TRR 391 Working Paper #12

June 2026

Factor-based IVX Predictive Regression

Fabian Schmidt[a]

Suggested citation:
Schmidt, F. (2026). “Factor-based IVX Predictive Regression”. *TRR 391 Working Paper #12*. DOI: 10.17877/DE290R-26657.

Version 1.0, June 2026

[a] Department of Statistics, TU Dortmund University
Corresponding author: f.schmidt@statistik.tu-dortmund.de

Factor-based IVX Predictive Regression*

Fabian Schmidt[a]

[a] Department of Statistics, TU Dortmund University

June 9, 2026

Abstract

With the growing availability of financial data, new variables are constantly proposed to predict stock returns, although their incremental explanatory power is often limited because many capture overlapping information. While it suggests itself to extract latent factors summarizing the underlying information — *e.g.* consider common trends in bond yields across maturities — from these variables and to subsequently utilize these factors as predictors, the usual problems with the variables' unknown persistence and predictive regression endogeneity resulting in spurious predictability findings still apply. To address these issues, we combine factor extraction with the IVX framework of Kostakis *et al.* (2015), whose instrumental variable approach is able to resolve the endogeneity issue regardless of the particular degree of persistence. Monte Carlo simulations confirm that the proposed factor-based IVX regression approach achieves good size control and, in addition, strong power should predictability be present. The empirical relevance of the approach is illustrated using S&P 500 returns and a set of commonly used predictors.

Keywords: predictive regression; unknown regressor persistence; endogeneity; factor models.

JEL classifications: C12, C22, G17.

*The author is grateful for the computing time provided by the Institute for Statistics & Econometrics, Kiel University, and the Linux HPC cluster at TU Dortmund University (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as project 271512359. In addition, financial support from the German Research Foundation (DFG) within TRR 391: Spatio-temporal Statistics for the Transition of Energy and Transport (520388526) is gratefully acknowledged. Address correspondence to: Fabian Schmidt, Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany. Email: `f.schmidt@statistik.tu-dortmund.de`.

1 Motivation

An educated guess about future developments is an important aspect of research in many scientific disciplines; be it for maintenance reasons in engineering, to examine future sales potential in marketing, or to predict the effect of economic policy interventions. However, the idea of being able to predict changes in financial asset prices has always been particularly important, *e.g.* in portfolio optimization, causing a tremendous amount of literature in the field; see Fama (1981), Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988a,b), and Fama and French (1988, 1989) for early contributions to the field as well as the review by Goyal *et al.* (2024). In addition, financial crises such as the 2007–2008 global financial crisis and the European debt crisis led to new policies that increased the need to forecast prices of financial instruments and values of assets whose prices are not immediately available, as these are now required for risk management.

At the same time, the number of potential predictors rapidly grows thanks to more and more data becoming available; consider, for instance, the variety of key figures available for financial assets or the number of macroeconomic indicators that are available today. Because of this data abundance, there are far more potential predictors than one can reasonably include in a predictive regression model, *inter alia* due to multicollinearity; for example, bond yields are tightly linked across maturities, with a single yield adding little explanatory power to the model because of its low signal-to-noise ratio. In response, researchers have proposed pre-selecting variables (based on theory or information criteria), or building indices such as the macroeconomic and ESG indices of Dai *et al.* (2025) and Li and Qin (2026), respectively, in advance. Furthermore, combining predictions from several approaches or employing shrinkage techniques as in Zhang *et al.* (2020) has been proposed to better harness the information each available variable contains. However, these approaches do not necessarily exploit all the cross-sectional information in the predictor set, potentially leaving out valuable information.

Meanwhile, one might expect standard ordinary least squares [OLS] to provide an easy, off-the-shelf tool for predictive regression, but with strongly persistent predictors, standard OLS inference has been shown to be unreliable. Applying OLS to strongly persistent predictors whose innovations are contemporaneously correlated with those of the dependent variable (both characteristics are frequently observed in financial predictors) yields test statistics whose distribution does not obey standard OLS asymptotics. As a result, tests can be considerably oversized, resulting in spurious predictability findings; cf. Nelson and Kim (1993) and Stambaugh (1999), for example.

But simply transforming each variable so that standard weakly persistent data asymptotics apply directly, as McCracken and Ng (2016) do for FRED-MD variables, may reduce predictive power for variables with already low signal-to-noise ratios. To resolve the uncertainty about equity premium predictability findings in the face of strongly persistent candidate predictors, alternative testing procedures have been proposed; see, among others, Cavanagh *et al.* (1995), Campbell and Yogo (2006), Jansson and Moreira (2006), and Elliott *et al.* (2015). However, earlier predictability findings tend to be weak, cf. Welch and Goyal (2008), or vanish once robust methods are applied; see Breitung and Demetrescu (2015) among others. Moreover, robustness against predictive regression endogeneity does not come as a free lunch. Often, the relevant kind of persistence has to be known in advance.

One approach that does not require *ex ante* knowledge of the persistence type of predictors is the extended IV regression approach [IVX] of Kostakis *et al.* (2015). Its core idea, based on the work of both Magdalinos and Phillips (2009) and Phillips and Magdalinos (2009), is to utilize an instrumental variable whose persistence is controlled and lower than that of the original variable by construction. Afterwards, this instrument is used within standard instrumental variable [IV] regression. This considerably improves inferential performance compared to standard OLS inference, while remaining straightforward to apply, even in the case of multiple predictors.

Subsequent work has extended Kostakis *et al.* (2015) by considering windows of predictability, relaxing assumptions, or investigating other regression types such as quantile regression; see, among others, Demetrescu *et al.* (2022), Demetrescu *et al.* (2023), and Lee (2016). Meanwhile, Xu and Guo (2024) show that IVX inference performance worsens as the number of putative predictors increases. The higher the number of predictors, the more off is the size of the tests, that is, performing significance tests (individually and jointly) leads to spurious predictability findings when many predictors are considered.

Given the observable predictors' multicollinear nature and their low signal-to-noise ratios, pre-selecting just a few variables might not be ideal. For example, Demetrescu and Hillmann (2022) find strong indications that model selection has a significant detrimental effect on predictability when signal-to-noise ratios are small. Therefore, we suggest resorting to a preliminary dimensional reduction step. The idea is to extract new predictors that condense the information from the many observable variables into a few factors. These factors shall then capture the latent comovement driving the observable variables; *e.g.*, cf. Bai (2004) and the references therein. Recall, for instance, the bond yields and spreads that may be summarizable across different maturities using one common component. Since replacing observable variables with extracted factors improves forecast performance (see Stock and Watson (1998, 1999, 2002a) for early work on diffusion index forecasting or Bai and Ng (2002) and its succeeding literature on factor models), we use extracted factors, with the extraction performed by means of principal component analysis, in place of observable variables in the predictive regression. Hence, the extracted factors are directly employed in the subsequent IVX regression step. Like in Dai *et al.* (2025), who construct a macroeconomic indicator from a long-term trend component of stock returns to use it in their forecast, our goal is to extract latent common trends in financial and macroeconomic variables. These latent common trends then hopefully carry, thanks to reduced noise, stronger predictability signals than the original variables.

By reducing the effective number of variables within the predictive regression, standard IVX estimation and inference becomes feasible again. Hence, the factor model allows to consider more predictors without increasing the effective number of variables in the regression model to an extent that would deteriorate the inferential performance.

As demonstrated by our simulation study, inference performance can be further refined in practice by adapting the residual wild bootstrap of Demetrescu *et al.* (2023), which jointly resamples residuals from the predictive regression model and from a parametric autoregressive model fitted to the predictors, to a factor setup, with the factor-setup extension drawing on ideas from Gonçalves and Perron (2014).

In the remainder, we proceed as follows. In section 2, we specify our model and the assumptions that come with it. Afterwards, we describe the estimation approach and the test statistics, as well as their asymptotic properties, in section 3. Section 4 demonstrates the finite-sample performance by means of Monte Carlo simulations, finding that our factor-based IVX approach provides both correct empirical size under the null and the ability to discover predictability under the alternative; especially when employing the bootstrap algorithm we outline to refine inference in finite samples. Lastly, we apply our factor-based IVX predictive regression approach to predict the S&P 500 equity premium in section 5. After extending the datasets of Welch and Goyal (2008) and Goyal *et al.* (2024) with additional financial and macroeconomic variables from the FRED-MD database of McCracken and Ng (2016), we find that factors can be extracted from data that has not been transformed in advance, that these extracted factors relate to factors obtained from transformed data, nonetheless, and that they can (episodically) predict the equity premium. Finally, section 6 concludes. Proofs of our main theoretical results, accompanying auxiliary findings, and additional results from both the simulation and the empirical study are provided in an online supplementary appendix.

2 Factor-based Predictive Regression Model

In our setup, y_t is predictable using a set of N lagged predictors x_{t-1}. However, as noted above, conceptual similarity, strong multicollinearity, and low signal-to-noise ratios of $\mathbf{x}_t$ complicate the analysis of predictability using the entire set of predictors when N is large, while simultaneously motivating common factors in the predictors. We therefore assume that the predictors admit a low-dimensional latent factor structure and that the predictive information contained in $\mathbf{x}_{t-1}$ is captured by some latent common factors $\mathbf{F}_{t-1}$. Thus,

$$y_t = \mu_y + \mathbf{F}'_{t-1}\boldsymbol{\beta} + u_t, \quad t = 2, \dots, T \tag{1}$$

while for the set of predictors $\mathbf{x}_t$ it holds that

$$\mathbf{x}_t = \mathbf{\Lambda} \cdot \mathbf{F}_t + \mathbf{e}_t, \quad t = 1, \dots, T. \tag{2}$$

In total, there are $r \ll N$ latent factors $\mathbf{F}$ with r fixed and, although (1) permits a nonzero unconditional mean, we henceforth set $\mu_y = 0$ w.l.o.g. because this mean can be eliminated by centering the data *ex ante*.[1]

However, since we only observe the $N \times 1$ vector $\mathbf{x}_t$, we have to estimate the common components $\mathbf{F}_{t-1}$ before their explanatory power regarding y_t can be examined. This immediately gives rise to the idea of principal component regression [PCR]. However, the validity of PCR-based inference then crucially hinges on the predictors' specific type of persistence.

