments in the linear case, we provide here arguments that the IVX instrument (already employed in linear regression by Kostakis et al., 2015) is the more suitable choice as a type-I instrument in the nonlinear setup. The choice for a type-II instrument is not affected by the potential nonlinearity, and we resort to the deterministic instrument motivated by Phillips (1998) and used by Breitung and Demetrescu (2015); this is a sine transformation of the scaled time index. In the limit, the 2SLS combination of these two instruments puts all weight on the instrument suitable for the underlying data generating process without requiring user input. At the same time, the limiting distribution is chi square in all cases. Thus, the 2SLS IV procedure has the nice feature that it does not require any additional data or information about the persistence of the regressors.

Section 4 shows our procedure to perform well against competing methods in finite samples, in particular so against nonparametric methods. It should be emphasized that, although we favour

[4]Following Breitung and Demetrescu (2015), optimality refers here to the rate achievable by the OLS estimator for the concrete persistence of the data generating process.

[5]Breitung and Demetrescu (2015) show that several type-I instruments may be used for the 2SLS procedure in principle, but only one of the possible type-II instruments; in practice they find that one instrument of each kind is sufficient to exploit the potential of the procedure.

the 2SLS IV procedure, the point we make is more general and refers to the power losses of nonparametric methods compared to linear methods, even if locally misspecified.

We then examine the possibly nonlinear predictability of S&P 500 stock returns in Section 5 using logarithmized dividend-price ratio [log(D/P)] and logaritmized earnings-price ratio [log(E/P)] as predictor variables. Following Lettau and Van Nieuwerburgh (2008), we further adjust these predictors for possible permanent breaks in their mean to improve predictability. To find the breaks, we resort to a testing procedure due to Perron and Yabu (2009), which is designed to work in the presence of time-varying volatility and uncertain persistence of the examined series; this results in different break dates compared with Lettau and Van Nieuwerburgh (2008). After finding significant predictability, driven in particular by log earnings-price ratio, we find evidence of nonlinear or, for log(E/P), even nonmonotonic relations. It appears, however, that the nonmonotonicity is driven by the 2008 crisis. Robstness checks further find some evidence of predictability when using long-term rates of return as predictor.

The final section concludes, and the proofs have been gathered in the Appendix, which also contains details on all robustness checks performed.

## 2 Setup

For simplicity, we start by setting the intercept $\beta_0$ to equal zero. We include it of course in our final proposal (see Proposition 3), but an intercept is of secondary importance for the preliminary discussions of this section. Let us now discuss the functional form under the alternative of predictability.

**Assumption 1** *Let $f$ be asymptotically homogenous of some order $\alpha > 0$ in the sense that $f(\cdot) \equiv H_\alpha(\cdot) + I(\cdot)$ where $H_\alpha(sx) = s^\alpha H_\alpha(x)$ for any $x \in \mathbb{R}$ and $s \geq 0$, and $I(\cdot)$ is Lipschitz-continuous and integrable, $\int_{-\infty}^{\infty} |I(x)| \, \mathrm{d}x < \infty$. Let furthermore $I(0) = 0$ and assume that $f$ is monotonic.*

The centering condition $I(0) = 0$ implies that $f(0) = 0$ since $H_\alpha(0) = 0$ for $\alpha > 0$. Because we allow for an intercept in Section 3, which can be set to equal $\beta_0 + \beta_1 f(0)$, the centering condition is not restrictive.

The assumption excludes purely integrable functions. For integrated regressors, integrable transformations were analyzed by Chang and Park (2010) and Shi and Phillips (2012); see also Marmer (2008). Such functions, however, which must converge to zero as their argument goes to plus or minus infinity, imply that a predictor would lose predictive power as it moves away from some equilibrium value or region. Since we do not find such a restriction reasonable, we do not consider purely integrable regression functions here. Monotonicity, which we do find plausible for the typical predictive regression with stock returns and financial ratios as potential predictors, plays an important role under regressor stationarity guaranteeing nontrivial local power; see Proposition 2 and the discussion following it, as well as Section 3 for more details. (In Section 3, we also discuss how to deal with violations of the monotonicity requirement in practice.) Actually, monotonicity could completely be disposed of under persistence. Still, we require it in both cases for coherence of the exposition.

**Assumption 2** *The series $y_t$ and $x_t$ are generated according to Equations* (1) *and* (2) *with $e_t$ a linear process with 1-summable Wold coefficients, i.e. $e_t = \sum_{j\geq 0} b_j v_{t-j}$ such that $\sum_{j\geq 0} j\,|b_j| < \infty$, where $\lambda = \sum_{j\geq 0} b_j > 0$ and the shocks $(u_t, v_t)'$ are serially uncorrelated as specified below.*

**Assumption 3** *Let $v_t = \sigma_{vt}\nu_t$ and $u_t = \gamma_t v_t + \sigma_{\varepsilon t}\varepsilon_t$ where $(\varepsilon_t, \nu_t)'$ is a strictly stationary and ergodic martingale difference [md] sequence w.r.t. the natural filtration, with finite moments of some order $\delta > \max\{4, 4\alpha\}$. Let furthermore $\gamma_t = \gamma\left({}^t/_T\right)$, $\sigma_{vt} = \sigma_v\left({}^t/_T\right)$ and $\sigma_{ut} = \sigma_u\left({}^t/_T\right) = \sqrt{\gamma^2\left({}^t/_T\right)\sigma_v^2\left({}^t/_T\right) + \sigma_\varepsilon^2\left({}^t/_T\right)}$ where $\gamma(\cdot)$, $\sigma_\varepsilon(\cdot)$ and $\sigma_v(\cdot)$ are piecewise Lipschitz functions. Finally, let $\sup_{t\in\mathbb{Z}}\left|\mathrm{E}\left(\left(\nu_t^2 - \mathrm{E}\left(\nu_t^2\right)\right)\nu_{t-j}\nu_{t-k}\right)\right| \leq C\,(jk)^{-1/2-\eta/2}$ and $\sup_{t\in\mathbb{Z}}\left|\mathrm{E}\left(\left(\varepsilon_t^2 - \mathrm{E}\left(\varepsilon_t^2\right)\right)\varepsilon_{t-j}\varepsilon_{t-k}\right)\right| \leq C\,(jk)^{-1/2-\eta/2}$ $\forall j, k > 0$.*

The assumption allows for conditional heteroskedasticity as well as time-varying unconditional variance and covariance. This makes $(u_t, v_t)'$ a uniformly modulated (Priestley, 1988, p. 165) or locally stationary process (see e.g. Dahlhaus, 2012, for a recent review). Indeed, such stylized facts have been reported in the literature; see e.g. Amado and Teräsvirta (2014) and the references therein. The finite kurtosis requirement is standard in the literature and plausible for monthly or quarterly stock returns, while the bounds on $\mathrm{E}\left(\left(\nu_t^2 - \mathrm{E}\left(\nu_t^2\right)\right)\nu_{t-j}\nu_{t-k}\right)$ and $\mathrm{E}\left(\left(\varepsilon_t^2 - \mathrm{E}\left(\varepsilon_t^2\right)\right)\varepsilon_{t-j}\varepsilon_{t-k}\right)$ restrict the serial dependence in the second-order moments of the innovations; cf. also Assumption 1 in Breitung and Demetrescu (2015), which is only slightly more general. The assumption allows e.g. for asymmetric responses in the conditional heteroskedasticity.

Under Assumptions 2 and 3, the following limiting behaviour arises for suitably normalized partial sums of the innovations and is relevant for the near-integrated case.

**Lemma 1** *Let $\bar\omega_u^2 = \int_0^1 \sigma_u^2(s)\,\mathrm{d}s$ be the average variance of $u_t$ and $\bar\omega_v^2 = \int_0^1 \sigma_v^2(s)\,\mathrm{d}s$ be the average variance of $v_t$. Under Assumptions 2 and 3, it holds as $T \to \infty$ that*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\begin{pmatrix} u_t \\ v_t \end{pmatrix} \Rightarrow \int_0^s \begin{pmatrix} \sigma_\varepsilon(s) & \sigma_v(s)\gamma(s) \\ 0 & \sigma_v(s) \end{pmatrix} \mathrm{d}\mathbf{W}(s) := \begin{pmatrix} \bar\omega_u W_{\sigma_u}(s) \\ \bar\omega_v V_{\sigma_v}(s) \end{pmatrix}$$

*where "$\Rightarrow$" stands for weak convergence in a suitable space of cadlag functions, $\mathbf{W}(s) = (W_1(s), W_2(s))'$ is a two-dimensional vector of independent Wiener processes, $(W_{\sigma_u}(s), V_{\sigma_v}(s))'$ is a vector of so-called time-transformed Wiener processes also exhibiting with time-varying correlation.*

*Moreover, for $\rho = 1 - {}^c/_T$,*

$$\frac{1}{\sqrt{T}}x_{[sT]} \Rightarrow \lambda\,\bar\omega_v J_{c,\sigma_v}(s)$$

*where convergence is joint and $J_{c,\sigma_v}(s)$ is the Ornstein-Uhlenbeck type process with mean reversion parameter $c$ driven by the time-transformed Wiener process $V_{\sigma_v}(s)$, $J_{c,\sigma_v}(s) = V_{\sigma_v}(s) - c\int_0^s e^{-c(s-r)}V_{\sigma_v}(r)\,\mathrm{d}r$.*

**Proof:** *see Appendix C.*

For the examination of the low-persistence case where $|\rho| < 1$ fixed and hence $x_t$ a stable autoregression, where sample averages of nonlinear transformations of locally stationary processes are involved, we define for some continuous function $h(\cdot)$ the functional

$$\mathcal{E}_m[h] = \int_0^1 \mathrm{E}\left(h\left(m + \sigma_v(s)\,\tilde{x}_t\right)\right) \mathrm{d}s$$

where $\tilde{x}_t = (1-\rho L)^{-1}\left(\sum_{j\geq 0} b_j L^j\right)\nu_t$ with $L$ the lag operator is strictly stationary for $|\rho| < 1$ fixed. We shall also require a weighted version thereof,

$$\mathcal{E}_m^*[h] = \int_0^1 \mathrm{E}\left(h\left(m + \sigma_v(s)\,\tilde{x}_t\right)\left(\gamma(s)\,v_t + \sigma_\varepsilon(s)\,\varepsilon_t\right)^2\right) \mathrm{d}s.$$

Note that, in the case where the regressor is a stable autoregression, the nonhomogenous component of $h$ is not asymptotically negligible anymore and $\mathcal{E}$ or $\mathcal{E}^*$ depend on it too. Also, both $\mathcal{E}_m[h]$ and $\mathcal{E}_m^*[h]$ are positive if $h(\cdot)$ is positive and negative if $h(\cdot)$ is negative. The functionals are then used to express the probability limit of sample averages, as shown in the following

**Lemma 2** *Under Assumptions 2 and 3, we have for any function $h$ satisfying Assumption 1 and $|\rho| < 1$ fixed that*

$$\frac{1}{T}\sum_{t=1}^{T} h\left(x_{t-1}\right) \xrightarrow{p} \mathcal{E}_\mu[h]$$

*and*

$$\frac{1}{T}\sum_{t=1}^{T} h\left(x_{t-1}\right) u_t^2 \xrightarrow{p} \mathcal{E}_\mu^*[h]$$

*as $T \to \infty$.*

**Proof:** *see Appendix C.*

Let us now examine the unfeasible OLS estimator assuming known shape of the regression function $f$ to assess what is achievable in terms of power. Denote by $t_\beta^{ls}$ the $t$ statistic from a regression of $y_t$ on $f(x_{t-1})$ with the usual standard errors computed for simplicity under the null hypothesis $\beta_1 = 0$, $\hat{\sigma}_u^2 = {}^1\!/\!_T \sum_{t=1}^{T} y_t^2$.[6] The asymptotic behaviour of $t_\beta^{ls}$ is summarized in the following

**Proposition 1** *Under Assumptions 1, 2 and 3, we have the following limiting behavior as $T \to \infty$.*

1. *If $\rho = 1 - {}^c\!/\!_T$ and $\beta = \frac{b}{T^{(\alpha+1)/2}}$, then*

$$t_\beta^{ls} \xrightarrow{d} \frac{\int_0^1 H_\alpha\left(J_{c,\sigma_v}(s)\right) \mathrm{d}W_{\sigma_u}(s)}{\sqrt{\int_0^1 H_\alpha^2\left(J_{c,\sigma_v}(s)\right) \mathrm{d}s}} + b\,\frac{\lambda^\alpha \bar{\omega}_v^\alpha}{\bar{\omega}_u}\sqrt{\int_0^1 H_\alpha^2\left(J_{c,\sigma_v}(s)\right) \mathrm{d}s};$$

2. *If $|\rho| < 1$ fixed and $\beta = \frac{b}{T^{1/2}}$, then*

$$t_\beta^{ls} \xrightarrow{d} \mathcal{Z}\sqrt{\frac{\mathcal{E}_\mu^*[f^2]}{\bar{\omega}_u^2 \mathcal{E}_\mu[f^2]}} + b\,\frac{\sqrt{\mathcal{E}_\mu[f^2]}}{\bar{\omega}_u}$$

[6] The usual residual variance estimator can easily be shown to work under the considered local alternative as well, in spite of the linear misspecification; we omit the details.

*where $\mathcal{Z}$ is a standard normal variate and $\mathcal{E}^*_\mu\left[f^2\right] = \bar{\omega}_u^2 \mathcal{E}_\mu\left[f^2\right]$ when $u_t$ is conditionally and unconditionally homoskedastic.*

**Proof:** *see Appendix C.*

Under persistence and endogeneity, the null distribution is nonstandard and depends on nuisance parameters, so corrective action would have been required even when the regression had been feasible. The null distribution under regressor stationarity is standard normal only if there is no time-varying volatility and no conditional heteroskedasticity. But, as can be seen from the proof and from Section 3, Eicker-White heteroskedasticity-robust standard errors would correct for this problem in the low-persistence case, while, in the near-integrated case, Eicker-White standard errors would not affect the limiting results; see Proposition 3.

Armed with knowledge about what the upper bounds for the local power are, let us examine the local power of tests based on a misspecified linear regression. Denote by $t_\beta^{lin}$ the resulting $t$ statistic with residual variance computed again under the null. We then obtain the following

**Proposition 2** *The following limiting behavior results as $T \to \infty$ under the assumptions of Proposition 1:*

1. *If $\rho = 1 - c/T$ and $\beta = \frac{b}{T^{(\alpha+1)/2}}$, then*

$$t_\beta^{lin} \xrightarrow{d} \frac{\int_0^1 J_{c,\sigma_v}(s)\,\mathrm{d}W_{\sigma_u}(s)}{\sqrt{\int_0^1 J^2_{c,\sigma_v}(s)\,\mathrm{d}s}} + b\,\frac{\lambda^\alpha \bar{\omega}_v^\alpha \int_0^1 J_{c,\sigma_v}(s)\,H_\alpha\left(J_{c,\sigma_v}(s)\right)\mathrm{d}s}{\bar{\omega}_u \sqrt{\int_0^1 J^2_{c,\sigma_v}(s)\,\mathrm{d}s}};$$

2. *If $|\rho| < 1$ fixed and $\beta = \frac{b}{T^{1/2}}$, then, with $i$ the identity function, $i(s) = s\ \forall s$,*

$$t_\beta^{lin} \xrightarrow{d} \mathcal{Z}\sqrt{\frac{\mathcal{E}^*_\mu\left[i^2\right]}{\bar{\omega}_u^2 \mathcal{E}_\mu\left[i^2\right]}} + b\,\frac{\mathcal{E}_\mu\left[f \cdot i\right]}{\bar{\omega}_u \sqrt{\mathcal{E}_\mu\left[i^2\right]}}$$

*where $\mathcal{Z}$ is a standard normal variate and $\mathcal{E}^*_\mu\left[i^2\right] = \bar{\omega}_u^2 \mathcal{E}_\mu\left[i^2\right]$ when $u_t$ is conditionally and unconditionally homoskedastic.*

**Proof:** *see Appendix C.*

This result serves to pin down the bounds on power offered by linear models. We note that parametric local alternatives are achieved in spite of the nonlinearity; this is not the case with nonparametric procedures, who have power equal to size against such alternatives by construction. Since the limiting distribution depends on nuisance parameters, a test based on this result is infeasible, of course. Predictability testing in linear models with regressors of uncertain persistence is however well understood and we may resort to existing procedures; see Section 3.

Note that regressor endogeneity always leads to a non-standard distribution of the $t$ statistic if the regressor is highly persistent, but the test has local power in optimal neighbourhoods (optimal in the sense that the rates of the unfeasible OLS test are achieved). Should the regressor $x_{t-1}$ be stable, however, the test has nontrivial power only when $\mathcal{E}_\mu [f \cdot i] \neq 0$.[7] Having assumed that $f$ is monotonic with $f(0) = 0$, the condition is fulfilled in our setup. One could actually require $\mathcal{E}_\mu [f \cdot i] \neq 0$ as identifying condition, which is considerably more general than the monotonicity requirement on $f$. We stick to monotonicity, though, and do so for several reasons: first, it is a reasonable requirement for a regression function many setups, including stock return predictability, and second, it does not involve the distributional properties of $x_t$ and is thus easier to check, at least in principle. Third, should monotonicity be violated, simply adding $x_{t-1}^2$ as regressor in the predictive regression would side-step the problem (recall that any function can be written as the sum of an odd and an even function, so $x_{t-1}^2$ would correlate with the even component of $f(x_{t-1})$).

Summing up, inference in the misspecified linear model has the potential to outperform nonparametric procedures, provided of course that size control is given. To achieve both goals, we discuss the local power of a test based on overidentified IV estimation following Breitung and Demetrescu (2015) and show that it is not affected by the local nonlinearity of our setup in a critical manner.

# 3 A robust combination test

In the linear case, Breitung and Demetrescu (2015) argue that simple IV-based tests using carefully chosen instruments for the regressor $x_{t-1}$ may be used to obtain a test statistic that on the one hand possesses a standard limiting distribution (chi square) irrespective of the type of regressor dynamics, and on the other hand has power in the optimal neighbourhood of the null, again irrespective of whether the regressors have low or high persistence. Concretely, they recommend the combination of instruments with different properties via 2SLS. Their so-called type-I instruments are allowed to be endogenous, but must have lower persistence than the nearly integrated regressor, while the so-called type-II instruments are trending in an essentially deterministic manner (to exclude endogeneity) but may exhibit high persistence.[8]

Intuitively, type-II instruments share the trending behaviour with the time-varying OU process (the limit of the regressor in the highly persistent case), so they are not weak instruments when the regressor is near integrated; the correlation with the instrumented variable is random, however; see Phillips (1998). Thus, type-II instruments achieve the optimal rate of the OLS estimator in the linear model, while still implying a standard normal null distribution of the corresponding test statistic, since they are exogenous by construction. The 2SLS statistic advocated by Breitung and Demetrescu (2015) has the nice feature that it puts, asymptotically, all the weight on the instrument optimal for the degree of persistence of the regressor given in the data generating process, without having to actually specify the persistence.

So let us examine this procedure from a "nonlinear alternatives" perspective. In this respect, one contribution of the paper is to show that the IV-based test using a type-II instrument (concretely, the sine of $\frac{\pi t}{2T}$ recommended by Breitung and Demetrescu, 2015) has power against the same sequences of local alternatives as the unfeasible test based on $t_\beta^{ls}$ with knowledge of the true shape

[7] Since it is not specified whether $f$ is increasing or decreasing, two-sided testing would be required even when the sign of $b$ is known under the alternative.

[8] Breitung and Demetrescu (2015) also discuss the so-called Cauchy instrument as a type-II instrument, which is employed by Choi et al. (2016) in conjunction with a time transformation approach to account for time-varying volatility.

of the relation. Moreover, it has a null limiting distribution not depending on the persistence of a near-integrated regressor, namely Gaussian. But this is only the case for near integration: deterministic functions are not valid instruments for low-persistence variables, the latter not being (stochastically) trending. See Proposition 3 below and its proof.

While type-II instruments do fulfill their role under persistence and endogeneity, we need a type-I instrument to account for the possibility of stable regressors under our assumptions. Breitung and Demetrescu (2015) discuss type-I instruments that are a causal filter of $\Delta x_{t-1}$, less persistent than the levels $x_{t-1}$. This would typically lead to local power in $1/\sqrt{T}$ neighbourhoods of the null in the stable linear case. Such filters require some care in the nonlinear case; to see why, consider as a (counter-)example the simplest instrument $z_{t-1} = \Delta x_{t-1}$. In our case, its application would require the covariance $\mathrm{E}\left(f\left(x_{t-1}\right)\Delta x_{t-1}\right)$ to be non-zero. Uncorrelatedness of $f\left(x_{t-1}\right)$ and $z_{t-1}$ is unfortunately a case which we cannot always plausibly exclude for any type-I instrument, since $f$ is not known in advance. The validity of the type-I instruments discussed by Breitung and Demetrescu (2015) therefore requires instrument-specific double-checking.

We therefore show in Proposition 3 that the IVX instrument employed by Kostakis et al. (2015) does meet validity requirements in the stable nonlinear case.[9] The IVX instrument is constructed as $w_t = \sum_{j=0}^{t-2} \varrho^j \Delta x_{t-j}$ where $\varrho = 1 - \frac{a}{T^\eta}$ for some $a > 0$ and $\eta \in (0,1)$ and has, under near-integration, less persistence than $x_t$. The nice feature of IVX under stability is that $\varrho$ is close enough to unity to practically undo the differencing, and hence $w_t$ is a valid instrument whenever $x_t$ is itself a valid instrument in the nonlinear case, which is the case in our setup according to Proposition 2.

Then, we only need to operationalize the test by including deterministic components, by considering both instruments at the same time, and by accounting for time-varying volatility.

So let us now examine the test statistic with demeaning; denote by $\tilde{\cdot}$ the demeaned variables, i.e. $\tilde{x}_{t-1} = x_{t-1} - \bar{x}$ etc. Then, with Eicker-White standard errors, computed for simplicity under the null, the proposed test statistic results as

$$t_\beta^{2S} = \frac{\sum_{t=2}^T \tilde{x}_{t-1}\tilde{\boldsymbol{z}}'_{t-1}\left(\sum_{t=2}^T \tilde{\boldsymbol{z}}_{t-1}\tilde{\boldsymbol{z}}'_{t-1}\right)^{-1}\sum_{t=2}^T \tilde{\boldsymbol{z}}_{t-1}\tilde{y}_t}{\sqrt{\sum_{t=2}^T \tilde{x}_{t-1}\tilde{\boldsymbol{z}}'_{t-1}\left(\sum_{t=2}^T \tilde{\boldsymbol{z}}_{t-1}\tilde{\boldsymbol{z}}'_{t-1}\right)^{-1}\sum_{t=2}^T \tilde{\boldsymbol{z}}_{t-1}\tilde{\boldsymbol{z}}'_{t-1}\tilde{y}_t^2\left(\sum_{t=2}^T \tilde{\boldsymbol{z}}_{t-1}\tilde{\boldsymbol{z}}'_{t-1}\right)^{-1}\sum_{t=2}^T \tilde{\boldsymbol{z}}_{t-1}\tilde{x}_{t-1}}},$$

where

$$\boldsymbol{z}_t = \begin{pmatrix} \sin\frac{\pi t}{2T} \\ w_t \end{pmatrix}$$

and $w_t$ is the IVX instrument. Kostakis et al. (2015) propose not to demean $w_t$ (while still demeaning regressor and dependent variable), since demeaning $w_t$ is asymptotically negligible in the IVX test statistic, and not demeaning reduces the endogeneity bias in finite samples; this allows to pick $\eta$ closer to unity such that the local power of IVX-based tests is improved. Let hence $\tilde{w}_t = w_t$, and note that $\tilde{w}_t$ does not depend on $\mu$ in either stable or near-integrated cases, since $w_t$ is obtained by filtering $\Delta x_t$ which washes out a nonzero mean of $x_t$.

Upon squaring, $t_\beta^{2S}$ follows a chi-square distribution with one degree of freedom under the null

[9] Breitung and Demetrescu (2015) recommend a fractionally integrated filter, but it is not clear that the fractional instrument is a valid instrument under nonlinearity and stability, unlike the IVX instrument.

hypothesis, and has local power in the respective optimal neighbourhood (corresponding to the actual persistence of the regressor $x_{t-1}$); we prove this in Proposition 3 below. More importantly, specifying whether $x$ is (near-)integrated or low-persistence is not necessary, the 2SLS procedure automatically picks the "correct" instrument, as can be seen in

**Proposition 3** *Under Assumptions 1, 2 and 3, we have the following limiting behavior as $T \to \infty$.*

1. *If the regressor $x_{t-1}$ is nearly integrated and $\beta_1 = \frac{b}{T^{(\alpha+1)/2}}$, then*
$$\left(t_\beta^{2S}\right)^2 \xrightarrow{d} \chi^2_{1,\kappa^2}$$
*with noncentrality parameter* $\kappa^2 = b^2 \frac{\lambda^{2\alpha}\bar{\omega}_v^{2\alpha}\left(\int_0^1\left(\sin\frac{\pi s}{2}-\frac{1}{2}\right)H_\alpha(J_{c,\sigma_v}(s))\mathrm{d}s\right)^2}{\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)^2\sigma_u^2(s)\mathrm{d}s}$.

2. *If $x_t$ is stable and $\beta_1 = \frac{b}{T^{0.5}}$, then*
$$\left(t_\beta^{2S}\right)^2 \xrightarrow{d} \chi^2_{1,\kappa^2}$$
*with noncentrality parameter* $\kappa^2 = b^2 \frac{(\mathcal{E}_\mu[fi]-\mathcal{E}_\mu[f]\mathcal{E}_\mu[i])^2}{\mathcal{E}_0^*[i^2]}$.

**Proof:** *see the Appendix.*

The result in 1 involves some abuse of notation, since the noncentrality parameter $\kappa^2$ is random and, for $b \neq 0$, the distribution is not "the textbook" noncentral $\chi^2$ distribution; see the proof for details. Concretely, $\left(t_\beta^{2S}\right)^2 \xrightarrow{d} (Z+\kappa)^2$, where $Z$ is standard normal but not necessarily independent of $\kappa$, which is zero under the null and nonzero, but random, under the alternative. Under the null, the limiting distribution in item *1* of the proposition is therefore $\chi^2_1$, like in item *2*; there, however, the noncentral $\chi^2$ distribution is genuine, so we use this notation to give a summary of the limiting behavior of $t_\beta^{2S}$ in both studied cases.

In the case of a multivariate predictive regression with $K$ potential predictors, we stick to the additive model and write $y_t = \beta_0 + \sum_{k=1}^K \beta_k f_k(x_{kt-1}) + u_t$. To test, one simply resorts to several linear independent functions of the time, one for each regressor, as type-II instruments, while type-I instruments are easily built as in the single-regressor case; see Subsection 3.3 of Breitung and Demetrescu (2015). The resulting limiting null distribution of the Wald-type statistic is $\chi^2_K$ and the test has power against the same types of local alternatives. We omit the technical details since they are straightforward multivariate extensions of the proof of the above proposition.

Finally, should $f(x_t)$ be orthogonal to $x_t$ in the weakly persistent case, one may resort to the Taylor expansion argument of Luukkonen et al. (1988) and employ a predictive regression with $x_{t-1}$ and $x_{t-1}^2$ as potential predictors. To deal with this situation, we simply use the squared IVX instrument and a second sine frequency as instruments intended to work for $x_{t-1}^2$. Note that building an IVX instrument on the basis of the differences of $x_{t-1}^2$ would lead to technical difficulties, but $w_{t-1}^2$ could be dealt with using existing IVX results; we omit however the technical details here. Furthermore, we find the procedure to work quite well in finite samples; see the following section.

# 4 Finite-sample comparison of parametric and nonparametric methods

In this section we compare Juhl's (2014) nonparametric U test and the $\hat{F}_{sum}$ test based on the Nadaraya-Watson estimator (Kasparis et al., 2015) with the IVX procedure of Kostakis et al. (2015) and the combination test discussed in the previous section.[10] Finally, we include a size-corrected version of the OLS based test to have some idea about the power level, although this test is of course not feasible. Given the theoretical results, we expect the power advantage of linear methods to increase as $T$ grows and the differences in local power become more evident.

Juhl's (2014) $U$-statistic can be expressed as the ratio $Z/\sqrt{2W}$ with

$$Z = \sum_{\substack{t=1 \\ t\neq s}}^{T} \sum_{s=1}^{T} K\left(\frac{x_{t-1} - x_{s-1}}{h}\right) \tilde{y}_t \tilde{y}_s \quad \text{and}$$

$$W = \sum_{\substack{t=1 \\ t\neq s}}^{T} \sum_{s=1}^{T} K\left(\frac{x_{t-1} - x_{s-1}}{h}\right)^2 \tilde{y}_t^2 \tilde{y}_s^2,$$

where $K(\cdot)$ is a kernel, which we choose here to be the Gaussian one, and $\tilde{y}_t$ denotes the demeaned observations. Note that there is a misprint in Theorem 4.1 of Juhl (2014); cf. Wang and Phillips (2012, Eq. (2.2)) and Fan and Li (1999, Eq. (1)). We resort to Juhl's bandwidth choices $h = \hat{\sigma}_x T^{-0.2}$ and $h = \hat{\sigma}_{\Delta x} T^{-0.2}$.

The Nadaraya-Watson based test of Kasparis et al. (2015, Eq. 16), i.e. the $\hat{F}_{sum}$ test statistic, is computed with two bandwidth choices, concretely $h = \hat{\sigma}_v T^{-0.1}$ and $h = \hat{\sigma}_v T^{-0.2}$ following the implementation in Section 6 of Kasparis et al. (2015, p. 478). The standard deviation $\hat{\sigma}_v$ of the residuals of an AR(1) fit for $x_t$ is used to ensure approximate scale invariance of the properties of $\hat{F}_{sum}$.

The IVX procedure is implemented without instrument demeaning, i.e. $\tilde{w}_t = w_t$ where the IVX instrument $w_t$ is computed using $\varrho = 1 - {}^{1}/{}_{T^{0.95}}$ as proposed by Kostakis et al. (2015). Finally, the $t_{\beta}^{2S}$ statistic from Section 3 is computed with the same choice $\varrho = 1 - {}^{1}/{}_{T^{0.95}}$ as for the pure IVX instrument.

