Out of Sample Predictability in Predictive Regressions with Many Predictor Candidates

Supplementary Material

Jesús Gonzalo
Department of Economics, Universidad Carlos III de Madrid
and
Jean-Yves Pitarakis
Department of Economics, University of Southampton

February 17, 2022

This supplementary appendix contains two sections. Section A provides the technical proofs of the Propositions stated in the main text. Section B provides additional simulations that further illustrate the finite sample properties of our test and key player estimator. Emphasis is placed on various robustness considerations and a broader range of DGP parameterisations.

A PROOFS

LEMMA A1: Let

$$\mathcal{N}_n(m_0) = \sqrt{n-k_0}\left(\frac{1}{2}\left(\frac{\sum_{t=k_0}^{k_0+m_0-1}\eta_{t+1}}{m_0} + \frac{\sum_{t=k_0+m_0}^{n-1}\eta_{t+1}}{n-k_0-m_0}\right) - \frac{\sum_{t=k_0}^{n-1}\eta_{t+1}}{n-k_0}\right) \tag{A.1}$$

with $\eta_t = (u_t^2 - E[u_t^2])$ and $m_0 = [(n-k_0)\mu_0]$. The long run variance of $\mathcal{N}_n(m_0)$ is given by

$$\omega^2 = \frac{(1-2\mu_0)^2}{4\mu_0(1-\mu_0)}\phi^2 \tag{A.2}$$

where $\phi^2 = \sum_{s=-\infty}^{\infty} \gamma_\eta(s)$.

PROOF OF LEMMA A1. First, let us rewrite (A.1) as

$$\mathcal{N}_n(m_0) = \sqrt{n-k_0}\left(\frac{1}{2}\left(\frac{n-k_0}{m_0}\frac{\sum_{t=k_0}^{k_0+m_0-1}\eta_{t+1}}{n-k_0} + \frac{n-k_0}{n-k_0-m_0}\frac{\sum_{t=k_0+m_0}^{n-1}\eta_{t+1}}{n-k_0}\right) - \frac{\sum_{t=k_0}^{n-1}\eta_{t+1}}{n-k_0}\right) \tag{A.3}$$

Using $m_0 = [(n-k_0)\mu_0]$ we have

$$\mathcal{N}_n(\mu_0) = \sqrt{n-k_0}\left(\frac{1}{2}\left(\frac{1}{\mu_0}\frac{\sum_{t=k_0}^{k_0+m_0-1}\eta_{t+1}}{n-k_0} + \frac{1}{1-\mu_0}\frac{\sum_{t=k_0+m_0}^{n-1}\eta_{t+1}}{n-k_0}\right) - \frac{\sum_{t=k_0}^{n-1}\eta_{t+1}}{n-k_0}\right) + o(1). \tag{A.4}$$

Next, let $I_{1t} \equiv I(k_0 \leq t < k_0+m_0)$ and $I_{2t} \equiv I(k_0+m_0 \leq t < n)$ and define

$$\begin{aligned} Z_t &= \eta_{t+1}\left(\frac{1}{2\mu_0}I_{1t} + \frac{1}{2(1-\mu_0)}I_{2t} - 1\right) \\ &\equiv \eta_{t+1}g_t \end{aligned} \tag{A.5}$$

so that (A.4) can be reformulated as

$$\mathcal{N}_n(\mu_0) = \sqrt{n-k_0}\;\overline{Z}_n + o(1) \tag{A.6}$$

where $\overline{Z}_n = \sum_{t=k_0}^{n-1} Z_t/(n-k_0)$. Standard algebra now leads to

$$\begin{aligned} V\left[\sum_{t=k_0}^{n-1} Z_t\right] &= \sum_{t=k_0}^{n-1} V[Z_t] + \sum_{t\neq s} Cov[Z_t, Z_s] \\ &= (n-k_0)\gamma_Z(0) + 2\sum_{s=1}(n-k_0-s)\gamma_Z(s) \end{aligned} \tag{A.7}$$

where $\gamma_Z(0) = E[\eta_{t+1}^2 g_t^2]$ and $\gamma_Z(s) = E[\eta_{t+1}\eta_{t+1-s}g_t g_{t-s}]$. Recalling the expression of the deterministic term g_t from (A.5) it now suffices to note that $E[\eta_{t+1}^2 g_t^2] = E[g_t^2]E[\eta_{t+1}^2]$ and $E[\eta_{t+1}\eta_{t+1-s}g_t g_{t-s}] = E[g_t^2]E[\eta_{t+1}\eta_{t+1-s}]$ with

$$E[g_t^2] = \frac{(1-2\mu_0)^2}{4\mu_0(1-\mu_0)} \tag{A.8}$$

from which it immediately follows that as $n \to \infty$

$$V[\mathcal{N}_n(\mu_0)] \to \frac{(1-2\mu_0)^2}{4\mu_0(1-\mu_0)}\;\phi^2 \tag{A.9}$$

for $\phi^2 = \sum_{s=-\infty}^{\infty} \gamma_\eta(s)$. ■

PROOF OF PROPOSITION 1. Under the null hypothesis Assumption 1(iii) implies that for given μ_0 we can write $\overline{\mathcal{D}}_n(m_0)$ as

$$\overline{\mathcal{D}}_n(m_0) = \frac{\sqrt{n-k_0}}{\hat{\omega}_n} \left(\frac{1}{2} \left(\frac{\sum_{t=k_0}^{k_0+m_0-1} (u_{t+1}^2 - \sigma_u^2)}{m_0} + \frac{\sum_{t=k_0+m_0}^{n-1} (u_{t+1}^2 - \sigma_u^2)}{n-k_0-m_0} \right) - \frac{\sum_{t=k_0}^{n-1} (u_{t+1}^2 - \sigma_u^2)}{n-k_0} \right) + o_p(1). \tag{A.10}$$

It now follows directly from Assumptions 1(i)-(ii) and the continuous mapping theorem that as $n \to \infty$

$$\overline{\mathcal{D}}_n(\mu_0) \xrightarrow{d} \frac{1}{\omega} \left(\frac{\phi}{2} \left(\frac{W(\mu_0)}{\mu_0} + \frac{W(1) - W(\mu_0)}{1-\mu_0} \right) - \phi W(1) \right). \tag{A.11}$$