In practice, the persistence of the putative predictors is unknown and, depending on its form, can cause considerable size distortions for tests in endogenous settings. Therefore, we assume

$$\mathbf{F}_t = \mathrm{diag}(\boldsymbol{\rho}) \cdot \mathbf{F}_{t-1} + \mathbf{w}_t, \quad \mathbf{\Psi}(L)\mathbf{w}_t = \mathbf{v}_t, \quad t = 1, \dots, T \tag{3}$$

with the coefficient matrix $\mathrm{diag}(\boldsymbol{\rho})$ governing persistence via the individual ρ_i of each factor. This admits the latent factors to be either strongly or weakly persistent in the following sense:

Assumption 1 *Let* $\mathrm{diag}(\boldsymbol{\rho}) = \mathrm{diag}(\rho_1, \dots, \rho_r) := \left(\mathbf{I}_r - \mathbf{C}_r/T^{\theta-1}\right)$ *where* $\mathbf{C}_r$ *depends on the degree of persistence.*

1. ***Weak persistence****: Let* $\theta = 1$ *and constant* $c_i > 0$ *such that* $\mathbf{C}_r = \mathrm{diag}(c_1, \dots, c_r)$ *yields* ρ_i *being fixed and bounded away from 1, i.e.* $|\rho_i| < 1$ *for* $i = 1, \dots, r$.

2. ***Strong persistence****: Alternatively, let* $\theta = 2$ *and* $c_i \geq 0$ *for* $i = 1, \dots, r$ *in the case of strong persistence.*

Examples of such strongly persistent variables could be the group of dividend- and earnings-related variables that we will consider later. For instance, the dividend yield, the dividend-price ratio, the earnings-price ratio, and the dividend-earnings ratio all display AR(1) coefficients of 0.985 or larger, *i.e.* close to 1. On the other hand, variables like inflation whose estimated AR(1) coefficient is 0.6 can be considered weakly persistent; cf. section 5 below. Finally, it is important to emphasize that we restrict the vector-autoregressive structure in $\mathbf{F}_t$ and its innovations $\mathbf{w}_t$ in such a way that the factors may not be cointegrated.

Remark 1 An alternative functional form of near-integrated processes is sometimes used to model strong persistence; see, for instance, Hansen (1992) and Onatski and Wang (2025) who adopt an exponential representation $\rho_i = \exp\left[-\frac{\phi_i}{T}\right]$. However, an appropriate parameterization allows us to turn their functional form into ours and vice versa; by choosing c_i and

[1] Similarly, unconditional means could be added to equation (2).

$\phi_i = -T \ln\left(1 - \frac{c_i}{T}\right)$, we end up with our definition from Assumption 1. Hence, both approaches can be consolidated allowing for unit roots by picking $c_i, \phi_i = 0$, and local-to-unit roots by picking $c_i, \phi_i > 0$ appropriately. $\square$

To account for the weak predictability signals usually encountered in equity premium prediction, cf. Welch and Goyal (2008) among others, our slope parameters $\boldsymbol{\beta}$ are modeled as local-to-zero. In particular:

Assumption 2 *Let the elements β_i of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)'$ in (1) satisfy $\beta_i := \frac{b_i}{\sqrt{T}}$ for $b_i \in \mathbb{R}$ under Assumption 1.1. Alternatively, let $\beta_i := \frac{b_i}{T^{1/2+\eta/2}}$, where $b_i \in \mathbb{R}$ and $\eta \in (0,1)$ is the tuning parameter employed in IVX instrument generation (cf. equation (6) below), under Assumption 1.2.*

The next crucial feature of our data is the predictive regression endogeneity resulting from contemporaneously correlated innovations driving both the dependent variable and the latent putative predictors. Depending on the persistence of the latter variables, this endogeneity may very well result in spurious predictability findings if one does not employ persistence- and endogeneity-robust inference such as the IVX approach of Kostakis *et al.* (2015), which can accommodate both strongly and weakly persistent predictors in a unified fashion. But it is not just the innovations' correlation that poses a substantial confounding effect to inference. In addition, the data may display various kinds of (un-)conditional heteroskedasticity over time. Therefore, we proceed with the assumptions of Demetrescu *et al.* (2023), imposing assumptions on the innovation vector $(u_t, \mathbf{v}_t)'$ in equations (1) and (3) that allow for various types of heteroskedasticity that are frequently observed in (financial) data. Both u_t and $\mathbf{v}_t$ are serially uncorrelated martingale difference [MD] sequences, but the factor innovations $\mathbf{v}_t$ may enter the model through separate stable finite-order AR(p) processes summarized in $\boldsymbol{\Psi}(L)\mathbf{w}_t = \mathbf{v}_t$ where we denote $\boldsymbol{\Psi}(1)^{-1}$ by $\boldsymbol{\omega}$. In more detail, u_t and $\mathbf{v}_t$ satisfy the following assumptions:

Assumption 3 *Let*

$$\begin{pmatrix} u_t \\ \mathbf{v}_t \end{pmatrix} := \mathbf{H}\left(\frac{t}{T}\right) \begin{pmatrix} \varepsilon_{u,t} \\ \boldsymbol{\varepsilon}_{\mathbf{v},t} \end{pmatrix}.$$

1. *Now, matrix $\mathbf{H}(\cdot) := \begin{pmatrix} h_{11}(\cdot) & \mathbf{h}_{12}(\cdot) \\ \mathbf{h}_{21}(\cdot) & \mathbf{h}_{22}(\cdot) \end{pmatrix}$ consists of piecewise Lipschitz-continuous bounded functions on $(-\infty, 1]$ and is of full rank at all but a finite number of points. Furthermore, $\mathbf{H}(\cdot)$ is defined such that $\Omega_{\mathbf{ww}} = \lim_{T\to\infty} T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathrm{E}\left[\mathbf{w}_t \mathbf{w}_s'\right]$ is positive definite for $\mathbf{w}_t$ arising from stable finite-order AR(p) processes.*

2. *For the vector $(\varepsilon_{u,t}, \boldsymbol{\varepsilon}_{\mathbf{v},t}')'$, we assume that it is a L_4-bounded stationary and ergodic MD sequence satisfying $\mathrm{E}[(\varepsilon_{u,t}, \boldsymbol{\varepsilon}_{\mathbf{v},t}')'(\varepsilon_{u,t}, \boldsymbol{\varepsilon}_{\mathbf{v},t}')] = \mathbf{I}_{r+1}$ where $\mathbf{I}_k$ denotes a $k \times k$ identity matrix.*

Moreover, let

$$\mathrm{E}\left[\left\|\mathrm{E}_0\left[\sum_{t=1}^{T}\left((\varepsilon_{u,t},\boldsymbol{\varepsilon}'_{\mathbf{v},t})'(\varepsilon_{u,t},\boldsymbol{\varepsilon}'_{\mathbf{v},t})-\mathbf{I}_{r+1}\right)\right]\right\|^2\right]=O\left(T^{2\psi}\right)$$

where $\psi < 1/2$ *and where* $\mathrm{E}_0[\cdot]$ *denotes the expectation conditional on the* σ*-algebra generated by* $\{(\varepsilon_{u,t-i},\boldsymbol{\varepsilon}'_{\mathbf{v},t-i})'\}_{i=0}^{\infty}$.

The first part of Assumption 3 allows for various types of unconditional changes in the variance and covariance of u_t and $\mathbf{v}_t$, with the unconditional covariance matrix given by $\mathbf{H}(t/T)\mathbf{H}(t/T)'$. This includes (multiple) shifts, broken trends, and smooth transitions. Furthermore, note that positive definiteness of the average long-run covariance matrix $\Omega_{\mathbf{ww}}$ excludes cointegration among the factors in $\mathbf{F}$. Meanwhile, the martingale difference setup in the second part allows for conditional (stochastic) heteroskedasticity (Demetrescu *et al.*, 2023, p. 6).[2] In summary, our model allows for a wide range of heteroskedastic scenarios of which many are considered to be stylized facts of macroenomic and financial data; see, *inter alia*, Stock and Watson (2002b) and Clark (2009).

Assumptions 1 and 3 have a very important implication, that is, the limit of $\mathbf{F}'\mathbf{F}/T^{\theta}$ for $T, N \to \infty$, which we will need to estimate the factors, depends on the degree of persistence. First,

$$T^{-1}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}_t' \to \boldsymbol{\Sigma}_{\mathbf{F}} \quad \text{as} \quad T\to\infty$$

where the $r \times r$ matrix $\boldsymbol{\Sigma}_{\mathbf{F}}$ is deterministic and positive definite whenever Assumption 1.1 holds and the matrix lag polynomial $\boldsymbol{\Psi}(L)$ is stable; which it is by assumption. Then, $\boldsymbol{\Sigma}_{\mathbf{F}}$ can be found by writing the model in state-space form and solving the system for $\boldsymbol{\Sigma}_{\mathbf{F}}$ with the help of the discrete Lyapunov equation. This is no longer true when Assumption 1.2 holds instead of Assumption 1.1. Now, $T^{-1/2}\mathbf{F}_{\lfloor t/T\rfloor} \Rightarrow \boldsymbol{\omega}\mathbf{J}_{\mathbf{c},\mathbf{H}}(t/T)$, *i.e.* $T^{-1/2}\mathbf{F}_{\lfloor t/T\rfloor}$ weakly converges to a vector of Ornstein-Uhlenbeck processes with $\mathbf{J}_{\mathbf{c},\mathbf{H}}(t/T)$ denoting the vector-wise counterpart of the univariate Ornstein-Uhlenbeck process $J_{c,H}(\tau) := \int_0^{\tau} e^{-c(\tau-s)}\mathrm{d}M_v(s)$ in Demetrescu *et al.* (2023, p. 8); note that $T^{-1/2}\sum_{t=1}^{\lfloor\tau T\rfloor} v_t \Rightarrow M_v(\tau)$ with the limiting process $M_v(\tau)$ being a variance-transformed Brownian motion. As a result, the convergence limit of $\mathbf{F}'\mathbf{F}/T^{\theta}$ is random for local-to-unit root factors and

$$T^{-2}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}_t' \Rightarrow \int_0^1 \boldsymbol{\omega}\mathbf{J}_{\mathbf{c},\mathbf{H}}(s)\mathbf{J}_{\mathbf{c},\mathbf{H}}(s)'\boldsymbol{\omega}'\mathrm{d}s \tag{4}$$

as $T \to \infty$ given Assumption 1.2 with $c_i \geq 0$, $i = 1, \ldots, r$.[3]

---

[2] We refer the interested reader to Demetrescu *et al.* (2022, Remark 4) and Demetrescu *et al.* (2023, Remark 3) for more details regarding the previous assumption and its implications.

[3] Sufficient conditions for the covergence of this expression can be found in Hansen (1992).

Beyond the implications of Assumptions 1 and 3, we require additional assumptions to identify the factors. Letting $\|\mathbf{X}\| = \sqrt{\mathrm{tr}(\mathbf{X}'\mathbf{X})}$ denote the norm of some matrix $\mathbf{X}$ and letting C denote a finite positive constant independent of T and N, we adopt the following standard assumptions of Bai and Ng (2002) and Bai (2004) regarding the factor model:

Assumption 4 *Let*

1. *an initial* $\mathbf{F}_0$ *satisfy* $\mathrm{E}[\mathbf{F}_0] = 0$ *and* $\mathrm{E}[\|\mathbf{F}_0\|^4] \leq C$ *regardless of the degree of persistence.*

2. *Moreover, the factor loadings satisfy* $\|\boldsymbol{\lambda}_i\| \leq C$ *and* $(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})/N \to \boldsymbol{\Sigma}_\Lambda$ *as* $N \to \infty$ *for some deterministic* $r \times r$ *positive definite matrix* $\boldsymbol{\Sigma}_\Lambda$ *irrespective of the degree of persistence.*

Here, $\boldsymbol{\Lambda}$ represents the $N \times r$ loading matrix from equation (2), which is independent of the degree of persistence and converts our r latent factors to the N observable variables.