We consider the following data generating process ($t = 1, \ldots, T$) for $T = 250$:

$$y_t = \beta \,\mathrm{sgn}(x_{t-1})\, |x_{t-1}|^{\alpha} + u_t,$$
$$x_t = \rho x_{t-1} + v_t,$$

and

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim iid\mathcal{N}\left(\mathbf{0},\, \sigma_t^2 \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}\right)$$

where $\rho$ is either near-integrated, $\rho = 1 - {}^{c}/{}_{T}$ for $c \in \{0, 5, 15\}$, or stable $\rho \in \{0.7, 0.8, 0.9\}$, and $\beta$ is chosen in the relevant neighbourhood of the null: $\beta = b/\sqrt{T}$ for stable regressors and $\beta = b/T^{(\alpha+1)/2}$ for highly persistent regressors.

[10] The alternative $\hat{F}_{max}$-test statistic discussed by Kasparis et al. (2015) yielded the lowest power of all tests in our simulations so we do not report the corresponding results.

The baseline simulation setup is based on constant variance, $\sigma^2_\varepsilon(s) = \sigma^2_\nu(s) = 1$. Without loss of generality, we consider negative correlations $\delta$ between the disturbances $u_t$ and $v_t$. We set the correlation $\delta$ causing endogeneity to $-0.95$, and rely on 5000 replications for each parameter setting.

The top panel of Table 1 displays rejection frequencies for the case $\alpha = 0.5$, i.e. when the influence of $x_{t-1}$ on $y_t$ grows less than linearly. The empirical sizes of the tests are displayed in the row $\beta = 0$. The U test is severely undersized for both suggested bandwidth choices, while the $\hat{F}_{sum}$-test is oversized especially for one of its bandwidths at $c = 0$. This oversizedness of $\hat{F}_{sum}$ almost vanishes as $c$ increases. The IVX test as well as the robust combination test tend to be undersized, more so for smaller autoregressive coefficient $\rho$. Moving away from the null, the U tests almost always have the lowest power which is not too surprising since they are undersized. Although the $\hat{F}_{sum}$-tests are oversized, they have lower power than the robust combination test. The IVX test performs similar to the robust combination test for stable regressors while the robust combination test has more power for nearly-integrated regressors.

The results change to some extent when the influence of the regressor on returns grows more than linearly in $x_t$. Rejection rates under this scenario are displayed in the bottom panel of Table 1 for $\alpha = 1.5$. The size of all tests does not depend on $\alpha$ and therefore stays the same as for $\alpha = 0.5$. The power however turns out to depend on $\alpha$ in our setup. The U tests still have the lowest power, followed by the $\hat{F}_{sum}$-test. IVX improves its power relative to the nonparametric approaches compared to the case $\alpha = 0.5$. The robust combination test has again the largest power but its advantage over IVX decreases. The growth in power while moving away from the null is slower for $\alpha = 1.5$ than for $\alpha = 0.5$ for all tests. The reason for the power drop is that the function $f$ has zero slope at the origin, where more observations tend to be available, and the "effective" alternative is thus closer to the null. All tests are equally affected.

The purely linear case (the middle panel of Table 1) checks the advantage of exploiting (correct) information about the structure of the model. The performance of all tests in terms of power lies between the cases $\alpha = 0.5$ and $\alpha = 1.5$, and the ranking of the examined tests does not change.

We then provide results for $T = 1000$ (Table 2). We note that power drops for the nonparametric procedures, but not for the linear ones. This illustrates the lower, nonparametric, convergence rates of the kernel-based procedures. The power even tends to increase a bit for linear procedures. Furthermore we observe that the oversizedness of the IVX test of Kasparis et al. (2015) decreases but it remains marginally oversized.

Our simulation results suggest that the IVX test and especially the robust combination test are able to outperform the nonparametric approaches in terms of power in spite of the linear missspecification, while controlling size.

Finally, we study three cases which are not covered by our monotonicity assumptions, $f(x) = |x|^{0.5}$, $f(x) = |x|$ and $f(x) = |x|^{1.5}$ for $T = 250$ (Table 3). We find in line with Juhl (2014) that the linear procedures are not able to compete with the nonparametric ones in such cases. One exception is 2SLS for $c = 0$ and small $b$, i.e. alternatives very close to the null. The power of the linear procedures is much lower for stable regressors especially when the persistence of the regressor decreases. We included further instruments to deal with this problem. For IVX$^2$ we added the square of the regular IVX instrument and in 2S$^2$ we furthermore added the sine of $3\frac{\pi t}{2T}$ as additional type-II instrument. It can be seen that IVX$^2$ and 2S$^2$ are able to compete with the nonparametric approaches in terms

Table 1: Size and local power of nonparametric and misspecified linear procedures: $T = 250$

| b | OLS | U1 | UD1 | F1 | F2 | IVX | 2SLS | OLS | U1 | UD1 | F1 | F2 | IVX | 2SLS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | $f(x) = \lvert x\rvert^{0.5}\,\mathrm{sgn}\,x$ | | | | | | | |
| | | | Local alternative: $\beta = \frac{b}{T^{(1+\alpha)/2}}$ | | | | | | | Local alternative: $\beta = \frac{b}{\sqrt{T}}$ | | | | |
| | | | $c = 0$ ($\rho = 1$) | | | | | | | | $\rho = 0.9$ | | | |
| 0 | 5.0 | 1.3 | 3.6 | 9.4 | 6.5 | 4.9 | 6.7 | 5.0 | 1.2 | 2.0 | 6.7 | 6.2 | 3.3 | 3.4 |
| 2 | 10.0 | 3.7 | 6.1 | 19.7 | 11.6 | 18.0 | 24.5 | 35.5 | 9.4 | 8.1 | 35.2 | 26.8 | 33.2 | 32.0 |
| 5 | 69.5 | 39.2 | 18.4 | 58.9 | 35.0 | 57.0 | 77.9 | 99.9 | 91.8 | 78.3 | 99.4 | 96.8 | 99.1 | 99.2 |
| 10 | 98.7 | 93.5 | 70.7 | 97.4 | 85.9 | 84.1 | 98.5 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 100.0 | 98.6 | 100.0 | 99.9 | 97.4 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 5$ ($\rho = 0.98$) | | | | | | | | $\rho = 0.8$ | | | |
| 0 | 5.0 | 0.9 | 3.0 | 6.6 | 4.9 | 3.3 | 3.5 | 5.0 | 1.6 | 1.9 | 7.0 | 7.2 | 3.5 | 3.3 |
| 2 | 9.8 | 2.1 | 3.8 | 13.5 | 8.9 | 11.7 | 11.5 | 32.1 | 10.5 | 10.2 | 37.6 | 31.2 | 32.8 | 31.9 |
| 5 | 57.4 | 15.5 | 11.1 | 45.3 | 28.8 | 50.7 | 56.4 | 99.6 | 89.0 | 82.5 | 99.4 | 98.0 | 98.8 | 98.7 |
| 10 | 99.2 | 90.7 | 67.3 | 97.9 | 90.1 | 95.5 | 98.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 99.9 | 98.7 | 100.0 | 99.9 | 99.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 15$ ($\rho = 0.94$) | | | | | | | | $\rho = 0.7$ | | | |
| 0 | 5.0 | 1.1 | 2.6 | 6.5 | 5.6 | 3.7 | 3.9 | 5.0 | 1.1 | 1.4 | 6.2 | 6.8 | 2.9 | 2.8 |
| 2 | 8.5 | 2.1 | 3.2 | 11.4 | 8.6 | 8.6 | 8.5 | 36.4 | 11.2 | 11.1 | 40.0 | 34.1 | 34.3 | 33.7 |
| 5 | 38.1 | 10.2 | 8.1 | 35.1 | 25.1 | 37.1 | 35.3 | 99.5 | 88.4 | 85.2 | 99.4 | 98.6 | 98.5 | 98.5 |
| 10 | 98.5 | 73.5 | 50.1 | 94.4 | 82.9 | 94.5 | 96.5 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | | | | | $f(x) = \lvert x\rvert^{1.0}\,\mathrm{sgn}\,x$ | | | | | | | |
| | | | Local alternative: $\beta = \frac{b}{T^{(1+\alpha)/2}}$ | | | | | | | Local alternative: $\beta = \frac{b}{\sqrt{T}}$ | | | | |
| | | | $c = 0$ ($\rho = 1$) | | | | | | | | $\rho = 0.9$ | | | |
| 0 | 5.0 | 1.3 | 3.6 | 9.4 | 6.5 | 4.9 | 6.7 | 5.0 | 1.2 | 2.0 | 6.7 | 6.2 | 3.3 | 3.4 |
| 2 | 11.2 | 4.8 | 5.8 | 19.0 | 10.9 | 11.3 | 22.4 | 45.8 | 7.0 | 6.7 | 38.4 | 28.7 | 42.1 | 40.8 |
| 5 | 62.0 | 33.4 | 16.7 | 55.4 | 30.4 | 56.4 | 77.6 | 100.0 | 97.2 | 80.3 | 100.0 | 99.9 | 99.9 | 100.0 |
| 10 | 99.4 | 92.6 | 62.6 | 98.1 | 87.0 | 93.4 | 99.6 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 100.0 | 99.7 | 100.0 | 100.0 | 99.7 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 5$ ($\rho = 0.98$) | | | | | | | | $\rho = 0.8$ | | | |
| 0 | 5.0 | 0.9 | 3.0 | 6.6 | 4.9 | 3.3 | 3.5 | 5.0 | 1.6 | 1.9 | 7.0 | 7.2 | 3.5 | 3.3 |
| 2 | 5.4 | 1.2 | 3.5 | 9.5 | 6.3 | 6.8 | 6.4 | 42.5 | 8.5 | 8.5 | 41.7 | 34.7 | 40.8 | 40.2 |
| 5 | 19.2 | 3.1 | 4.7 | 19.0 | 12.2 | 19.3 | 20.1 | 100.0 | 92.9 | 83.7 | 100.0 | 99.9 | 99.9 | 100.0 |
| 10 | 90.9 | 24.8 | 13.9 | 65.9 | 40.1 | 78.2 | 87.4 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 99.8 | 89.0 | 100.0 | 99.8 | 99.3 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 15$ ($\rho = 0.94$) | | | | | | | | $\rho = 0.7$ | | | |
| 0 | 5.0 | 1.1 | 2.6 | 6.5 | 5.6 | 3.7 | 3.9 | 5.0 | 1.1 | 1.4 | 6.2 | 6.8 | 2.9 | 2.8 |
| 2 | 3.6 | 1.2 | 2.6 | 7.1 | 6.2 | 4.8 | 4.6 | 46.4 | 9.7 | 9.2 | 44.8 | 38.5 | 43.4 | 42.5 |
| 5 | 9.5 | 1.8 | 3.4 | 12.5 | 8.9 | 10.3 | 10.0 | 100.0 | 90.7 | 85.6 | 100.0 | 99.9 | 99.8 | 99.9 |
| 10 | 34.9 | 5.8 | 5.9 | 29.0 | 20.0 | 33.7 | 32.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 99.8 | 50.3 | 28.6 | 94.5 | 78.6 | 96.7 | 98.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | | | | | $f(x) = \lvert x\rvert^{1.5}\,\mathrm{sgn}\,x$ | | | | | | | |
| | | | Local alternative: $\beta = \frac{b}{T^{(1+\alpha)/2}}$ | | | | | | | Local alternative: $\beta = \frac{b}{\sqrt{T}}$ | | | | |
| | | | $c = 0$ ($\rho = 1$) | | | | | | | | $\rho = 0.9$ | | | |
| 0 | 5.0 | 1.3 | 3.6 | 9.4 | 6.5 | 4.9 | 6.7 | 5.0 | 1.2 | 2.0 | 6.7 | 6.2 | 3.3 | 3.4 |
| 2 | 15.0 | 8.4 | 6.6 | 20.8 | 11.6 | 8.4 | 22.9 | 70.9 | 7.4 | 7.1 | 51.5 | 37.9 | 62.2 | 61.2 |
| 5 | 51.7 | 36.4 | 22.2 | 54.4 | 35.8 | 51.2 | 68.3 | 100.0 | 99.8 | 94.3 | 100.0 | 100.0 | 100.0 | 100.0 |
| 10 | 93.5 | 74.9 | 56.1 | 90.1 | 75.8 | 95.4 | 98.6 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 98.2 | 89.7 | 100.0 | 98.9 | 99.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 5$ ($\rho = 0.98$) | | | | | | | | $\rho = 0.8$ | | | |
| 0 | 5.0 | 0.9 | 3.0 | 6.6 | 4.9 | 3.3 | 3.5 | 5.0 | 1.6 | 1.9 | 7.0 | 7.2 | 3.5 | 3.3 |
| 2 | 3.7 | 1.0 | 3.3 | 7.9 | 5.5 | 5.3 | 4.9 | 65.2 | 8.7 | 8.5 | 55.7 | 46.1 | 60.0 | 59.8 |
| 5 | 9.2 | 1.4 | 3.8 | 11.9 | 7.8 | 10.0 | 9.8 | 100.0 | 99.4 | 96.1 | 100.0 | 100.0 | 100.0 | 100.0 |
| 10 | 38.4 | 5.7 | 6.2 | 29.8 | 18.4 | 34.9 | 37.8 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 99.4 | 62.8 | 34.8 | 95.1 | 76.4 | 93.6 | 98.2 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 15$ ($\rho = 0.94$) | | | | | | | | $\rho = 0.7$ | | | |
| 0 | 5.0 | 1.1 | 2.6 | 6.5 | 5.6 | 3.7 | 3.9 | 5.0 | 1.1 | 1.4 | 6.2 | 6.8 | 2.9 | 2.8 |
| 2 | 2.8 | 1.2 | 2.6 | 6.7 | 5.7 | 3.9 | 3.7 | 67.0 | 10.0 | 9.6 | 58.0 | 50.4 | 60.5 | 60.3 |
| 5 | 4.7 | 1.2 | 2.9 | 7.8 | 6.2 | 5.3 | 4.9 | 100.0 | 98.8 | 96.7 | 100.0 | 100.0 | 100.0 | 100.0 |
| 10 | 8.9 | 1.7 | 3.5 | 11.5 | 8.6 | 9.3 | 8.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 37.6 | 4.2 | 4.9 | 29.7 | 19.7 | 35.2 | 33.6 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

*Notes:* The tables displays rejection rates for the unfeasible size-corrected OLS test (OLS), the U test of Juhl (2014) (U1 and UD1 denote two different bandwidth choices), the $\hat{F}_{sum}$ test of Kasparis et al. (2015) (F1 and F2 denote two different bandwidth choices), the IVX test of Kostakis et al. (2015) (IVX) and for the combination/IV test (2SLS). Results for different values of the autoregressive coefficient $\rho$ are displayed in the corresponding panels. For further details see the text.

Table 2: Size and local power of nonparametric and misspecified linear procedures: $T = 1000$

| b | OLS | U1 | UD1 | F1 | F2 | IVX | 2SLS | OLS | U1 | UD1 | F1 | F2 | IVX | 2SLS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | $f(x) = \|x\|^{0.5} \operatorname{sgn} x$ | | | | | | |
| | | Local alternative: $\beta = \frac{b}{T^{(1+\alpha)/2}}$ | | | | | | | | Local alternative: $\beta = \frac{b}{\sqrt{T}}$ | | | | |
| | | | $c = 0$ $(\rho = 1)$ | | | | | | | | $\rho = 0.9$ | | | |
| 0 | 5.0 | 1.0 | 4.3 | 5.9 | 5.9 | 5.1 | 7.1 | 5.0 | 1.7 | 3.1 | 7.3 | 5.2 | 3.4 | 3.2 |
| 2 | 9.5 | 3.3 | 5.1 | 10.8 | 8.5 | 20.2 | 26.7 | 33.2 | 9.6 | 7.9 | 30.9 | 19.3 | 33.5 | 32.8 |
| 5 | 69.9 | 35.8 | 11.2 | 29.1 | 15.1 | 59.0 | 79.6 | 99.5 | 86.6 | 71.7 | 98.5 | 92.1 | 98.9 | 99.0 |
| 10 | 99.0 | 93.2 | 52.1 | 81.4 | 46.2 | 86.5 | 99.1 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 100.0 | 93.4 | 99.7 | 90.9 | 97.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 5$ $(\rho = 0.995)$ | | | | | | | | $\rho = 0.8$ | | | |
| 0 | 5.0 | 1.3 | 3.5 | 4.2 | 4.1 | 3.7 | 3.9 | 5.0 | 1.6 | 2.0 | 6.4 | 5.1 | 3.8 | 3.5 |
| 2 | 9.9 | 2.2 | 4.4 | 6.4 | 5.5 | 12.0 | 11.7 | 31.7 | 10.5 | 9.1 | 36.5 | 26.5 | 33.2 | 33.5 |
| 5 | 57.1 | 13.7 | 7.6 | 21.8 | 12.4 | 53.4 | 59.5 | 98.8 | 85.1 | 77.9 | 98.6 | 94.9 | 98.3 | 98.5 |
| 10 | 99.3 | 89.1 | 39.6 | 83.2 | 46.9 | 96.3 | 99.1 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 99.8 | 96.1 | 99.8 | 97.6 | 99.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 15$ $(\rho = 0.985)$ | | | | | | | | $\rho = 0.7$ | | | |
| 0 | 5.0 | 1.1 | 4.3 | 5.7 | 5.3 | 4.2 | 4.9 | 5.0 | 1.6 | 1.9 | 7.1 | 7.2 | 3.7 | 3.6 |
| 2 | 10.8 | 3.0 | 4.6 | 9.2 | 7.4 | 11.1 | 16.7 | 30.5 | 10.6 | 10.1 | 37.3 | 28.5 | 32.9 | 32.4 |
| 5 | 56.6 | 27.8 | 9.3 | 24.5 | 14.0 | 49.9 | 63.9 | 98.6 | 85.0 | 80.9 | 99.1 | 96.9 | 97.8 | 98.0 |
| 10 | 89.8 | 69.0 | 40.5 | 67.1 | 38.0 | 84.1 | 92.2 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 99.7 | 95.2 | 76.1 | 92.5 | 78.6 | 98.4 | 99.7 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | | | | | | $f(x) = \|x\|^{1.0} \operatorname{sgn} x$ | | | | | | |
| | | Local alternative: $\beta = \frac{b}{T^{(1+\alpha)/2}}$ | | | | | | | | Local alternative: $\beta = \frac{b}{\sqrt{T}}$ | | | | |
| | | | $c = 0$ $(\rho = 1)$ | | | | | | | | $\rho = 0.9$ | | | |
| 0 | 5.0 | 1.0 | 4.3 | 5.9 | 5.9 | 5.1 | 7.1 | 5.0 | 1.7 | 3.1 | 7.3 | 5.2 | 3.4 | 3.2 |
| 2 | 9.6 | 3.8 | 5.2 | 10.2 | 8.1 | 13.2 | 23.2 | 44.3 | 7.5 | 6.3 | 34.2 | 21.6 | 43.3 | 43.0 |
| 5 | 61.3 | 30.0 | 10.4 | 25.2 | 13.3 | 59.8 | 80.4 | 100.0 | 87.9 | 67.6 | 99.8 | 97.9 | 100.0 | 99.9 |
| 10 | 99.3 | 89.6 | 41.4 | 77.4 | 38.0 | 95.2 | 99.7 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 100.0 | 94.4 | 100.0 | 92.9 | 99.7 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 5$ $(\rho = 0.995)$ | | | | | | | | $\rho = 0.8$ | | | |
| 0 | 5.0 | 1.3 | 3.5 | 4.2 | 4.1 | 3.7 | 3.9 | 5.0 | 1.6 | 2.0 | 6.4 | 5.1 | 3.8 | 3.5 |
| 2 | 5.0 | 1.5 | 4.3 | 4.6 | 4.9 | 7.2 | 6.8 | 42.8 | 8.5 | 7.9 | 41.5 | 29.9 | 44.3 | 44.2 |
| 5 | 19.0 | 3.2 | 4.9 | 9.6 | 7.1 | 20.5 | 21.2 | 100.0 | 86.1 | 75.9 | 99.8 | 98.8 | 99.8 | 99.9 |
| 10 | 92.2 | 20.4 | 9.2 | 30.4 | 15.9 | 81.7 | 90.5 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 99.8 | 59.7 | 99.3 | 70.4 | 99.7 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 15$ $(\rho = 0.985)$ | | | | | | | | $\rho = 0.7$ | | | |
| 0 | 5.0 | 1.7 | 4.0 | 4.4 | 4.2 | 3.2 | 3.3 | 5.0 | 1.6 | 1.9 | 7.1 | 7.2 | 3.7 | 3.6 |
| 2 | 4.4 | 1.4 | 4.2 | 4.7 | 4.1 | 4.8 | 4.7 | 41.2 | 8.8 | 8.2 | 43.1 | 32.0 | 43.2 | 43.0 |
| 5 | 10.4 | 2.1 | 4.6 | 7.4 | 5.7 | 10.3 | 9.7 | 100.0 | 84.6 | 79.2 | 99.8 | 99.1 | 99.8 | 99.8 |
| 10 | 36.8 | 4.5 | 5.1 | 14.1 | 8.6 | 33.2 | 32.8 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 41.7 | 15.0 | 66.5 | 32.2 | 97.5 | 98.4 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | | | | | | $f(x) = \|x\|^{1.5} \operatorname{sgn} x$ | | | | | | |
| | | Local alternative: $\beta = \frac{b}{T^{(1+\alpha)/2}}$ | | | | | | | | Local alternative: $\beta = \frac{b}{\sqrt{T}}$ | | | | |
| | | | $c = 0$ $(\rho = 1)$ | | | | | | | | $\rho = 0.9$ | | | |
| 0 | 5.0 | 1.0 | 4.3 | 5.9 | 5.9 | 5.1 | 7.1 | 5.0 | 1.7 | 3.1 | 7.3 | 5.2 | 3.4 | 3.2 |
| 2 | 13.2 | 7.1 | 5.4 | 10.9 | 8.4 | 9.9 | 23.9 | 66.7 | 7.7 | 6.4 | 45.8 | 28.9 | 63.7 | 64.0 |
| 5 | 51.9 | 34.4 | 14.8 | 29.1 | 15.4 | 56.9 | 71.7 | 100.0 | 98.6 | 82.5 | 100.0 | 100.0 | 100.0 | 100.0 |
| 10 | 91.4 | 71.6 | 41.4 | 66.9 | 40.0 | 96.8 | 98.8 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 100.0 | 97.4 | 79.6 | 97.3 | 81.3 | 99.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 5$ $(\rho = 0.995)$ | | | | | | | | $\rho = 0.8$ | | | |
| 0 | 5.0 | 1.3 | 3.5 | 4.2 | 4.1 | 3.7 | 3.9 | 5.0 | 1.6 | 2.0 | 6.4 | 5.1 | 3.8 | 3.5 |
| 2 | 3.7 | 1.3 | 4.3 | 4.0 | 4.6 | 5.2 | 5.2 | 63.1 | 8.8 | 8.0 | 54.8 | 40.2 | 63.2 | 63.7 |
| 5 | 9.4 | 1.9 | 4.6 | 7.0 | 5.9 | 10.6 | 10.2 | 100.0 | 95.7 | 88.4 | 100.0 | 100.0 | 100.0 | 100.0 |
| 10 | 38.6 | 5.9 | 5.8 | 14.0 | 9.2 | 36.9 | 41.1 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 99.6 | 58.4 | 20.3 | 64.5 | 31.0 | 95.1 | 98.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | $c = 15$ $(\rho = 0.985)$ | | | | | | | | $\rho = 0.7$ | | | |
| 0 | 5.0 | 1.7 | 4.0 | 4.4 | 4.2 | 3.2 | 3.3 | 5.0 | 1.6 | 1.9 | 7.1 | 7.2 | 3.7 | 3.6 |
| 2 | 3.2 | 1.3 | 4.2 | 4.5 | 3.9 | 3.7 | 3.8 | 62.2 | 9.2 | 8.4 | 56.2 | 43.9 | 61.8 | 63.0 |
| 5 | 4.9 | 1.7 | 4.3 | 5.0 | 4.6 | 5.2 | 5.0 | 100.0 | 94.1 | 90.2 | 100.0 | 100.0 | 100.0 | 100.0 |
| 10 | 9.8 | 1.7 | 4.1 | 7.0 | 5.1 | 9.0 | 8.7 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 39.5 | 3.9 | 4.8 | 14.3 | 9.0 | 35.6 | 34.6 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

*Notes:* See Table 1.

Table 3: Size and local power of nonparametric and misspecified linear procedures: $T = 250$