As we operate under a given μ_0 it is now straightforward to observe that the variance of the expression between brackets in (A.11) is given by $\phi^2(1-2\mu_0)^2/(4\mu_0(1-\mu_0))$. As $\hat{\omega} \xrightarrow{p} \omega = \phi(1-2\mu_0)/\sqrt{4\mu_0(1-\mu_0)}$ it follows from Slutsky's theorem that $\overline{\mathcal{D}}_n(\mu_0) \xrightarrow{d} N(0,1)$ as stated. ■

Before proceeding with the proofs of Propositions 2A, 2B and 2C we introduce a series of intermediate results and further notation that will be used throughout. As we operate under the hypothesis of at least one active predictor the true specifications under our three scenarios A, B and C are understood to be given by

$$y_{t+1} = \sum_{i \in \mathcal{I}^*} (\beta_i^*/n^{1/4})\ x_{it} + u_{t+1} \tag{A.12}$$

$$y_{t+1} = \sum_{i \in \mathcal{I}^*} (\beta_i^*/n^{(1+2\alpha)/4})\ x_{it} + u_{t+1} \tag{A.13}$$

and

$$y_{t+1} = \sum_{i \in \mathcal{I}_1^*} (\beta_i^*/n^{1/4})\ x_{it} + \sum_{i \in \mathcal{I}_2^*} (\beta_i^*/n^{(1+2\alpha)/4})\ x_{it} + u_{t+1} \tag{A.14}$$

respectively. We also recall that the fitted specification involving one predictor at a time is here given by

$$y_{t+1} = \beta_j x_{jt} + u_{t+1} \qquad j = 1, \ldots, p \tag{A.15}$$

so that using (A.12) and (A.13) we can write the recursively estimated slope parameters as

$$\hat{\beta}_{jt} = \frac{\sum_{i \in \mathcal{I}^*} \beta_i^* (\sum_{s=1}^t x_{is} x_{js})}{n^{\gamma} \sum_{s=1}^t x_{js}^2} + \frac{\sum_{s=1}^t x_{js} u_{s+1}}{\sum_{s=1}^t x_{js}^2} \tag{A.16}$$

where $\gamma = 1/4$ under scenario A and $\gamma = (1 + 2\alpha)/4$ under scenario B. For the mixed predictor scenario C and using (A.14) we have instead

$$\hat{\beta}_{jt} = \frac{\sum_{i \in \mathcal{I}_1^*} \beta_i^* (\sum_{s=1}^t x_{is} x_{js})}{n^{\gamma_1} \sum_{s=1}^t x_{js}^2} + \frac{\sum_{i \in \mathcal{I}_2^*} \beta_i^* (\sum_{s=1}^t x_{is} x_{js})}{n^{\gamma_2} \sum_{s=1}^t x_{js}^2} + \frac{\sum_{s=1}^t x_{js} u_{s+1}}{\sum_{s=1}^t x_{js}^2}. \tag{A.17}$$

The specifications in (A.12)-(A.14) are the DGPs under the local alternatives of interest and the $\hat{\beta}_{jt}$'s in (A.16)-(A.17) are the slope parameters estimated via recursive least squares when fitting (A.15). As for notational convenience we have abstracted from the inclusion of an intercept in the above specifications it is naturally understood that the forecasts under the null model will be taken as $\hat{y}_{0,t+1|t} = 0$ instead of $\sum_{j=1}^t y_j/t$. This has no bearing on any of the asymptotic results presented in Propositions 2A-2C. We can now write the forecast errors as

$$\begin{aligned} \hat{e}_{0,t+1|t} &= y_{t+1} - 0 \\ \hat{e}_{j,t+1|t} &= y_{t+1} - \hat{\beta}_{jt} x_{jt} \end{aligned} \tag{A.18}$$

with y_{t+1} given by either (A.12), (A.13) or (A.14).

LEMMA A2. Under Assumption 2A, $\hat{\beta}_{jt}$ as in (A.16) and $\forall j \in \{1, \ldots, p\}$ we have as $n \to \infty$

(i) $\sup_{r\in[\pi_0,1]}\left|n^{1/4}\hat{\beta}_{j,[nr]} - \frac{1}{E[x_{jt}^2]}\sum_{i\in\mathcal{I}^*}\beta_i^* E[x_{it}x_{jt}]\right| = o_p(1)$

(ii) $\sup_{k_0\le t\le n}\left|\frac{\sum_{\ell=k_0}^t \hat{\beta}_{j\ell}x_{j\ell}u_{\ell+1}}{\sqrt{n-k_0}}\right| = o_p(1)$

(iii) $\sup_{k_0\le t\le n}\left|\frac{\sum_{\ell=k_0}^t \hat{\beta}_{j\ell}^2 x_{j\ell}^2}{\sqrt{n-k_0}} - \frac{\sqrt{1-\pi_0}}{E[x_{jt}^2]}\left(\sum_{i\in\mathcal{I}^*}\beta_i^* E[x_{it}x_{jt}]\right)^2\right| = o_p(1)$

(iv) $\sup_{k_0\le t\le n}\left|\frac{\beta_i^*}{n^{1/4}}\frac{\sum_{\ell=k_0}^t \hat{\beta}_{j\ell}x_{i\ell}x_{j\ell}}{\sqrt{n-k_0}} - \sqrt{1-\pi_0}\ \beta_i^*\frac{E[x_{it}x_{jt}]}{E[x_{jt}^2]}(\sum_{i\in\mathcal{I}^*}\beta_i^* E[x_{it}x_{jt}])\right| = o_p(1)$

PROOF OF LEMMA A2. (i) From (A.16) we have

$$n^{1/4}\hat{\beta}_{jt} = \frac{\sum_{i\in\mathcal{I}^*}\beta_i^*(\sum_{s=1}^t x_{is}x_{js})}{\sum_{s=1}^t x_{js}^2} + n^{1/4}\frac{\sum_{s=1}^t x_{js}u_{s+1}}{\sum_{s=1}^t x_{js}^2} \quad (A.19)$$

and

$$n^{1/4}\sup_t|\hat{\beta}_{jt}| \le \sup_t\left|\frac{\sum_{i\in\mathcal{I}^*}\beta_i^*(\sum_{s=1}^t x_{is}x_{js})}{\sum_{s=1}^t x_{js}^2}\right| + n^{1/4}\sup_t\left|\frac{\sum_{s=1}^t x_{js}u_{s+1}}{\sum_{s=1}^t x_{js}^2}\right|. \quad (A.20)$$