Now, only the idiosyncratic error vector $\mathbf{e}_t$ is left, whose ith entry we denote by e_{it}. Again, we follow the framework of Bai (2003, 2004), imposing:

Assumption 5 *Let*

1. $\mathrm{E}\left[e_{it}\right] = 0,\ \mathrm{E}\left[e_{it}^8\right] \leq C$.

2. $\mathrm{E}\left[\mathbf{e}_s'\mathbf{e}_t/N\right] = \gamma_N(s,t)$ *where* $|\gamma_N(s,s)| \leq C$ *for all* s*. In addition, let* $\sum_{s=1}^{T} |\gamma_N(s,t)| \leq C$ *for all* $t \leq T$*.*

3. $\mathrm{E}\left[e_{it}e_{jt}\right] = \tau_{ij,t}$ *where* $|\tau_{ij,t}| \leq |\tau_{ij}|$ *holds for all* t *and some finite* τ_{ij}*. Moreover, let* $\sum_{k=1}^{N} |\tau_{ki}| \leq C$ *for all* $i \leq N$*.*

4. $\mathrm{E}\left[e_{it}e_{js}\right] = \tau_{ij,ts}$ *where* $(NT)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |\tau_{ij,ts}| \leq C$*.*

5. $\mathrm{E}\left[\left|N^{-1/2} \sum_{i=1}^{N} \left(e_{is}e_{it} - \mathrm{E}\left[e_{is}e_{it}\right]\right)\right|^4\right] \leq C$ *for every* (t,s)*.*

Motivated by past empirical observations, Assumption 5 permits heteroskedasticity in both the time and the cross-section dimension as well as weak autocorrelation of idiosyncratic errors by Assumption 5.2 and limited cross-correlation among the series by Assumptions 5.3 and 5.4 (Bai, 2004, p. 141). By allowing for weak cross-sectional correlation between the idiosyncratic errors, one does not obtain a strict factor model in the classical sense. This would have required the independence of the idiosyncratic component. Instead, one obtains an approximate factor structure in the sense of Chamberlain and Rothschild (1983). Lastly, it is worth noting that Assumption 5 does not impose explicit distributional assumptions, only moment restrictions.

We exclude integrated idiosyncratic errors to facilitate factor extraction in levels. According to Banerjee *et al.* (2017, Sec. 4.1, p. 1077), this is theoretically justifiable in economic data because I(1) errors would let the otherwise co-moving variables $X_{i,t}, \mathbf{F}_t$ drift apart in the long run, which in turn might contradict, for example, general equilibrium arguments. Furthermore, Banerjee *et al.* empirically support their claim by showing that I(1) idiosyncratic errors occur rarely in the FRED-MD dataset of McCracken and Ng (2016), which we will also use later, and thus are of marginal relevance.

Finally, consistency of the common component estimates will require independence of the loadings and the common and idiosyncratic innovations:

Assumption 6 *$\{\boldsymbol{\lambda}_i\}$, $\{\mathbf{v}_t\}$, and $\{e_{it}\}$ represent three groups of mutually independent (stochastic) variables. Moreover, let u_t be independent of the idiosyncratic errors and loadings.*

3 Factor-based IVX Inference

Given these assumptions, we extend the IVX framework by factor extraction to test for predictability in (1), but only after having reviewed IVX predictive regression without latent factors first.

3.1 Review of IVX Inference

Leaving out the latent factor structure for a moment and assuming that there is only one directly observable candidate predictor simplifies the data-generating process [DGP] in (1)–(3) to:

$$y_t = \mu_y + \beta F_{t-1} + u_t. \tag{5}$$

We are still interested in answering the question whether F_{t-1} is a significant predictor of our dependent variable. Probably the most straightforward approach would be a Wald test applied to an OLS estimate $\hat{\beta}$. If Assumption 1.1 holds, *i.e.* the data is weakly persistent, the resulting test statistic will be asymptotically χ^2_1 under the null if heteroskedasticity-robust standard errors such as Eicker-White standard errors are applied. If, however, Assumption 1.2 would hold, the Wald statistic's limiting distribution depends on the local-to-unity parameter c and is therefore non-standard. With the exception of F_t being strictly exogenous with respect to u_t, this results in spurious predictability findings; cf. Campbell and Yogo (2006) and the references therein.

Based on the results of Phillips and Magdalinos (2009), Kostakis *et al.* (2015) suggest an instrumental variable framework to resolve this endogeneity issue. The idea is to instrument each presumed predictor by an instrument whose degree of persistence can be controlled and that is lower than the original predictor's degree of persistence. Within the IVX framework, such instruments can be generated from the predictor time series themselves. At time t, the instrument is obtained as follows:

$$z_{i,t} = \sum_{j=0}^{t-1} \varrho^j \Delta F_{t-j}, \quad t = 1, \dots, T, \text{ and } z_0 = 0. \tag{6}$$

In (6), $\varrho = 1 - a/T^{\eta}$ ensures that the instrument is less persistent than the original putative predictor, but the specific degree of persistence depends on the tuning parameter choice $a > 0$ and $\eta \in (0, 1)$. This choice is entirely up to the practitioner; but a widely used recommendation, which enforces reduced persistence of the instrument, is that of Kostakis *et al.* (2015) who suggest $a = 1$ and $\eta = 0.95$. As a result, the instrument z_t derived from a strongly persistent predictor F_t satisfying Assumption 1.2 approximately follows a mildly integrated process rather than a nearly integrated one. Alternatively, whenever F_t satisfies Assumption 1.1, that is, F_t is weakly dependent instead of strongly dependent, $z_t \approx F_t$. This results from ϱ being "close" enough to 1, so the differencing is basically undone by cumulating the first differences.

Based on these self-generated instruments, applying standard IV estimation and obtaining the corresponding Eicker-White standard errors allows us to test $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$, where the instrument does not need to be demeaned prior to its use because the slope estimator of β in (5) is invariant to the location of z_t (Kostakis *et al.*, 2015, p. 1514). According to Demetrescu *et al.* (2023), who relaxed the assumptions of Kostakis *et al.* (2015), the resulting Wald test statistic is then χ^2_1-distributed in the limit under the null, irrespective of whether the predictor is weakly (Assumption 1.1) or strongly (Assumption 1.2) persistent.

Remark 2 Kostakis *et al.* (2015, p. 1514 ff.) recommend a finite-sample correction for standard error estimation to mitigate the impact of intercept estimation in finite samples; therefore, we apply it in all subsequent simulations and in the empirical illustration. □

3.2 IVX Inference for Estimated Factors

Now, we return to our original setup. Ultimately, predictability testing is concerned with pairs of hypotheses such as

$$H_0 : \beta_1 = \ldots = \beta_r = 0 \text{ vs. } H_1 : \neg H_0 \quad \text{or} \quad H_0 : \beta_i = 0 \text{ vs. } H_1 : \neg H_0.$$

However, the candidate predictors are not observable. Therefore, we first estimate them, generate the corresponding instruments for use in the predictive IV regression next, and then turn to inference. However, note that the exact number of factors is currently unknown. We will therefore state results in terms of an arbitrary dimension $k < \min(T, N)$ and address how many factors to extract later.

Common factors in the data can be obtained by minimizing the sum of squared residuals

$$V(k) = \min_{\boldsymbol{\Lambda}^k, \mathbf{F}^k} \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(x_{it} - \boldsymbol{\lambda}_i^k \mathbf{F}_t^k \right)^2, \tag{7}$$

resulting from the reconstruction of $\mathbf{x}_t$ using a lower-dimensional representation based on factors, where $\boldsymbol{\lambda}_i$ represents a row of our loading matrix $\boldsymbol{\Lambda}$. But since we have to estimate both $\boldsymbol{\Lambda}$ and

$\mathbf{F}_t$ while having just one restriction so far, we have to impose an additional normalization; either $\mathbf{F}^{k\prime}\mathbf{F}^k/T^\theta = \mathbf{I}_k$ or $\mathbf{\Lambda}^{k\prime}\mathbf{\Lambda}^k/N = \mathbf{I}_k$. After using the former normalization and concentrating $\mathbf{\Lambda}^k$ out of equation (7), the estimate matrix $\tilde{\mathbf{F}}^k$ corresponds to $T^{\theta/2}$ times the eigenvectors belonging to the k largest eigenvalues of the $T \times T$-dimensional outer product $\mathbf{X}\mathbf{X}'$. Afterwards, the corresponding loading matrix can be estimated as $\tilde{\mathbf{\Lambda}}^{k\prime} = \tilde{\mathbf{F}}^{k\prime}\mathbf{X}/T^\theta$. Thus, this problem can be solved by principal component analysis [PCA], and as such, the aforementioned solution is not unique, although the residual sum of squares is. Alternatively, we could use $\left(\bar{\mathbf{F}}^k, \bar{\mathbf{\Lambda}}\right)$ based on the inner product $\mathbf{X}'\mathbf{X}$ instead of the outer product earlier. Then, $\bar{\mathbf{\Lambda}}$ corresponds to $N^{1/2}$ times the eigenvectors that belong to the k largest eigenvalues and $\bar{\mathbf{F}}^k = \mathbf{X}\bar{\mathbf{\Lambda}}^k/N$. This approach is computationally less costly when $N < T$ as it is computationally less demanding to find the eigenvalues of the $N \times N$ matrix $\mathbf{X}'\mathbf{X}$ than of the $T \times T$ matrix $\mathbf{X}\mathbf{X}'$ (and *vice versa*) (Bai, 2004, p. 142).