| b | OLS | U1 | F2 | IVX | IVX² | 2SLS | 2S² | OLS | U1 | F2 | IVX | IVX² | 2SLS | 2S² |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | symmetric $f(x) = \|x\|^{0.5}$ | | | | | | | |
| | Local alternative: $\beta = \frac{b}{T^{(1+\alpha)/2}}$ | | | | | | | Local alternative: $\beta = \frac{b}{\sqrt{T}}$ | | | | | | |
| | $c = 0$ ($\rho = 1$) | | | | | | | $\rho = 0.9$ | | | | | | |
| 0 | 5.0 | 1.1 | 6.2 | 4.9 | 6.8 | 6.5 | 4.0 | 5.0 | 1.4 | 6.2 | 3.4 | 2.8 | 3.5 | 2.5 |
| 2 | 3.9 | 1.5 | 7.1 | 5.2 | 8.8 | 8.1 | 5.2 | 2.7 | 1.7 | 7.7 | 3.8 | 4.2 | 3.8 | 3.4 |
| 5 | 13.4 | 6.4 | 10.8 | 9.3 | 18.7 | 20.6 | 14.8 | 4.9 | 10.2 | 21.1 | 6.3 | 14.7 | 6.6 | 11.1 |
| 10 | 36.7 | 37.8 | 27.8 | 32.3 | 44.1 | 50 | 44.7 | 11.9 | 56.6 | 71.0 | 13.0 | 57.2 | 13.7 | 51.1 |
| 20 | 61.3 | 79.9 | 73.8 | 63.0 | 79.2 | 74.9 | 80.8 | 31.2 | 99.8 | 100.0 | 31.1 | 95.5 | 32.3 | 97.4 |
| | $c = 5$ ($\rho = 0.98$) | | | | | | | $\rho = 0.8$ | | | | | | |
| 0 | 5.0 | 0.9 | 4.6 | 3.4 | 2.7 | 3.5 | 2.0 | 5.0 | 1.1 | 6.6 | 3.0 | 2.8 | 3.3 | 2.4 |
| 2 | 2.3 | 1.2 | 5.3 | 3.6 | 2.9 | 3.8 | 2.0 | 2.9 | 2.6 | 10.2 | 3.5 | 5.0 | 3.8 | 3.7 |
| 5 | 4.6 | 2.6 | 7.0 | 5.9 | 5.9 | 6.4 | 4.1 | 4.0 | 9.4 | 24.1 | 4.6 | 15.4 | 4.8 | 11.3 |
| 10 | 12.3 | 11.7 | 15.8 | 14.2 | 16.8 | 16.8 | 15.3 | 7.4 | 55.7 | 75.2 | 7.8 | 58.2 | 7.8 | 51.4 |
| 20 | 33.2 | 65.6 | 61.5 | 36.1 | 58.5 | 41.1 | 57.7 | 20.5 | 99.9 | 100.0 | 18.6 | 96.7 | 18.8 | 98.4 |
| | $c = 15$ ($\rho = 0.94$) | | | | | | | $\rho = 0.7$ | | | | | | |
| 0 | 5.0 | 1.2 | 5.5 | 3.5 | 2.7 | 3.6 | 2.3 | 5.0 | 1.3 | 7.2 | 3.4 | 3.2 | 3.2 | 2.7 |
| 2 | 2.5 | 1.6 | 5.7 | 3.5 | 2.8 | 3.6 | 2.3 | 2.6 | 2.3 | 10.5 | 3.4 | 4.5 | 3.2 | 3.2 |
| 5 | 2.9 | 2.7 | 7.0 | 4.0 | 4.4 | 4.3 | 3.2 | 3.5 | 9.9 | 26.2 | 4.1 | 16.2 | 4.1 | 12 |
| 10 | 4.3 | 6.7 | 13.6 | 6.3 | 10.6 | 6.9 | 8.2 | 5.4 | 56.1 | 77.7 | 6.1 | 59.7 | 6.1 | 52.8 |
| 20 | 13.4 | 39.9 | 49.1 | 15.0 | 42.6 | 16.1 | 35.9 | 13.2 | 99.9 | 100.0 | 12.6 | 96.4 | 12.9 | 98.8 |
| | | | | | | | symmetric $f(x) = \|x\|^{1.0}$ | | | | | | | |
| | Local alternative: $\beta = \frac{b}{T^{(1+\alpha)/2}}$ | | | | | | | Local alternative: $\beta = \frac{b}{\sqrt{T}}$ | | | | | | |
| | $c = 0$ ($\rho = 1$) | | | | | | | $\rho = 0.9$ | | | | | | |
| 0 | 5.0 | 1.1 | 6.1 | 4.9 | 6.8 | 6.5 | 4.0 | 5.0 | 1.6 | 6.7 | 3.2 | 2.8 | 3.3 | 2.7 |
| 2 | 5.7 | 2.4 | 7.6 | 5.3 | 11.2 | 10.9 | 7.5 | 3.7 | 3.9 | 12.4 | 4.7 | 9.5 | 4.8 | 7.0 |
| 5 | 23.8 | 17.4 | 15.6 | 16.0 | 29.1 | 33.0 | 26.1 | 12.8 | 33.8 | 54.0 | 13.5 | 50.7 | 14.1 | 42.6 |
| 10 | 47.7 | 50.7 | 42 | 47.1 | 58.2 | 59.7 | 56.9 | 31.8 | 96.8 | 99.4 | 30.1 | 93.7 | 30.8 | 95.8 |
| 20 | 71.1 | 83.4 | 81.1 | 73.1 | 87.4 | 80.3 | 85.9 | 53.9 | 100.0 | 100.0 | 47.6 | 98.8 | 49.2 | 99.3 |
| | $c = 5$ ($\rho = 0.98$) | | | | | | | $\rho = 0.8$ | | | | | | |
| 0 | 5.0 | 1.0 | 4.5 | 3.4 | 2.5 | 3.5 | 2.1 | 5.0 | 1.2 | 7.2 | 3.1 | 3.0 | 3.2 | 2.5 |
| 2 | 3.4 | 1.4 | 5.5 | 4.0 | 3.3 | 4.2 | 2.4 | 3.5 | 4.0 | 15.3 | 3.8 | 10.1 | 4.1 | 7.6 |
| 5 | 4.7 | 2.8 | 6.8 | 5.9 | 4.7 | 6.7 | 4.5 | 7.5 | 33.4 | 59.6 | 7.9 | 52.6 | 7.8 | 44.3 |
| 10 | 15.1 | 10.0 | 14.3 | 15.6 | 18.7 | 18.6 | 16.5 | 20.8 | 97.4 | 99.8 | 17.3 | 95.7 | 18.4 | 97.6 |
| 20 | 39.0 | 54.9 | 54.1 | 39.9 | 60.2 | 44.4 | 58.4 | 42.1 | 100.0 | 100.0 | 32.5 | 99.6 | 33.5 | 99.7 |
| | $c = 15$ ($\rho = 0.94$) | | | | | | | $\rho = 0.7$ | | | | | | |
| 0 | 5.0 | 1.4 | 5.1 | 3.3 | 2.7 | 3.3 | 2.1 | 5.0 | 1.4 | 6.9 | 3.4 | 3.1 | 3.1 | 2.5 |
| 2 | 2.9 | 1.6 | 5.3 | 3.4 | 2.9 | 3.3 | 2.4 | 3.3 | 4.3 | 16.3 | 3.7 | 9.9 | 3.8 | 7.2 |
| 5 | 3.2 | 1.9 | 6.6 | 3.6 | 3.3 | 3.9 | 3.0 | 5.5 | 32.9 | 62.7 | 5.8 | 54.2 | 5.9 | 45.0 |
| 10 | 5.0 | 3.8 | 9.7 | 5.6 | 7.7 | 6.1 | 6.5 | 15.0 | 98.0 | 99.8 | 12.3 | 95.8 | 12.4 | 97.6 |
| 20 | 11.6 | 16.2 | 26.6 | 13.2 | 27.9 | 13.6 | 22.6 | 32.9 | 100.0 | 100.0 | 22.7 | 99.6 | 23.4 | 99.8 |
| | | | | | | | symmetric $f(x) = \|x\|^{1.5}$ | | | | | | | |
| | Local alternative: $\beta = \frac{b}{T^{(1+\alpha)/2}}$ | | | | | | | Local alternative: $\beta = \frac{b}{\sqrt{T}}$ | | | | | | |
| | $c = 0$ ($\rho = 1$) | | | | | | | $\rho = 0.9$ | | | | | | |
| 0 | 5.0 | 1.1 | 6.1 | 4.9 | 6.8 | 6.5 | 4.0 | 5.0 | 1.6 | 6.7 | 3.2 | 2.8 | 3.3 | 2.7 |
| 2 | 8.6 | 5.0 | 8.3 | 5.5 | 14.1 | 13.9 | 11.0 | 6.5 | 8.0 | 21.1 | 7.4 | 20.6 | 7.6 | 16.2 |
| 5 | 28.8 | 23.8 | 21.1 | 22.6 | 34.6 | 37.1 | 31.5 | 26.2 | 66.5 | 86.5 | 25.0 | 82.9 | 25.7 | 80.8 |
| 10 | 51.3 | 51.3 | 47.6 | 52.1 | 60.5 | 61.6 | 58.5 | 48.4 | 99.9 | 100.0 | 42.5 | 97.9 | 43.7 | 98.7 |
| 20 | 71.0 | 78.3 | 77.7 | 74.2 | 86.2 | 79.0 | 82.4 | 65.2 | 100.0 | 100.0 | 54.0 | 99.3 | 54.9 | 99.5 |
| | $c = 5$ ($\rho = 0.98$) | | | | | | | $\rho = 0.8$ | | | | | | |
| 0 | 5.0 | 1.0 | 4.5 | 3.4 | 2.5 | 3.5 | 2.1 | 5.0 | 1.2 | 7.2 | 3.1 | 3.0 | 3.2 | 2.5 |
| 2 | 3.4 | 1.3 | 5.4 | 3.9 | 3.2 | 4.1 | 2.3 | 5.1 | 7.4 | 25.9 | 5.4 | 21.9 | 5.5 | 16.7 |
| 5 | 3.9 | 2.3 | 6.2 | 5.4 | 4.4 | 6.0 | 3.9 | 16.2 | 67.4 | 91.2 | 14 | 86.5 | 14.3 | 84.5 |
| 10 | 12.1 | 6.5 | 10.8 | 12.8 | 14.4 | 15.1 | 12.4 | 37.0 | 100.0 | 100.0 | 27.5 | 99.0 | 28.5 | 99.5 |
| 20 | 33.7 | 34.6 | 38.0 | 34.2 | 47.5 | 38.0 | 44.2 | 54.5 | 100.0 | 100.0 | 37.4 | 99.9 | 38.2 | 99.8 |
| | $c = 15$ ($\rho = 0.94$) | | | | | | | $\rho = 0.7$ | | | | | | |
| 0 | 5.0 | 1.4 | 5.1 | 3.3 | 2.7 | 3.3 | 2.1 | 5.0 | 1.4 | 6.9 | 3.4 | 3.1 | 3.1 | 2.5 |
| 2 | 2.9 | 1.6 | 5.2 | 3.3 | 2.8 | 3.3 | 2.4 | 4.4 | 7.5 | 27.3 | 4.5 | 22.3 | 4.5 | 16.3 |
| 5 | 2.9 | 1.5 | 6.0 | 3.4 | 2.6 | 3.7 | 2.5 | 11.7 | 68.5 | 92.6 | 10.0 | 87.5 | 10.0 | 85.8 |
| 10 | 3.8 | 2.0 | 6.9 | 4.5 | 4.9 | 4.8 | 4.3 | 28.6 | 100.0 | 100.0 | 19.1 | 99.0 | 19.5 | 99.3 |
| 20 | 6.5 | 5.7 | 13.4 | 7.6 | 12.8 | 8.4 | 10.2 | 46.2 | 100.0 | 100.0 | 27.3 | 99.9 | 27.8 | 99.9 |

*Notes:* See Table 1.

of power in spite of being seriously undersized; in terms of size control, $IVX^2$ might be preferred.

## 5 Nonlinear predictability of S&P 500 stock returns

We now turn our attention to the issue of whether stock returns may be predicted, be it in a linear or a nonlinear fashion. We take advantage of the superior performance of linear methods (represented here by IV-based tests) in terms of power to detect the alternative of predictability. We test the null hypothesis of no predictability of S&P 500 stock returns by applying the tests compared in the previous sections, i.e. the tests by Juhl (2014), Kasparis et al. (2015) and Kostakis et al. (2015), as well as the two-stage least squares test from Section 3. For a particular emphasis of the nonlinearity aspect, we also include quadratic terms and use the statistics $\mathbf{IVX^2}$ and $\mathbf{2S^2}$; see Section 4 for implementation details. For the nonparametric approaches we only include results for one bandwidth each; we include the U1 test with bandwidth $h = \hat{\sigma}_x T^{-0.2}$ and the F2 test with $h = \hat{\sigma}_v T^{-0.2}$ as the two performed better in simulations.

### 5.1 Data processing

The analysis is conducted using monthly data provided on the webpages of Amit Goyal and Robert Shiller.[11] We focus on log dividend-price and earnings-price ratios as possible predictors, as is common in the literature (see e.g. Campbell and Yogo, 2006). The log dividend-price ratio [log(D/P)] is computed as difference of log moving one-year average dividends and log prices of the S&P 500 index. The log earnings-price ratio [log(E/P)] is defined analogously but using log moving one-year average earnings. The dependent variable stock returns is computed including dividends. For further details to the data we refer to Welch and Goyal (2008). The considered dataset includes the most recent update (as to August 2017), with data from January 1926 to December 2016 and a total of 1092 observations (see Figure 1 again).

In what concerns the stability of the financial ratios, their logs are seen to have dropped to lower levels towards the turn of the millenium. This is an issue, since Lettau and Van Nieuwerburgh (2008) point out that persistent changes in the mean of valuation ratios can have a substantial impact on the inference of return forecasting regressions; in fact, they find stable in-sample stock return predictability when log(D/P) is adjusted for one or two breaks (but not for log(D/P) itself).

We therefore account for changes in the level of the predictor series, but choose not to adopt the break dates suggested by Lettau and Van Nieuwerburgh. Rather, we conduct a new break analysis in the mean of log(D/P) and log(E/P). We do this for two reasons. First, the method employed by Lettau and Van Nieuwerburgh (2008) to identify breaks is not valid under uncertain persistence, while we apply a procedure proposed by Perron and Yabu (2009) which is designed to be robust. Second, Lettau and Van Nieuwerburgh (2008) use a different data set and the longer time span of the data used here allows for a more precise timing of the breaks, even if we had used the original break detection tool. Indeed, we find differences in the time of the breaks; see below. After identifying structural changes in the mean of log(D/P) and log(E/P), we adjust the series for permanent shifts accordingly.

[11] We would like to thank Amit Goyal and Robert Shiller for making the data freely available on their webpages `http://www.hec.unil.ch/agoyal/` and `http://www.econ.yale.edu/~shiller/data.htm`.

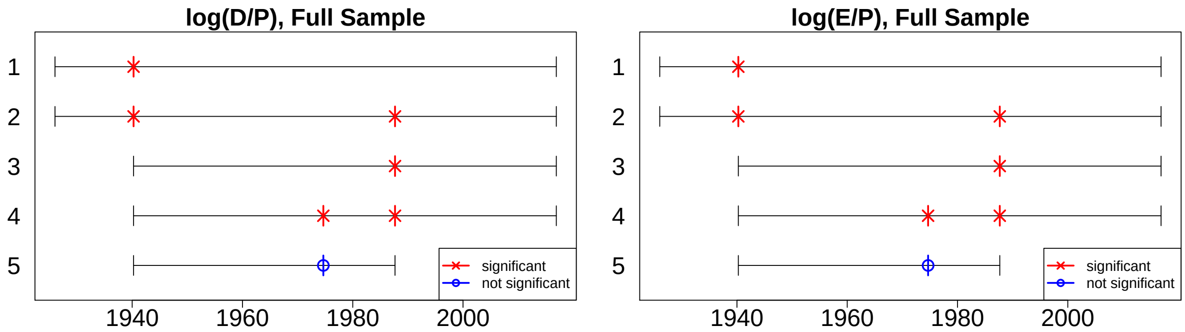


Figure 3: Sequential break identification, five steps in total; Left: log(D/P), right: log(E/P); see the text for more details and Table 4 for exact figures.

Table 4: Sequential break identification, test results of the five consecutive steps to identify structural breaks in the full sample

| | Step | Start | 1st | 2nd | End | $Exp - W_{FS}$ | $CV_{0.95}$ | Significant |
|---|---|---|---|---|---|---|---|---|
| | 1 | 1926:M01 | 1940:M04 | | 2016:M12 | 6.65 | 1.74 | ✓(0.990) |
| | 2 | 1926:M01 | 1940:M04 | 1987:M09 | 2016:M12 | 10.75 | 1.69 | ✓(0.990) |
| log(D/P) | 3 | 1940:M04 | 1987:M09 | | 2016:M12 | 10.79 | 1.74 | ✓(0.990) |
| | 4 | 1940:M04 | 1974:M09 | 1987:M09 | 2016:M12 | 11.48 | 1.69 | ✓(0.990) |
| | 5 | 1940:M04 | 1974:M09 | | 1987:M09 | 1.16 | 1.74 | |
| | 1 | 1926:M01 | 1940:M04 | | 2016:M12 | 3.07 | 1.74 | ✓(0.975) |
| | 2 | 1926:M01 | 1940:M04 | 1987:M09 | 2016:M12 | 4.97 | 1.69 | ✓(0.990) |
| log(E/P) | 3 | 1940:M04 | 1987:M09 | | 2016:M12 | 4.69 | 1.74 | ✓(0.990) |
| | 4 | 1940:M04 | 1974:M09 | 1987:M09 | 2016:M12 | 4.35 | 1.69 | ✓(0.990) |
| | 5 | 1940:M04 | 1974:M09 | | 1987:M09 | 1.22 | 1.74 | |

*Notes:* For details see the text.

Regarding implementation details of the Perron and Yabu (2009) procedure (concretely, we use the $Exp - W_{FS}$ test statistic of Perron and Yabu with parameter $\varepsilon = 0.15$.), Perron and Yabu only provide critical values to test for either one or two breaks which makes testing for more breaks less straightforward. This is why we apply five consecutive steps to identify structural breaks in the full sample. As a result, we identify identical break dates for log(D/P) and log(E/P) as follows. In the first two steps we test for one and for two breaks in the full sample. We find one significant break in April 1940 and two jointly significant breaks in April 1940 and September 1987. The third step serves as robustness check of the break in 1987. We test the subperiod beginning in April 1940 for one break and confirm the break in September 1987. In the fourth step we search for two breaks in the aforementioned subperiod and find jointly significant breaks in September 1974 and September 1987. The last step is a robustness check for the break in 1974. On the subperiod from April 1940 to September 1987 we find again a break in 1974 but it turns out to be insignificant. The findings are summarized in Figure 3 and the precise test results can be found in Table 4.

Thus, we find one significant break in April 1940 and a second significant break in September 1987. A third break can be found in September 1974. The latter break turns out to be only jointly significant with the break in 1987 while it is marginally insignificant if taken alone. For comparison, Lettau and Van Nieuwerburgh (2008), report one break in the early 1990s or two breaks around 1954 and 1994 for log(D/P) but they apply a different method and use an annual sample from 1927 to 2004.

The observed break dates roughly correspond to critical economic events. While the break identified

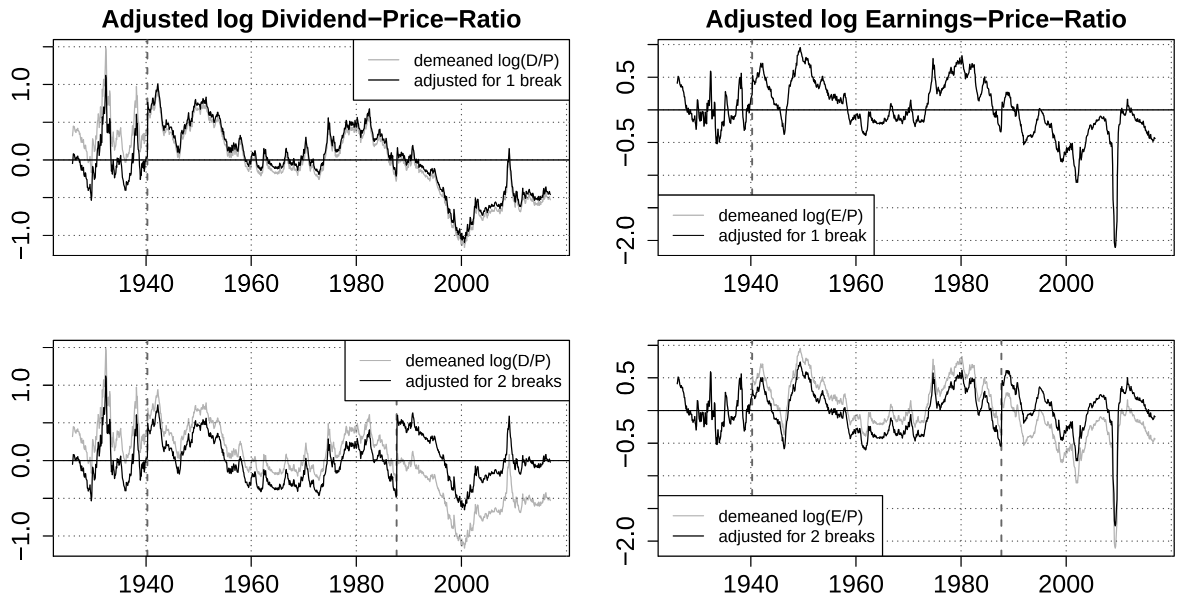


Figure 4: Demeaned and break-adjusted predictor variables; left: log(D/P), right: log(E/P), upper panels are adjusted for one break, lower panels adjusted for two breaks.

in April 1940 might be a late indicator for the Great Depression in the early 1930s, it is actually more likely that it is tied to the beginning of World War II. The break in September 1987 is only one month prior to the Black Monday. We do not consider the break in September 1974 since it is insignificant, although it is a plausible break date around the First Oil Crisis. The demeaned and break adjusted series log(D/P) and log(E/P) are displayed in Figure 4. We do not treat the peak in log(D/P) and the drop in log(E/P) in 2009 as permanent shifts but as outliers since the values for Dividends, Earnings and Prices return to levels comparable to 2008 until 2010. The reason for the outliers is that Dividends, Earnings as well as Prices decreased in 2008-2009 during the financial crisis. Dividends, however, decreases much less than Earnings since Dividends are long dated while Earnings react immediately. The effect on Prices is smaller than for Earnings but larger than for Dividends which leads to an increase in the ratio log(D/P) and a decrease in the ratio log(E/P). In the following analysis, we shall use the original series, the series adjusted for one break, as well as the series adjusted for two breaks as putative predictors, for both log(D/P) and log(E/P).

Before proceeding to the analysis, note that adjusting for permanent shifts in the mean tends to reduce the degree of persistence of a series. This can be shown by fitting the best (as indicated by Akaike's Information Criterion) AR(p) process to the data and adding up the estimated autoregressive parameters; the cumulated autoregressive coefficients serve as an (indirect) indicator for the degree of persistence of a series; see Cochrane (1988). Table 5 suggests that the more breaks are adjusted for, the less persistent the series are. The decrease in persistence is of course not an issue here since the considered tests cope with stable as well as nearly integrated regressors. However, the drop in persistence is not very large in absolute terms.

## 5.2 Data analysis

The starting step of the analysis is to fit a local polynomial regression of stock returns on the lagged financial ratios from above to get an idea about the shape of the predictive relations; this also

Table 5: Sum of autoregressive parameters of autoregressive model fits for series adjusted for 0, 1 or 2 breaks

| Data | $p_{aic}$ | $\sum_{i=1}^{p_{aic}} \phi_i$ | Data | $p_{aic}$ | $\sum_{i=1}^{p_{aic}} \phi_i$ |
|---|---|---|---|---|---|
| $\log(D/P)_{FS;0}$ | 22 | 0.9935 | $\log(E/P)_{FS;0}$ | 12 | 0.9831 |
| $\log(D/P)_{FS;1}$ | 6 | 0.9891 | $\log(E/P)_{FS;1}$ | 12 | 0.9831 |
| $\log(D/P)_{FS;2}$ | 6 | 0.9723 | $\log(E/P)_{FS;2}$ | 6 | 0.9634 |

*Notes:* Model order selection conducted via AIC.

allows for a quick check of whether the monotonicity assumptions are fulfilled. The nonparametric regression curves are computed as local quadratic regressions using the closest 75% data points at each $x$-value with tricubic weighting; the local regressions also give the pointwise asymptotic confidence bands.

The findings are largely consistent with the preliminary discussion in Figure 2, even after adjusting for breaks; see Figure 5. This suggests that nonlinear predictive power, should it be confirmed statistically, is no artifact of the persistence of the predictors. A pointwise confidence band is computed as fitted conditional mean plus/minus two times standard error of the fits. The fact that it is a pointwise band implies that inferece on predictability cannot be based on it; see Juhl (2014). The monotonicity assumption appears to hold for log(D/P) while it is apparently violated for log(E/P). For this reason, we shall also discuss the outcomes of tests relying on IV regressions with quadratic terms, i.e. IVX² and 2S² from the previous section.

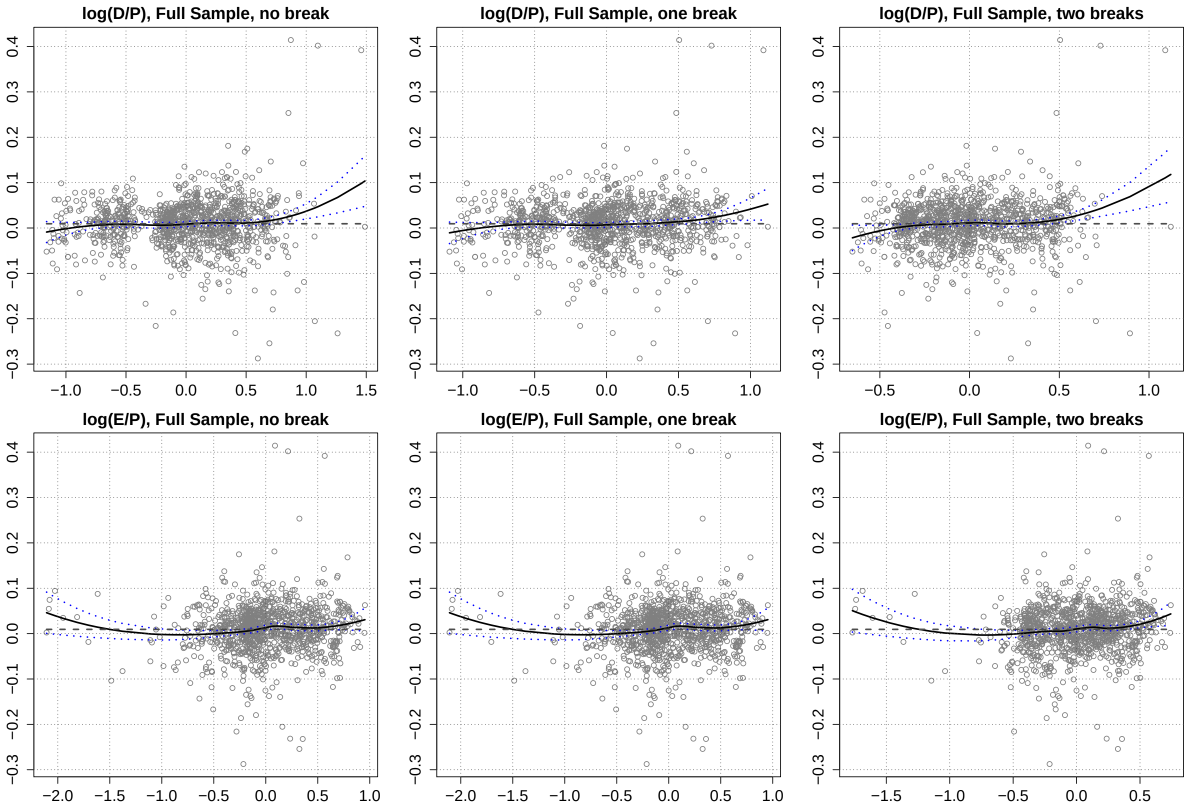


Figure 5: Stock returns against lagged financial ratios in full sample, including pointwise confidence band computed as fit plus/minus two times standard error, dashed line is mean of regressor, see the text for computation details; left to right: demeaned, adjusted for one break, adjusted for two breaks, top: log(D/P), bottom: log(E/P).

Table 6: Test significance of nonparametric and linear test procedures, log(D/P) and log(E/P) with 0, 1 or 2 breaks as predictors; full sample (1926:M01 - 2016:M12).

| Predictor | U1 | F2 | IVX | $IVX^2$ | 2SLS | $2S^2$ |
|---|---|---|---|---|---|---|
| $\log(D/P)_0$ | 0.612 | 0.738 | 0.111 | 0.223 | 0.316 | 0.442 |
| $\log(D/P)_1$ | 0.336 | 0.227 | 0.056 (*) | 0.157 | 0.045 (**) | 0.118 |
| $\log(D/P)_2$ | 0.936 | 0.424 | 0.049 (**) | 0.054 (*) | 0.047 (**) | 0.033 (**) |
| $\log(E/P)_0$ | 0.192 | 0.081 (*) | 0.037 (**) | 0.084 (*) | 0.038 (**) | 0.029 (**) |
| $\log(E/P)_1$ | 0.200 | 0.062 (*) | 0.037 (**) | 0.085 (*) | 0.039 (**) | 0.030 (**) |
| $\log(E/P)_2$ | 0.070 (*) | 0.003 (***) | 0.028 (**) | 0.054 (*) | 0.020 (**) | 0.003 (***) |

*Notes:* Significance: (*) $p \leq 0.10$ , (**) $p \leq 0.05$, (***) $p \leq 0.01$; for further details see the text.

We conduct a series of tests for predictability of stock returns with null hypothesis of no predictability. The predictor variables are lagged log(D/P) and lagged log(E/P). The series log(D/P) and log(E/P) are demeaned and adjusted for either zero, one or two breaks.

Table 6 summarizes the resulting p-values for all of these tests of predictability. (We display the p-values instead of test statistics to maintain comparability since the test statistics are of different magnitude and have different critical values.)

The test according to Juhl (2014) shows almost no predictability of S&P 500 stock returns in the full sample, neither for log(D/P) nor for log(E/P). This result is not too surprising since the test is undersized and has low power as seen in the previous section. The tests based on Kasparis et al. (2015) show no predictability for log(D/P) but some predictability for log(E/P) in the full sample. One has to keep in mind that the test tends to be a bit oversized, though.

In contrast to the nonparametric approaches, both linear procedures show substantial rejections of the null hypothesis of no predictability. It shows that log(D/P) is able to predict stock returns when adjusted for permanent shifts in its mean: both linear procedures yield significant predictability (at the 5% level). The series log(E/P) on the other hand is able to significantly predict stock returns in this sample regardless of whether it is adjusted for breaks or not.

Both $IVX^2$ and $2S^2$ show less predictability than IVX and 2SLS for log(D/P); however, for log(E/P) it looks the other way round. This is in line with the derivations in Chapter 4 since the monotonicity assumptions are violated for log(E/P) but appear to hold for log(D/P). We furthermore summarize the individual $t$-statistics for the parameters of the linear and the quadratic terms in $IVX^2$ and $2S^2$ in Table 7 to check whether the quadratic terms are significant or not. Significance of the quadratic term would point towards a U-shaped predictive relation, but they are only significant for the log(E/P) series adjusted for two breaks, indicating a weak, at most moderate, U-shaped relation as hinted upon by Figure 5.

Note that, in line with the results of our Monte Carlo experiments, IVX and 2SLS are able to outperform the nonparametric approaches on the log(D/P) samples where the monotonicity assumption appears to hold. IVX and 2SLS still perform quite well on the log(E/P) samples even if the monotonicity assumption appears to be violated. Adding quadratic terms ($IVX^2$ and $2S^2$) leads to further refinements, with $2S^2$ indicating a U-shaped predictive relation for the log(E/P) series.

Summing up, we find evidence of stock return predictability. Adjusting for permanent shifts in

Table 7: Summary of $t$-statistics for individual parameters of quadratic predictive regressions estimated via IVX$^2$ and 2S$^2$

| Predictor | $t_1^{IVX2}$ | $t_2^{IVX2}$ | $t_1^{2S2}$ | $t_2^{2S2}$ |
|---|---|---|---|---|
| $\log(D/P)_0$ | 0.948 | 0.569 | 1.094 | 0.215 |
| $\log(D/P)_1$ | 1.841 (*) | 0.786 | 2.011 (**) | 0.571 |
| $\log(D/P)_2$ | 2.412 (***) | 0.364 | 2.072 (**) | 0.575 |
| $\log(E/P)_0$ | 2.168 (**) | 1.316 | 2.052 (**) | 0.439 |
| $\log(E/P)_1$ | 2.160 (**) | 1.315 | 2.044 (**) | 0.436 |
| $\log(E/P)_2$ | 2.373 (***) | 1.730 (*) | 2.486 (***) | 0.785 |

*Notes:* $t_1$ denotes the $t$-statistic associated to the linear term, $t_2$ the $t$-statistic associated to the quadratic term.

the mean always leads to more significant results for the linear procedures. For log(E/P), there is evidence of a nonmonotonic predictive relation.

### 5.3 Refinements

We now conduct a detailed subsample analysis. The purpose of this further discussion is to refine the results coming from the full-sample analysis. We also become more precise on the differences in outcomes when using different inferential procedures, and do this in a compact manner that is not biased by choosing a "suitable" subsample by data snooping, as argued by Scheithauer (2008).[12] Concretely, we test for predictability in each possible subperiod beginning in January and ending in December. Therefore, we obtain 4186 subperiods in total (91 subperiods of a one year length, 90 subperiods of two year length, etc.).

The results for all nonparametric and linear tests, periods and predictor variables are summarized in Figures 6 and 7. Following Scheithauer (2008), the results are shown in colored upper triangular matrices with 4186 cells each (so-called $p$-value surfaces). Every cell represents a different test period where the rows and columns denote different starting and end points. A white cell denotes no significant predictability of S&P 500 stock returns for the corresponding period on a 10% level; otherwise, the higher the significance, the darker the cell color. Significance is not to be taken at face value because there are multiple tests conducted to set up the matrices: we simply use the full-scale subsample analysis to pin down the periods of stronger predictability which are likely to have driven the significant full-sample findings.

Even in the subsample analysis, the test according to Juhl (2014) shows almost no predictability of S&P 500 stock returns, neither for dividend price ratio nor for earnings price ratio (except for the series adjusted for one break and longer subsamples starting before the war). This result is not too striking since the test is undersized and has less local power, as seen in the previous section. The tests based on Kasparis et al. (2015) show predictability for several periods but one has to keep in mind that the test tends to be a bit oversized. Predictability is featured for different periods, depending on the utilized predictor variable (log(E/P) or log(D/P)).

The two linear procedures yield stronger evidence for predictability. For each predictor variable, the $p$-value surfaces show predictability for roughly the same periods, whereby the evidence for two-stage least squares is, expectedly, slightly stronger than for IVX. The results for log(E/P) and

[12]See Hansen and Timmermann (2012) for a discussion of the effects of different subsample splits in the context of pseudo out-of-sample forecasting evaluations.