We can now note that

$$n^{1/4}\sup_t\left|\frac{\sum_{s=1}^t x_{js}u_{s+1}}{\sum_{s=1}^t x_{js}^2}\right| \le \sup_t\left|\frac{t}{\sum_{s=1}^t x_{js}^2}\right|\frac{n^{1/4}}{t}\sup_t\left|\sum_{s=1}^t x_{js}u_{s+1}\right| \xrightarrow{p} 0 \quad (A.21)$$

which follows directly from Assumption 2A(iii). This latter assumption now also leads to

$$\sup_t\left|\frac{\sum_{i\in\mathcal{I}^*}\beta_i^*(\sum_{s=1}^t x_{is}x_{js})}{\sum_{s=1}^t x_{js}^2} - \sum_{i\in\mathcal{I}^*}\beta_i^*\frac{E[x_{it}x_{jt}]}{E[x_{jt}^2]}\right| = o_p(1) \quad (A.22)$$

as required. (ii) We write

$$\sup_{k_0\le t\le n}\left|\frac{\sum_{\ell=k_0}^t \hat{\beta}_{j\ell}x_{j\ell}u_{\ell+1}}{\sqrt{n-k_0}}\right| = \frac{1}{\sqrt{1-\pi_0}}\frac{1}{n^{1/4}}\sup_{r\in[\pi_0,1]}\left|\frac{\sum_{l=k_0}^{[nr]}(n^{1/4}\hat{\beta}_{jt})x_{j\ell}u_{\ell+1}}{\sqrt{n}}\right| + o_p(1). \quad (A.23)$$

The result in part (i) combined with Assumption 2A(iii) allows us to appeal to Theorem 3.3 in Hansen (1993) from which the statement in (ii) follows. For part (iii) it is sufficient to focus on

$$\frac{1}{\sqrt{n-k_0}}\sum_{\ell=k_0}^{n-1}\hat{\beta}_{j\ell}^2 x_{j\ell}^2 = \frac{1}{\sqrt{1-\pi_0}}\frac{1}{n}\sum_{\ell=k_0}^{n}(\sqrt{n}\hat{\beta}_{j\ell}^2)x_{j\ell}^2 + o(1) \quad (A.24)$$

for which part (i) combined with Assumptions 2A(iii) ensures that

$$\frac{1}{\sqrt{n-k_0}} \sum_{\ell=k_0}^{n-1} \hat{\beta}_{j\ell}^2 x_{j\ell}^2 \xrightarrow{p} \sqrt{1-\pi_0} \left(\sum_{i \in \mathcal{I}^*} \beta_i^* \frac{E[x_{it} x_{jt}]}{\sqrt{E[x_{jt}^2]}} \right)^2 . \tag{A.25}$$

Part (iv) follows identical lines to part (iii) and its details are therefore omitted. ■

PROOF OF PROPOSITION 2A. Using y_{t+1} as in (A.12) in $\hat{e}^2_{0,t+1|t} = y^2_{t+1}$ from (A.18), we have

$$\begin{aligned}
\frac{\sum_{t=k_0}^{k_0+m_0-1} \hat{e}^2_{0,t+1|t}}{\sqrt{n-k_0}} &= \frac{\sum_{t=k_0}^{k_0+m_0-1} u^2_{t+1}}{\sqrt{n-k_0}} + \frac{\sum_{t=k_0}^{k_0+m_0-1} (\sum_{i \in \mathcal{I}^*} \beta_i^* x_{it})^2}{\sqrt{n}\sqrt{n-k_0}} \\
&+ \frac{2}{n^{1/4}} \sum_{i \in \mathcal{I}^*} \beta_i^* \left(\frac{\sum_{t=k_0}^{k_0+m_0-1} x_{it} u_{t+1}}{\sqrt{n-k_0}} \right) \\
&= \frac{\sum_{t=k_0}^{k_0+m_0-1} u^2_{t+1}}{\sqrt{n-k_0}} + \frac{\sum_{t=k_0}^{k_0+m_0-1} (\sum_{i \in \mathcal{I}^*} \beta_i^* x_{it})^2}{\sqrt{n}\sqrt{n-k_0}} + o_p(1) \\
&= \frac{\sum_{t=k_0}^{k_0+m_0-1} u^2_{t+1}}{\sqrt{n-k_0}} + \mu_0 \sqrt{1-\pi_0} \; E \left[\sum_{i \in \mathcal{I}^*} \beta_i^* x_{it} \right]^2 + o_p(1) \qquad (A.26)
\end{aligned}$$

where we made repeated use of Assumption 2A(iii). Proceeding as above it also follows that

$$\frac{\sum_{t=k_0+m_0}^{n-1} \hat{e}^2_{0,t+1|t}}{\sqrt{n-k_0}} = \frac{\sum_{t=k_0+m_0}^{n-1} u^2_{t+1}}{\sqrt{n-k_0}} + (1-\mu_0)\sqrt{1-\pi_0} \; E \left[\sum_{i \in \mathcal{I}^*} \beta_i^* x_{it} \right]^2 + o_p(1). \tag{A.27}$$

Next, we focus on $\hat{e}^2_{j,t+1|t}$ given by (A.18) with y_{t+1} as in (A.12). We have

$$\begin{aligned}
\frac{\sum_{t=k_0}^{n-1} \hat{e}^2_{j,t+1|t}}{\sqrt{n-k_0}} &= \frac{\sum_{t=k_0}^{n-1} u^2_{t+1}}{\sqrt{n-k_0}} + \frac{1}{\sqrt{n(n-k_0)}} \sum_{t=k_0}^{n-1} (\sum_{i \in \mathcal{I}^*} \beta_i^* x_{it})^2 \\
&+ \frac{2}{n^{1/4}\sqrt{n-k_0}} \sum_{i \in \mathcal{I}^*} \beta_i^* (\sum_{t=k_0}^{n-1} x_{it} u_{t+1}) + \frac{1}{\sqrt{n-k_0}} \sum_{t=k_0}^{n-1} \hat{\beta}_{jt}^2 x_{jt}^2 \\
&- \frac{2}{n^{1/4}\sqrt{n-k_0}} \sum_{i \in \mathcal{I}^*} \beta_i^* (\sum_{t=k_0}^{n-1} \hat{\beta}_{jt} x_{jt} x_{it}) \\
&- \frac{2}{\sqrt{n-k_0}} \sum_{t=k_0}^{n-1} \hat{\beta}_{jt} x_{jt} u_{t+1}. \qquad (A.28)
\end{aligned}$$