Remark 3 The aforementioned rotations are not exhaustive. The main issue with any such factor model is that $\mathbf{\Lambda} \cdot \mathbf{F}_t = \left(\mathbf{\Lambda}\tilde{\mathbf{S}}\right) \cdot \left(\tilde{\mathbf{S}}^{-1}\mathbf{F}_t\right)$ holds for any non-singular rotation matrix $\tilde{\mathbf{S}}$. Thus, the model with $\mathbf{\Lambda}$ and $\mathbf{F}_t$ on the left-hand side and the model with $(\mathbf{\Lambda}\tilde{\mathbf{S}})$ and $(\tilde{\mathbf{S}}^{-1}\mathbf{F}_t)$ on the right-hand side are observationally equivalent. For this reason, more restrictive assumptions are imposed from time to time. For example, Stock and Watson (2002a) obtain a more constrained rotation matrix where $\tilde{\mathbf{S}}$ is diagonal with entries ± 1 by imposing additional restrictions on the factor covariance matrix in a setup with stationary factors. However, this reduces the generality of the model, which is why we stick to our more flexible assumptions. □

Subsequently to defining rescaled versions

$$\hat{\mathbf{F}}^k = \bar{\mathbf{F}}^k \left(\bar{\mathbf{F}}^{k\prime}\bar{\mathbf{F}}^k/T^\theta\right)^{1/2} \quad \text{and} \quad \hat{\mathbf{\Lambda}}^k = \bar{\mathbf{\Lambda}}^k \left(\bar{\mathbf{F}}^{k\prime}\bar{\mathbf{F}}^k/T^\theta\right)^{-1/2}$$

of $\left(\bar{\mathbf{F}}_t, \bar{\mathbf{\Lambda}}\right)$, Lemma 1 summarizes consistency results of the common components' PCA estimates given both weakly or strongly persistent predictors. Despite the potentially strongly persistent nature of the latent factors, it is possible to extract uniformly consistent (up to rotation) factor estimates:

Lemma 1 *Let Assumptions 1, 3–6 hold and let* $T/\sqrt{N} \to 0$ *as* $N, T \to \infty$. *Furthermore, let* r *be fixed. Then, there exists a matrix* $\mathbf{S}^k$ *such that*

1. $\max_{1\le t\le T} \left\|\hat{\mathbf{F}}_t - \mathbf{S}^{k\prime}\mathbf{F}_t\right\| = O_p\left(T^{-3/4}\right) + O_p\left(\sqrt{T/N}\right)$ *given Assumption 1.1;*
2. $\max_{1\le t\le T} \left\|\hat{\mathbf{F}}_t - \mathbf{S}^{k\prime}\mathbf{F}_t\right\| = O_p\left(T^{-1}\right) + O_p\left(\sqrt{T/N}\right)$ *given Assumption 1.2.*

The matrix $\mathbf{S}^k$, with its particular form given in the proof of Lemma 1, cf. section S.1 of the supplementary material, exists for any fixed $k \ge 1$, is of dimension $r \times k$, and satisfies $\text{rk}(\mathbf{S}^k) = \min(k, r)$. Hence, our estimated factors turn out to be uniformly consistent estimates of the common components in the data as long as N grows faster than T.

Remark 4 This result crucially depends on the factor number r being fixed. Onatski and Wang (2025, Sec. 5) show that spurious correlation between latent strongly persistent factors may otherwise begin to dominate any genuine cross-sectional structure once r itself tends to infinity. □

Remark 5 The result in Lemma 1 does not require $T/\sqrt{N} \to 0$ as $N, T \to \infty$ since letting $N, T \to \infty$ such that $N \geq C\sqrt{T}$ for some generic constant C already suffices; cf. the proof in section S.1 of the supplementary material. However, since we will ultimately need this rate condition for our main result in Proposition 2 below and it trivially fulfills the condition $N \geq C\sqrt{T}$ asymptotically, we already introduce it here. □

Remark 6 Note further that the result in Lemma 1.2 can also be sharpened by, for example, imposing more restrictive moment conditions on the idiosyncratic components. For instance, this approach could yield $O_p\left(\ln(T)/\sqrt{N}\right)$ instead of $O_p\left(\sqrt{T/N}\right)$ as the second summand for nearly-integrated factors (Bai, 2004, p. 143) □

Given common component estimates, the corresponding IVX instruments are then obtained by applying the filter given in (6) separately to each series $\hat{\mathbf{F}}_i$. Employing the same a, η in $\varrho = 1 - a/T^\eta$ throughout, this yields

$$\hat{z}_{i,t} = \sum_{j=0}^{t-1} \varrho^j \Delta \hat{F}_{i,t-j}, \quad t \in \{1, \ldots, T\},\ i \in \{1, \ldots, k\},\ \text{and } \hat{z}_{i,0} = 0$$

with the resulting instrument estimates being collected in the $k \times 1$ vector $\hat{\mathbf{z}}_t = (\hat{z}_{1,t}, \ \ldots, \ \hat{z}_{k,t})'$. Similarly, instruments $z_{i,t}$ based on the true factors $F_{i,t}$ (if they were hypothetically known) satisfy

$$z_{i,t} = \sum_{j=0}^{t-1} \varrho^j \Delta F_{i,t-j}, \quad t \in \{1, \ldots, T\},\ i \in \{1, \ldots, k\},\ \text{and } z_{i,0} = 0$$

and are pooled in $\mathbf{z}_t = (z_{1,t}, \ldots, z_{k,t})$. The instrument-generating filter recolors the first differences using an exponentially decaying weight. As a result, the same convergence rates that applied to factor estimates also apply to instrument estimates; cf. Lemma S.2 in the supplementary material for a statement regarding $\max_{1 \leq t \leq T} \|\hat{\mathbf{z}}_t - \mathbf{S}^{k'} \mathbf{z}_t\|$ that is similar to Lemma 1.

Now that we have factor estimates as well as the corresponding instruments at our disposal, all necessary ingredients are available to estimate the coefficient vector using standard IV regression. Thus,

$$\hat{\boldsymbol{\beta}}_F = \left(\sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} \hat{\mathbf{F}}'_{t-1} \right)^{-1} \left(\sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} y_t \right) \tag{8}$$

where we define $\boldsymbol{\beta}_F = \left(\mathbf{S}^k\right)^{-1}\boldsymbol{\beta}$ with the matrix $\mathbf{S}^k$ being invertible for $k = r$; how to suitably choose k will be discussed in section 4.1 below. The coefficient estimates are then consistent up to that rotation:

Proposition 1 *Consider the model in* (1)–(3) *with Assumptions 1–6. Moreover, let* $\eta \geq 0.5$ *and let* $T/\sqrt{N} \to 0$ *as* $N, T \to \infty$*. Then, regardless of whether Assumption 1.1 or 1.2 holds,*

$$\hat{\boldsymbol{\beta}}_F \xrightarrow{\text{p}} \left(\mathbf{S}^k\right)^{-1}\boldsymbol{\beta}.$$

Intermediate results regarding the summation terms involved in Propositions 1 and 2 (see below) are provided in Lemmas S.3 and S.4 in the supplementary appendix, section S.1.

Remark 7 Although the coefficient estimates in (8) converge in probability only to a rotated version of the true parameter vector $\boldsymbol{\beta}$ according to Proposition 1, they can nevertheless be used for prediction. This is since predictions utilize the product of factors and coefficients, which is rotation-invariant because the pairs $(\mathbf{F}_t, \boldsymbol{\beta})$ and $\left(\mathbf{S}^{k\prime}\mathbf{F}_t, \boldsymbol{\beta}_F\right)$ are observationally equivalent, *i.e.* $\mathbf{F}_{t-1}'\boldsymbol{\beta} = \mathbf{F}_{t-1}'\mathbf{S}^k\boldsymbol{\beta}_F$. Therefore, utilizing $\hat{\mathbf{F}}_{t-1}'\hat{\boldsymbol{\beta}}_F$ as an estimate of the predictive component in (1) does not adversely affect the prediction, as it still converges to $\mathbf{F}_{t-1}'\boldsymbol{\beta}$ due to the aforementioned rotation invariance. □

Remark 8 The auxiliary results in Lemmas S.3 and S.4 in section S.1 of the supplementary appendix also provide precursors to analyze the limiting distribution of $\hat{\boldsymbol{\beta}}_F$. However, given $\hat{\boldsymbol{\beta}}_F$'s non-straightforward interpretation due to its rotational indeterminacy, we do not pursue this further and focus on inference instead. □

Our main inferential interest is whether the latent factors predict the dependent variable. Because the factors are centered, this leads to the following pair of hypotheses:

$$H_0\colon\ \boldsymbol{\beta} = \mathbf{0}_r \quad \text{vs.} \quad H_1\colon\ \neg H_0. \tag{9}$$

Although the coefficient estimates converge to a rotated version of the true parameter vector due to the rotational indeterminacy of the factor representation, the hypothesis can still be tested because the null is rotation-invariant. If $\boldsymbol{\beta} = \mathbf{0}_r$, then we also have $\boldsymbol{\beta}_F = (\mathbf{S}^k)^{-1}\boldsymbol{\beta} = \mathbf{0}_k$. Hence, testing $H_0 : \boldsymbol{\beta} = \mathbf{0}_r$ is equivalent to testing $H_0 : \boldsymbol{\beta}_F = \mathbf{0}_k$, and the rotation does not affect the test; see the proof in section S.1. The resulting IVX Wald test statistic thus reads as

$$\mathcal{W}^{EW} := \hat{\boldsymbol{\beta}}_F'\left(\widehat{\text{Cov}}\left(\hat{\boldsymbol{\beta}}_F\right)\right)^{-1}\hat{\boldsymbol{\beta}}_F,$$

indicating a rejection of the null hypothesis for "large" values. Employing the Eicker-White sandwich estimator $\widehat{\text{Cov}}\left(\hat{\boldsymbol{\beta}}_F\right) = \boldsymbol{A}_T^{-1}\boldsymbol{D}_T(\boldsymbol{A}_T')^{-1}$, where $\boldsymbol{A}_T = \sum_{t=2}^T \hat{\mathbf{z}}_{t-1}\hat{\mathbf{F}}_{t-1}'$, $\boldsymbol{B}_T = \sum_{t=2}^T \hat{\mathbf{z}}_{t-1}y_t$, and $\boldsymbol{D}_T = \sum_{t=2}^T \hat{\mathbf{z}}_{t-1}\hat{\mathbf{z}}_{t-1}'\hat{u}_t^2$, then yields $\mathcal{W}^{EW} = \boldsymbol{B}_T'\boldsymbol{D}_T^{-1}\boldsymbol{B}_T$ for a joint test of the hypothesis in (9).

Now, let the null hypothesis be true, so that $y_t = u_t$. In this case, the residuals can be estimated using $\hat{u}_t = y_t - \bar{y}$ given either kind of predictor persistence in Assumption 1. Thanks to the Eicker-White standard errors the test statistic $\mathcal{W}^{EW}$ then obeys the following standard limiting result:

Proposition 2 *Consider the model in* (1)–(3) *given the null hypothesis* H_0 *of no predictability in equation* (9). *Moreover, let Assumptions 1–6 hold and let* $T/\sqrt{N} \to 0$ *as* $N, T \to \infty$. *Then,*

$$\mathcal{W}^{EW} \xrightarrow{\text{d}} \chi^2_k$$

regardless of whether Assumption 1.1 or 1.2 holds.