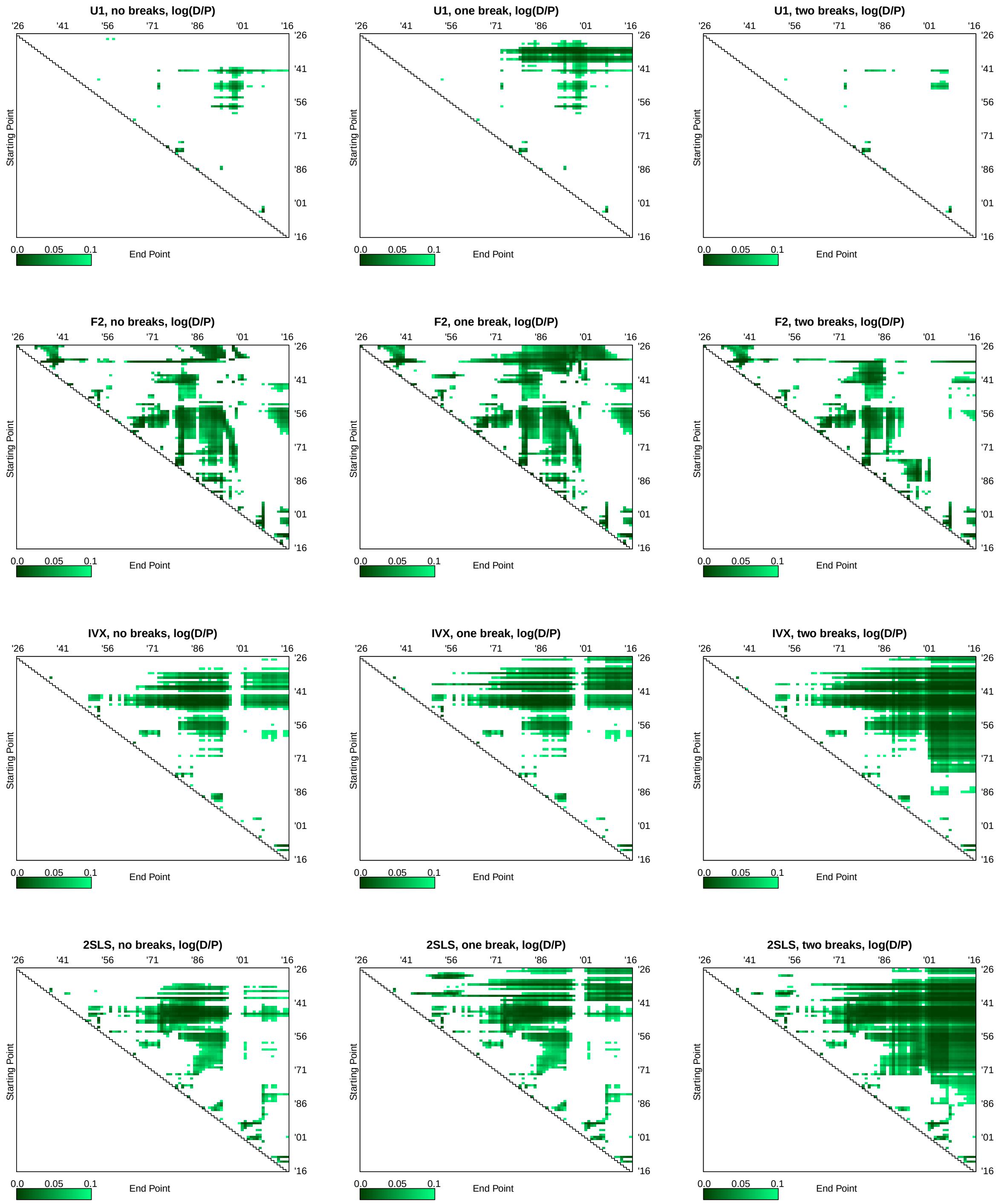


Figure 6: Subsample-wise tests of predictive power of log(D/P): p-values for the U1, the $\hat{F}_{sum}$, the IVX and the combination/IV tests for all possible subsamples starting (ending) in January (December). For further details see the text

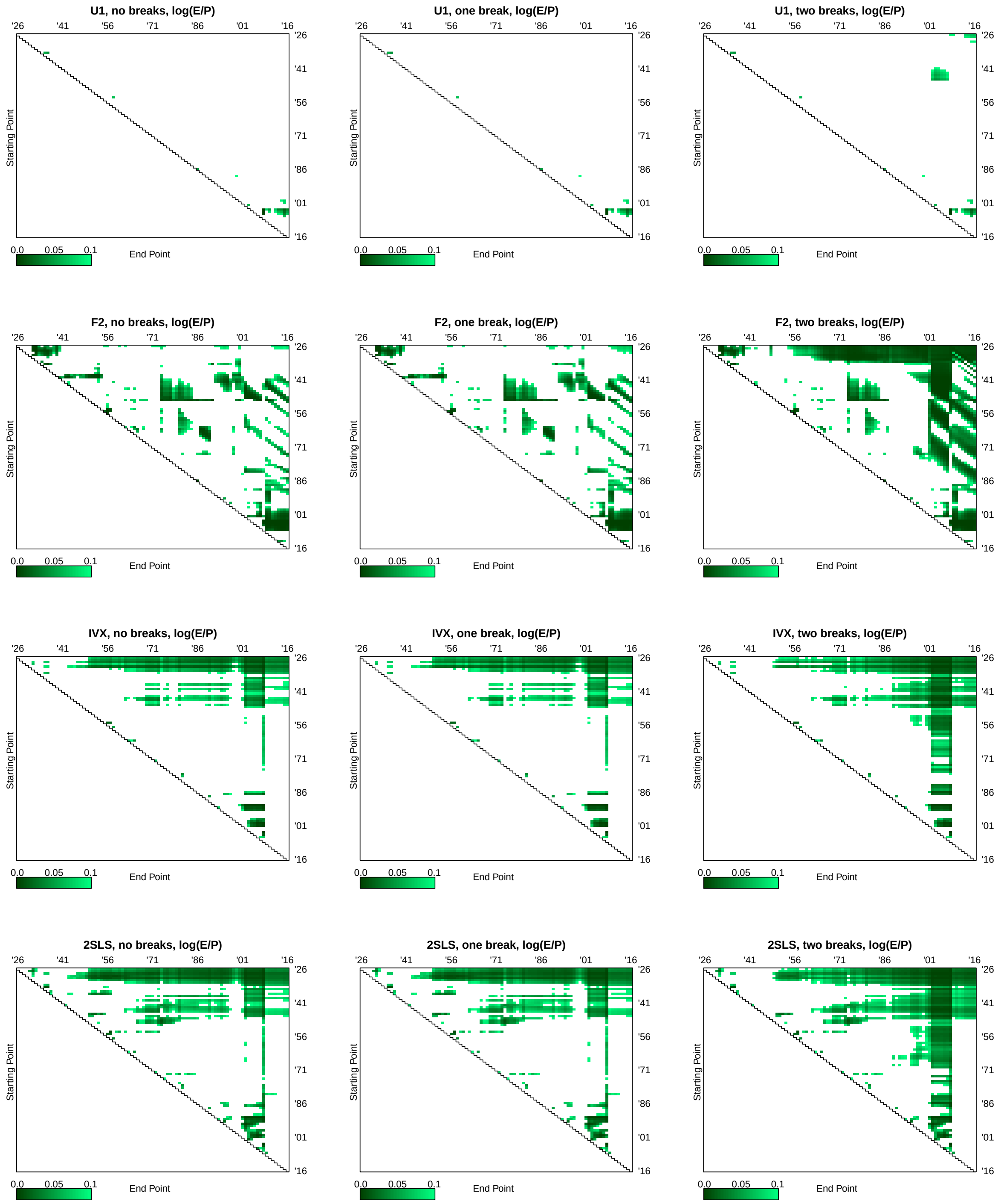


Figure 7: Subsample-wise tests of predictive power of log(E/P); for details see Figure 6

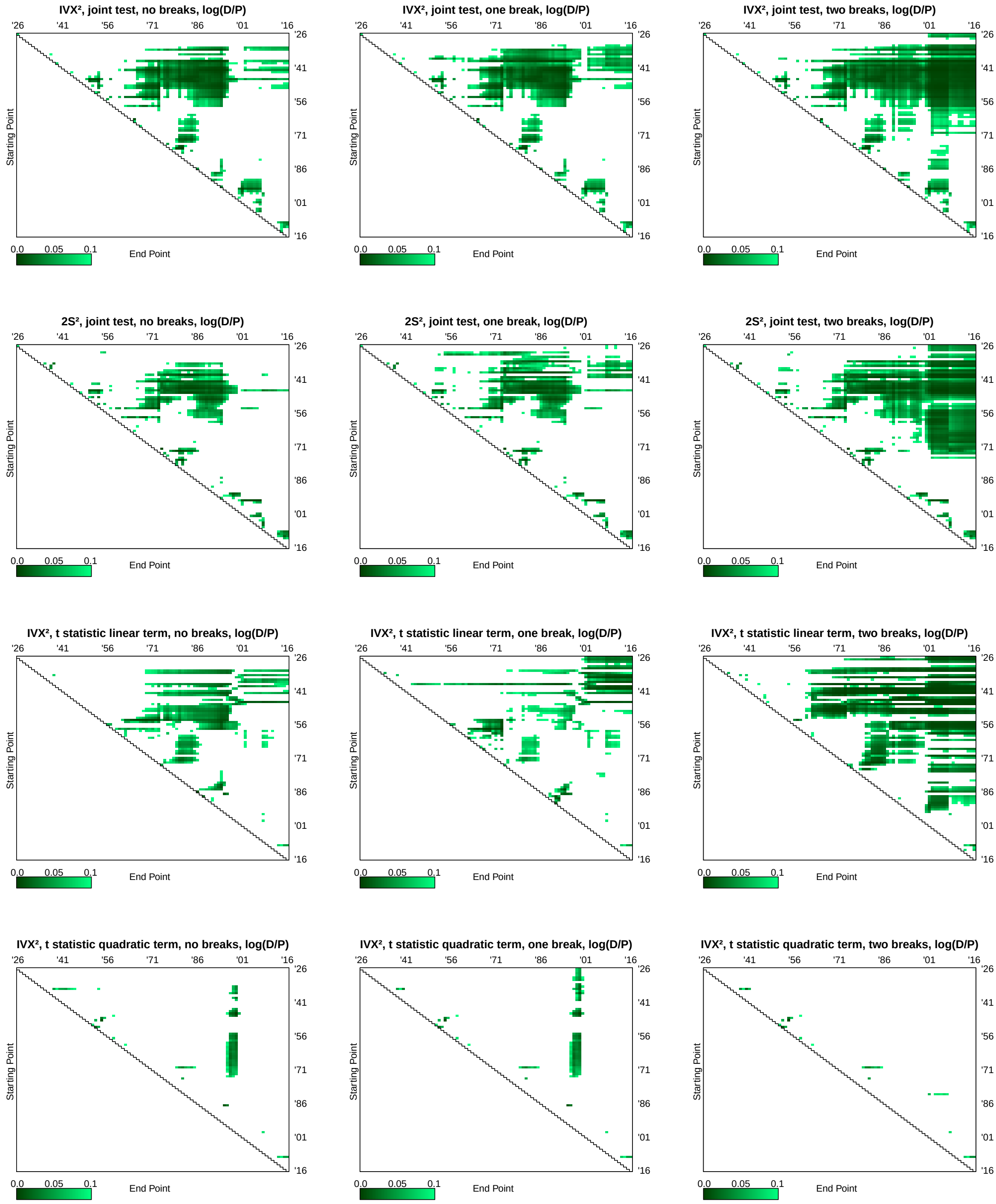


Figure 8: Subsample-wise results for quadratic predictive regressions with log(D/P); p-values for the IVX², the 2S² and the individual $t$-statistics of the linear and the quadratic term in IVX²

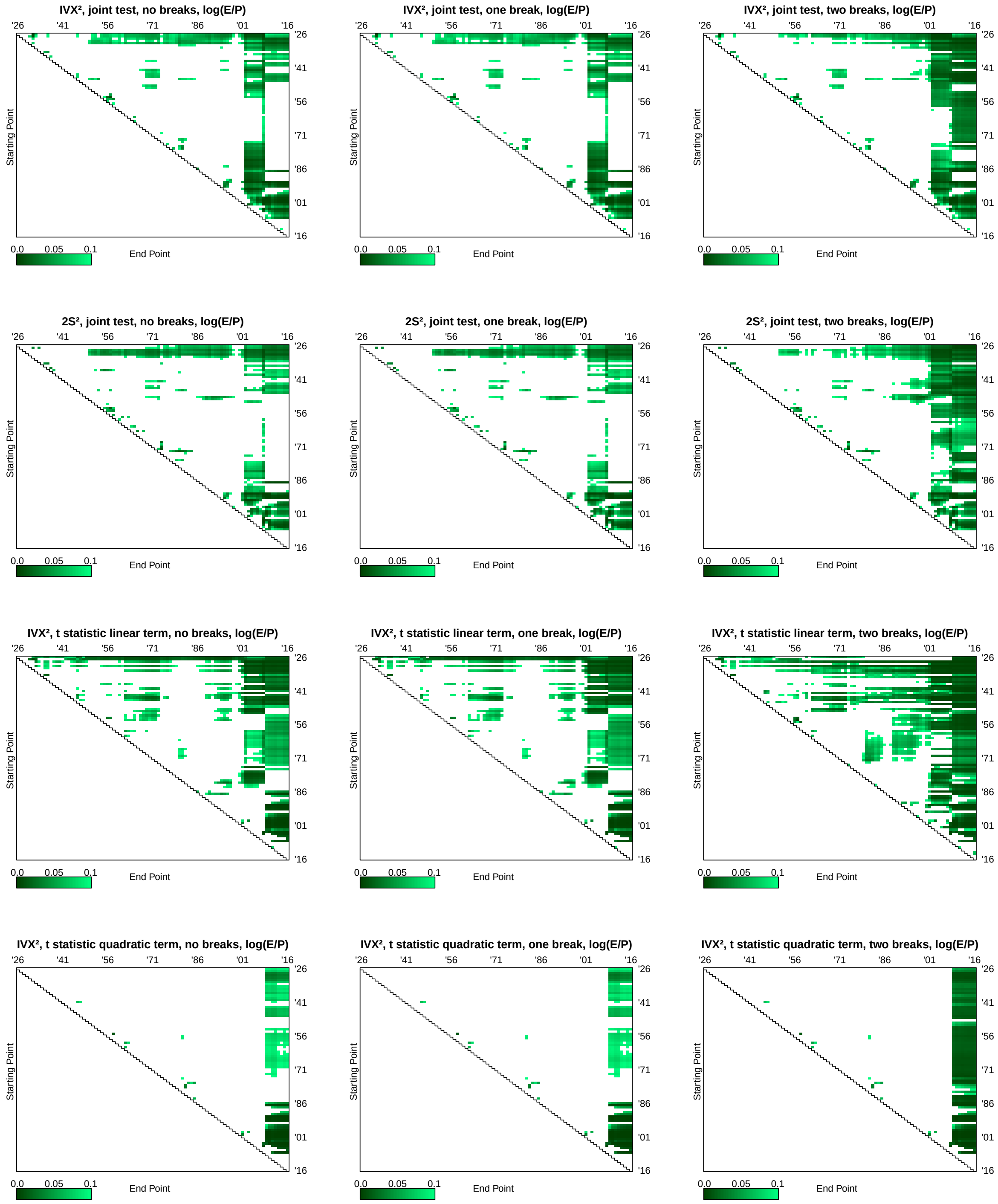


Figure 9: Subsample-wise results for quadratic predictive regressions with log(E/P); for details see Figure 8

log(D/P) do not fully coincide for the linear tests, like in the nonparametric discussion. Some predictive power is indicated for log(D/P) between the mid-thirties and mid-nineties, while log(E/P) would rather predict in periods staring prior to the mid-forties and ending after 1950; again, there are differences in timing between nonparametric and linear procedures. After adjusting log(D/P) for one break, there seems to be additional predictability in samples starting before the fifties and ending past 2000. Adjusting for two breaks leads to predictability in most samples with length of at least 30 years. For log(E/P) there appears to be predictability in samples starting prior to the fifties. Adjusting the data for one break has no large impact here. After adjusting for a second break, also the samples ending between 2002 and 2008 indicate a relation. In comparison, the $\hat{F}_{sum}$ test (as the more powerful nonparametric test) finds predictability in relatively different times and overall less often when compared to the linear tests.

To discuss the type of nonlinearity (monotonic vs. nonmonotonic), we give in Figures 8 and 9 the subsample $t$-statistics for the linear and quadratic terms to see where nonmonotonicity is strong. (Recall, the linear term captures a monotonic relation, while the quadratic term indicates a possible U-shaped predictive relation.) There is little difference in the periods of predictability when considering different numbers of breaks adjusted for; if anything, predictability increases slightly with more breaks adjusted for, but just for log(D/P). The quadratic terms for log(D/P) are practically never significant, confirming that the predictive relation is monotonic. In fact, examining Figure 5, it even looks reasonably close to linearity, with slightly different slopes for very large or very small values of log(D/P). For log(E/P), the quadratic term is always significant for, and only for, subsamples starting before 2008 and ending after 2009, strongly indicating that the U-shaped relation is largely driven by the large drop in log(E/P) during and in the aftermath of the 2008 crisis. The changing predictability of the log(E/P) ratio during the crisis lends support to the findings of Gupta et al. (2013), who find that stock markets are less efficient in incorporating firm-specific information in stock prices. This is likely because a significant reduction in risk arbitrage activities of investors during the crisis leads to an overall reduction in the flow and transparency of information.

## 5.4 Robustness checks

First, conducting the analysis for D/P and E/P in levels rather than logs lead to similar results. One particularity is that stability tests tend to find more breaks in the predictors; this is likely due to the presence of more spikes in the series, spikes that are dampened when taking logs. For two breaks we choose April 1940 and September 1974 and for four breaks additionally September 1987 and September 2008 (2009 for EP) as break dates. The results for the detailed subsample analysis for zero, two or four breaks can be found in Appendix A.1.

Second, we also considered interest rates and long term rates of return as possible predictors beyond the "classical" financial ratios. As proxy for interest rates we take treasury bills, since they have been found to have predictive power (e.g. Campbell and Yogo, 2006). The data analysis is given in Appendix A.2. One particularity of the treasury bill subsample analysis is that we excluded all subsamples starting and ending between 1942 and 1947 since treasury bills are constant in the mid fourties for a span of five years. For both, treasury bills and long-term rates, we find at most one break in mean. In short, the long-term rates provide some evidence in favour of predictability as well, e.g. for samples beginning between the thirties and the seventies, albeit weaker than the earnings-price ratio in its logarithmized form. The evidence for treasury bills is weak at best.

Finally, we also conducted a subsample analysis for linear predictive regressions involving bivariate combinations of log(D/P), log(E/P), treasury bills and long term rates of return for the linear models. We do not add more predictor variables and also stick to data without break adjustment to not inflate the amount of results. It turns out that combinations including long term rates of return seem to perform best in predicting stock returns; nevertheless, predictibility seems to have decreased since the seventies. The results of the subsample analysis with two predictor variables can be seen in Appendix A.3.

# 6 Concluding remarks

The paper argued in favor of the use of linear models to test the forecasting ability of possibly nonstationary regressors in potentially nonlinear models, on the grounds that the loss in local power caused by possible misspecification of the functional form is outweighed by the loss in power caused by resorting to nonparametric estimation and testing procedures. Estimation can be conducted in a flexible manner after having established the existence of a predictive relation.

To correct the size problem of linear procedures in predictive regressions with endogenous regressors of uncertain persistence, we recommend the 2SLS combination of two types of instruments, exogenous yet persistent ones working under persistent regressors only, and certain nonlinear transformations of the regressors, working under stable regressors only. Overall, this IV procedure fares better than alternative nonparametric statistics, under the null and under nonlinear local alternatives, provided that the regression curve is monotonic under the alternative. But if one chooses nonparametric models, say because the monotonicity requirement is violated in an obvious manner, one should rather use the $\hat{F}_{sum}$ test by Kasparis et al. (2015) than the U test by Juhl (2014). Also, adding a quadratic term in the IVX regression is a competitive solution for cases of nonmonotonicity.

The methodological and experimental findings are complemented by an analysis of the predictability of monthly S&P 500 returns. We find that the predictive power of log(D/P) is stronger between the 40s and the 90s, while that of log(E/P) diminished in the post-war period. For log(E/P) we find some evidence of an U-shaped regression function. This appears, however, to be driven by a few abnormal observations during and after the peak of the financial crisis in 2008-2009. Furthermore, long-term rates of return may also predict stock returns, with no evidence of U-shaped relations. As a byproduct, we find that linear methods are well-suited to detect even nonlinear predictability.

# Appendix

## A Tables and Plots for further predictors

### A.1 Financial ratios in levels

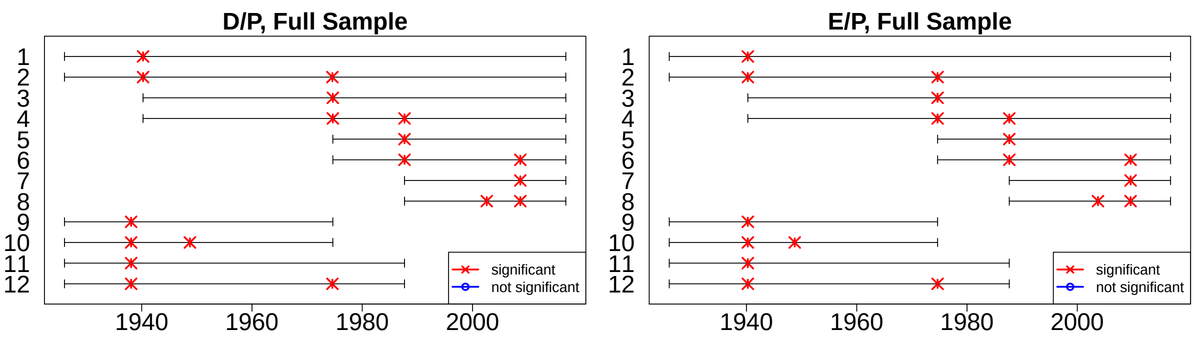


Figure 10: (In-) Significant breaks in D/P and E/P in the full sample based on a significance level of $\alpha = 0.05$.

Table 8: Sequential break identification, test results of twelve consecutive steps to identify structural breaks in the full sample

| | Step | Start | 1st | 2nd | End | $Exp - W_{FS}$ | $CV_{0.95}$ | Significant |
|---|---|---|---|---|---|---|---|---|
| D/P | 1 | 1926:M01 | 1940:M04 | | 2016:M12 | 7.62 | 1.74 | ✓(0.990) |
| | 2 | 1926:M01 | 1940:M04 | 1974:M08 | 2016:M12 | 7.39 | 1.69 | ✓(0.990) |
| | 3 | 1940:M04 | 1974:M09 | | 2016:M12 | 4.00 | 1.74 | ✓(0.990) |
| | 4 | 1940:M04 | 1974:M09 | 1987:M09 | 2016:M12 | 6.88 | 1.69 | ✓(0.990) |
| | 5 | 1974:M09 | 1987:M09 | | 2016:M12 | 8.71 | 1.74 | ✓(0.990) |
| | 6 | 1974:M09 | 1987:M09 | 2008:M09 | 2016:M12 | 9.38 | 1.69 | ✓(0.990) |
| | 7 | 1987:M09 | 2008:M09 | | 2016:M12 | 5.32 | 1.74 | ✓(0.990) |
| | 8 | 1987:M09 | 2002:M08 | 2008:M09 | 2016:M12 | 9.99 | 1.69 | ✓(0.990) |
| | 9 | 1926:M01 | 1938:M02 | | 1974:M09 | 7.88 | 1.74 | ✓(0.990) |
| | 10 | 1926:M01 | 1938:M02 | 1948:M10 | 1974:M09 | 7.33 | 1.69 | ✓(0.990) |
| | 11 | 1926:M01 | 1938:M02 | | 1987:M09 | 9.34 | 1.74 | ✓(0.990) |
| | 12 | 1926:M01 | 1938:M02 | 1974:M08 | 1987:M09 | 8.79 | 1.69 | ✓(0.990) |
| E/P | 1 | 1926:M01 | 1940:M04 | | 2016:M12 | 13.22 | 1.74 | ✓(0.990) |
| | 2 | 1926:M01 | 1940:M04 | 1974:M09 | 2016:M12 | 18.51 | 1.69 | ✓(0.990) |
| | 3 | 1940:M04 | 1974:M09 | | 2016:M12 | 9.56 | 1.74 | ✓(0.990) |
| | 4 | 1940:M04 | 1974:M09 | 1987:M09 | 2016:M12 | 12.61 | 1.69 | ✓(0.990) |
| | 5 | 1974:M09 | 1987:M09 | | 2016:M12 | 4.51 | 1.74 | ✓(0.990) |
| | 6 | 1974:M09 | 1987:M09 | 2009:M09 | 2016:M12 | 6.12 | 1.69 | ✓(0.990) |
| | 7 | 1987:M09 | 2009:M09 | | 2016:M12 | 6.42 | 1.74 | ✓(0.990) |
| | 8 | 1987:M09 | 2003:M10 | 2009:M09 | 2016:M12 | 6.47 | 1.69 | ✓(0.990) |
| | 9 | 1926:M01 | 1940:M04 | | 1974:M09 | 9.29 | 1.74 | ✓(0.990) |
| | 10 | 1926:M01 | 1940:M04 | 1948:M10 | 1974:M09 | 12.36 | 1.69 | ✓(0.990) |
| | 11 | 1926:M01 | 1940:M04 | | 1987:M09 | 9.00 | 1.74 | ✓(0.990) |
| | 12 | 1926:M01 | 1940:M04 | 1974:M09 | 1987:M09 | 12.03 | 1.69 | ✓(0.990) |

*Notes:* Test results of consecutive steps to identify structural breaks in the full sample.