Appealing to Assumption 2A(iii) and using Lemma A2(ii)-(iii) in (A.28) also allows us to write

$$\frac{\sum_{t=k_0}^{n-1} \hat{e}_{j,t+1|t}^2}{\sqrt{n-k_0}} = \frac{\sum_{t=k_0}^{n-1} u_{t+1}^2}{\sqrt{n-k_0}} + \sqrt{1-\pi_0}\; E\left[\sum_{i\in\mathcal{I}^*} \beta_i^* x_{it}\right]^2 - \frac{\sqrt{1-\pi_0}}{E[x_{jt}^2]}\left(\sum_{i\in\mathcal{I}^*} \beta_i^* E[x_{it}x_{jt}]\right)^2 + o_p(1). \tag{A.29}$$

Next, using (A.26)-(A.29) in $\mathcal{D}_n(m_0,j)$ now gives

$$\mathcal{D}_n(m_0,j) = \frac{1}{\omega(\mu_0)}\left(\frac{1}{2}\left(\frac{n-k_0}{m_0}\frac{\sum_{t=k_0}^{k_0+m_0-1} u_{t+1}^2}{\sqrt{n-k_0}} + \frac{n-k_0}{n-k_0-m_0}\frac{\sum_{t=k_0+m_0}^{n-1} u_{t+1}^2}{\sqrt{n-k_0}}\right) - \frac{\sum_{t=k_0}^{n-1} u_{t+1}^2}{\sqrt{n-k_0}}\right) + \sqrt{1-\pi_0}\frac{1}{\omega(\mu_0)}\left(\sum_{i\in\mathcal{I}^*} \beta_i^* \frac{E[x_{it}x_{jt}]}{\sqrt{E[x_{jt}^2]}}\right)^2 + o_p(1) \tag{A.30}$$

leading to the desired result. ■

LEMMA B1. Under Assumption 2B, $\hat{\beta}_{jt}$ as in (A.16) and $\forall j \in \{1,\ldots,p\}$ we have as $n\to\infty$

(i) $\sup_{r\in[\pi_0,1]}\left| n^{(1+2\alpha)/4}\hat{\beta}_{j,[nr]} - \sum_{i\in\mathcal{I}^*}\beta_i^* \frac{\sigma_{v_iv_j}}{\sigma_{v_j}^2}\left(\frac{2c_j}{c_i+c_j}\right)\right| = o_p(1)$

(ii) $\sup_{k_0\le t\le n}\left|\frac{\sum_{\ell=k_0}^{t}\hat{\beta}_{j\ell}x_{j\ell}u_{\ell+1}}{\sqrt{n-k_0}}\right| = o_p(1)$

(iii) $\sup_{k_0\le t\le n}\left|\frac{\sum_{\ell=k_0}^{t}\hat{\beta}_{j\ell}^2 x_{j\ell}^2}{\sqrt{n-k_0}} - \sqrt{1-\pi_0}\left(\sum_{i\in\mathcal{I}^*}\beta_i^*\frac{\sigma_{v_iv_j}}{\sqrt{\sigma_{v_j}^2}}\frac{\sqrt{2c_j}}{c_i+c_j}\right)^2\right| = o_p(1),$

(iv) $\sup_{k_0\le t\le n}\left|\frac{\beta_i^*}{n^{(1+2\alpha)/4}}\frac{\sum_{\ell=k_0}^{t}\hat{\beta}_{j\ell}x_{i\ell}x_{j\ell}}{\sqrt{n-k_0}} - \sqrt{1-\pi_0}\frac{2c_j\sigma_{v_iv_j}}{(c_i+c_j)\sigma_{v_j}^2}\left(\sum_{i\in\mathcal{I}^*}\beta_i^*\frac{\sigma_{v_iv_j}}{c_i+c_j}\right)\right| = o_p(1)$

PROOF OF LEMMA B1. For all four cases the results follow in an identical manner to Lemma A2(i)-(iv) with the use of Assumption 2A(iii) replaced with Assumption 2B(iii) and $n^{1/4}$ replaced with $n^{(1+2\alpha)/4}$. ■

PROOF OF PROPOSITION 2B. Using y_{t+1} as in (A.13) in $\hat{e}_{0,t+1|t} = y_{t+1}^2$ from (A.18) we have

$$\begin{aligned}
\frac{\sum_{t=k_0}^{k_0+m_0-1} \hat{e}_{0,t+1|t}^2}{\sqrt{n-k_0}} &= \frac{\sum_{t=k_0}^{k_0+m_0-1} u_{t+1}^2}{\sqrt{n-k_0}} + \frac{\sum_{t=k_0}^{k_0+m_0-1} (\sum_{i\in\mathcal{I}^*} \beta_i^* x_{it})^2}{n^{(1+2\alpha)/2}\sqrt{n-k_0}} \\
&\quad + \frac{2}{n^{(1+2\alpha)/4}} \sum_{i\in\mathcal{I}^*} \beta_i^* \left(\frac{\sum_{t=k_0}^{k_0+m_0-1} x_{it}u_{t+1}}{\sqrt{n-k_0}} \right) \\
&= \frac{\sum_{t=k_0}^{k_0+m_0-1} u_{t+1}^2}{\sqrt{n-k_0}} + \frac{\sum_{t=k_0}^{k_0+m_0-1} (\sum_{i\in\mathcal{I}^*} \beta_i^* x_{it})^2}{n^{(1+2\alpha)/2}\sqrt{n-k_0}} + o_p(1) \qquad \text{(A.31)}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\sum_{t=k_0+m_0}^{n-1} \hat{e}_{0,t+1|t}^2}{\sqrt{n-k_0}} &= \frac{\sum_{t=k_0+m_0}^{n-1} u_{t+1}^2}{\sqrt{n-k_0}} + \frac{\sum_{t=k_0+m_0}^{n-1} (\sum_{i\in\mathcal{I}^*} \beta_i^* x_{it})^2}{n^{(1+2\alpha)/2}\sqrt{n-k_0}} \\
&\quad + \frac{2}{n^{(1+2\alpha)/4}} \sum_{i\in\mathcal{I}^*} \beta_i^* \left(\frac{\sum_{t=k_0+m_0}^{n-1} x_{it}u_{t+1}}{\sqrt{n-k_0}} \right) \\
&= \frac{\sum_{t=k_0+m_0}^{n-1} u_{t+1}^2}{\sqrt{n-k_0}} + \frac{\sum_{t=k_0+m_0}^{n-1} (\sum_{i\in\mathcal{I}^*} \beta_i^* x_{it})^2}{n^{(1+2\alpha)/2}\sqrt{n-k_0}} + o_p(1). \qquad \text{(A.32)}
\end{aligned}$$