Finally, it might also be of interest in practice to test a particular *extracted* factor $\hat{F}_i$ if it was meaningfully interpretable. However, note that such a test does not test predictability of a true latent factor since they are only identified up to rotation during extraction. Instead,

$$H_0\colon\ [\boldsymbol{\beta}_F]_i = \left[\left(\mathbf{S}^k\right)^{-1}\boldsymbol{\beta}\right]_i = 0 \quad \text{vs.} \quad H_1\colon \neg H_0, \quad i \in \{1, \ldots, k\}. \tag{10}$$

Based on this, testing whether the ith extracted factor has predictive power with respect to the dependent variable is straightforward via a t-type test

$$\left(t_i^{EW}\right)^2 := \frac{\left(\left[\hat{\boldsymbol{\beta}}_F\right]_i\right)^2}{\hat{\text{se}}^2\left(\left[\hat{\boldsymbol{\beta}}_F\right]_i\right)}.$$

Corollary 1 *Consider the model in* (1)–(3) *given the null hypothesis* H_0 *of no predictability in equation* (10). *Moreover, let Assumptions 1–6 hold and let* $\eta \geq 0.5$. *Then, regardless of whether Assumption 1.1 or 1.2 holds,*

$$\left(t_i^{EW}\right)^2 \xrightarrow{\text{d}} \chi^2_1$$

as $T/\sqrt{N} \to 0$ *for* $N, T \to \infty$.

4 Performance in Finite Samples

Following our formal discussion, we now examine the finite-sample performance of our proposed factor-based IVX regression framework, which we henceforth abbreviate by *IPA* as shorthand for *IVX-PCR (Asymptotics)*, by means of Monte Carlo simulations, evaluating the performance of our approach under both the null and alternative.

4.1 Implementation Remarks

But before that, we address several outstanding issues, starting with the question of how many factors to extract. Various estimators have been proposed to answer this question. For example, Bai and Ng (2002) propose information criteria built around penalty functions, while Onatski (2010) and Kapetanios (2010) focus instead on the eigenvalues of the estimated covariance matrix of the observables. We base our factor number choice on the latter idea using the *GR* criterion of Ahn and Horenstein (2013). *GR* detects the threshold between the eigenvalues driven by the factors and those driven by the idiosyncratic errors, which vanish asymptotically, by utilizing the extracted eigenvalues and relating them to the growth rates of residual variances when fewer principal components are included. Preliminary simulations showed that both the *GR* criterion and Ahn and Horenstein (2013)'s alternative *ER* criterion, which picks the factor number based on the ratio of two adjacent eigenvalues, perform satisfactorily irrespective of the degree of persistence. Furthermore, they outperformed the criteria of Bai and Ng (2002) in said simulations regardless of whether there was a factor structure present in the data or the predictors' type of persistence.

We try to further improve the reliability of our factor number selection following Gonzalo and Pitarakis (2021). Studying PCA for high-dimensional, highly persistent data, they argue that spurious detections are less frequent when using differenced data, even if some stationary variables are over-differenced. They conclude that reducing false findings outweighs the costs of over-differencing, a result corroborated by our preliminary simulations with respect to factor-number estimation.

Remark 9 This approach is further justified by the arguments of Morico and Stauskas (2025, p. 21 f.) in the context of their cross-sectional average-based factor extraction, as well as by the theoretical discussion and simulation results in Corona *et al.* (2017). Corona *et al.* analyze factor-number selection criteria (among them the *ER* and *GR* criteria proposed by Ahn and Horenstein (2013)) in a setting with non-stationary data and preparatory stationarity-inducing transformations of the data, finding that *GR* nevertheless performs reliable. □

Moreover, our Monte Carlo results in section 4.3 below document that applying asymptotic-based inference to a finite sample is only approximately accurate. As shown previously, IVX inference in finite samples can be improved using resampling techniques; see, for example, Demetrescu *et al.* (2022) and Demetrescu *et al.* (2023). This applies in particular to our factor-based IVX setup, since the preliminary factor estimation introduces another source of estimation uncertainty that may influence the performance in finite samples. For example, Gonçalves and Perron (2014) show that the finite sample performance of asymptotic approximations may be poor in factor-augmented regression setups if N was not yet large enough relative to T. As a way out, they advocate relying on bootstrap inference instead to improve reliability.

For this reason, we propose adding the factor estimation step as an additional layer to the residual wild bootstrap of Demetrescu *et al.* (2023). Inspired by the work of Gonçalves and Perron (2014), our bootstrap algorithm proceeds as follows:

Residual Wild Bootstrap

Step 1: Obtain residual estimates $\hat{u}_t$, $t = 2, \ldots, T$, using $\hat{\mathbf{F}}$ and $\hat{\boldsymbol{\beta}}_F$.

Step 2: Next, estimate the predictors' innovations $\mathbf{w}_t$, $t = 1, \ldots, T$, by obtaining $\hat{\boldsymbol{\rho}}_i$ using separate autoregressions (of order p) applied to $\hat{F}_{i,t}$, $i = 1, \ldots, k$, and then combine these estimates into $\hat{\mathbf{w}}_t$. As a starting condition, set $\mathbf{w}_1 = \mathbf{0}$ and let $\mathbf{w}_{t,2}, \ldots, \mathbf{w}_{t,p}$ be the residuals from the respective AR$(\cdot)$ models.

Step 3: Estimate the idiosyncratic errors $\hat{\mathbf{e}}_t$, $t = 1, \ldots, T$ via

$$\hat{\mathbf{e}}_t = \mathbf{x}_t - \hat{\boldsymbol{\Lambda}} \cdot \hat{\mathbf{F}}_t$$

Step 4: Next, create bootstrap innovations for $t = 1, \ldots, T$, starting with

$$u_t^* = \hat{u}_t \cdot \kappa_t$$
$$\mathbf{w}_t^* = \hat{\mathbf{w}}_t \cdot \kappa_t$$

where it is crucial to multiply both $\hat{u}_t$ and $\hat{\mathbf{w}}_t$ by the same random variable κ_t which is scalar, independent of the sample, $iid(0,1)$ and satisfies $\mathrm{E}\left[\kappa_t^4\right] < \infty$. Otherwise, the contemporaneous correlation might be altered or removed altogether so that the bootstrap would no longer account for this confounding effect. Afterwards, obtain

$$e_{j,t}^* = \hat{e}_{j,t} \cdot \kappa_{j,t}, \quad t = 1, \ldots, T, \; j = 1, \ldots, N.$$

where $\kappa_{j,t}$ is again $iid(0,1)$ across (j,t) and independent of κ_t used on $(\hat{u}_t, \hat{\mathbf{w}}_t')'$.

Step 5: Now, generate the bootstrap sample from the parameter estimates and the bootstrap innovations. Start from $\mathbf{F}_1^* = \mathbf{w}_1^*$ and recursively obtain

$$F_{i,t}^* = \sum_{j=1}^{p} \hat{\rho}_{i,j} F_{i,t-j}^* + w_{i,t}^*, \quad t = 2, \ldots, T, \; i = 1, \ldots, r,$$

implicitly assuming $\mathbf{F}_0, \mathbf{F}_{-1}, \ldots, \mathbf{F}_{-p+1} = \mathbf{0}$. To complete the bootstrap sample, define

$$\mathbf{x}_t^* = \hat{\boldsymbol{\Lambda}} \cdot \mathbf{F}_t^* + \mathbf{e}_t^*, \quad t = 1, \ldots, T$$
$$y_t^* = u_t^*, \quad t = 1, \ldots, T$$

since the test statistics shall be resampled under H_0.

Step 6: Subsequently, apply the estimation and testing procedure, namely factor extraction, instrument generation (re-using our earlier η), parameter estimation, and testing to the resulting bootstrap sample consisting of T observations $(y_t^*, (\mathbf{F}_t^*)', (\mathbf{x}_t^*)')'$.

Step 7: Repeat steps 1–6 B times. Afterwards, obtain the respective bootstrap test statistic's quantiles under H_0 from the empirical quantile function applied to the B bootstrap statistics. Lastly, compare the test statistics $\mathcal{W}^{EW}$ and $t^{EW} = \hat{\beta}_i / \hat{se}_{\beta_i}$ to the bootstrapped quantiles $\mathcal{W}^*_{1-\alpha}$ and $t^*_{j,\alpha/2}$, $t^*_{j,1-\alpha/2}$, respectively.

Choosing the order of the AR(p) processes and a distribution for $\boldsymbol{\kappa}$ then completes the bootstrap. In what follows, we refer to the resulting procedure as *IPR*, short for *IVX-PCR (Residual Wild Bootstrap)*. When selecting p, remember that the AR models are fitted to estimated rather than directly observable variables, so estimation uncertainty must be taken into account. Therefore, we resort to the BIC information criterion of Groen and Kapetanios (2013) who propose an extension of the standard BIC (and HQ) criterion that is supposed to account for the estimation error resulting from factor extraction. It remains to choose a distribution for κ_t. In line with the above requirements, we select the Rademacher distribution, though any distribution satisfying the moment restrictions mentioned in Step 4 would also work.

Finally, note that we normalize the observed data before extracting the factors as is commonly done in the literature.[4]

4.2 Simulation Setup

For all upcoming experiments, we generate our data according to the DGP given by equations (1)-(3) where $\mu_y = 0$ w.l.o.g. Now, the two key characteristics of our data will be predictive regression endogeneity and uncertain persistence of the predictors. Regarding predictive regression endogeneity, it has been established in the literature that the predictor's innovations can be strongly negatively correlated with those of the dependent variable; two examples include, *inter alia*, the dividend–price ratio and the book–to–market ratio (Demetrescu *et al.*, 2026, Suppl. Appx. Table S.2). Therefore, $(u_t, \mathbf{v}_t)'$ is drawn from an *i.i.d.* multivariate Normal distribution which is parameterized as follows

$$\begin{pmatrix} u_t \\ \mathbf{v}_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.9 & 0 & \dots \\ -0.9 & 1 & 0 & \dots \\ 0 & 0 & 1 & \ddots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}\right). \tag{11}$$

[4] As part of the normalization, the shrinkage covariance matrix estimator of Ledoit and Wolf (2004) is used to stabilize the estimate which in turn is supposed to improve reliability of subsequent factor extraction.

Due to correlated u_t and $v_{1,t}$, there is predictive regression endogeneity in the model that, depending on the type of persistence, may adversely affect inference. In the first step, all factors are equally persistent and we consider $\rho_i \in \{1, 0.995, 0.95, 0.9\}$, $i = 1, \dots, r$. Thus, factors can be I(1) ($\rho_i = 1$), local-to-unity ($\rho_i = 0.995$), or stationary (*e.g.* $\rho_i = 0.9$).