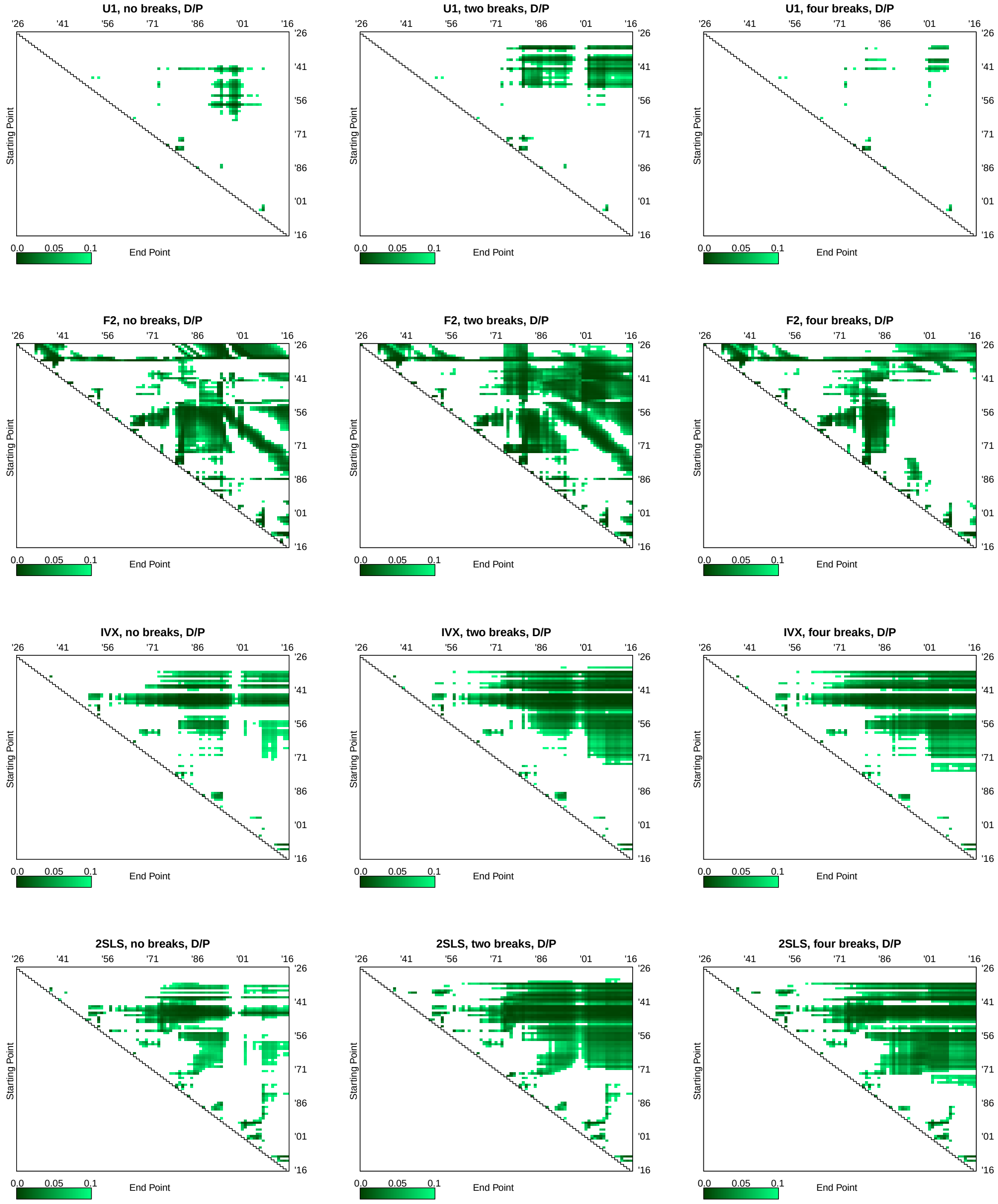


Figure 11: Subsample-wise tests of predictive power of D/P: p-values for the U1, the $\hat{F}_{sum}$, the IVX and the combination/IV tests for all possible subsamples starting (ending) in January (December). For further details see Figure 6

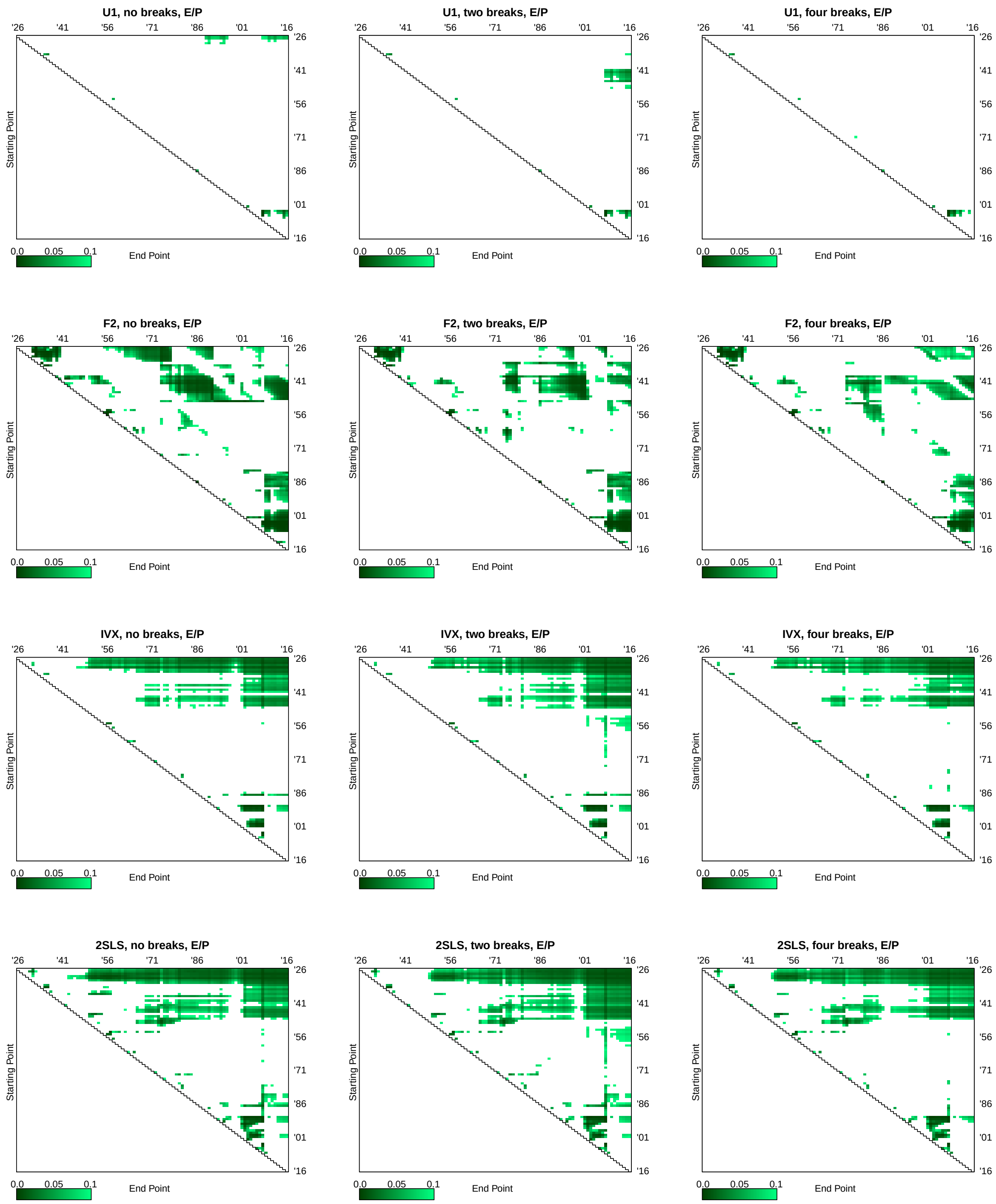


Figure 12: Subsample-wise tests of predictive power of E/P; for details see Figure 11

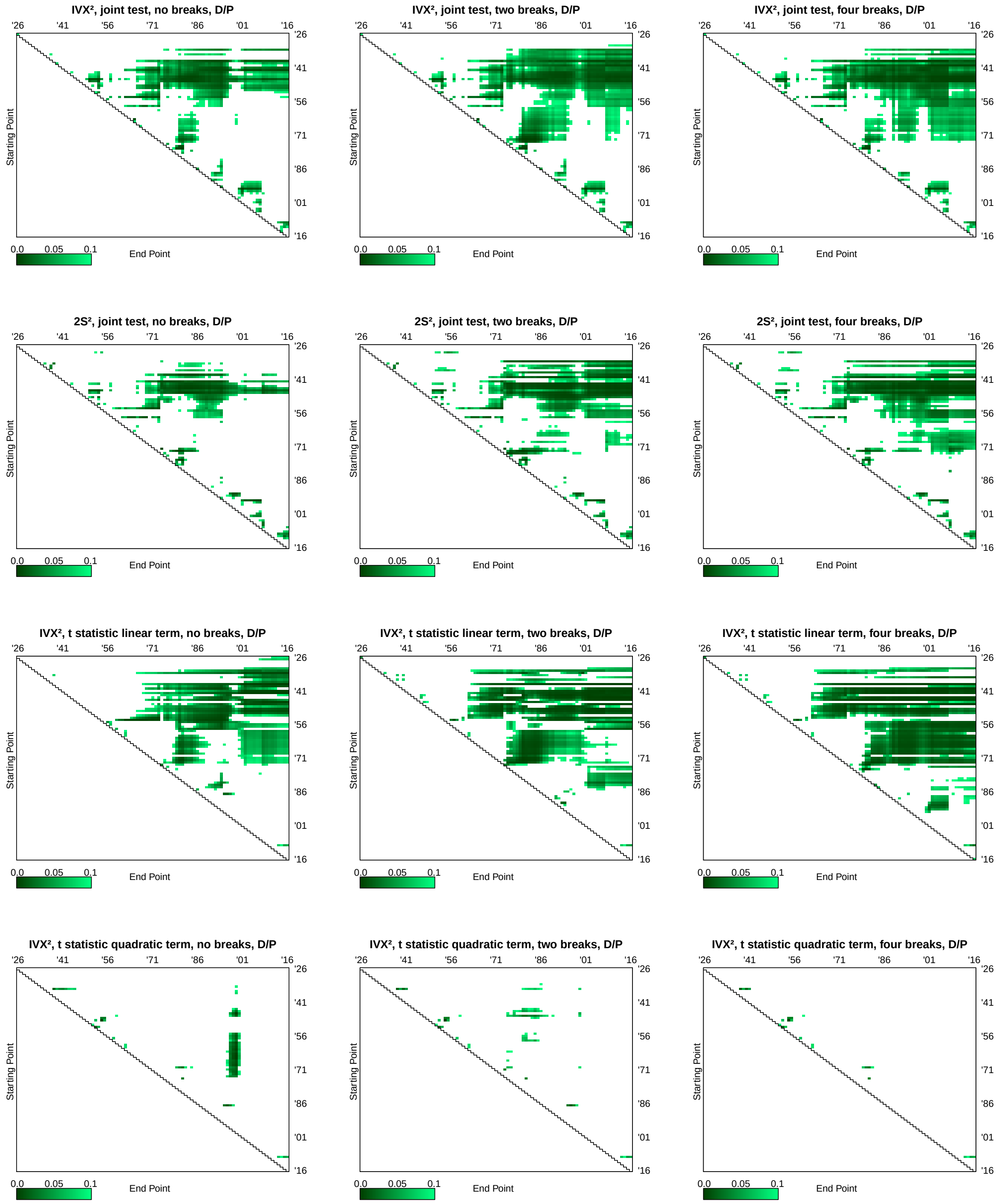


Figure 13: Subsample-wise results for quadratic predictive regressions with D/P; p-values for the $IVX^2$, the $2S^2$ and the individual $t$-statistics of the linear and the quadratic term in $IVX^2$

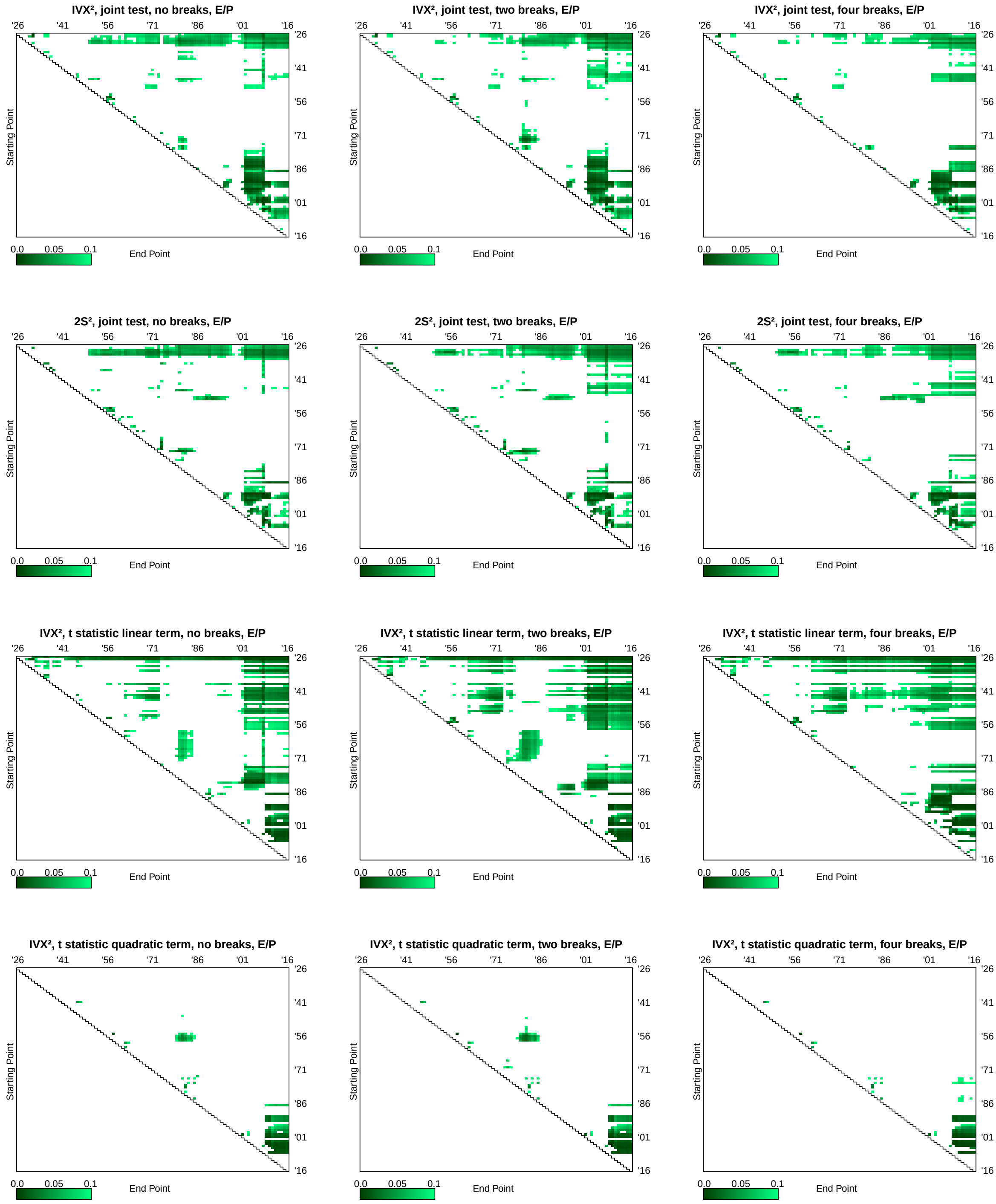


Figure 14: Subsample-wise results for quadratic predictive regressions with E/P; for details see Figure 13

Table 9: Test results on nonparametric and linear test procedures, D/P and E/P; full sample (1926:M01 - 2016:M12) with 0, 2 or 4 breaks.

| Predictor | U1 | F2 | IVX | IVX$^2$ | 2SLS | 2S$^2$ |
|---|---|---|---|---|---|---|
| D/P$_{FS;0}$ | 0.600 | 0.119 | 0.214 | 0.362 | 0.311 | 0.514 |
| D/P$_{FS;2}$ | 0.469 | 0.070 (*) | 0.130 | 0.208 | 0.120 | 0.137 |
| D/P$_{FS;4}$ | 0.932 | 0.014 (**) | 0.174 | 0.269 | 0.167 | 0.202 |
| E/P$_{FS;0}$ | 0.062 (*) | 0.146 | 0.020 (**) | 0.051 (*) | 0.017 (**) | 0.062 (*) |
| E/P$_{FS;2}$ | 0.112 | 0.156 | 0.019 (**) | 0.041 (**) | 0.014 (**) | 0.044 (**) |
| E/P$_{FS;4}$ | 0.187 | 0.053 (*) | 0.032 (**) | 0.095 (*) | 0.021 (**) | 0.040 (**) |

*Notes:* Significance: (*) $p \leq 0.10$ , (**) $p \leq 0.05$, (***) $p \leq 0.01$; for further details see the text.

## A.2 Treasury bills and long-term rates of return

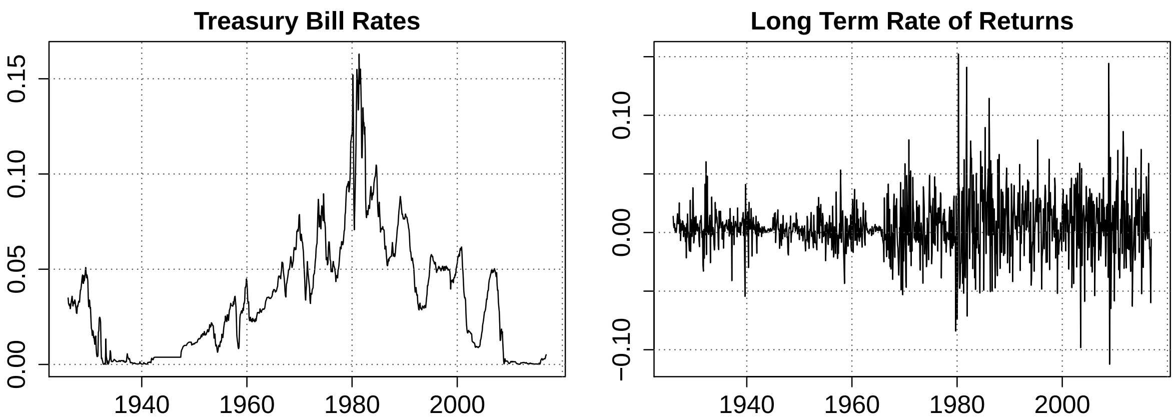


Figure 15: Treasury Bill Rate and Long Term Rate of Returns – monthly observations from December 1926 to December 2016; see Section 5.4 for details

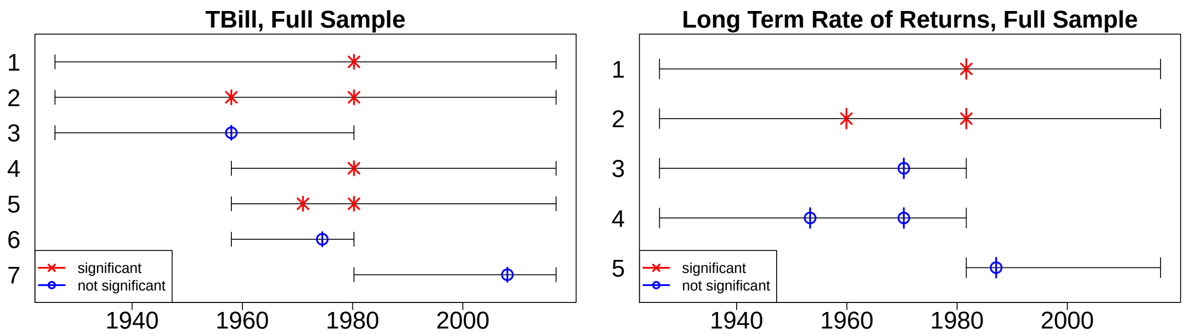


Figure 16: (In-) Significant breaks in Treasury Bill Rate and Long Term Rate of Returns in the full sample based on a significance level of $\alpha = 0.05$.

Table 10: Sequential break identification, test results of five (and seven, respectively) consecutive steps to identify structural breaks in the full sample

| | Step | Start | 1st | 2nd | End | $Exp-W_{FS}$ | $CV_{0.95}$ | Significant |
|---|---|---|---|---|---|---|---|---|
| TBill | 1 | 1926:M01 | 1980:M04 | | 2016:M12 | 86.66 | 1.74 | ✓(0.990) |
| | 2 | 1926:M01 | 1958:M01 | 1980:M04 | 2016:M12 | 86.20 | 1.69 | ✓(0.990) |
| | 3 | 1926:M01 | 1958:M01 | | 1980:M04 | 0.38 | 1.74 | |
| | 4 | 1958:M01 | 1980:M04 | | 2016:M12 | 62.38 | 1.74 | ✓(0.990) |
| | 5 | 1958:M01 | 1971:M01 | 1980:M04 | 2016:M12 | 61.60 | 1.69 | ✓(0.990) |
| | 6 | 1958:M01 | 1974:M07 | | 1980:M04 | 1.37 | 1.74 | ✓(0.900) |
| | 7 | 1980:M04 | 2008:M02 | | 2016:M12 | -0.22 | 1.74 | |
| LTR | 1 | 1926:M01 | 1981:M09 | | 2016:M12 | 2.49 | 1.74 | ✓(0.975) |
| | 2 | 1926:M01 | 1959:M12 | 1981:M09 | 2016:M12 | 2.38 | 1.69 | ✓(0.975) |
| | 3 | 1926:M01 | 1970:M05 | | 1981:M09 | 0.06 | 1.74 | |
| | 4 | 1926:M01 | 19753M05 | 1970:M05 | 1981:M09 | -1.12 | 1.69 | |
| | 5 | 1981:M09 | 1987:M02 | | 2016:M12 | -0.12 | 1.74 | |

*Notes:* Test results of consecutive steps to identify structural breaks in the full sample.

Table 11: Test results on nonparametric and linear test procedures, Treasury Bill Rate and Long Term Rate of Returns with 0 or 1 breaks; full sample (1926:M01 - 2016:M12).

| Predictor | U1 | F2 | IVX | IVX² | 2SLS | 2S² |
|---|---|---|---|---|---|---|
| $\text{TBill}_{FS;0}$ | 0.501 | 0.642 | 0.801 | 0.701 | 0.824 | 0.785 |
| $\text{TBill}_{FS;1}$ | 0.395 | 0.348 | 0.707 | 0.520 | 0.722 | 0.615 |
| $\text{LTR}_{FS;0}$ | 0.273 | 0.190 | 0.115 | 0.240 | 0.115 | 0.260 |
| $\text{LTR}_{FS;2}$ | 0.256 | 0.459 | 0.128 | 0.265 | 0.130 | 0.288 |

*Notes:* Significance: (*) $p \leq 0.10$ , (**) $p \leq 0.05$, (***) $p \leq 0.01$; for further details see the text.

Table 12: Summary of $t$-statistics for individual parameters of quadratic predictive regressions estimated via $\mathbf{IVX^2}$ and $\mathbf{2S^2}$

| Predictor | $t_1^{IVX2}$ | $t_2^{IVX2}$ | $t_1^{2S2}$ | $t_2^{2S2}$ |
|---|---|---|---|---|
| $\text{TBill}_{FS;0}$ | 0.168 | -0.711 | -0.317 | -0.676 |
| $\text{TBill}_{FS;1}$ | 0.035 | -1.025 | -0.493 | -0.958 |
| $\text{LTR}_{FS;0}$ | 1.654 (*) | -0.655 | 1.522 | -0.238 |
| $\text{LTR}_{FS;2}$ | 1.562 | -0.576 | 1.463 | -0.301 |

*Notes:* $t_1$ denotes the $t$-statistic associated to the linear term, $t_2$ the $t$-statistic associated to the quadratic term.

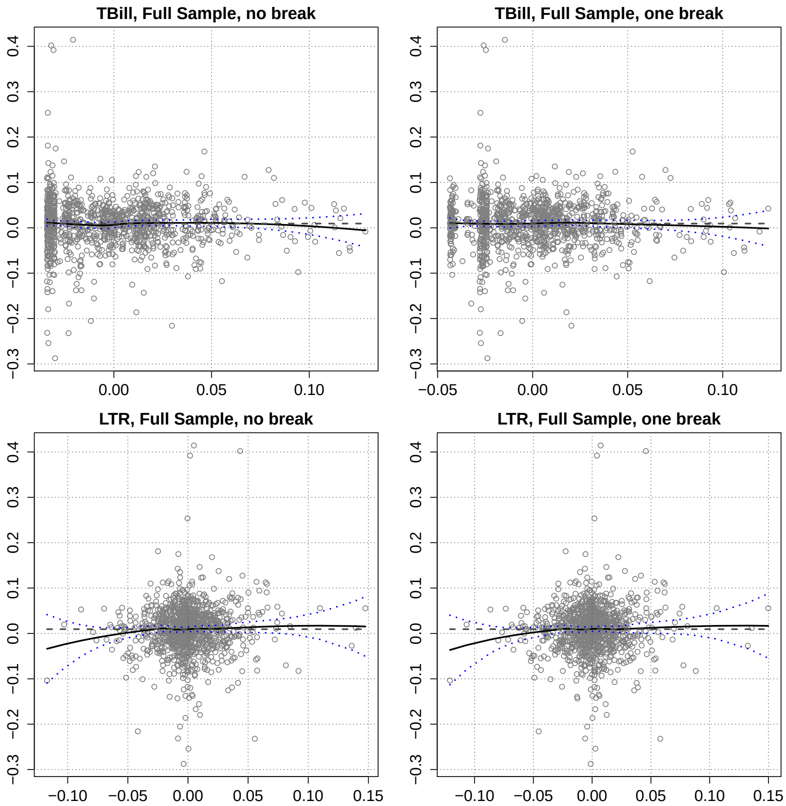


Figure 17: Stock returns against lagged financial ratios in full sample, including pointwise confidence band computed as fit plus/minus two times standard error, dashed line is mean of regressor, left to right: demeaned, adjusted for one break, top: Treasury Bill Rate, bottom: Long Term Rate of Returns.

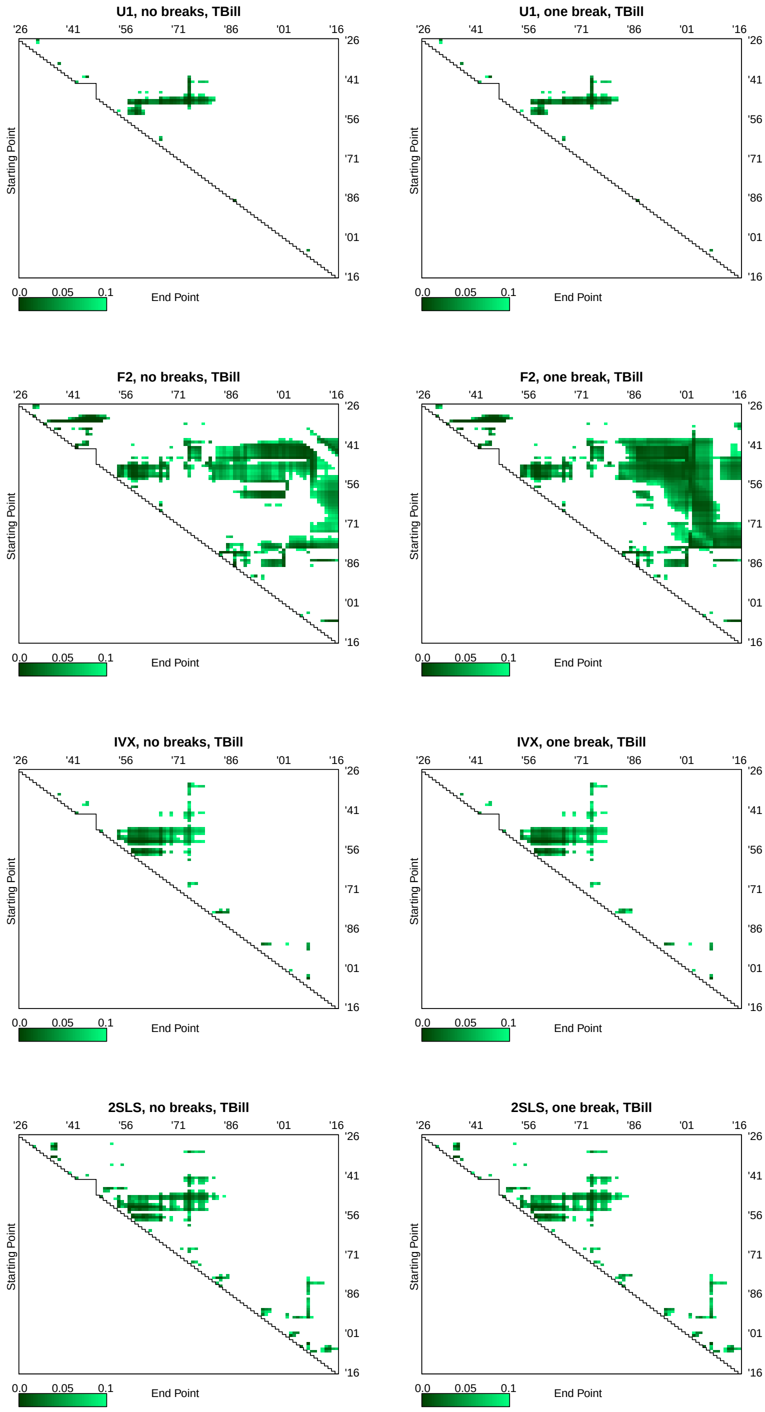


Figure 18: Subsample-wise tests of predictive power of Treasury Bill Rate: p-values for the U1, the $\hat{F}_{sum}$, the IVX and the combination/IV tests for all possible subsamples starting (ending) in January (December), subsamples starting and ending between 1942 and 1947 are excluded since treasury bills are constant in mid fourties. For further details see Figure 6 and Section 5.4

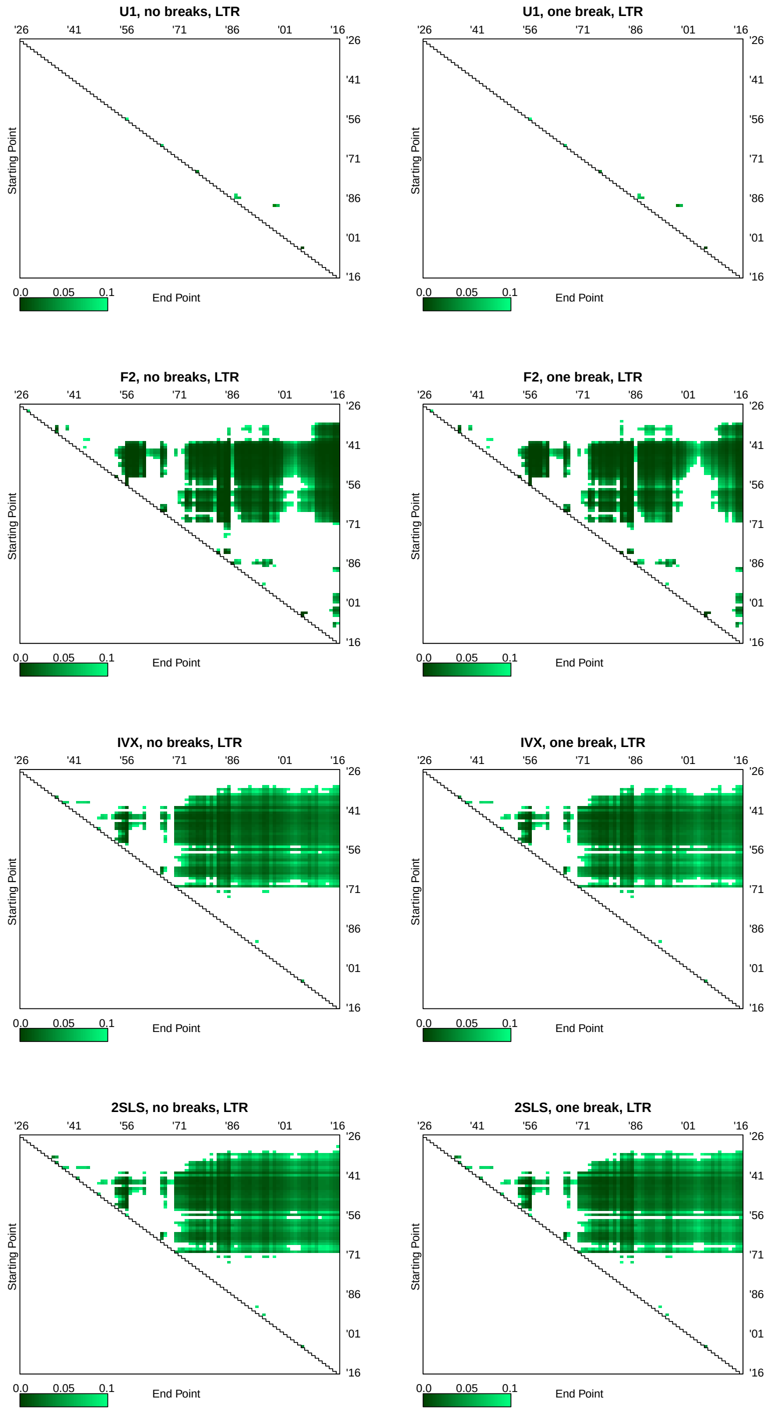


Figure 19: Subsample-wise tests of predictive power of Long Term Rate of Returns; for details see Figure 18