Next, for $\hat{e}_{j,t+1|t}^2$ we have

$$\begin{aligned}
\frac{\sum_{t=k_0}^{n-1} \hat{e}_{j,t+1|t}^2}{\sqrt{n-k_0}} &= \frac{\sum_{t=k_0}^{n-1} u_{t+1}^2}{\sqrt{n-k_0}} + \frac{1}{n^{(1+2\alpha)/2}\sqrt{(n-k_0)}} \sum_{t=k_0}^{n-1} (\sum_{i\in\mathcal{I}^*} \beta_i^* x_{it})^2 \\
&\quad - \sqrt{1-\pi_0}\, \frac{2c_j}{\sigma_{v_j}^2} \left(\sum_{i\in\mathcal{I}^*} \beta_i^* \frac{\sigma_{v_i v_j}}{c_i + c_j} \right)^2 + o_p(1). \qquad \text{(A.33)}
\end{aligned}$$

Using (A.31)-(A.33) in $\mathcal{D}_n(m_0, j)$ and rearranging gives

$$\begin{aligned}
\mathcal{D}_n(m_0, j) &= \frac{1}{\omega(\mu_0)} \left(\frac{1}{2} \left(\frac{n-k_0}{m_0} \frac{\sum_{t=k_0}^{k_0+m_0-1} u_{t+1}^2}{\sqrt{n-k_0}} + \frac{n-k_0}{n-k_0-m_0} \frac{\sum_{t=k_0+m_0}^{n-1} u_{t+1}^2}{\sqrt{n-k_0}} \right) - \frac{\sum_{t=k_0}^{n-1} u_{t+1}^2}{\sqrt{n-k_0}} \right) \\
&\quad + \sqrt{1-\pi_0} \frac{1}{\omega(\mu_0)} \left(\sum_{i\in\mathcal{I}^*} \beta_i^* \frac{\sigma_{v_i v_j}}{\sqrt{\sigma_{v_j}^2}} \sqrt{\frac{2c_j}{(c_i+c_j)^2}} \right)^2 + o_p(1) \qquad \text{(A.34)}
\end{aligned}$$

leading to the result in Proposition 2B. ■

LEMMA C1. Under Assumption 2C, $\hat{\beta}_{jt}$ as in (A.17) and $\forall j \in \{1, \ldots, p\}$ we have as $n \to \infty$

(i) $\sup_{r \in [\pi_0, 1]} \left| n^{1/4} \hat{\beta}_{j,[nr]} - \frac{1}{E[x_{jt}^2]} \sum_{i \in \mathcal{I}_1^*} \beta_i^* E[x_{it} x_{jt}] \right| = o_p(1) \quad \text{for } j \in \mathcal{I}_1^*$

(ii) $\sup_{r \in [\pi_0, 1]} \left| n^{(1+2\alpha)/4} \hat{\beta}_{j,[nr]} - \sum_{i \in \mathcal{I}_2^*} \beta_i^* \frac{\sigma_{v_i v_j}}{\sigma_{v_j}^2} \left(\frac{2c_j}{c_i + c_j} \right) \right| = o_p(1) \quad \text{for } j \in \mathcal{I}_2^*.$

PROOF OF LEMMA C1. (i) and (ii) are obtained following the same derivations as Lemma A2(i) and LemmaA B1(i) and details are therefore omitted. It is here useful to note the distinct behaviour of the slope estimates obtained from the *one predictor at a time* regressions depending on whether the fitted predictor belongs to $\mathcal{I}_1^*$ or $\mathcal{I}_2^*$. This result is driven by the well known phenomenon of asymptotic independence between persistent and stationary predictors. ■

PROOF OF PROPOSITION 2C. The result in (21) is obtained following identical derivations to (14) and (18) and details are therefore omitted. ■

PROOF OF PROPOSITION 3. The result in part (i) is a direct consequence of Assumption 1(iii) which under the null hypothesis holds for both $\ell = 0$ and $\ell \in \{1, \ldots, p\}$. The result in part (ii) follows exact same steps as in the proofs of Propositions 2A-2C. ■

PROOF OF PROPOSITION 4. For part (i) of Proposition 4 we focus solely on the case of a single stationary active predictor in the DGP as the remaining scenarios follow identical lines. It is useful to first note that the argmax of $\mathcal{D}_n(m_0, j)$ will be equivalent to $\arg\min_j \mathcal{S}_n(j)$ where

$$\mathcal{S}_n(j) = \frac{\sum_{t=k_0}^{n-1} (\hat{e}_{j,t+1|t}^2 - u_{t+1}^2)}{\sqrt{n - k_0}} \quad j = 1, \ldots, p. \tag{A.35}$$