Choosing a loading matrix $\mathbf{\Lambda}$ to link factors and observable variables is non-trivial in our simulations. In addition to Wald tests that assess all extracted factors jointly, we also examine t-tests applied to individual extracted factors. Because factors are identified only up to rotation and are subject to sampling variation, the first extracted factor in Monte Carlo sample 1 does not need to match the first one in samples 2, 3, 4, etc. This is because, first, different samples can lead to different rotations, and, second, even with the same factors, their order can change when factor variances are of similar magnitude. Thus, even if the test result was correct for each factor, sampling variation may still reorder them, so a factor that truly predicts the dependent variable might appear as the first, second, etc. estimated factor. In practice, this is not an issue because we obtain the test results just once. Hence, it does not matter whether the first or second factor is the statistically significant one in a real-world application. However, when we simulate the same DGP multiple times in our Monte Carlo simulation and store the results of the t-test, the predictability signal should appear in the same position. Otherwise, all t-test power curves would wrongly suggest that every extracted factor is a predictor, simply because the true predictor appears as the first, second, third, *etc.* factor across different Monte Carlo samples.

For this reason, our loading matrix is patterned so that each observable variable mainly loads on one factor with a share of N/r variables per factor:

$$\mathbf{\Lambda} = \begin{pmatrix} 0.8 & \lambda_{1,2} & \lambda_{1,3} & \dots \\ \vdots & \lambda_{2,2} & \lambda_{2,3} & \dots \\ 0.8 & \lambda_{3,2} & \lambda_{3,3} & \dots \\ \lambda_{4,1} & 0.8 & \lambda_{4,3} & \dots \\ \lambda_{5,1} & \vdots & \lambda_{5,3} & \dots \\ \lambda_{6,1} & 0.8 & \lambda_{6,3} & \dots \\ \lambda_{7,1} & \lambda_{7,2} & 0.8 & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{12}$$

where $\sum_{j=1}^{r} \mathbf{\Lambda}_{i,j} = 1, \ i = 1, \dots, N$. Hence, each observable variable mainly loads on one particular factor, but we also introduce some variation to the loadings by letting the variables load a little bit on the remaining factors, too. We do this by drawing the remaining $\lambda_{i,j}$ *i.i.d.* half-normally distributed and subsequently standardizing them so that the weights in each row add up to 1, *i.e.* $\sum_{j=1}^{r} \mathbf{\Lambda}_{i,j} = 1$ for all i. This enables the application of an Equamax rotation to the estimated loading matrix and to sort the estimated factors to match $\mathbf{\Lambda}$ in the sense that the first few observable variables load the most on the first factor, the next couple of variables load the most on the second, etc. This lets us visualize the inferential performance of testing individual extracted factors through meaningful t-test power curves.

Our simulation study examines both H_0 and H_1 by setting $\beta_1 \in \{0, 0.02, 0.04, \ldots, 0.1\}$ and keeping it fixed over the sample, while the remaining coefficients, $\beta_2 = \ldots = \beta_r$, are always 0. Thereby, power curves of the factors that predict the dependent variable can be compared with those of the factors that are not informative with respect to y_t.

Since the performance of *IPA* and *IPR* cannot be assessed in isolation, we will consider the following competitors. First, we apply standard Principal Component Regression (henceforth *PCR*) using Eicker-White standard errors and standard OLS asymptotics. Next, we adopt the factor-based predictive regression approach suggested by Beckmann *et al.* (2025). Building on the PANIC methodology of Bai and Ng (2004), they propose to extract the factors from data that has been first-differenced so that stationarity of the data is ensured. Once an appropriate number of factors has been extracted from the differenced data (here, we also use the *GR* criterion), the obtained factors are recumulated over time. Hence, the estimated predictors are integrated back to the observed variables' original order of integration in some sense, and the differencing step is undone before using the estimated variables as predictors. Afterwards, these estimated predictors are employed within standard IVX predictive regression.

Remark 10 The approach of Beckmann *et al.* (2025) aligns with the reasoning of Gonzalo and Pitarakis (2021), who argue that the risk of extracting spurious components from highly persistent data outweighs the cost of overdifferencing some I(0) variables. They therefore recommend first taking differences before applying PCA when faced with unknown persistence. □

However, note that we deviate from the original proposal of Beckmann *et al.* (2025) by not considering dynamic factors but instead using PCA. We name this method *IVX-PCR (Differences Integrated)* and abbreviate it by *IPDI* subsequently. Lastly, the proposal of Beckmann *et al.* (2025) motivates another competitor that is standard in the literature. Once the factors are extracted from the differenced data, they are directly applied as predictors instead of recumulating them beforehand. This should give an idea how stationarity-inducing transformations affect the predictive performance of the data. In line with our previously introduced shorthand notation, *IPDI*, we call this method *IVX-PCR (Differences)* and abbreviate it by *IPD*.

In particular, we subject *IPA*, *IPR* and the competing methods *IPD*, *IPDI*, and *PCR* to the following DGPs:

> **DGP1 (baseline model)**: As our baseline, we consider homoskedastic innovations from (11), *i.e.* the factor innovation vector satisfies $\mathbf{w}_t = \mathbf{v}_t$. Predictive regression endogeneity can then be considerable for highly persistent predictors due to u_t and v_{1t} being strongly negatively correlated. In DGP 1, we simulate $T \in \{250, 500\}$ observations of $r = 4$ latent factors and $N \in \{200, 400\}$ observable variables (50

or 100 per factor), linked to the factors through simulated loading matrices $\mathbf{\Lambda}$ as described in (12).

DGP2 (mixed degrees of persistence): In DGP 1, all factors were equally persistent, which is unrealistic in practice. We therefore rerun the simulations using the DGP 1 parameterization but assign the four latent factors distinct autoregressive parameters from $\{1, 0.995, 0.95, 0.9\}$. The first factor always has predictive power (if any) for the dependent variable, and we circularly shift the vector of autoregressive parameters so that each ρ value appears once in the predictive factor: starting with $\boldsymbol{\rho} = (1, 0.995, 0.95, 0.9)$, then $\boldsymbol{\rho} = (0.995, 0.5, 0.9, 1)$, *etc.* This relaxes the equal-persistence assumption and captures settings where some factors are strongly persistent while others are weakly persistent. Although we do not formalize this, we conjecture that mixtures of strongly and weakly persistent factors can be handled in practice by treating the latter as strongly persistent with a "large" c. Moreover, we restrict ourselves to $T = 250, N = 200$, to limit the computational effort, henceforth.[5]

DGP3 & DGP 4 (unconditional volatility breaks): One presumed key advantage of *IPA* and *IPR* is their robustness to (un-)conditional volatility changes. DGP 3, which is otherwise similar to DGP 2, investigates unconditional upward volatility shifts: midway through the sample, the residual variance σ_u^2 and the innovation variances σ_v^2 each quadruple. These shifts are likely larger than what we usually see in practice, especially with monthly data, but they let us test whether our approach is robust to changes of this magnitude, and thus, by implication, to smaller variance changes as well. Afterwards, DGP 4 examines a shift in the opposite direction, *i.e.* one where the volatility suddenly decreases from $\sigma_{u,t}^2 = \sigma_{v,t}^2 = 4$, $t \leq 0.5T$, to $\sigma_{u,t}^2 = \sigma_{v,t}^2 = 1$, $t > 0.5T$, halfway through the sample.

DGP 5 (conditional heteroskedasticity): In practice, volatility does not only change unconditionally. Conditional heteroskedasticity and volatility clustering are well-known stylized facts of financial data, too. Following Demetrescu *et al.* (2023), we maintain the (innovations' unconditional covariance) setup from DGP 2 and then introduce conditional heteroskedasticity in u_t and $\mathbf{v}_t$ by modeling their conditional variances using separate GARCH(1,1) processes. Hence, the innovations u_t and $\mathbf{v}_t$ are rescaled using standard deviations based on

$$\sigma_{\cdot,t}^2 = (1 - \alpha - \beta) + \alpha\varepsilon_{\cdot,t-1}^2 + \beta\sigma_{\cdot,t-1}^2$$

[5] Bootstrapping critical values is computationally intensive. For DGP 1 with $(T = 250, N = 200)$, about 400 cores of a HPC cluster needed roughly one hour, whereas for DGP 1 with $(T = 500, N = 400)$ it took the same setup nearly five hours. Since the performance is qualitatively similar across small and large samples (as shown below), we henceforth focus on the parameterization $T = 250, N = 200$.

where, for $i = 1, \ldots, r$, $\mathrm{E}[u_t^2] = \mathrm{E}\left[v_{i,t}^2\right] = \frac{(1-\alpha-\beta)}{1-\alpha-\beta} = 1$ if $\alpha, \beta \geq 0$ and $\alpha + \beta < 1$. Setting $\alpha = 0.1, \beta = 0.85$ then yields stationary GARCH(1,1) processes where $\mathrm{E}[u_t^2] = \mathrm{E}\left[v_{i,t}^2\right] = 1$.

DGP 6 (predictive regression endogeneity outside the predictor): DGP 6 sets $\mathrm{Corr}(u, v_1) = 0$ and $\mathrm{Corr}(u, v_2) = -0.9$, but otherwise maintains the setup of DGP 2. So far, we only assumed that the innovations of the factor predicting the dependent variable are correlated with the innovations of the dependent variable. However, in reality, u_t can be correlated with any factor's innovations $v_{i,t}$ which is why we investigate this setup to unravel the effect of predictive regression endogeneity on the power of our tests

DGP 7 (correlated factor innovations): Moreover, u_t need not be correlated with any factor innovation $v_{i,t}$. The factor innovations themselves, *e.g.* $v_{1,t}$ and $v_{2,t}$, may be correlated. In that case, predictive regression endogeneity disappears and most approaches should be correctly sized. However, correlation between factor innovations may complicate factor (number) extraction. DGP 7 addresses this by $\mathrm{Corr}(u, v_1) = 0$, $\mathrm{Corr}(v_1, v_2) = 0.5$, while otherwise retaining the DGP 2 setup.

DGP 8 (pattern-free loading matrix): In DGP 8, we drop the patterned loading structure and allow variables to load freely on all factors while keeping the parameterization from DGP 2 other than that.

By default, our tests use a 5 % significance level, we generate $N_{MC} = 2000$ Monte Carlo samples and use $B = 500$ bootstrap samples in each simulation. The data is generated using MATLAB 2024a and its Threefry 4×64 random number generator.

4.3 Discussion of Results

We start by discussing key findings from our baseline scenario DGP 1 while referring the interested reader to the complete DGP 1 results in Tables 5 and 6 ($T = 250$), and in Tables S.1 and S.2 in the supplementary material ($T = 500$).

As it turns out, *GR* reliably detects the true number of factors regardless of the degree of persistence throughout all T–N combinations; cf. the average numbers of extracted factors in the aforementioned tables.