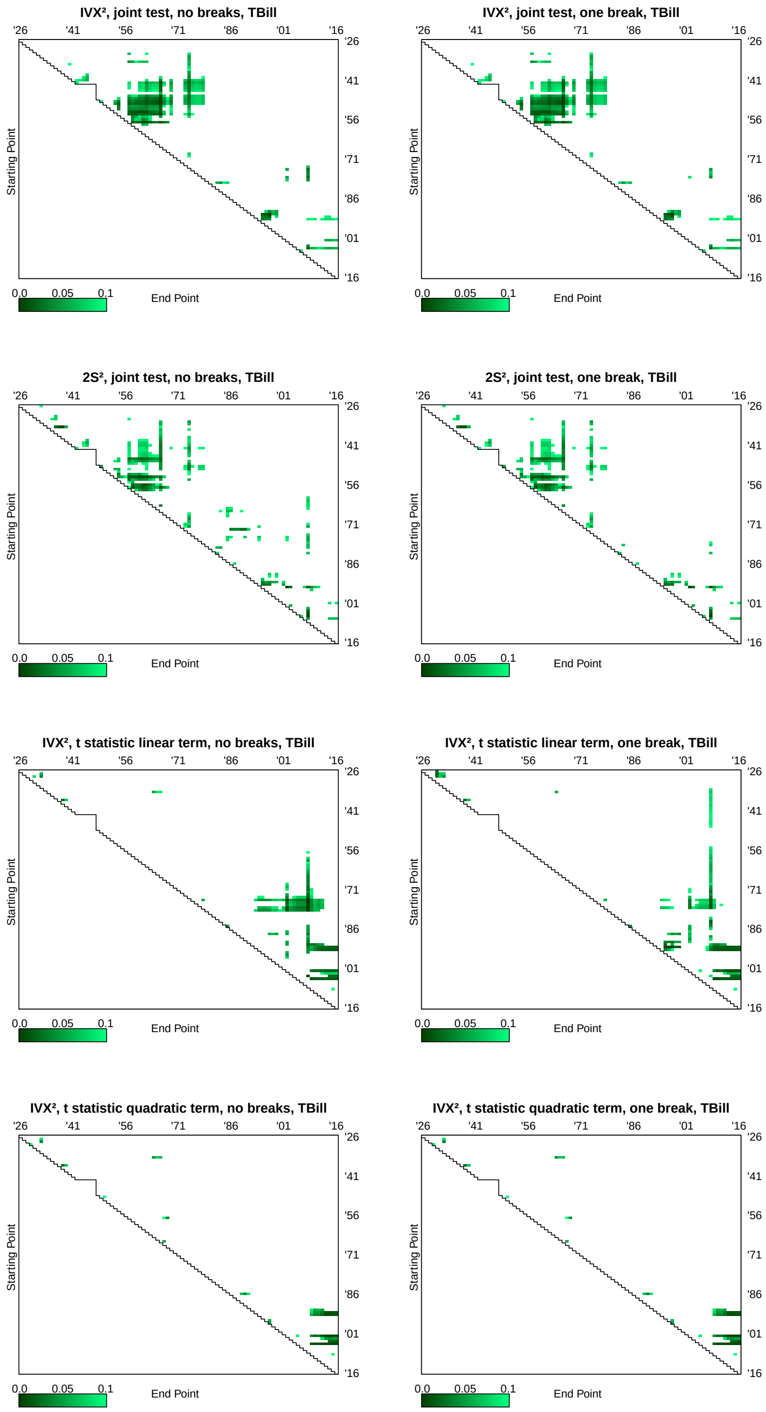


Figure 20: Subsample-wise results for quadratic predictive regressions with Treasury Bill Rate; p-values for the IVX², the 2S² and the individual $t$-statistics of the linear and the quadratic term in IVX². Subsamples starting and ending between 1942 and 1947 are excluded since treasury bills are constant in mid fourties

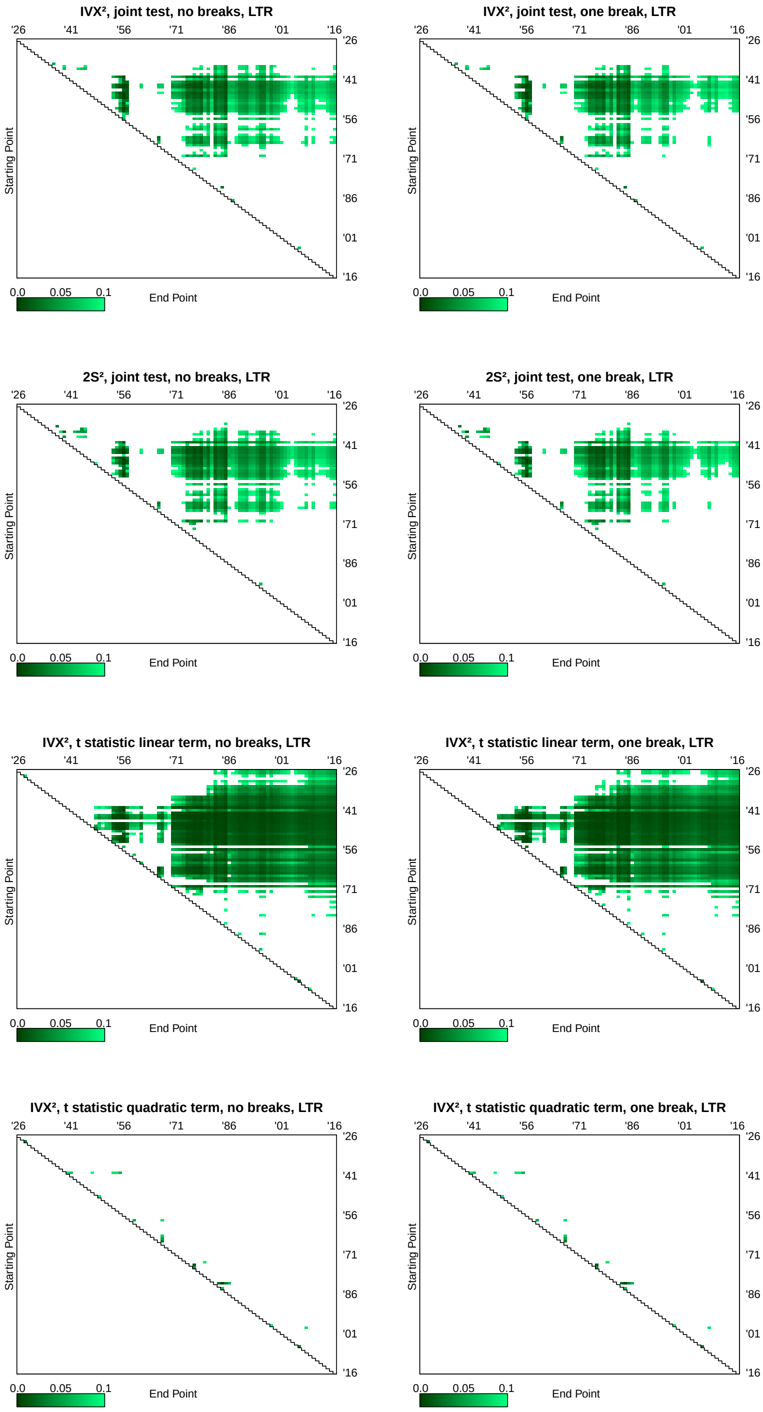


Figure 21: Subsample-wise results for quadratic predictive regressions with Long Term Rate of Returns; for details see Figure 20

## A.3 Combination of two predictors

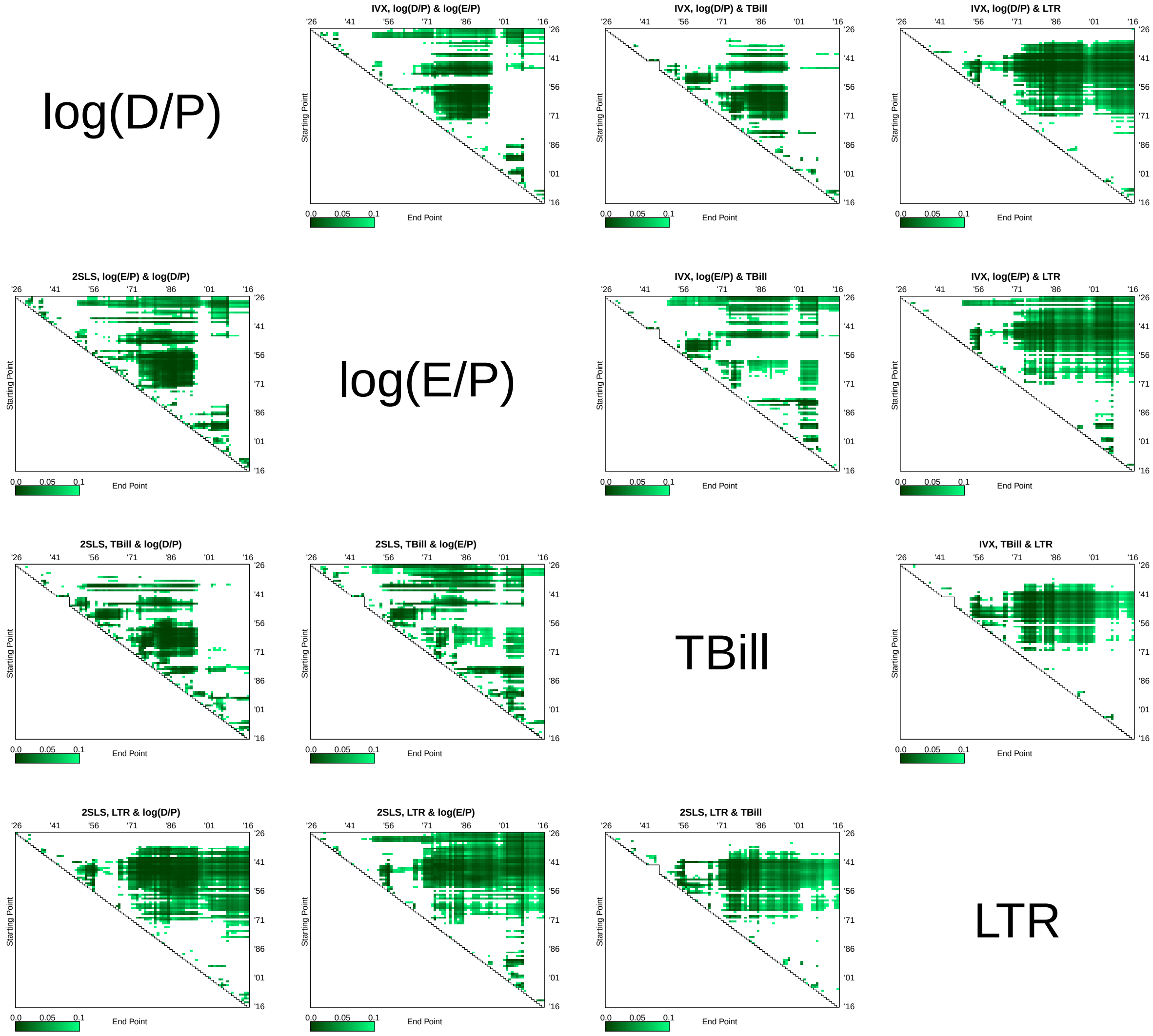


Figure 22: Subsample-wise tests of predictive power for bivariate combinations of log(D/P), log(E/P), Treasury Bill Rate and Long Term Rate of Returns: upper triangle shows p-values for IVX, lower triangle for the combination/IV tests for all possible subsamples starting (ending) in January (December). Subsamples starting and ending between 1942 and 1947 are excluded if tests includes treasury bills since they are constant in mid fourties. For further details see the text

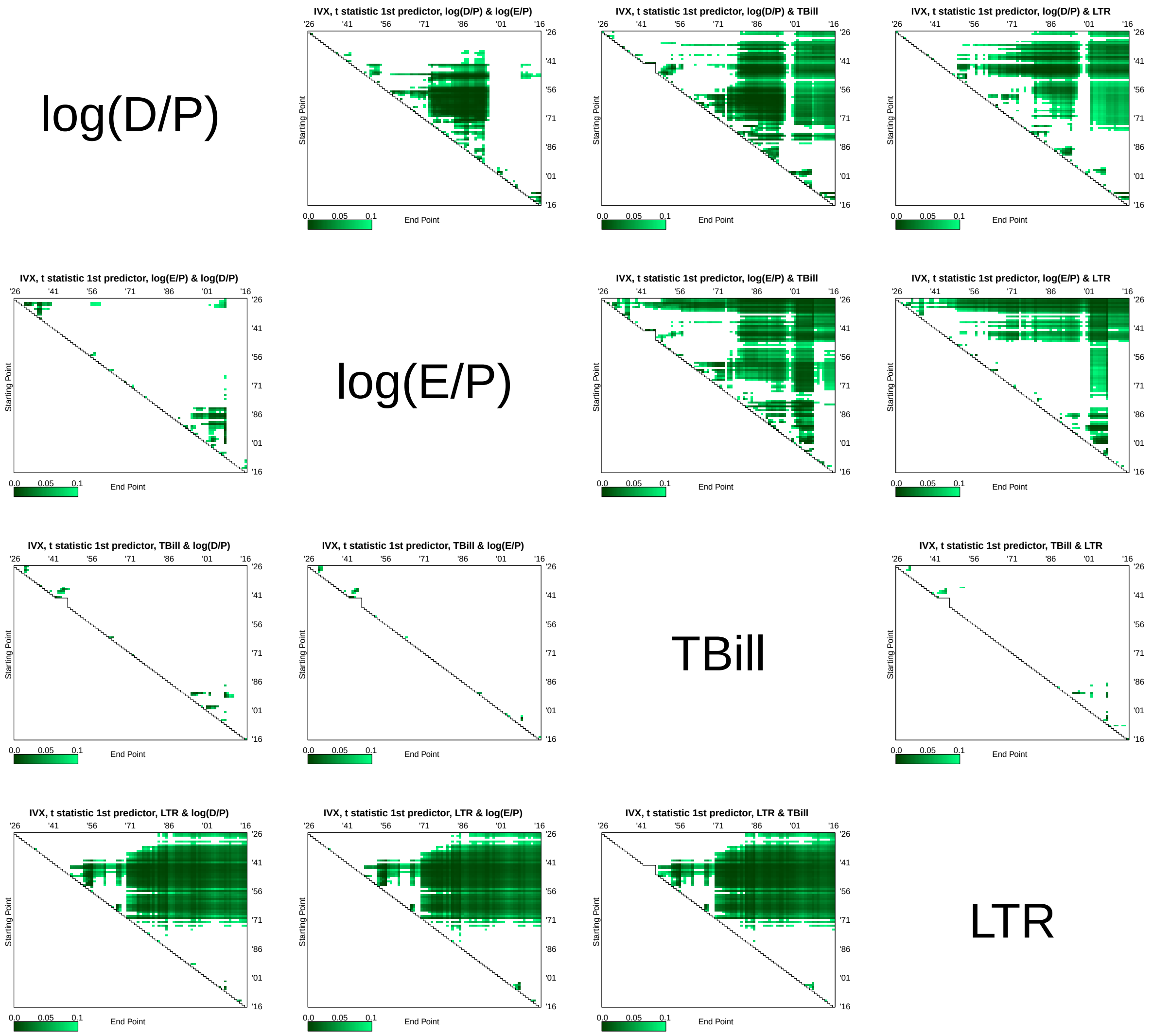


Figure 23: Subsample-wise tests of predictive power for bivariate combinations of log(D/P), log(E/P), Treasury Bill Rate and Long Term Rate of Returns: individual $t$-statistics of IVX for the first instrument which is defined by the corresponding row (individual $t$-statistics of second instruments can be achieved by transposing this figure). Subsamples starting and ending between 1942 and 1947 are excluded if tests includes treasury bills since they are constant in mid fourties

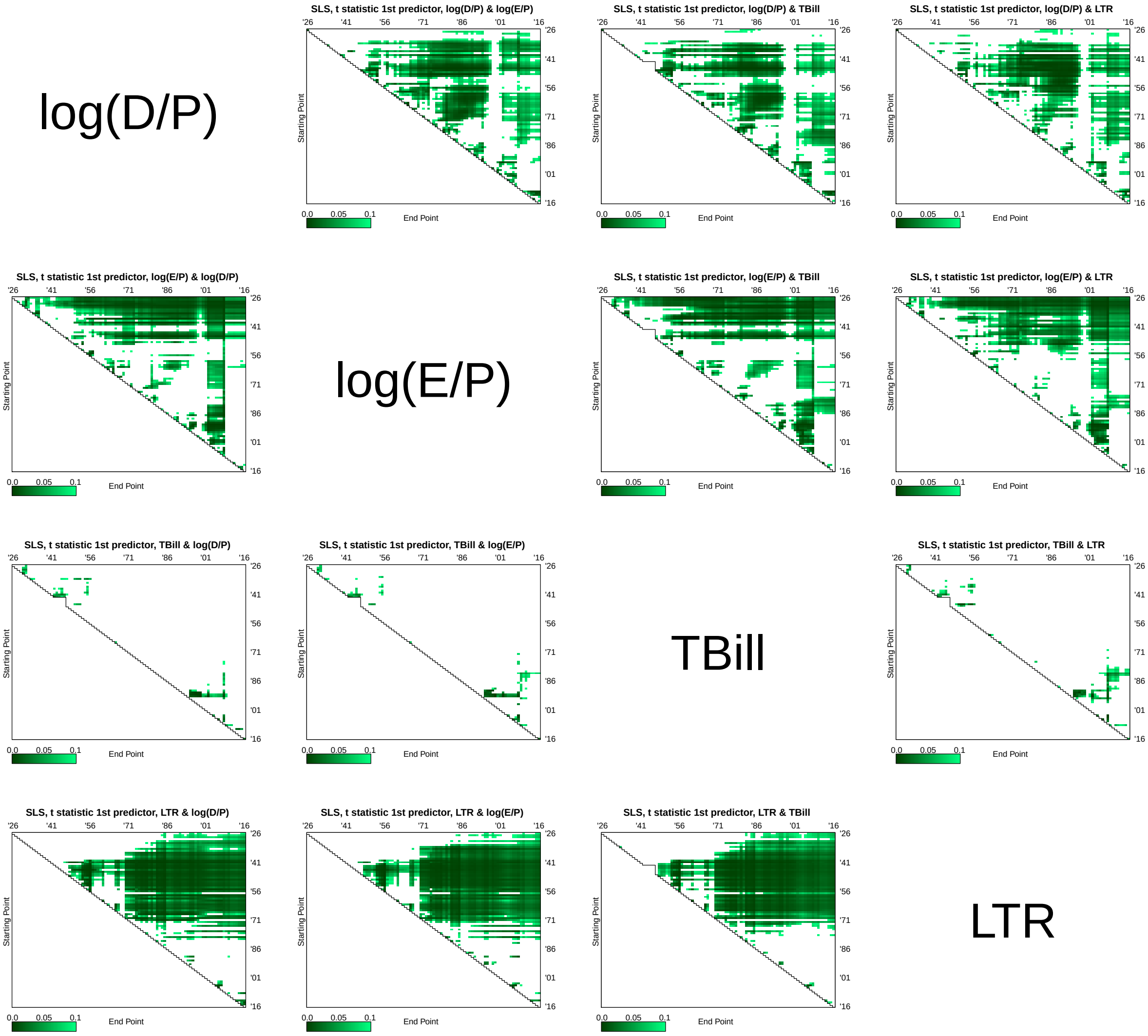


Figure 24: Subsample-wise tests of predictive power for bivariate combinations of log(D/P), log(E/P), Treasury Bill Rate and Long Term Rate of Returns: individual $t$-statistics of the combination/IV tests for the first instrument which is defined by the corresponding row (individual $t$-statistics of second instruments can be achieved by transposing this figure). Subsamples starting and ending between 1942 and 1947 are excluded if tests includes treasury bills since they are constant in mid fourties

# B Auxiliary results

Throughout, $C$ denotes a generic constant whose value may change from occurrence to occurrence. We also make use of $\sum_{j=0}^{t} \varrho^{kj} = \frac{1-\varrho^{k(t+1)}}{1-\varrho^k} \leq CT^{\eta}$ for all $t = 2, \ldots, T$ and $k = 1, 2$. The $L_p$ norm of a random variable is denoted by $\|\cdot\|_p = \sqrt[p]{\mathrm{E}\left(|\cdot|^p\right)}$.

**Lemma 3** *Under the Assumptions of Proposition 3, we have for near-integrated $x_t$ as $T \to \infty$ that*

1. *For $t = 2, \ldots, T$, $w_{t-1} = \lambda \bar{w}_{t-1} + r_{t-1}$ where $\bar{w}_{t-1} = \sum_{j=0}^{t-3} \varrho^j v_{t-1-j}$ for $t \geq 3$ and $0$ for $t = 2$, with $\sup_t \|\bar{w}_{t-1}\|_4 = O\left(T^{\eta/2}\right)$, and $\sup_t \|r_{t-1}\|_4 = o\left(T^{\eta/2}\right)$;*
2. $\frac{1}{\lambda^\alpha \bar{\omega}_v^\alpha T^{\alpha/2}} f\left(x_{[sT]-1}\right) \Rightarrow H_\alpha\left(J_{c,\sigma_v}(s)\right)$;
3. $\frac{1}{\lambda^{2\alpha} \bar{\omega}_v^{2\alpha} T^{\alpha+1}} \sum_{t=2}^{T} f^2\left(x_{[sT]-1}\right) \Rightarrow \int_0^1 H_\alpha^2\left(J_{c,\sigma_v}(s)\right) \mathrm{d}s$;
4. $\frac{1}{\lambda^\alpha \bar{\omega}_v^\alpha T^{\alpha/2+0.5}} \sum_{t=2}^{T} f\left(x_{t-1}\right) u_t \Rightarrow \int_0^1 H_\alpha\left(J_{c,\sigma_v}(s)\right) \mathrm{d}W_{\sigma_u}(s)$;
5. $\frac{1}{T^{1+\eta/2}} \sum_{t=2}^{T} w_{t-1} = o_p(1)$;
6. $\frac{1}{T^{3/2+\eta/2}} \sum_{t=2}^{T} x_{t-1} w_{t-1} = o_p(1)$;
7. $\frac{1}{T^{1+\eta/2}} \sum_{t=2}^{T} \left(\sin\frac{\pi t}{2T} - \overline{\sin}\right) w_{t-1} = o_p(1)$ *where* $\overline{\sin} = \frac{1}{T^1} \sum_{t=2}^{T} \sin\frac{\pi t}{2T}$;
8. $\frac{1}{T^{1+\eta/2}} \sum_{t=2}^{T} \left(\sin\frac{\pi t}{2T} - \overline{\sin}\right) w_{t-1} \tilde{y}_t^2 = O_p(1)$;
9. $\frac{1}{T^{\alpha/2+1+\eta/2}} \sum_{t=2}^{T} w_{t-1} f\left(x_{t-1}\right) = O_p(1)$;
10. $\frac{1}{T^{1+\eta}} \sum_{t=2}^{T} w_{t-1}^2 = \Theta_p(1)$ *with $\Theta_p$ denoting an exact order of magnitude;*
11. $\frac{1}{T^{1+\eta}} \sum_{t=2}^{T} w_{t-1}^2 \tilde{y}_t^2 = O_p(1)$;
12. $\frac{1}{T} \sum_{t=2}^{T} \left(\sin\frac{\pi t}{2T} - \overline{\sin}\right)^2 \tilde{y}_t^2 \Rightarrow \int_0^1 \left(\sin\frac{\pi s}{2} - \frac{2}{\pi}\right)^2 \sigma_u^2(s) \mathrm{d}s$.

**Proof:** *see Appendix C.*

**Lemma 4** *Under the Assumptions of Proposition 3, we have as $T \to \infty$ for $|\rho| < 1$ fixed that*

1. $w_{t-1} = x_{t-1} - \mu - \varrho^{t-3}\left(x_1 - \mu\right) + r_{t-1}$ *for $t = 2, \ldots, T$, where $T^{\eta/2} r_{t-1}$ is uniformly $L_4$-bounded,* $\left\|T^{\eta/2} r_{t-1}\right\|_4 < C\ \forall t$;
2. $\frac{1}{T} \sum_{t=2}^{T} \tilde{x}_{t-1} \sin\frac{\pi t}{2T} = o_p(1)$;
3. $\frac{1}{T} \sum_{t=2}^{T} w_{t-1}\left(\sin\frac{\pi t}{2T} - \overline{\sin}\right) = o_p(1)$;
4. $\frac{1}{T} \sum_{t=2}^{T} w_{t-1}\left(\sin\frac{\pi t}{2T} - \overline{\sin}\right) \tilde{y}_t^2 = O_p(1)$;
5. $\frac{1}{T} \sum_{t=2}^{T} w_{t-1}^2 \xrightarrow{p} \mathcal{E}_0\left[i^2\right]$;
6. $\frac{1}{T} \sum_{t=2}^{T} w_{t-1}^2 \tilde{y}_t^2 \xrightarrow{p} \mathcal{E}_0^*\left[i^2\right]$;
7. $\frac{1}{\sqrt{T}} \sum_{t=2}^{T} w_{t-1} \tilde{u}_t \xrightarrow{d} \mathcal{N}\left(0, \mathcal{E}_0^*\left[i^2\right]\right)$;
8. $\frac{1}{T} \sum_{t=2}^{T} \left(\sin\frac{\pi t}{2T} - \overline{\sin}\right)^2 \tilde{y}_t^2 \Rightarrow \int_0^1 \left(\sin\frac{\pi s}{2} - \frac{2}{\pi}\right)^2 \sigma_u^2(s) \mathrm{d}s$.

**Proof:** *see Appendix C.*

# C Proofs

## Proof of Lemma 1

Note that, upon division by $\sqrt{T}$, the mean $\mu$ is asymptotically negligible. The result follows e.g. with the arguments of Cavaliere et al. (2010).

## Proof of Lemma 2

Split the sample in $B$ blocks of length $n_T$ and write

$$\frac{1}{T}\sum_{t=2}^{T} h\left(x_{t-1}\right) = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{n_T}\sum_{\tau=1}^{n_T} h\left(x_{(b-1)n_T+\tau}\right)$$

where $T-1 = B\cdot n_T$ for simplicity. Let $\beta_j$ denote the coefficients of the lag polynomial $(1-\rho L)^{-1}\left(\sum_{j\geq 0} b_j L^j\right)$ and note that they also fulfill the 1-summability condition fulfilled by $b_j$ since $(1-\rho L)^{-1}$ has exponentially decaying coefficients. Examining now the $B$ block sums individually, we have that

$$\frac{1}{n_T}\sum_{\tau=1}^{n_T} h\left(x_{(b-1)n_T+\tau}\right) = \frac{1}{n_T}\sum_{\tau=1}^{n_T} h\left(\mu + \sum_{j\geq 0}\beta_j \sigma_v\left(\frac{(b-1)\,n_T+\tau-j}{T}\right)\nu_{(b-1)n_T+\tau-j}\right).$$

Using Proposition 1 from Demetrescu and Sibbertsen (2014), note that, for all $b$ and $\tau$,

$$\left\|\sum_{j\geq 0}\beta_j\sigma_v\left(\frac{(b-1)\,n_T+\tau-j}{T}\right)\nu_{(b-1)n_T+\tau-j} - \sigma_v\left(\frac{bn_T}{T}\right)\sum_{j\geq 0}\beta_j\nu_{(b-1)n_T+\tau-j}\right\|_2 \leq \frac{C}{T}$$

with $\|\cdot\|_2$ the $L_2$ norm of a random variable. Let $q_{b,\tau} = \sum_{j\geq 0}\beta_j\sigma_v\left(\frac{(b-1)n_T+\tau-j}{T}\right)\nu_{(b-1)n_T+\tau-j}$ and $\tilde{q}_{b,\tau} = \sigma_v\left(\frac{bn_T}{T}\right)\sum_{j\geq 0}\beta_j\nu_{(b-1)n_T+\tau-j}$ and note that both quantities are uniformly $L_\delta$-bounded since $|\rho|<1$ is fixed and $\sigma_v$ is bounded.

We now examine the approximation error $h(\mu+q_{b,\tau}) - h(\mu+\tilde{q}_{b,\tau})$. To this end it suffices to focus on homogenous functions $h$, since the nonhomogenous part is Lipschitz and shall not affect the following derivations. We distinguish two cases, $\alpha>1$ and $1\geq\alpha>0$.

For $\alpha>1$, use the mean-value theorem to obtain

$$h(\mu+q_{b,\tau}) = h(\mu+\tilde{q}_{b,\tau}) + h'(\xi_{b,\tau})(q_{b,\tau}-\tilde{q}_{b,\tau}),$$

where $\xi_{b,\tau}$ lies between $\mu+\tilde{q}_{b,\tau}$ and $\mu+q_{b,\tau}$ so $|\xi_{b,\tau}| \leq |\mu| + \max\{|\tilde{q}_{b,\tau}|;|q_{b,\tau}|\}$. Since both $\tilde{q}_{b,\tau}$ and $q_{b,\tau}$ are uniformly $L_\delta$-bounded, $|\xi_{b,\tau}|$ must itself be $L_\delta$-bounded where, recall, $\delta > \max\{4, 4\alpha\}$.