The main result now follows by establishing that $\mathcal{S}_n(j)$ converges to a deterministic limit that is uniquely minimized at $j = j_0$. We continue to operate under the DGP given by (A.12) with $|\mathcal{I}^*| = 1$ (i.e. there is a single active predictor) and with no loss of generality we set that predictor to be x_{1t}. Recalling that $\hat{e}_{j,t+1|t} = y_{t+1} - \hat{\beta}_{jt}x_{jt}$ and using Lemma A2 it immediately follows that for $j = j_0 = 1$ we have $\mathcal{S}_n(j = 1) \xrightarrow{p} 0$ while for $j \neq j_0 = 1$ and using Lemma A2 we have

$$\mathcal{S}_n(j) \xrightarrow{p} (\beta_1^*)^2\sqrt{1-\pi_0}E[x_{1t}^2](1-\rho_{1j}^2) \quad \forall j \neq j_0 \tag{A.36}$$

which is strictly positive for any predictor different from x_{1t}, thus leading to the required result. (ii) For part (ii) of Proposition 4 we consider the DGP given by (A.14) and that consists of predictors with mixed persistence properties. We operate with a pool of p_1 stationary predictors and $p - p_1 \equiv p_2$ persistent predictors and with no loss of generality take $j = 1, \ldots, p_1$ to index the stationary predictors and $j = p_1 + 1, \ldots, p$ the persistent predictors. We assume two active predictors given by $x_{at} = x_{1t}$ and $x_{bt} = x_{p_1+1,t}$ respectively. Using the results in Lemmas A1, B1 and C1 and standard algebra gives

$$\mathcal{S}_n(j = 1) \xrightarrow{p} \sqrt{1-\pi_0}(\beta_{p_1+1}^*)^2\frac{\sigma_{v_{p_1+1}}^2}{2c_{p_1+1}} \tag{A.37}$$

$$\mathcal{S}_n(j \in \{2, \ldots, p_1\}) \xrightarrow{p} \sqrt{1-\pi_0}(\beta_{p_1+1}^*)^2\frac{\sigma_{v_{p_1+1}}^2}{2c_{p_1+1}} + \sqrt{1-\pi_0}(\beta_1^*)^2E[x_{1t}^2](1-\rho_{1j}^2) \tag{A.38}$$

$$\mathcal{S}_n(j = p_1 + 1) \xrightarrow{p} \sqrt{1-\pi_0}(\beta_1^*)^2E[x_{1t}^2] \tag{A.39}$$

$$\begin{aligned}\mathcal{S}_n(j \in \{p_1 + 2, \ldots, p\}) \xrightarrow{p} &\sqrt{1-\pi_0}(\beta_1^*)^2E[x_{1t}^2] + \sqrt{1-\pi_0}(\beta_{p_1+1}^*)^2\frac{\sigma_{v_{p_1+1}}^2}{2c_{p_1+1}} \\ &\times \left(1 - \frac{(\sigma_{p_1+1,j}/(c_{p_1+1}+c_j))^2}{(\sigma_{p_1+1}^2/2c_{p_1+1})(\sigma_j^2/2c_j)}\right)\end{aligned} \tag{A.40}$$

Comparing (A.37) with (A.38) and (A.39) with (A.40) implies that $\hat{j}_n$ will asymptotically point to either $j = 1$ or $j = p_1 + 1$ (i.e. one of the two true predictors) as stated. ■

B Further Experimental Properties

This section provides additional simulation based results extending the size/power based outcomes presented in Section 7 of the main paper.

B.1 Empirical Size

The supplementary size based simulations use the same DGP as in Section 7 (Table 1) and aim to illustrate the influence of alternative variance normalisers on size (i.e., using the formulation of $\hat{\omega}_n^{2,a}$ in (26) that is based on residuals under the null hypothesis instead of $\hat{\omega}_n^{2,b}$ in (27) that was used in Table 1 of the main paper).

To highlight the role played by the size-neutral power enhancing transformation introduced in Section 5 we also present corresponding empirical size outcomes based on the unadjusted $\mathcal{D}_n(\mu_0)$ statistic. For notational purposes we refer to the power enhanced/size-neutral test statistic evaluated using the variance normalisers $\hat{\omega}_n^{2,a}$ and $\hat{\omega}_{n,j}^{2,b}$ as $\mathcal{D}_n^{d,a}(\mu_0)$ and $\mathcal{D}_n^{d,b}(\mu_0)$ respectively. Their non-enhanced versions are denoted $\mathcal{D}_n^{a}(\mu_0)$ and $\mathcal{D}_n^{b}(\mu_0)$. All our simulations in the main text have been obtained using $\mathcal{D}_n^{d,b}(\mu_0)$ so that in what follows we compare outcomes with $\mathcal{D}_n^{d,a}(\mu_0)$ (adjusted, null residuals), $\mathcal{D}_n^{a}(\mu_0)$ (unadjusted, null residuals) and $\mathcal{D}_n^{b}(\mu_0)$ (unadjusted, residuals under alternative).

Table B1 presents empirical size outcomes based on the power enhanced statistic as in the main text but using a variance normaliser based on the null residuals (i.e., $\hat{\omega}_n^2 = \hat{\omega}_n^{2,a}$) and can be compared with Table 1 in the main text which was based on $\hat{\omega}_n^2 = \hat{\omega}_{n,j}^{2,b}$. We note virtually identical outcomes across all DGP scenarios suggesting that when it comes to the size of our proposed test statistic the use of either $\hat{\omega}_n^{2,a}$ or $\hat{\omega}_{n,j}^{2,b}$ makes little practical difference. All empirical size outcomes match their nominal counterparts very closely and often nearly perfectly regardless of which variance normaliser is used.

Tables B2 and B3 have repeated the same exercise using tha "raw" (unadjusted) version of our test statistic. Although here we continue to note very little difference in outcomes based on either $\hat{\omega}_n^{2,a}$ or $\hat{\omega}_{n,j}^{2,b}$ the main message that comes across these two tables is the importance and effectiveness of our proposed adjustment. The size properties of the unadjusted statistics clearly deteriorates as $\mu_0 \to 0.5$ (i.e., the variance degeneracy bound) despite remaining unaffected by the number of predictors under consideration. Under $p = 500$ for instance empirical size is in the vicinity of 8% for $\mu_0 = 0.35$ but drops to about 3% for $\mu_0 = 0.45$. Our adjusted statistic is not subject to such distortions and shows a remarkably effective ability to align itself with its nominal counterparts regardless of DGP parameterisations. Although our proposed adjustment is solely designed to enhance power it is also clear that it helps maintain good finite sample size, a feature motivated in Remark 4 in the main text.