Turning to the inferential performance of our tests, we observe that the empirical size of Wald tests examining the joint null in (9) using *IPR*, *i.e.* bootstrap-based inference, is not distorted. To be more precise, *IPR*'s empirical size is within reasonable limits ranging from 3.5 to 7.75 % given the theoretical size of 5 %. Furthermore, *IPR* exhibits considerable power; see Figure 1, which depicts the Wald test power curves for $T = 250$ and $\rho_i \in \{1, 0.995, 0.95, 0.9\}$ that increase monotonically in β_1. Only in panel 1 does *IPR* exhibit moderate oversizedness at 7.75 %, which

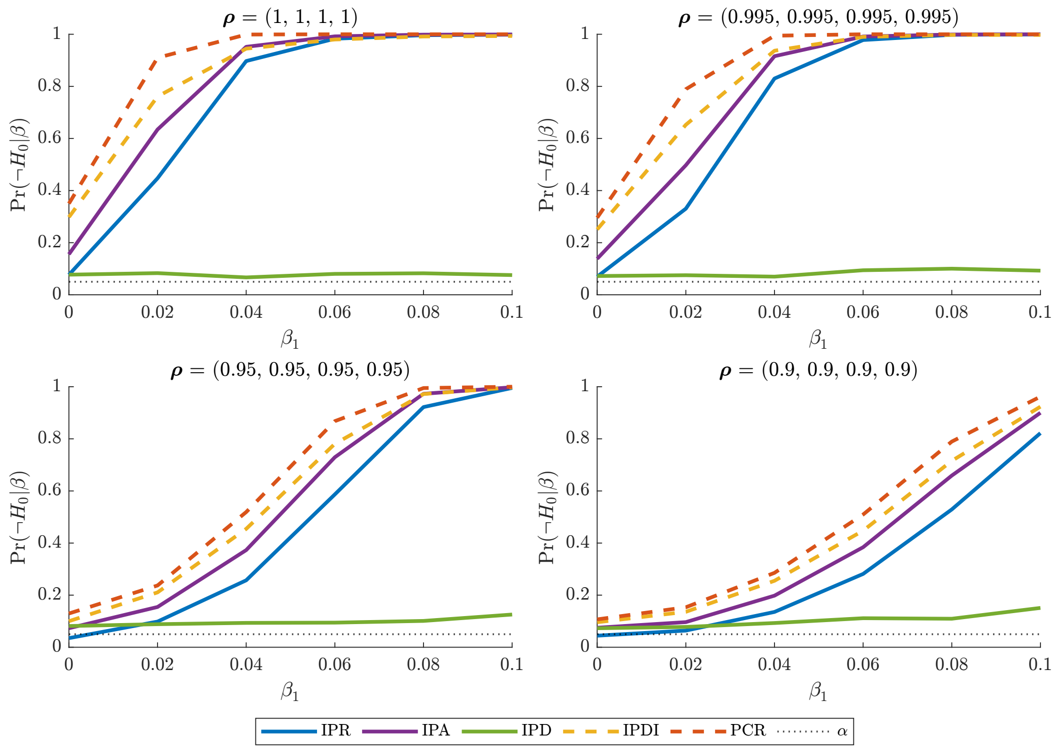

Figure 1: Wald test power curves of jointly testing the extracted factors at 5% significance level. The panels are separated by the autoregressive parameter value ρ and the data is generated from DGP 1, $T = 250$, $N = 200$.

is not surprising because all latent factors are I(1) there, making size control nontrivial, so some oversizedness was expected.

Most other approaches are clearly oversized, especially as factor persistence increases, with the least robust method, *PCR*, performing worst. Notably, *IPA* performs relatively well, second-best overall and markedly better than *PCR*, but it remains oversized, indicating that asymptotic critical values are unreliable in finite samples when predictors are strongly persistent. Only in the situation where all factors are more or less weakly persistent, see, for example, the lower right panel of Figure 1, *IPA*s empirical size is within acceptable levels, but so is that of the other methods. As ρ_i decreases, *i.e.* as the factors become less persistent, the more predictive regression endogeneity diminishes, resulting in smaller size distortions of non-robust approaches. This can be observed regardless of whether the test is bootstrap-based, asymptotics-based, or differences-based. Even *PCR* would have been acceptable if we had known in advance that the predictors were only weakly persistent, since inference would not suffer from predictive regression endogeneity and *PCR* would then offer the highest power.

Meanwhile, both approaches where the factors, not just their number, are obtained from differenced data perform poorly. First, observe that *IPD* exhibits little to no power regardless of predictor persistence. Furthermore, observe that *IPDI* exhibits severe oversizedness when

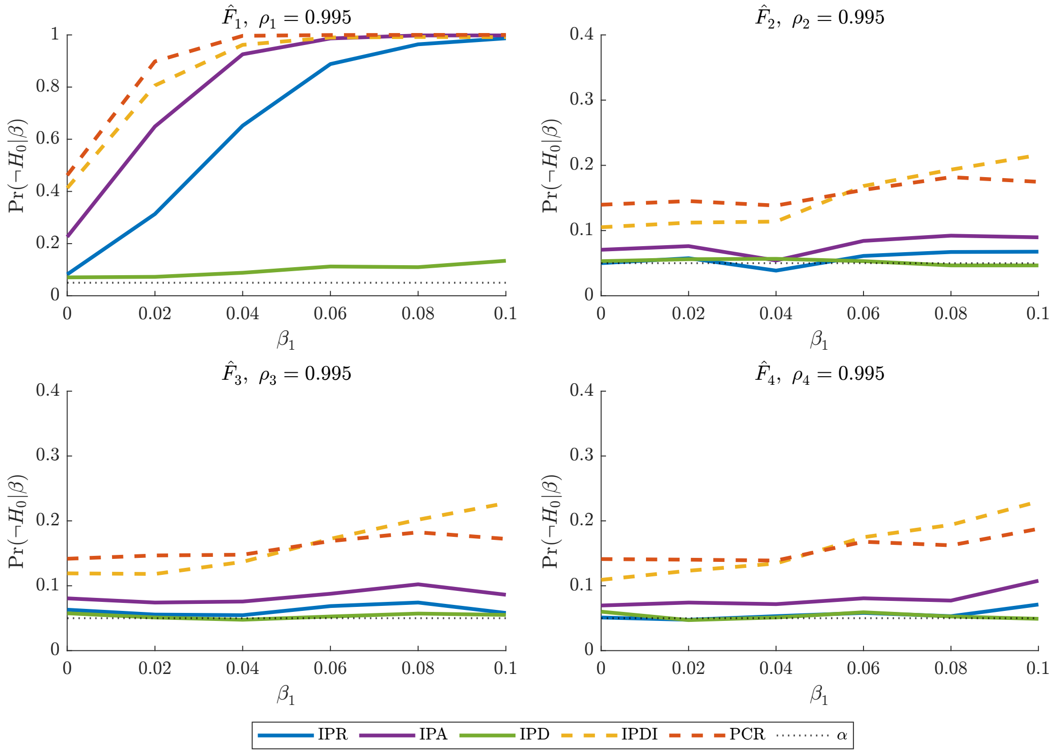

Figure 2: t-test power curves of individually testing the extracted factors (two-sided) at 5% significance level. Data generated from DGP 1; $T = 250$, $N = 200$, and $\rho_1 = \ldots = \rho_4 = 0.995$.

the predictors are strongly persistent. Only as ρ decreases and all presumed predictors become weakly persistent, the oversizedness of *IPDI* vanishes.

Next, we consider individual two-sided tests for each extracted factor, since practitioners care not only whether all factors are informative, but also which ones are, especially when they become interpretable after a suitable rotation. To get an idea, we discuss exemplary results for $T = 250$ and $\rho_i = 0.995$, $i = 1, \ldots, 4$, with data coming from DGP 1. The other results belonging to $\rho_i \in \{1, 0.95, 0.9\}$ can be found in Table 6 and do not differ qualitatively. Similarly to the previous Wald test results, IVX combined with our residual wild bootstrap performs best, while standard PCR performs worst in terms of size, as can also be seen in Figure 2. In this figure, we plot the size and power of our extracted factors as β_1 varies while $\beta_2 = \beta_3 = \beta_4 = 0$. We find that using *IPR* to test the first extracted factor yields an empirical size of approximately 5 %, and we also observe that the empirical power increases to 1 as β_1 increases. Hence, *IPR* correctly identifies the first extracted factor as a statistically significant predictor of our dependent variable y_t, while exhibiting sizes between 5 and 6.5 % throughout the other extracted factors, *i.e.* correctly identifying them as insignificant.

Not surprisingly, *IPA* performs second best again. The power curves resemble those of *IPR* with a slight upward shift. However, *IPA*'s oversizedness reiterates the need for a proper finite-sample adjustment. Moreover, we observe that the empirical size of *IPR* and *IPA* increases for factors 2–4 once β_1 becomes large and $T = 500$; empirical t-test sizes range from 5 to 10%.

However, additional simulations suggest that these distortions are due to the fact that the true factors are not always recovered. When the loading structure is simplified so that each variable loads on exactly one latent factor, the oversizedness disappears.

At the individual factor level, the differences-based approaches remain mostly oversized or uninformative. As shown in Figure 2, *IPD* is correctly sized and even exhibits some power when $\beta_1 = 0.1$, indicating at least some power at the individual factor level with the power increasing as ρ decreases and T increases; cf. Tables 6 and S.2, where *IPD* reaches its highest power at $T = 500, \rho = 0.9$. However, both *IPD*'s and *IPDI*'s performance remains far from being comparable to that of *IPR* with *IPDI* not only being oversized, but t-tests becoming more oversized for the extracted factors 2, 3, and 4 the larger β_1 becomes. Apparently, *IPDI* does not properly identify which factor predicts y_t, indicating that the cumulated factor structure based on the first difference factors is not an ideal representation of the latent structure in the original data. This finding supports our claim that one should work with data that has been preprocessed as little as possible and refrain from using transformations to stationarize data as often as possible. However, note that the size distortion decreases for weakly persistent factors, as with the Wald tests.

The last performance dimension of interest is forecast accuracy. We assess it by comparing 1-step-ahead forecast errors of predictions based on extracted factors to those based on (expanding-window) historical average, a standard benchmark in the literature. Let $\overline{y}_T = \frac{1}{T}\sum_{j=1}^{T} y_j$ denote the benchmark and $\widehat{y}_{T+1}$ the predictive regression forecast of y_{T+1}, both based on data up to time T. We then compare the two forecasts to the out-of-sample observation [OOS] y_{T+1}, calculating

$$R^2_{OOS} = 1 - \frac{\sum_{i=1}^{N_{MC}} \left(y_{T+1} - \widehat{y}_{T+1}\right)^2}{\sum_{i=1}^{N_{MC}} \left(y_{T+1} - \overline{y}_T\right)^2}. \tag{13}$$

R^2_{OOS} in equation (13) is an adaptation of the R^2_{OS} metric of Campbell and Thompson (2008) because it averages the forecast errors across the Monte Carlo samples and not over time when evaluating forecast accuracy in our Monte Carlo study. But still, R^2_{OOS} scales any performance differences (measured in terms of 1-step-ahead forecast squared errors) to the interval $(-\infty, 1]$, so that it indicates a superior (inferior) performance of the predictive regression relative to the benchmark by values greater (smaller) than 0.