The homogeneity of order $\alpha>1$ of $h$ implies $h'$ to be homogenous of order $\alpha-1>0$ such that $\frac{h'(x)}{x^{\alpha-1}} = O(1)$ as $x\to\pm\infty$. Hence, $h'(\xi_t)$ is uniformly $L_{\frac{\delta}{\alpha-1}}$-bounded where $\frac{\delta}{\alpha-1} > 2$, such that the Cauchy-Schwarz inequality applies, leading to

$$\|h(\mu+q_{b,\tau}) - h(\mu+\tilde{q}_{b,\tau})\|_1 \leq \frac{C}{T}.$$

Then, ergodicity of $\nu_t$ implies that

$$\frac{1}{n_T}\sum_{\tau=1}^{n_T} h\left(\mu+\sigma_v\left(\frac{bn_T}{T}\right)\tilde{x}_{(b-1)n_T+\tau}\right) \xrightarrow{p} \mathrm{E}\left(h\left(\mu+\sigma_v\left(\frac{bn_T}{T}\right)\tilde{x}_1\right)\right);$$

given uniform integrability of $h(\mu+\tilde{q}_t)$ this implies $L_1$ convergence. Moreover, strict stationarity and boundedness of $\sigma_v$ imply that convergence must take place at the same rate, so

$$\max_b \mathrm{E}\left(\frac{1}{n_T}\sum_{\tau=1}^{n_T} h\left(\mu+\sigma_v\left(\frac{bn_T}{T}\right)\tilde{x}_{(b-1)n_T+\tau}\right) - \mathrm{E}\left(h\left(\mu+\sigma_v\left(\frac{bn_T}{T}\right)\tilde{x}_1\right)\right)\right) \to 0.$$

Thus,

$$\begin{aligned}
&\left\| \frac{1}{B}\sum_{b=1}^{B}\frac{1}{n_T}\sum_{\tau=1}^{n_T} h\left(x_{(b-1)n_T+\tau}\right) - \frac{1}{B}\sum_{b=1}^{B}\mathrm{E}\left(h\left(\mu+\sigma_v\left(\frac{bn_T}{T}\right)\tilde{x}_1\right)\right)\right\|_1 \\
&\leq \frac{1}{B}\sum_{b=1}^{B}\frac{1}{n_T}\sum_{\tau=1}^{n_T}\left\| h\left(\mu+q_{b,\tau}\right) - h\left(\mu+\tilde{q}_{b,\tau}\right)\right\|_1 \\
&\quad + \frac{1}{B}\sum_{b=1}^{B}\left\|\frac{1}{n_T}\sum_{\tau=1}^{n_T} h\left(\mu+\sigma_v\left(\frac{bn_T}{T}\right)\tilde{x}_{(b-1)n_T+\tau}\right) - \mathrm{E}\left(h\left(\mu+\sigma_v\left(\frac{bn_T}{T}\right)\tilde{x}_1\right)\right)\right\|_1 \\
&\to 0.
\end{aligned}$$

For $1 \geq \alpha > 0$, note that $h(s) = s^{\alpha} h(1)$ for $s > 0$ and $h(s) = s^{\alpha} h(-1)$ for $s < 0$ so $h$ satisfies a uniform Hölder condition of order $\alpha$ such that

$$\left| h\left(\mu+q_{b,\tau}\right) - h\left(\mu+\tilde{q}_{b,\tau}\right)\right| \leq C\left|q_{b,\tau}-\tilde{q}_{b,\tau}\right|^{\alpha}.$$

With $\alpha \leq 1$, Jensen's inequality implies that

$$\mathrm{E}\left(\left|q_{b,\tau}-\tilde{q}_{b,\tau}\right|^{\alpha}\right) \leq \left(\mathrm{E}\left(\left|q_{b,\tau}-\tilde{q}_{b,\tau}\right|\right)\right)^{\alpha}$$

such that

$$\left\| h\left(\mu+q_{b,\tau}\right) - h\left(\mu+\tilde{q}_{b,\tau}\right)\right\|_1 \leq \frac{C}{T^{\alpha}};$$

this way,

$$\left\| \frac{1}{B}\sum_{b=1}^{B}\frac{1}{n_T}\sum_{\tau=1}^{n_T} h\left(x_{(b-1)n_T+\tau}\right) - \frac{1}{B}\sum_{b=1}^{B}\mathrm{E}\left(h\left(\mu+\sigma_v\left(\frac{bn_T}{T}\right)\tilde{x}_1\right)\right)\right\|_1 \to 0$$

for $1 \geq \alpha > 0$ as well.

Finally,

$$\frac{1}{B}\sum_{b=1}^{B}\mathrm{E}\left(h\left(\mu+\sigma_v\left(\frac{bn_T}{T}\right)\tilde{x}_1\right)\right) \to \mathcal{E}_{\mu}[h]$$

thanks to the integrability of $\sigma_v$, leading to the desired result. Analog convergence to $\mathcal{E}^*_{\mu}[h]$ follows along the same lines.

## Proof of Lemma 3

1. Use the Phillips-Solo decomposition to write $e_t = \lambda v_t + \Delta\bar{v}_t$ where $\bar{v}_t$ is a linear process in $v_t$ with absolutely summable coefficients. It holds that

$$\begin{aligned}
w_{t-1} &= \sum_{j=0}^{t-3}\varrho^j \Delta x_{t-1-j} = \sum_{j=0}^{t-3}\varrho^j\left(e_{t-1-j} - \frac{c}{T}x_{t-2-j}\right) \\
&= \lambda\bar{w}_{t-1} + \sum_{j=0}^{t-3}\varrho^j\left(\Delta\bar{v}_{t-1-j} - \frac{c}{T}x_{t-2-j}\right).
\end{aligned}$$

   Also, we have

$$\mathrm{E}\left(\bar{w}_{t-1}^4\right) = \sum_{j=0}^{t-3}\sum_{k=0}^{t-3}\sum_{l=0}^{t-3}\sum_{m=0}^{t-3}\varrho^j\varrho^k\varrho^l\varrho^m\,\mathrm{E}\left(v_{t-1-j}v_{t-1-k}v_{t-1-l}v_{t-1-m}\right)$$

which, upon exploiting the md property of $v_t$, gives

$$\begin{aligned}
\mathrm{E}\left(\bar{w}_{t-1}^4\right) &= \sum_{j=0}^{t-3} \varrho^{4j}\,\mathrm{E}\left(v_{t-1-j}^4\right) + 3\sum_{\substack{j=0\\ j\neq k}}^{t-3}\sum_{k=0}^{t-3} \varrho^{3j}\varrho^{k}\,\mathrm{E}(v_{t-j}^3 v_{t-k}) \\
&\quad + 3\sum_{\substack{j=0\\ j\neq k}}^{t-3}\sum_{k=0}^{t-3} \varrho^{2j}\varrho^{2k}\,\mathrm{E}\left(v_{t-1-j}^2 v_{t-1-k}^2\right) + 6\sum_{\substack{j=0\\ j\neq k\neq l}}^{t-3}\sum_{k=0}^{t-3}\sum_{l=0}^{t-3} \varrho^{2j}\varrho^{k}\varrho^{l}\,\mathrm{E}\left(v_{t-1-j}^2 v_{t-1-k} v_{t-1-l}\right).
\end{aligned}$$

Now, $\left|\mathrm{E}(v_{t-j}^3 v_{t-k})\right| \leq \mathrm{E}(\left|v_{t-j}^3 v_{t-k}\right|) \leq \left\|v_{t-j}^3\right\|_{4/3}\left\|v_{t-k}\right\|_4$ thanks to the Hölder inequality, where the norms are uniformly bounded, so

$$\left|\sum_{\substack{j=0\\ j\neq k}}^{t-3}\sum_{k=0}^{t-3} \varrho^{3j}\varrho^{k}\,\mathrm{E}(v_{t-j}^3 v_{t-k})\right| \leq C\sum_{\substack{j=0\\ j\neq k}}^{t-3}\sum_{k=0}^{t-3} \varrho^{3j}\varrho^{k} \leq C\left(\sum_{j=0}^{t-3}\varrho^{3j}\right)\left(\sum_{k=0}^{t-3}\varrho^{k}\right) \leq CT^{2\eta},$$

and a similar argument shows that $3\sum_{\substack{j=0\\ j\neq k}}^{t-3}\sum_{k=0}^{t-3} \varrho^{2j}\varrho^{2k}\,\mathrm{E}\left(v_{t-1-j}^2 v_{t-1-k}^2\right) = O\left(T^{2\eta}\right)$ as well. Furthermore, since $\mathrm{E}\left(v_{t-1-k}v_{t-1-l}\right) = 0$ and $\mathrm{E}\left(v_{t-1-j}^2 v_{t-1-k}v_{t-1-l}\right) = 0$ for $j \neq k \neq l$ whenever $j \geq l$ or $j \geq k$,

$$\begin{aligned}
\left|\mathrm{E}\left(v_{t-1-j}^2 v_{t-1-k}v_{t-1-l}\right)\right| &\leq \mathrm{Var}\left(v_{t-1-j}^2\right)\left|\mathrm{E}\left(v_{t-1-k}v_{t-1-l}\right)\right| \\
&\quad + \left|\mathrm{E}\left(\left(v_{t-1-j}^2 - \mathrm{Var}\left(v_{t-1-j}^2\right)\right) v_{t-1-k}v_{t-1-l}\right)\right| \\
&\leq \frac{C}{\sqrt{(k-j)^{1+\psi}\,(l-j)^{1+\psi}}}
\end{aligned}$$

for all $t$, and thus, with $\sigma_v$ uniformly bounded,

$$\begin{aligned}
&\left|\sum_{\substack{j=0\\ j\neq k\neq l}}^{t-3}\sum_{k=0}^{t-3}\sum_{l=0}^{t-3} \varrho^{2j}\varrho^{k}\varrho^{l}\,\mathrm{E}\left(v_{t-1-j}^2 v_{t-1-k}v_{t-1-l}\right)\right| \leq \\
&\leq C\sum_{j=0}^{t-3}\varrho^{4j}\sum_{\substack{k=j+1\\ k\neq l}}^{t-3}\sum_{l=j+1}^{t-3}\frac{\varrho^{k-j}\varrho^{l-j}}{\sqrt{(k-j)\,(l-j)}} \leq C\sum_{j=0}^{t-3}\varrho^{4j}\sum_{\substack{k=j+1\\ k\neq l}}^{t-3}\sum_{l=j+1}^{t-3}\frac{\varrho^{k-j}\varrho^{l-j}}{\sqrt{(k-j)^{1+\psi}\,(l-j)^{1+\psi}}} \\
&\leq C\sum_{j=0}^{t-3}\varrho^{4j}\sum_{k=j+1}^{t-3}\sum_{l=j+1}^{t-3}\frac{\varrho^{k-j}\varrho^{l-j}}{\sqrt{(k-j)\,(l-j)}} \leq C\sum_{j=0}^{t-3}\varrho^{4j}\left(\sqrt{\sum_{k=1}^{t-j-3}\varrho^{2k}}\sqrt{\sum_{k=1}^{t-j-3}\frac{1}{k^{1+\psi}}}\right)^2, \\
&\leq C\left(\sum_{j=0}^{T-1}\varrho^{4j}\right)\left(\sum_{k=0}^{T-1}\varrho^{2k}\right)\left(\sum_{k=1}^{T-1}\frac{1}{k^{1+\psi}}\right)
\end{aligned}$$

where the positivity of the summands was used in the last step. Summing up, we have for all $t = 2, \ldots, T$ that

$$\mathrm{E}\left(\bar{w}_{t-1}^4\right) \leq CT^{2\eta}$$

as required. Examining now $r_{t-1} = \sum_{j=0}^{t-3}\varrho^j\Delta\bar{v}_{t-1-j} - \frac{c}{T}\sum_{j=0}^{t-3}\varrho^j x_{t-2-j}$, we have first that

$$\sum_{j=0}^{t-3}\varrho^j\Delta\bar{v}_{t-1-j} = \bar{v}_{t-1} - \varrho^{t-3}\bar{v}_1 - (1-\varrho)\sum_{j=0}^{t-4}\varrho^j\bar{v}_{t-2-j}$$

where $\bar{v}_t$ is uniformly $L_4$-bounded since its coefficients are absolutely summable and its shocks are (at least) uniformly $L_4$-bounded, which leads to

$$\sup_{t=2,\ldots,T}\left\|\bar{v}_{t-1}\right\|_4 = C < CT^{\eta/2} \qquad \text{and} \qquad \sup_{t=2,\ldots,T}\left\|\varrho^{t-3}\bar{v}_1\right\|_4 = C\varrho^{t-3} < CT^{\eta/2};$$

furthermore,

$$\left\| \sum_{j=0}^{t-4} \varrho^j \bar{v}_{t-2-j} \right\|_4 \leq \sum_{j=0}^{t-4} \varrho^j \left\| \bar{v}_{t-2-j} \right\|_4 \leq CT^{\eta}$$

hence $(1-\varrho)\sum_{j=0}^{t-4} \varrho^j \bar{v}_{t-2-j}$ is uniformly $L_4$ bounded as required. To complete the result, note that

$$\left\| \frac{c}{T} \sum_{j=0}^{t-3} \varrho^j x_{t-2-j} \right\|_4 \leq C \frac{1}{T} \sum_{j=0}^{t-3} \varrho^j \left\| x_{t-2-j} \right\|_4 \leq CT^{\eta - 1/2}$$

since $T^{-1/2} x_t$ is uniformly $L_4$ bounded (which can be shown along the lines of showing that $T^{-\eta/2} \bar{w}_{t-1}$ is uniformly $L_4$ bounded). The bounds on the $L_4$ norms of the summands imply also that $T^{-\eta/2} w_{t-1}$ is itself uniformly $L_4$ bounded.

2. Follows with the continuous mapping theorem [CMT] after noting that the integrable component of $f$ is a bounded function (as it is Lipschitz with vanishing tails) so $I(x_{t-1})$ vanishes upon division by $T^{\alpha}$ uniformly in $t$.
3. Follows from item 2 with the CMT.
4. Follows by the arguments of Kurtz and Protter (1991) from joint convergence in item 2 and Lemma 1, since $u_t$ has the martingale difference property.
5. Using item 1 of the lemma, we obtain immediately with the Markov inequality that

$$\frac{1}{T^{1+\eta/2}} \sum_{t=2}^{T} w_{t-1} = \frac{\lambda}{T^{1+\eta/2}} \sum_{t=2}^{T} \bar{w}_{t-1} + \frac{1}{T} \sum_{t=2}^{T} \frac{r_{t-1}}{T^{\eta/2}} = \frac{\lambda}{T^{1+\eta/2}} \sum_{t=2}^{T} \bar{w}_{t-1} + o_p(1);$$

to arrive at the desired result, note that, after re-arranging the summands of $\sum_{t=2}^{T} \bar{w}_{t-1}$,

$$\frac{1}{T^{1+\eta/2}} \sum_{t=2}^{T} \bar{w}_{t-1} = \frac{1}{T^{1+\eta/2}} \sum_{t=2}^{T-1} \frac{1-\varrho^{T-t}}{1-\varrho} v_t = T^{\eta/2 - 1/2} \frac{1}{\sqrt{T}} \sum_{t=2}^{T-1} \left(1-\varrho^{T-t}\right) v_t$$

where $T^{\eta/2-1/2} \to 0$ and $v_t$ is a md sequence and $0 < \left|1-\varrho^{T-t}\right| < 1$, so $\frac{1}{\sqrt{T}} \sum_{t=2}^{T-1} \left(1-\varrho^{T-t}\right) v_t = O_p(1)$.

6. Let $S_{t-1} = \sum_{j=1}^{t-1} w_j$ and, with item 1, write

$$\left\| S_{t-1} \right\|_4 = |\lambda| \left\| \sum_{j=1}^{t-1} \bar{w}_j \right\|_4 + \sum_{j=1}^{t-1} \left\| r_j \right\|_4,$$

where $\sum_{j=1}^{t-1} \|r_j\|_4 \leq C(t-1) \sup_{2\leq t \leq T} \|r_{t-1}\|_4 = o\left(T^{1+\eta/2}\right)$ and

$$\left\| \sum_{j=1}^{t-1} \bar{w}_j \right\|_4 = \left\| \sum_{j=1}^{t-2} \frac{1-\varrho^{t-j-1}}{1-\varrho} v_{j+1} \right\|_4 = T^{\eta} \left\| \sum_{j=1}^{t-2} \left(1-\varrho^{t-j-1}\right) v_{j+1} \right\|_4.$$

Now, $0 < \left|1-\varrho^{t-j-1}\right| < 1$ so the arguments in the proof of item 1 establish uniform boundedness of $\left\| \frac{1}{\sqrt{T}} \sum_{j=1}^{t-2} \left(1-\varrho^{t-j-1}\right) v_{j+1} \right\|_4$, such that

$$\sup_{2\leq t\leq T} \left\| S_{t-1} \right\|_4 = O\left(T^{\eta+1/2}\right) = o\left(T^{1+\eta/2}\right).$$

Then, recall that $\sup_{2\leq t\leq T} \|x_{t-1}\|_4 = O\left(\sqrt{T}\right)$ and the result follows with

$$\frac{1}{T^{3/2+\eta/2}} \sum_{t=2}^{T} x_{t-1} w_{t-1} = \frac{1}{T^{3/2+\eta/2}} S_{T-1} x_{T-1} - \frac{1}{T^{3/2+\eta/2}} \sum_{t=2}^{T-1} S_{t-1} v_t + \frac{c}{T^{5/2+\eta/2}} \sum_{t=2}^{T-1} S_{t-1} x_{t-1}$$

and the md property of $v_t$ (for the evaluation of the 2nd term on the r.h.s.).

7. Write
$$\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\left(\sin\frac{\pi t}{2T}-\overline{\sin}\right)w_{t-1}=\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}w_{t-1}-\overline{\sin}\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}w_{t-1}$$
where the second term on the r.h.s. is $o_p(1)$ thanks to item 5 of this lemma; for the first term we have that
$$\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}w\sin\frac{\pi t}{2T}w_{t-1}=\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}\bar{w}_{t-1}+\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}r_{t-1}.$$
Since $\bar{w}_{t-1}=\sum_{j=0}^{t-3}\varrho^j v_{t-1-j}$, we have that $0<\mathrm{E}(\bar{w}_{t-1}\bar{w}_{s-1})$ and consequently
$$\begin{aligned}\mathrm{Var}\left(\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}\bar{w}_{t-1}\right) &= \frac{1}{T^{2+\eta}}\sum_{t=2}^{T}\sum_{s=2}^{T}\sin\frac{\pi t}{2T}\sin\frac{\pi s}{2T}\mathrm{E}(\bar{w}_{t-1}\bar{w}_{s-1})\\ &\leq \frac{1}{T^{2+\eta}}\sum_{t=2}^{T}\sum_{s=2}^{T}\mathrm{E}(\bar{w}_{t-1}\bar{w}_{s-1}).\end{aligned}$$
Then, since $\bar{w}_{t-1}$ is an AR(1) process with uniformly $L_2$bounded md innovations, it follows that
$$\mathrm{E}(\bar{w}_{t-1}\bar{w}_{s-1})\leq C\varrho^{|s-t|}\sum_{j=0}^{\min\{s,t\}-3}\varrho^{2j}\leq CT^{\eta}\varrho^{|s-t|},$$
and, consequently,
$$\begin{aligned}\mathrm{Var}\left(\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}\bar{w}_{t-1}\right) &\leq \frac{C(T-1)}{T^2}+2C\frac{1}{T}\sum_{h=1}^{T-2}\frac{T-h}{T}\varrho^h\leq\frac{C}{T}+2C\frac{1}{T}\sum_{h=1}^{T-2}\varrho^h\\ &\leq CT^{\eta-1}.\end{aligned}$$
The result follows with Markov's inequality if $\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}r_{t-1}=o_p(1)$. But this is indeed the case since
$$\mathrm{E}\left(\left|\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}r_{t-1}\right|\right)\leq\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\mathrm{E}(|r_{t-1}|)\leq CT^{-\eta/2}\sup_{t=1,\ldots,T}\|r_{t-1}\|_2\to 0$$
thanks to item 1 of this lemma.

8. Note first that $\max_{t=1,\ldots,T}|x_{t-1}|=O_p\left(T^{1/2}\right)$ thanks to weak converge of $T^{-1/2}x_{[sT]}$ to a process bounded in probability on $[0,1]$; with $\bar{u}=\frac{1}{T}\sum_{t=2}^{T}u_t=O_p\left(T^{-1/2}\right)$, $\overline{f(x)}=\frac{1}{T}\sum_{t=2}^{T}f(x_{t-1})=T^{\alpha/2}$ (cf. item 3 of this Lemma) and $\beta_1=O\left(T^{-(\alpha+1)/2}\right)$, it follows that
$$\max_t|\tilde{y}_t-u_t|=O_p\left(T^{-1/2}\right).$$
Then,
$$\begin{aligned}\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}w_{t-1}\tilde{y}_t^2 &= \frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}w_{t-1}u_t^2+\frac{2}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}w_{t-1}u_t(\tilde{y}_t-u_t)\\ &\quad+\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}w_{t-1}(\tilde{y}_t-u_t)^2\end{aligned}$$
where
$$\mathrm{E}\left(\left|\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}w_{t-1}u_t^2\right|\right)\leq\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\mathrm{E}\left(|w_{t-1}u_t^2|\right)\leq\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\|w_{t-1}\|_2\left\|u_t^2\right\|_2=O(1)$$
thanks to the Cauchy-Schwarz inequality (recall, both $u_t$ and $T^{-\eta/2}w_{t-1}$ are uniformly $L_4$ bounded). Moreover, $\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}|w_{t-1}||u_t|=O_p(1)$ thanks to Markov's and the cauchy-Schwarz inequalities

again, so
$$\left|\frac{2}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}w_{t-1}u_t\left(\tilde{y}_t-u_t\right)\right|\leq C\frac{\max_t\left|\tilde{y}_t-u_t\right|}{T^{1+\eta/2}}\sum_{t=2}^{T}\left|w_{t-1}\right|\left|u_t\right|\overset{p}{\to}0,$$
just like $\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\sin\frac{\pi t}{2T}w_{t-1}\left(\tilde{y}_t-u_t\right)^2$, and the result follows.

9. We have that
$$\left|\frac{1}{T^{\alpha/2+1+\eta/2}}\sum_{t=2}^{T}w_{t-1}f\left(x_{t-1}\right)\right|\leq\sup_{t=2,\dots,T}\left|f\left(\frac{x_{t-1}}{\sqrt{T}}\right)\right|\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\left|w_{t-1}\right|.$$
Then, with $T^{-1/2}x_{t-1}$ uniformly bounded in probability thanks to the weak convergence in Lemma 1, we have $\sup_t\left|f\left(\frac{x_{t-1}}{\sqrt{T}}\right)\right|=O_p(1)$ thanks to the continuity of $f$. Then, $T^{-\eta/2}w_{t-1}$ is uniformly $L_2$-bounded, such that
$$\mathrm{E}\left(\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T}\left|w_{t-1}\right|\right)=O(1);$$
the result then follows with Markov's inequality.

10. Write
$$\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}w_{t-1}^2=\frac{\lambda^2}{T^{1+\eta}}\sum_{t=2}^{T}\bar{w}_{t-1}^2+\frac{2\lambda}{T^{1+\eta}}\sum_{t=2}^{T}\bar{w}_{t-1}r_{t-1}+\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}r_{t-1}^2$$
To establish the result, it suffices to show that $0<C_1<\mathrm{E}\left(\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\bar{w}_{t-1}^2\right)<C_2$ for suitable constants $C_{1,2}$ and that $\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}r_{t-1}^2=O_p(1)$. We have indeed
$$\mathrm{E}\left(\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\bar{w}_{t-1}^2\right)=\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\mathrm{E}\left(\bar{w}_{t-1}^2\right)<C_2$$
since $T^{-\eta/2}\bar{w}_{t-1}$ is uniformly $L_2$-bounded. It is also easily shown that that $\mathrm{E}\left(T^{-\eta}\bar{w}_{t-1}^2\right)$ is bounded away from zero. To complete the result, recall that $T^{-\eta/2}r_{t-1}$ is uniformly $L_2$-bounded and apply Markov's inequality.

11. Following the arguments of item 8, we first have that
$$\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}w_{t-1}^2\tilde{y}_t^2=\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}w_{t-1}^2u_t^2+o_p(1).$$
Then,
$$\mathrm{E}\left(\left|\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}w_{t-1}^2u_t^2\right|\right)\leq\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\mathrm{E}\left(\left|w_{t-1}^2u_t^2\right|\right)$$
and, exploiting the uniform $L_4$ boundedness of $T^{-\eta/2}w_{t-1}$ and $u_t$, the result follows with the Cauchy-Schwarz and Markov's inequalities.

12. Analogously to the proof of item 11,
$$\frac{1}{T}\sum_{t=2}^{T}\left(\sin\frac{\pi t}{2T}-\overline{\sin}\right)^2\tilde{y}_t^2=\frac{1}{T}\sum_{t=2}^{T}\left(\sin\frac{\pi t}{2T}-\overline{\sin}\right)^2u_t^2+o_p(1),$$
and the result follows with arguments analog to the ones in the proof of Lemma 2.

## Proof of Lemma 4

1. Begin by noting that
$$w_{t-1}=\left(x_{t-1}-\mu\right)-\varrho^{t-3}\left(x_1-\mu\right)+\left(\varrho-1\right)\sum_{j=0}^{t-4}\varrho^j\left(x_{t-2-j}-\mu\right),$$

where

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

with $d_{t-2}$ zero-mean mildly integrated. With arguments as in the proof of item 1 Lemma 3, it is straightforward to show that $T^{-\eta/2}d_{t-2}$ is uniformly $L_4$-bounded.

2. Write

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}\tilde{x}_{t-1}\sin\frac{\pi t}{2T} &= \frac{1}{T}\sum_{t=2}^{T}\left((x_{t-1}-\mu)-(\overline{x}-\mu)\right)\sin\frac{\pi t}{2T}\\ &= \frac{1}{T}\sum_{t=2}^{T}(x_{t-1}-\mu)\sin\frac{\pi t}{2T}-\frac{\overline{x}-\mu}{T}\sum_{t=2}^{T}\sin\frac{\pi t}{2T},\end{aligned}$$

where the second term on the r.h.s is $o_p(1)$. The expected value of the first term is zero and for its variance it holds

$$\operatorname{Var}\left(\frac{1}{T}\sum_{t=2}^{T}(x_{t-1}-\mu)\sin\frac{\pi t}{2T}\right) \leq \frac{1}{T^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\left|\mathrm{E}\left((x_t-\mu)(x_s-\mu)\right)\right|.$$

We can rewrite $(x_t-\mu)=B(L)\left(1-\varrho L\right)^{-1}v_t=B(\tilde{L})v_t$ with $\tilde{b}_j$ 1-summable. It follows that

$$\begin{aligned}&\frac{1}{T^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\left|\mathrm{E}\left(\left(\sum_{j\geq 0}\tilde{b}_j v_{t-j}\right)\left(\sum_{k\geq 0}\tilde{b}_k v_{s-k}\right)\right)\right|=\\ &=\frac{1}{T^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\sum_{j\geq 0}\sum_{k\geq 0}\left|\tilde{b}_j\tilde{b}_k\sigma_{v,t-j}\sigma_{v,s-k}\right|\left|\mathrm{E}\left(\nu_{t-j}\nu_{s-k}\right)\right|\\ &\leq \max_t\left(\sigma_{vt}^2\right)\frac{C}{T^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\sum_{j\geq 0}\sum_{k\geq 0}\left|\tilde{b}_j\tilde{b}_k\right|=\max_t\left(\sigma_{vt}^2\right)\frac{C}{T^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\left(\sum_{j\geq 0}\left|\tilde{b}_j\right|\right)^2.\end{aligned}$$

Exploiting the absolute summability of $\tilde{b}_j$, $\operatorname{Var}\left(\frac{1}{T}\sum_{t=2}^{T}(x_{t-1}-\mu)\sin\frac{\pi t}{2T}\right)$ is thus seen to be bounded as required.

3. Rewriting the original term yields

$$\frac{1}{T}\sum_{t=2}^{T}w_{t-1}\left(\sin\frac{\pi t}{2T}-\overline{\sin}\right)=\frac{1}{T}\sum_{t=2}^{T}w_{t-1}\sin\frac{\pi t}{2T}-\frac{\overline{\sin}}{T}\sum_{t=2}^{T}w_{t-1},$$

where the second term on the r.h.s expands to

$$\frac{\overline{\sin}}{T}\sum_{t=2}^{T}w_{t-1}=\frac{\overline{\sin}}{T}\sum_{t=2}^{T}(x_{t-1}-\mu)-\frac{x_1-\mu}{T}\overline{\sin}\sum_{t=2}^{T}\varrho^{t-3}+\frac{\overline{\sin}}{T^{1+\eta/2}}\sum_{t=2}^{T}r_{t-1}T^{\eta/2}.$$

It is easy to see that the first two terms on the r.h.s are $o_p(1)$. The third term is $o_p(1)$ as well since $r_{t-1}T^{\eta/2}$ is uniformly $L_4$-bounded as shown in item 1 of this Lemma. Regarding the remaining term

$$\frac{1}{T}\sum_{t=2}^{T}w_{t-1}\sin\frac{\pi t}{2T}=\frac{1}{T}\sum_{t=2}^{T}(x_{t-1}-\mu)\sin\frac{\pi t}{2T}-\frac{x_1-\mu}{T}\sum_{t=2}^{T}\varrho^{t-3}\sin\frac{\pi t}{2T}+\frac{1}{T}\sum_{t=2}^{T}r_{t-1}\sin\frac{\pi t}{2T},$$

it is again easy to see that the second part on the r.h.s is $o_p(1)$ and the first part is also $o_p(1)$ as previously shown in item 2 of this Lemma. Finally, with the uniformly $L_1$-boundedness of $r_{t-1}T^{\eta/2}$ follows

$$\mathrm{E}\left(\left|\frac{1}{T}\sum_{t=2}^{T}r_{t-1}\sin\frac{\pi t}{2T}\right|\right)\leq\frac{1}{T^{1+\eta/2}}\,\mathrm{E}\left(\sum_{t=2}^{T}\left|T^{\eta/2}r_{t-1}\right|\right)\leq\frac{C}{T^{\eta/2}}\to 0.$$

The result follows since all single terms are $o_p(1)$.

4. We have

$$\begin{aligned}
\mathrm{E}\left(\left|\frac{1}{T}\sum_{t=2}^{T} w_{t-1}\left(\sin\frac{\pi t}{2T}-\overline{\sin}\right)\tilde{y}_t^2\right|\right) &\leq \frac{1}{T}\sum_{t=2}^{T}\mathrm{E}\left(\left|w_{t-1}\right|\left|\sin\frac{\pi t}{2T}-\overline{\sin}\right|\left|\tilde{y}_t^2\right|\right)\\
&\leq \frac{1}{T}\sum_{t=2}^{T}\sqrt{\mathrm{E}\left(w_{t-1}^2\right)\mathrm{E}\left(\tilde{y}_t^4\right)} = \frac{1}{T}\sum_{t=2}^{T}\left\|w_{t-1}\right\|_2\left\|\tilde{y}_t\right\|_4^2.
\end{aligned}$$

Furthermore it holds that $w_{t-1} = (x_{t-1}-\mu) - \varrho^{t-3}(x_1-\mu) + r_{t-1}$ is uniformly $L_2$-bounded since $x_t$ as well as $T^{\eta/2} r_{t-1}$ are (at least) uniformly $L_2$-bounded. Moreover,

$$\begin{aligned}
\left\|\tilde{y}_t\right\|_4 &\leq \left|\beta_1\right|\left\|f\left(x_{t-1}\right)\right\|_4 + \left\|u_t\right\|_4 + \left|\beta_1\right|\left\|\overline{f\left(x\right)}\right\|_4 + \left\|\overline{u}\right\|_4,\\
\left\|f(x_{t-1})\right\|_4 &\leq \left|H_\alpha(1)\right|\left\|x_{t-1}\right\|_{4\alpha}^{\alpha} + \left\|I(x_{t-1})\right\|_4
\end{aligned}$$

With $H_\alpha(1)$ bounded, $I\left(x_{t-1}\right)$ a bounded function and $x_{t-1}$ uniformly $L_{4\alpha}$-bounded, $f(x_{t-1})$ turns out to be uniformly $L_4$-bounded. By applying Minkowski's inequality it follows that $\overline{f\left(x\right)}$ is also uniformly $L_4$-bounded and $\overline{u}$ is uniformly $L_4$-bounded since $u_t$ is. Hence $\tilde{y}_t$ is uniformly $L_4$-bounded and the result follows with Markov's inequality.