Table B1: Empirical Size of $\mathcal{D}_n^{d,a}(\mu_0)$ (10% Nominal)

μ_0	p=10	p=50	p=500	p=10	p=50	p=500	p=10	p=50	p=500
		A(i)			A(ii)			A(iii)	
0.35	0.106	0.103	0.103	0.107	0.104	0.104	0.106	0.105	0.104
0.40	0.108	0.104	0.104	0.109	0.106	0.106	0.107	0.106	0.106
0.45	0.110	0.093	0.100	0.115	0.094	0.101	0.115	0.094	0.101
		B(i)			B(ii)			B(iii)	
0.35	0.106	0.104	0.103	0.107	0.103	0.104	0.109	0.104	0.104
0.40	0.115	0.106	0.105	0.115	0.105	0.105	0.115	0.105	0.105
0.45	0.116	0.099	0.102	0.119	0.100	0.104	0.122	0.100	0.104
		C(i)			C(ii)			C(iii)	
0.35	0.107	0.104	0.103	0.106	0.103	0.103	0.107	0.103	0.103
0.40	0.111	0.105	0.104	0.110	0.107	0.104	0.110	0.107	0.104
0.45	0.116	0.093	0.101	0.118	0.098	0.102	0.119	0.099	0.102

Table B2: Empirical Size of $\mathcal{D}_n^a(\mu_0)$ (10% Nominal)

	A(i)			A(ii)			A(iii)		
μ_0	p=10	p=50	p=500	p=10	p=50	p=500	p=10	p=50	p=500
0.35	0.080	0.078	0.076	0.080	0.078	0.076	0.080	0.078	0.076
0.40	0.069	0.063	0.065	0.067	0.062	0.065	0.068	0.062	0.065
0.45	0.035	0.031	0.031	0.036	0.032	0.031	0.036	0.032	0.031
	B(i)			B(ii)			B(iii)		
0.35	0.074	0.073	0.069	0.074	0.073	0.070	0.073	0.073	0.070
0.40	0.061	0.054	0.059	0.059	0.054	0.059	0.060	0.053	0.059
0.45	0.026	0.023	0.024	0.030	0.024	0.024	0.030	0.024	0.024
	C(i)			C(ii)			C(iii)		
0.35	0.078	0.075	0.073	0.077	0.076	0.073	0.077	0.076	0.073
0.40	0.063	0.058	0.061	0.065	0.059	0.061	0.065	0.059	0.061
0.45	0.030	0.027	0.027	0.032	0.027	0.026	0.031	0.027	0.026

B.2 Empirical Power

Our supplementary power based simulations follow the same logical flow as above. They aim to illustrate the role of using an alternative variance normaliser, namely $\hat{\omega}_n^{2,a}$ instead of $\hat{\omega}_n^{2,b}$. As the former is based on the use of residuals under the null we expect potentially important finite sample differences in power performance between $\mathcal{D}_n^{d,b}(\mu_0)$ used in the main text and $\mathcal{D}_n^{d,a}(\mu_0)$ considered here. The simulations that follow also illustrate the important role played by our power-enhancing adjustment by repeating all experiments across adjusted and unadjusted versions of our test statistic. Although the outcomes presented below are based on the same DGPs as in the main text (numbered as (i), (ii-a) and (ii-b)) we also include an additional scenario (referred to as (iii) in what follows) that mixes both stationary and persistent predictors within (30). For this purpose we let $\beta_a^* \in \{(2,3,4,5)\}$, $\beta_b^* \in \{(5,6,7,8)\}$, $\beta_c^* \in \{(2,3,4,5)\}$, $\beta_d^* \in \{(5,6,7,8)\}$ so that the DGP as defined in (30)

Table B3: Empirical Size of $\mathcal{D}_n^b(\mu_0)$ (10% Nominal)

μ_0	p=10	p=50	p=500	p=10	p=50	p=500	p=10	p=50	p=500
		A(i)			A(ii)			A(iii)	
0.35	0.080	0.077	0.075	0.080	0.078	0.076	0.079	0.078	0.076
0.40	0.069	0.062	0.064	0.067	0.062	0.065	0.068	0.062	0.065
0.45	0.035	0.031	0.031	0.036	0.032	0.031	0.037	0.032	0.031
		B(i)			B(ii)			B(iii)	
0.35	0.074	0.072	0.068	0.074	0.072	0.069	0.073	0.072	0.069
0.40	0.060	0.054	0.058	0.059	0.054	0.059	0.059	0.053	0.059
0.45	0.026	0.023	0.024	0.030	0.024	0.025	0.030	0.024	0.024
		C(i)			C(ii)			C(iii)	
0.35	0.077	0.075	0.071	0.076	0.076	0.072	0.076	0.076	0.072
0.40	0.062	0.057	0.061	0.065	0.059	0.061	0.065	0.058	0.061
0.45	0.030	0.027	0.027	0.032	0.027	0.026	0.031	0.027	0.026

includes a total of four predictors, two of which are purely stationary and the remaining two persistent.

Comparing Tables B4-B5 with Table 2 in the main text highlights the unfavourable influence on power of using a variance normaliser based on the null residuals. Take for instance the case of DGP (i) with $\mu_0 = 0.40$ for which we have an empirical power estimate of 69.4% in Table B4 based on $\mathcal{D}_n^{d,a}(\mu_0)$. This can be compared with the estimate of 97.7% obtained using $\mathcal{D}_n^{d,b}(\mu_0)$ (Table 2 of main text). The fact that power is much superior when using $\hat{\omega}_{nj}^{2,b}$ as the variance normaliser and the DGP is driven by purely stationary predictors is a well known phenomenon that has been widely documented in contexts such as change-point detection with cusum type statistics. The mere fact that the variance normaliser takes into account information under the alternative acts as an important power booster.

Interestingly, these power differences arise solely in the context of DGP(i) that is driven solely by purely stationary predictors. If we compare Table B4 with Table 2 for DGPs (ii-a) and (ii-b) which contain solely persistent predictors we note very similar power estimates regardless of whether inferences are based on $\mathcal{D}_n^{d,a}(\mu_0)$ or $\mathcal{D}_n^{d,b}(\mu_0)$.