5. It holds

$$\begin{aligned}
\frac{1}{T}\sum_{t=2}^{T} w_{t-1}^2 &= \frac{1}{T}\sum_{t=2}^{T}\left(x_{t-1}-\mu\right)^2 + \frac{\left(x_1-\mu\right)^2}{T}\sum_{t=2}^{T}\left(\varrho^{t-3}\right)^2 + \frac{1}{T}\sum_{t=2}^{T} r_{t-1}^2\\
&- \frac{x_1-\mu}{T}\sum_{t=2}^{T}\left(x_{t-1}-\mu\right)\varrho^{t-3} + \frac{1}{T}\sum_{t=2}^{T}\left(x_{t-1}-\mu\right) r_{t-1} - \frac{x_1-\mu}{T}\sum_{t=2}^{T}\varrho^{t-3} r_{t-1}.
\end{aligned}$$

With Lemma 2 follows $\frac{1}{T}\sum_{t=2}^{T}\left(x_{t-1}-\mu\right)^2 \stackrel{p}{\to} \mathcal{E}_0\left[i^2\right]$. Each of the remaining terms is $o_p\left(1\right)$ thanks to the Markov inequality, since

$$\begin{aligned}
\mathrm{E}\left(\left|\frac{\left(x_1-\mu\right)^2}{T}\sum_{t=2}^{T}\left(\varrho^{t-3}\right)^2\right|\right) &\leq \frac{\mathrm{E}\left(\left(x_1-\mu\right)^2\right)}{T} C T^{\eta} = \mathrm{Var}\left(x_1\right)\frac{C}{T^{1-\eta}} \to 0\\
\mathrm{E}\left(\left|\frac{1}{T}\sum_{t=2}^{T} r_{t-1}^2\right|\right) &= \frac{1}{T^{\eta}}\frac{1}{T}\sum_{t=2}^{T}\left|T^{\eta} r_{t-1}^2\right| \leq \frac{C}{T^{\eta}} \to 0
\end{aligned}$$

with the crossproducts easily dealt with using the Cauchy-Schwarz inequality.

6. Note first that $\max_t \left\|\tilde{y}_t^2 - u_t^2\right\|_2 \to 0$ since

$$\tilde{y}_t^2 - u_t^2 = 2u_t\left(\beta_1 f\left(x_{t-1}\right) - \left(\overline{u} + \beta_1\overline{f\left(x\right)}\right)\right) + \left(\beta_1 f\left(x_{t-1}\right) - \left(\overline{u} + \beta_1\overline{f\left(x\right)}\right)\right)^2,$$

where, with $x_{t-1}$ uniformly $L_{4\alpha}$-bounded and thus $f\left(x_{t-1}\right)$ uniformly $L_4$-bounded,

$$\left\|u_t\beta_1 f\left(x_{t-1}\right)\right\|_2 \leq \left|\beta_1\right|\left\|u_t\right\|_4\left\|f\left(x_{t-1}\right)\right\|_4 \leq \frac{C}{\sqrt{T}};$$

furthermore, $\overline{f\left(x\right)}$ is itself uniformly $L_4$-bounded and $\left\|\overline{u}\right\|_4 \to 0$ (using the arguments from the proof of item 1, Lemma 3, it can be shown that $T^{-1/2}\sum_{t=1}^{T} u_t$ is uniformly $L_4$-bounded) so, uniformly in $t$,

$$\left\|u_t\left(\overline{u} + \beta_1\overline{f\left(x\right)}\right)\right\|_2 \leq \left\|u_t\right\|_4\left\|\overline{u}\right\|_4 + \left|\beta_1\right|\left\|u_t\right\|_4\left\|\overline{f\left(x\right)}\right\|_4 = o\left(1\right),$$

and, similarly,

$$\begin{aligned}
\left\|\left(\beta_1 f\left(x_{t-1}\right) - \left(\overline{u} + \beta_1\overline{f\left(x\right)}\right)\right)^2\right\|_2 &= \left\|\beta_1 f\left(x_{t-1}\right) - \left(\overline{u} + \beta_1\overline{f\left(x\right)}\right)\right\|_4^2\\
&\leq \left\|\beta_1 f\left(x_{t-1}\right)\right\|_4^2 + 2\left\|\beta_1 f\left(x_{t-1}\right)\right\|_4\left\|\overline{u} + \beta_1\overline{f\left(x\right)}\right\|_4 + \left\|\overline{u} + \beta_1\overline{f\left(x\right)}\right\|_4^2 = o\left(1\right)
\end{aligned}$$

uniformly in $t$. Then,

$$\frac{1}{T}\sum_{t=2}^{T} w_{t-1}^2 \tilde{y}_t^2 = \frac{1}{T}\sum_{t=2}^{T} w_{t-1}^2 u_t^2 + \frac{1}{T}\sum_{t=2}^{T} w_{t-1}^2 \left(\tilde{y}_t^2 - u_t^2\right)$$

where

$$\begin{aligned} \mathrm{E}\left(\left|\frac{1}{T}\sum_{t=2}^{T} w_{t-1}^2\left(\tilde{y}_t^2 - u_t^2\right)\right|\right) &\leq \frac{1}{T}\sum_{t=2}^{T}\mathrm{E}\left(\left|w_{t-1}^2\left(\tilde{y}_t^2 - u_t^2\right)\right|\right) \leq \frac{1}{T}\sum_{t=2}^{T}\left\|w_{t-1}\right\|_4^2\left\|\tilde{y}_t^2 - u_t^2\right\|_2 \\ &\leq \left(\max_t \left\|w_{t-1}\right\|_4\right)^2 \max_t \left\|\tilde{y}_t^2 - u_t^2\right\|_2 \to 0 \end{aligned}$$

since $w_{t-1}$ is uniformly $L_4$-bounded, see item 1 of this lemma, so $T^{-1}\sum_{t=2}^{T} w_{t-1}^2\left(\tilde{y}_t^2 - u_t^2\right) \xrightarrow{p} 0$. It is not difficult to show that $\max_t \left\|w_{t-1}^2 - x_{t-1}^2\right\|_2 \to 0$, so we obtain analogously that

$$\frac{1}{T}\sum_{t=2}^{T} w_{t-1}^2 u_t^2 = \frac{1}{T}\sum_{t=2}^{T} x_{t-1}^2 u_t^2 + o_p(1),$$

and the result follows with Lemma 2.

7. Note that, along the lines of the above items,

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T} w_{t-1}\tilde{u}_t = \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\left(x_{t-1}-\mu\right)u_t + o_p(1).$$

Then, with $T^{-1}\sum_{t=2}^{T}\left(x_{t-1}-\mu\right)^2 u_t^2 \xrightarrow{p} \mathcal{E}_0^*\left[i^2\right]$ according to Lemma 2 and $\left(x_{t-1}-\mu\right)u_t$ possessing the md property, it suffices to show that $\max_{1\leq t\leq T}\left|\left(x_{t-1}-\mu\right)u_t\right| = o_p\left(T^{1/2}\right)$ in order to be able to apply a CLT for md arrays (e.g. Davidson, 1994, Theorem 24.3) to establish the result. To this end, note that

$$\max_{1\leq t\leq T}\left|\left(x_{t-1}-\mu\right)u_t\right| \leq \max_{1\leq t\leq T}\left|x_{t-1}-\mu\right|\max_{1\leq t\leq T}\left|u_t\right|$$

where both $u_t$ and $x_{t-1}$ are uniformly $L_{4\alpha}$-bounded, so both maxima are easily shown to be $O_p\left(T^{1/4\alpha}\right) = o_p\left(T^{1/4}\right)$ as required.

8. Follows along the lines of item 12 of Lemma 3 and we omit the details.

## Proof of Proposition 1

The OLS-based $t$ statistic is given by

$$t_\beta^{ls} = \frac{\sum_{t=2}^{T} f\left(x_{t-1}\right)u_t}{\hat{\sigma}_u\sqrt{\sum_{t=2}^{T} f^2\left(x_{t-1}\right)}} + \frac{\beta}{\hat{\sigma}_u}\sqrt{\sum_{t=2}^{T} f^2\left(x_{t-1}\right)}.$$

Then, under near-integration,

$$\begin{aligned} t_\beta^{ls} &= \frac{\frac{1}{\bar{\omega}_v^\alpha T^{\alpha/2+0.5}}\sum_{t=2}^{T} f\left(x_{t-1}\right)u_t}{\hat{\sigma}_u\sqrt{\frac{1}{\bar{\omega}_v^{2\alpha}T^{\alpha+1}}\sum_{t=2}^{T} f^2\left(x_{t-1}\right)}} + \frac{T^{\alpha/2+0.5}\beta\lambda^\alpha\bar{\omega}_v^\alpha}{\hat{\sigma}_u}\sqrt{\frac{1}{\lambda^{2\alpha}\bar{\omega}_v^{2\alpha}T^{\alpha+1}}\sum_{t=2}^{T} f^2\left(x_{t-1}\right)} \\ &= \frac{\frac{1}{T^{0.5}}\sum_{t=2}^{T} f\left(\frac{1}{\bar{\omega}_v T^{0.5}}x_{t-1}\right)u_t}{\hat{\sigma}_u\sqrt{\frac{1}{T}\sum_{t=2}^{T} f^2\left(\frac{1}{\bar{\omega}_v T^{0.5}}x_{t-1}\right)}} + b\frac{\lambda^\alpha\bar{\omega}_v^\alpha}{\hat{\sigma}_u}\sqrt{\frac{1}{T}\sum_{t=2}^{T} f^2\left(\frac{1}{\lambda\bar{\omega}_v T^{0.5}}x_{t-1}\right)}; \end{aligned}$$

The result follows with the CMT and Lemma 3 if $\hat{\sigma}_u \xrightarrow{p} \bar{\omega}_u$ under the local alternative as well. We have that

$$\begin{aligned}\hat{\sigma}_u &= \frac{1}{T}\sum_{t=1}^{T} y_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left(u_t + \frac{b}{T^{(\alpha+1)/2}} f\left(x_{t-1}\right)\right)^2 \\ &= \frac{1}{T}\sum_{t=1}^{T} u_t^2 + \frac{2b}{T^{(\alpha+3)/2}}\sum_{t=1}^{T} f\left(x_{t-1}\right) u_t + \frac{b^2}{T^{\alpha+2}}\sum_{t=1}^{T} f^2\left(x_{t-1}\right) \\ &= \frac{1}{T}\sum_{t=1}^{T} u_t^2 + o_p(1)\end{aligned}$$

which leads to the desired result since $\frac{1}{T}\sum_{t=1}^{T} u_t^2 \xrightarrow{p} \bar{\omega}_u^2$ indeed (the latter convergence being a particular case of Lemma 2).

Under stationarity, we have similarly that

$$\hat{\sigma}_u^2 = \frac{1}{T}\sum_{t=1}^{T} u_t^2 + o_p(1) = \bar{\omega}_u^2 + o_p(1),$$

and the result follows with Lemma 2 and a CLT for martingale difference arrays (e.g. Davidson, 1994, Theorem 24.3).

## Proof of Proposition 2

Write

$$t_\beta^{lin} = \frac{\sum y_t x_{t-1}}{\hat{\sigma}_u\sqrt{\sum x_{t-1}^2}} = \frac{\sum x_{t-1} u_t}{\hat{\sigma}_u\sqrt{\sum x_{t-1}^2}} + \beta\,\frac{\sum x_{t-1} f\left(x_{t-1}\right)}{\hat{\sigma}_u\sqrt{\sum x_{t-1}^2}}$$

such that

$$t_\beta^{lin} = \frac{\frac{1}{\lambda\bar{\omega}_u\bar{\omega}_v T}\sum_{t=2}^{T} x_{t-1} u_t}{\frac{\hat{\sigma}_u}{\bar{\omega}_u}\sqrt{\frac{1}{\lambda^2\bar{\omega}_v^2 T^2}\sum_{t=2}^{T} x_{t-1}^2}} + \frac{T^{\alpha/2+1/2}\beta\lambda^\alpha\bar{\omega}_v^\alpha}{\hat{\sigma}_u}\,\frac{\frac{1}{\lambda^{\alpha+1}\bar{\omega}_v^{\alpha+1}T^{\alpha/2+3/2}}\sum_{t=2}^{T} x_{t-1} f\left(x_{t-1}\right)}{\sqrt{\frac{1}{\lambda^2\bar{\omega}_v^2 T^2}\sum_{t=2}^{T} x_{t-1}^2}}$$

and note that $x_{t-1} \equiv H_1\left(x_{t-1}\right)$; the result follows with Lemma 3 and the CMT. The proof is analogous to that of Proposition 1 in the stable case, where $\mathcal{E}_\mu\left[f \cdot i\right] \neq 0$ due to the monotonicity of $f$ and $f(0) = 0$.

## Proof of Proposition 3

Write the test statistic with Eicker-White heteroskedasticity-consistent covariance matrix estimators as

$$t_\beta^{2S} = \frac{A_T B_T}{\sqrt{A_T C_T A_T'}}$$

where

$$A_T = \sum_{t=2}^{T} \tilde{x}_{t-1}\tilde{\boldsymbol{z}}_{t-1}' D_T^{-1}\left(D_T^{-1}\sum_{t=2}^{T}\tilde{\boldsymbol{z}}_{t-1}\tilde{\boldsymbol{z}}_{t-1}' D_T^{-1}\right)^{-1},$$

$$B_T = D_T^{-1}\sum_{t=2}^{T}\tilde{\boldsymbol{z}}_{t-1}\tilde{y}_t \quad \text{and} \quad C_T = D_T^{-1}\sum_{t=2}^{T}\tilde{\boldsymbol{z}}_{t-1}\tilde{\boldsymbol{z}}_{t-1}'\tilde{y}_t^2 D_T^{-1}$$

with $D_T$ defined differently for the cases of near integration and stationarity as follows.

**Proof of item 1.** Under near-integration, let $D_T = \operatorname{diag}\left(T^{1/2}, T^{1/2+\eta/2}\right)$ and examine

$$t_\beta^{2S} = \frac{\frac{1}{T}A_T B_T}{\sqrt{\frac{1}{T^2}A_T C_T A_T'}}.$$

Then, with $\bar{J}_{c,\sigma_v} = \int_0^1 J_{c,\sigma_v}(s)\,\mathrm{d}s$ and $\overline{\sin} = 1/T \sum_{t=1}^{T} \sin \frac{\pi t}{2T}$,

$$\frac{1}{T}\sum_{t=2}^{T} \tilde{x}_{t-1}\tilde{\boldsymbol{z}}_{t-1}' D_T^{-1} \Rightarrow \left(\lambda\bar{\omega}_v \left(\int_0^1 J_{c,\sigma_v}(s) \sin\frac{\pi s}{2}\mathrm{d}s - \frac{2}{\pi}\bar{J}_{c,\sigma_v}\right), 0\right),$$

since $\overline{\sin} \to \int_0^1 \sin\frac{\pi s}{2}\mathrm{d}s = \frac{2}{\pi}$ and $\frac{1}{T^{3/2}}\sum_{t=2}^{T} x_{t-1} \Rightarrow \lambda\bar{\omega}_v \int_0^1 J_{c,\sigma_v}(s)\,\mathrm{d}s$, and, with Lemma 3 items 3, 5 and 6,

$$\frac{1}{T^{3/2+\eta/2}}\sum_{t=2}^{T} x_{t-1}w_{t-1} - \left(\frac{1}{T^{3/2}}\sum_{t=2}^{T} x_{t-1}\right)\left(\frac{1}{T^{1+\eta/2}}\sum_{t=2}^{T} w_{t-1}\right) = o_p(1).$$

Furthermore, we have with items 7 and 10 of Lemma 3 that

$$D_T^{-1}\sum_{t=2}^{T} \tilde{\boldsymbol{z}}_{t-1}\tilde{\boldsymbol{z}}_{t-1}' D_T^{-1} = \begin{pmatrix} \int_0^1 \left(\sin\frac{\pi s}{2} - \frac{2}{\pi}\right)^2 \mathrm{d}s + o_p(1) & 0 \\ 0 & \Theta_p(1) \end{pmatrix}$$

and

$$C_T = \begin{pmatrix} \int_0^1 \left(\sin\frac{\pi s}{2} - \frac{2}{\pi}\right)^2 \sigma_u^2(s)\,\mathrm{d}s + o_p(1) & O_p(1) \\ O_p(1) & O_p(1) \end{pmatrix}.$$

Note that it is not necessary to establish the precise limiting behavior of the $O_p(1)$ and $\Theta(1)$ terms since they are multiplied with 0 in $A_T$ and $A_T C_T A_T'$. Summing up,

$$\sqrt{\frac{1}{T^2}A_T C_T A_T'} \Rightarrow \frac{\lambda\bar{\omega}_v \left|\int_0^1 \left(J_{c,\sigma_v}(s)\sin\frac{\pi s}{2} - \frac{2}{\pi}\bar{J}_{c,\sigma_v}\right)\mathrm{d}s\right| \sqrt{\int_0^1 \left(\sin\frac{\pi s}{2} - \frac{2}{\pi}\right)^2 \sigma_u^2(s)\,\mathrm{d}s}}{\int_0^1 \left(\sin\frac{\pi s}{2} - \frac{2}{\pi}\right)^2 \mathrm{d}s}.$$

Then, examine

$$D_T^{-1}\sum_{t=2}^{T} \tilde{\boldsymbol{z}}_{t-1}\tilde{y}_t = D_T^{-1}\sum_{t=2}^{T} \tilde{\boldsymbol{z}}_{t-1}\tilde{u}_t + \beta_1 D_T^{-1}\sum_{t=2}^{T} \tilde{\boldsymbol{z}}_{t-1}\left(f(x_{t-1}) - \overline{f(x_{-1})}\right) \tag{3}$$

with $\overline{f(x_{-1})} = \frac{1}{T}\sum_{t=2}^{T} f(x_{t-1})$. Focusing on the first summand on the r.h.s., we observe that

$$D_T^{-1}\sum_{t=2}^{T} \tilde{\boldsymbol{z}}_{t-1}\tilde{u}_t = D_T^{-1}\sum_{t=2}^{T} \begin{pmatrix} \sin\frac{\pi t}{2T} - \frac{2}{\pi} \\ w_{t-1} - \bar{w}_{-1} \end{pmatrix} u_t = D_T^{-1}\sum_{t=2}^{T} \begin{pmatrix} \sin\frac{\pi t}{2T} - \frac{2}{\pi} \\ w_{t-1} \end{pmatrix} u_t + o_p(1)$$

since $\bar{w}_{-1} = \frac{1}{T}\sum_{t=2}^{T} w_{t-1} = O_p\left(T^{\eta/2}\right)$ is negligible (see Lemma 3). The first element of the vector converges to

$$\bar{\omega}_u \int_0^1 \left(\sin\frac{\pi s}{2} - \frac{2}{\pi}\right)\mathrm{d}W_{\sigma_u}(s)$$

which is normal with mean zero and variance $\int_0^1 \sin^2\frac{\pi s}{2}\sigma_u^2(s)\,\mathrm{d}s - \bar{\omega}_u^2\frac{4}{\pi^2}$. The second summand on the r.h.s. of (3) is given by

$$\frac{b}{T^{(\alpha+1)/2}} D_T^{-1}\sum_{t=2}^{T} \begin{pmatrix} \sin\frac{\pi t}{2T} - \frac{2}{\pi} \\ w_{t-1} \end{pmatrix} f(x_{t-1}) + o_p(1)$$

again since $\bar{w}_{-1}$ is negligible. Recall from Lemma 3 item 9 that $T^{-\alpha/2-1-\eta/2}\sum_{t=2}^{T} w_{t-1}f\left(x_{t-1}\right)=O_p\left(1\right)$, so we have that

$$\begin{aligned}
&\frac{1}{T}A_T B_T \Rightarrow \\
&\lambda\bar{\omega}_v\left(\int_0^1 J_{c,\sigma_v}\left(s\right)\sin\frac{\pi s}{2}\mathrm{d}s-\frac{2}{\pi}\bar{J}_{c,\sigma_v}\right)\left(\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)^2\mathrm{d}s\right)^{-1}\times \\
&\quad\times\bar{\omega}_u\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)\mathrm{d}W_{\sigma_u}\left(s\right) \\
&+b\,\lambda^{\alpha+1}\bar{\omega}_v^{\alpha+1}\left(\int_0^1 J_{c,\sigma_v}\left(s\right)\sin\frac{\pi s}{2}\mathrm{d}s-\frac{2}{\pi}\bar{J}_{c,\sigma_v}\right)\left(\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)^2\mathrm{d}s\right)^{-1}\times \\
&\quad\times\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)H_\alpha\left(J_{c,\sigma_v}\left(s\right)\right)\mathrm{d}s
\end{aligned}$$

leading to

$$\begin{aligned}
t_\beta^{2S} &= \operatorname{sgn}\left(\int_0^1 J_{c,\sigma_v}\left(s\right)\sin\frac{\pi s}{2}\mathrm{d}s-\frac{2}{\pi}\bar{J}_{c,\sigma_v}\right)\frac{\bar{\omega}_u\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)\mathrm{d}W_{\sigma_u}\left(s\right)}{\sqrt{\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)^2\sigma_u^2\left(s\right)\mathrm{d}s}} \\
&\quad+b\,\lambda^\alpha\bar{\omega}_v^\alpha\frac{\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)H_\alpha\left(J_{c,\sigma_v}\left(s\right)\right)\mathrm{d}s}{\sqrt{\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)^2\sigma_u^2\left(s\right)\mathrm{d}s}}\operatorname{sgn}\left(\int_0^1 J_{c,\sigma_v}\left(s\right)\sin\frac{\pi s}{2}\mathrm{d}s-\frac{2}{\pi}\bar{J}_{c,\sigma_v}\right);
\end{aligned}$$

this completes the result (under the null, the sign cancels out upon squaring and the ratio $\frac{\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)\mathrm{d}W_{\sigma_u}(s)}{\sqrt{\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)^2\sigma_u^2(s)\mathrm{d}s}}$ is standard normal).

**Proof of item 2.** Should $x_t$ be stable, redefine $D_T=\operatorname{diag}\left(T^{1/2},T^{1/2}\right)$ and consider a different standardization of numerator and denominator of $t_\beta^{2S}$,

$$t_\beta^{2S}=\frac{\frac{1}{\sqrt{T}}A_T B_T}{\sqrt{\frac{1}{T}A_T C_T A_T'}}.$$

The behaviour of the term $\sum_{t=2}^{T}\tilde{x}_{t-1}\tilde{\boldsymbol{z}}_{t-1}'D_T^{-1}$ is different:

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\tilde{x}_{t-1}\tilde{\boldsymbol{z}}_{t-1}'D_T^{-1}\to\left(0,\,\mathcal{E}_\mu\left[i^2\right]-\left(\mathcal{E}_\mu\left[i\right]\right)^2\right)$$

thanks to Lemmas 2 and 4. Comparing with item 1, it is seen that the sample covariance of $\tilde{x}_{t-1}$ and $\tilde{\boldsymbol{z}}_{t-1}$ acts as a selector (the limits in the two cases, stable and near integrated, are orthogonal); see Breitung and Demetrescu (2015). The sample covariance matrix of the instruments satisfies according to Lemma 4

$$D_T^{-1}\sum_{t=2}^{T}\tilde{\boldsymbol{z}}_{t-1}\tilde{\boldsymbol{z}}_{t-1}'D_T^{-1}\xrightarrow{p}\begin{pmatrix}\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)^2\mathrm{d}s & 0\\ 0 & \mathcal{E}_0\left[i^2\right]\end{pmatrix}$$

while

$$C_T=\begin{pmatrix}\int_0^1\left(\sin\frac{\pi s}{2}-\frac{2}{\pi}\right)^2\sigma_u^2\left(s\right)\mathrm{d}s+o_p\left(1\right) & O_p\left(1\right)\\ O_p\left(1\right) & \mathcal{E}_0^*\left[i^2\right]\end{pmatrix}.$$

Consequently,

$$\sqrt{\frac{1}{T}A_T C_T A_T'}\Rightarrow\frac{\mathcal{E}_\mu\left[i^2\right]-\left(\mathcal{E}_\mu\left[i\right]\right)^2}{\mathcal{E}_0\left[i^2\right]}\sqrt{\mathcal{E}_0^*\left[i^2\right]},$$

while, with $w_{t-1} = x_{t-1} - \mu + o_p(1)$ from Lemma 4 item 1, it is easily shown that uniform $L_2$-boundedness of the $o_p(1)$ term leads to

$$\begin{aligned} &b\frac{1}{T}\sum_{t=2}^{T} w_{t-1}\left(f(x_{t-1}) - \overline{f(x_{-1})}\right) = \\ &\quad b\frac{1}{T}\sum_{t=2}^{T}(x_{t-1} - \bar{x})\left(f(x_{t-1}) - \overline{f(x_{-1})}\right) - b\bar{x}\frac{1}{T}\sum_{t=2}^{T}\left(f(x_{t-1}) - \overline{f(x_{-1})}\right) + o_p(1) \\ &\quad \xrightarrow{p} b\left(\mathcal{E}_{\mu}[f \cdot i] - \mathcal{E}_{\mu}[f]\,\mathcal{E}_{\mu}[i]\right). \end{aligned}$$

The result follows with the suitable items of Lemma 4.

## References


Amado, C. and T. Teräsvirta (2014). Modelling Changes in the Unconditional Variance of Long Stock Return Series. *Journal of Empirical Finance 25*(1), 15–35.

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

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

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

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

Campbell, J. Y. and M. Yogo (2006). Efficient Tests of Stock Return Predictability. *Journal of Financial Economics 81*(1), 27–60.

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.

Chang, Y. and J. Y. Park (2010). Endogeneity in Nonlinear Regressions with Integrated Time Series. *Econometric Reviews 30*(1), 51–87.

Chen, Q. and Y. Hong (2010). Predictability of equity returns over different time horizons: a nonparametric approach. Technical report.

Choi, Y., S. Jacewitz, and J. Y. Park (2016). A reexamination of stock return predictability. *Journal of Econometrics 192*(1), 168–189.

Chung, P. Y. and Z.-g. Zhou (1996). The Predictability of Stock Returns: A nonparametric approach. *Econometric Reviews 15*(3), 299–330.

Cochrane, J. H. (1988). How big is the random walk in GNP? *Journal of Political Economy 96*(5), 893–920.

Dahlhaus, R. (2012). Locally Stationary Processes. In T. Subba Rao, S. Subba Rao, and C. R. Rao (Eds.), *Handbook of Statistics: Time Series Analysis: Methods and Applications*, Volume 30, pp. 351–413. North Holland.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press.

Demetrescu, M. and P. Sibbertsen (2014). Inference on the Long-Memory Properties of Time Series with Non-Stationary Volatility. Hannover Economic Papers 531, Leibniz Unviersity Hannover.

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.

Fan, Y. and Q. Li (1999). Central Limit Theorem for Degenerate U-Statistics of Absolutely Regular Processes with Applications to Model Specification Testing. *Journal of Nonparametric Statistics 10*(3), 245–271.

Gupta, K., C. Krishnamurti, and A. Tourani-Rad (2013). Is corporate governance relevant during the financial crisis? *Journal of International Financial Markets, Institutions and Money 23*, 85–110.

Hansen, P. R. and A. Timmermann (2012). Choice of sample split in out-of-sample forecast evaluation. EUI Working Papers ECO 2012/10, European University Institute.

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

Juhl, T. (2014). A Nonparametric Test of the Predictive Regression Model. *Journal of Business & Economic Statistics 32*(3), 387–394.

Kanas, A. (2005). Nonlinearity in the stock price–dividend relation. *Journal of International Money and Finance 24*(4), 583–606.

Kasparis, I., E. Andreou, and P. C. B. Phillips (2015). Nonparametric predictive regression. *Journal of Econometrics 185*(2), 468–494.

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

Kurtz, T. G. and P. Protter (1991). Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations. *Annals of Probability 19*(3), 1035–1070.

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

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988). Testing Linearity Against Smooth Transition Autoregressive Models. *Biometrika 75*(3), 491–499.

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.

Marmer, V. (2008). Nonlinearity, Nonstationarity, and Spurious Forecasts. *Journal of Econometrics 142*(1), 1–27.

McMillan, D. G. (2001). Nonlinear predictability of stock market returns: Evidence from nonparametric and threshold models. *International Review of Economics & Finance 10*(4), 353–368.

McMillan, D. G. (2003). Non-linear predictability of UK stock market returns. *Oxford Bulletin of Economics and Statistics 65*(5), 557–573.

Park, J. Y. and P. C. B. Phillips (1999). Asymptotics for Nonlinear Transformations of Integrated Time Series. *Econometric Theory 15*(3), 269–298.

Paye, B. S. and A. Timmermann (2006). Instability of return prediction models. *Journal of Empirical Finance 13*(3), 274–315.

Perron, P. and T. Yabu (2009). Testing for shifts in trend with an integrated or stationary noise component. *Journal of Business & Economic Statistics 27*(3), 369–396.

Phillips, P. C. B. (1998). New Tools for Understanding Spurious Regressions. *Econometrica 66*(6), 1299–1325.

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.

Priestley, M. B. (1988). *Non-Linear and Non-Stationary Time Series Analysis*. Academic Press London.

Scheithauer, J. (2008). P-value surfaces for time series econometrics–with an application to long-term interest rates. Mimeo.

Shi, X. and P. C. B. Phillips (2012). Nonlinear Cointegrating Regression Under Weak Identification. *Econometric Theory 28*(3), 509–547.

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

Viceira, L. M. (1997). Testing for structural change in the predictability of asset returns. Manuscript, Harvard University.

Wang, Q. and P. C. B. Phillips (2012). A specification test for nonlinear nonstationary models. *The Annals of Statistics 40*(2), 727–758.

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