Tables B5-B6 repeat the above power experiments using unadjusted test statistics. Table B5 is based on $\mathcal{D}_n^{a}(\mu_0)$ that uses null residuals while Table B6 is based on $\mathcal{D}_n^{b}(\mu_0)$ that uses residuals from the larger model. Outcomes strongly support the use of our power enhanced formulation. Comparing Table B6 with Table 2 in the main text for instance and focusing on their last columns we note that the power enhancement boosts empirical powers by more than 20%. A similar picture can also be observed from B7 based on DGP(iii) described above. Comparing outcomes on its mid-panel based on $\mathcal{D}_n^{b}(\mu_0)$ (unadjusted) with its bottom panel based on $\mathcal{D}_n^{d,b}(\mu_0)$ (adjusted) we note substantial spreads for low signal to noise parameterisations. As the signal to noise ratio increases these differences tend to progressively dissipate but remain non-negligible (e.g., 83.5% versus 99.9% under the most favourable signal to noise scenario).

Table B4: Empirical Power of $\mathcal{D}_n^{d,a}(\mu_0)$ under DGPs (i)-(ii)

	DGP (i)			
β_{an}	0.423	0.634	0.846	1.057
β_{bn}	1.057	1.269	1.480	1.692
$\mu_0 = 0.35$	0.487	0.530	0.572	0.589
$\mu_0 = 0.40$	0.694	0.753	0.793	0.819
$\mu_0 = 0.45$	0.977	0.992	0.996	0.998
	DGP (ii-a)			
β_{cn}	0.030	0.045	0.060	0.075
β_{dn}	0.075	0.090	0.106	0.121
$\mu_0 = 0.35$	0.171	0.224	0.292	0.329
$\mu_0 = 0.40$	0.211	0.286	0.368	0.434
$\mu_0 = 0.45$	0.337	0.479	0.611	0.707
	DGP (ii-b)			
β_{cn}	0.075	0.090	0.106	0.121
β_{dn}	0.121	0.136	0.151	0.166
$\mu_0 = 0.35$	0.317	0.372	0.427	0.451
$\mu_0 = 0.40$	0.427	0.491	0.554	0.590
$\mu_0 = 0.45$	0.704	0.779	0.839	0.870

Table B5: Empirical Power of $\mathcal{D}_n^a(\mu_0)$ under DGPs (i)-(ii)

	DGP (i)			
β_{an}	0.423	0.634	0.846	1.057
β_{bn}	1.057	1.269	1.480	1.692
$\mu_0 = 0.35$	0.224	0.241	0.258	0.258
$\mu_0 = 0.40$	0.276	0.299	0.333	0.337
$\mu_0 = 0.45$	0.451	0.520	0.577	0.592
	DGP (ii-a)			
β_{cn}	0.030	0.045	0.060	0.075
β_{dn}	0.075	0.090	0.106	0.121
$\mu_0 = 0.35$	0.093	0.124	0.161	0.177
$\mu_0 = 0.40$	0.089	0.125	0.164	0.192
$\mu_0 = 0.45$	0.070	0.118	0.179	0.231
	DGP (ii-b)			
β_{cn}	0.075	0.090	0.106	0.121
β_{dn}	0.121	0.136	0.151	0.166
$\mu_0 = 0.35$	0.169	0.209	0.243	0.261
$\mu_0 = 0.40$	0.180	0.225	0.265	0.295
$\mu_0 = 0.45$	0.221	0.281	0.336	0.377

Table B6: Empirical Power of $\mathcal{D}_n^b(\mu_0)$ under DGPs (i)-(ii)

	DGP (i)			
β_{an}	0.423	0.634	0.846	1.057
β_{bn}	1.057	1.269	1.480	1.692
$\mu_0 = 0.35$	0.407	0.519	0.613	0.679
$\mu_0 = 0.40$	0.577	0.739	0.834	0.896
$\mu_0 = 0.45$	0.918	0.983	0.997	1.000
	DGP (ii-a)			
β_{cn}	0.030	0.045	0.060	0.075
β_{dn}	0.075	0.090	0.106	0.121
$\mu_0 = 0.35$	0.093	0.128	0.167	0.187
$\mu_0 = 0.40$	0.091	0.128	0.172	0.204
$\mu_0 = 0.45$	0.073	0.128	0.198	0.262
	DGP (ii-b)			
β_{cn}	0.075	0.090	0.106	0.121
β_{dn}	0.121	0.136	0.151	0.166
$\mu_0 = 0.35$	0.178	0.221	0.261	0.286
$\mu_0 = 0.40$	0.195	0.241	0.291	0.323
$\mu_0 = 0.45$	0.252	0.323	0.395	0.445

Table B7: Empirical Power under DGP(iii)

β_{an}	0.423	0.634	0.846	1.057
β_{bn}	1.057	1.269	1.480	1.692
β_{cn}	0.030	0.045	0.060	0.075
β_{dn}	0.075	0.090	0.106	0.121
		$\mathcal{D}_n^a(\mu_0)$		
$\mu_0 = 0.35$	0.219	0.245	0.255	0.270
$\mu_0 = 0.40$	0.270	0.307	0.330	0.355
$\mu_0 = 0.45$	0.464	0.542	0.575	0.607
		$\mathcal{D}_n^b(\mu_0)$		
$\mu_0 = 0.35$	0.382	0.488	0.559	0.605
$\mu_0 = 0.40$	0.554	0.695	0.789	0.835
$\mu_0 = 0.45$	0.905	0.974	0.994	0.997
		$\mathcal{D}_n^{d,a}(\mu_0)$		
$\mu_0 = 0.35$	0.489	0.544	0.566	0.583
$\mu_0 = 0.40$	0.710	0.768	0.795	0.823
$\mu_0 = 0.45$	0.982	0.992	0.996	0.998
		$\mathcal{D}_n^{d,b}(\mu_0)$		
$\mu_0 = 0.35$	0.807	0.913	0.957	0.975
$\mu_0 = 0.40$	0.972	0.995	0.999	0.999
$\mu_0 = 0.45$	1.000	1.000	1.000	1.000