The results in Tables 5 and S.1 show that factor-based IVX and PCR-based predictions outperform the historic sample average; unsurprisingly, the more so the stronger the predictive signal and the larger the training sample. Meanwhile, neither the factors from first differences nor those obtained by switching between first differences and the original order of integration achieve similar forecast performance levels. Whilst *IPDI* yields at least some improvement compared to the benchmark, *IPD* does not result in any considerable improvement at all. Thus, extracting the factors from differenced data (and potentially aggregating them back to the original order of integration) has not paid off so far, neither in terms of inferential performance nor in terms of predictive performance.

Finally, note that increasing the cross-sectional dimension by doubling N to 400, *i.e.* 100 observable variables per factor, does not qualitatively change the results; see, in particular, Tables S.1 and S.2 in the supplementary material.

The same is true after having dismissed the assumption of equally persistent factors. As shown in Tables S.3 and S.4, the results for DGP 2 remain qualitatively similar, with the *GR* criterion almost perfectly recovering the true number of factors. Furthermore, both in the joint and in the individual tests for predictability, size, and power behave as in DGP 1 when we had $\rho_1 = \ldots = \rho_r$. *IPR* is not oversized while maintaining power, whereas its competitors struggle with their known weaknesses, be it oversizedness in the case of *IPA*, *IPDI*, or *PCR*, or not being powerful like *IPD*. One could even argue that the size of *IPR* slightly improves, since there are less I(1) or local-to-unity factors simultaneously, now, reducing the need for robustness against predictive regression endogeneity.

Probably the most notable finding of DGP 2 is that a stationary predictor can withstand a strongly persistent factor that does not predict the dependent variable. Figure 3 shows that the first, least persistent factor is correctly identified as the predictor of y_t, while the more persistent factors are not. In addition, the results in Table S.3 confirm that the factor model still offers an advantage over the benchmark in terms of forecast performance, although the effect diminishes for weakly persistent predictors. Thus, mixed persistence types appear to be manageable, allowing practitioners to consider both strongly and weakly persistent factors simultaneously.

Taken together, our homoskedastic baseline setups DGP 1 and 2 show that *IPR* performs reliably in terms of size, power, and FMSE, whereas its competitors *IPA*, *IPD*, *IPDI* and *PCR* all exhibit some shortcomings; for instance, *IPA*, *IPDI* and *PCR* being oversized or *IPD* being not powerful at all. For *PCR*, this observation is not surprising because of its lack of robustness, while *IPA*'s oversizedness highlights the need to properly take the effect of a finite sample into account, whereas the oversizedness of *IPDI* questions the value of going back and forth between the original order of integration and the first differences.

Continuing with the (un-)conditionally heteroskedastic DGPs 3–5, we note that unconditional changes in the volatility do not alter the results qualitatively. Although our variance shifts, both upward in DGP 3 and downward in DGP 4, are of considerable magnitude, we do not observe any qualitative changes compared to DGP 2 in any of our performance dimensions; cf. Tables S.3–S.6 in the supplementary material for details. Only when it comes to conditional heteroskedasticity in DGP 5, one observes that *GR* suggests extracting only 3.6 factors on average when innovation variances change due to GARCH(1,1) processes rather than volatility shifts. Nevertheless, empirical size and power remain qualitatively unchanged from the homoskedastic baseline, and employing extracted factors to predict y_t rather than to use its historic mean still substantially benefits the forecast performance.

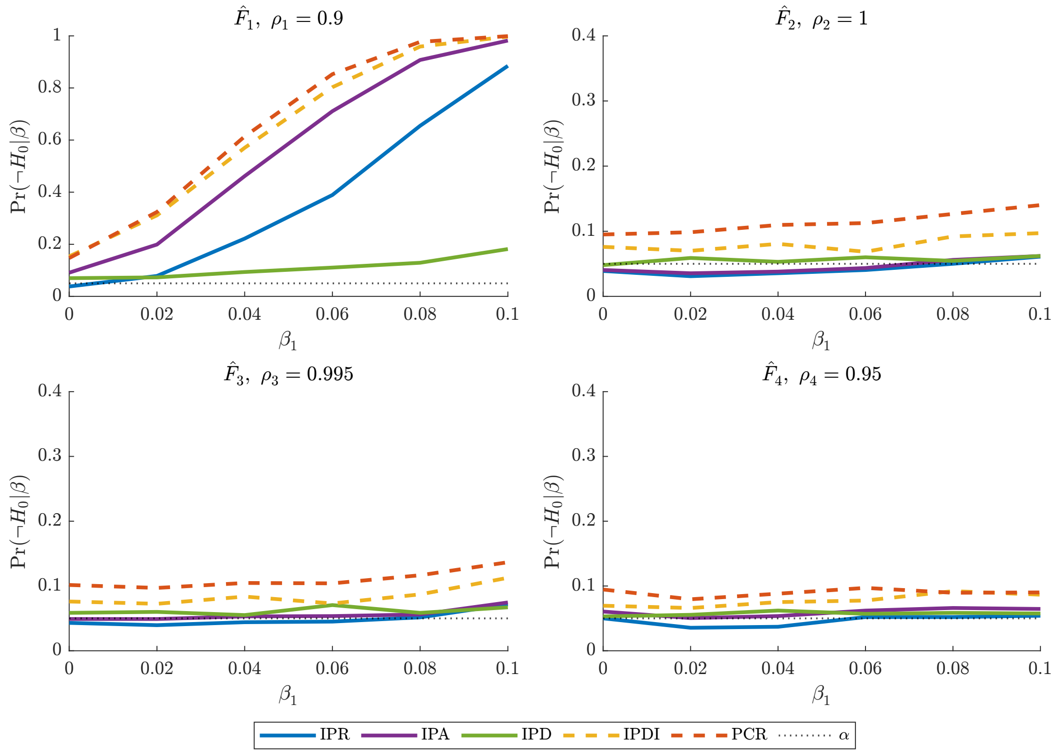

Figure 3: t-test power curves of individually testing the extracted factors (two-sided) at 5% significance level. Data generated from DGP 2; $T = 250$, $N = 200$.

Thus, these findings indicate that factor-based IVX, if combined with a suitable bootstrap as in *IPR*, constitutes a versatile approach that is able to handle a variety of heteroskedasticity patterns that are common in financial data, see the discussion in section 2, whereas its competitors are still subject to their aforementioned flaws.

Leaving heteroskedasticity behind, we turn to the results of DGP 6 where it is no longer the actual predictor that is plagued by predictive regression endogeneity, but one of the other factors. Comparing the DGP 6 results in Tables S.7 and S.8 to those of DGP 2, it is evident that *IPR* is still the only method that properly controls the size. A closer look at the t-test results shows that predictive regression endogeneity still causes oversizedness, but now in the second factor rather than the first; see, for example, Figure 4 that depicts the results for $\boldsymbol{\rho} = (0.995, 0.95, 0.9, 1)$. Endogeneity-induced spurious detections are no longer masked by the predictability signal of the first factor, so they are visible for all β_1 values.

The fact that there is no longer an overlap between the factor that predicts y_t and the one plagued by predictive regression endogeneity slightly reduces power. But even though this is true regardless of the approach we use, the particular β_1 value, or the degree of persistence, the tests still uncover predictability well; cf. Figure 4 and Tables S.7 and S.8. Finally, it should be noted that *PCR* is no longer the worst performing competitor, with *IPDI* often performing equally poorly. This reiterates our earlier conclusion that transforming the data may very well alter it in a way that clouds its predictive information.

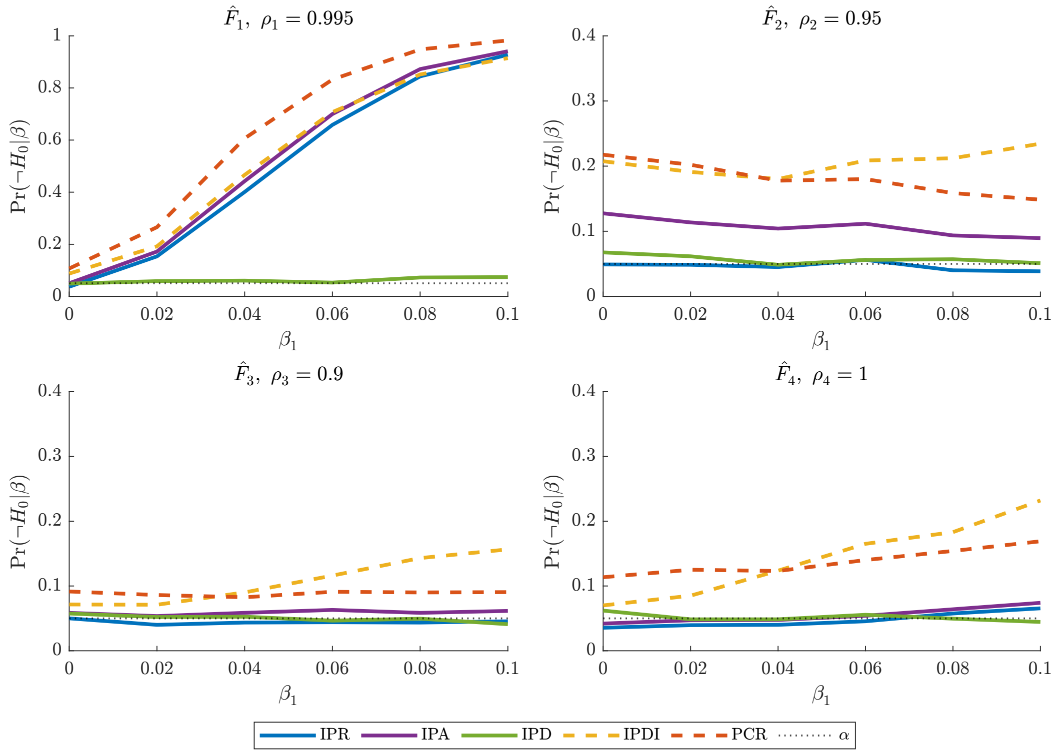

Figure 4: t-test power curves of individually testing the extracted factors (two-sided) at 5% significance level. Data generated from DGP 6; $T = 250$, $N = 200$.

Afterwards, DGP 7 replaces the predictive regression endogeneity-causing correlation between u_t and $v_{1,t}$ altogether by correlation between the factor innovations $v_{1,t}$ and $v_{2,t}$. Now, the Wald tests are (still) correctly sized regardless of the procedure since the data is no longer endogeneous; see Table S.7. However, the t-tests become less informative, which does not strike us as a surprise because we extract only about 3.3 factors on average. Consequently, our approach to rotate and sort the extracted factors works less well for correlated factor innovations. Moreover, the different types of persistence of the correlated factors make it harder for *IPR* to replicate the true DGP in bootstrap samples, especially if fewer than four factors are extracted. Therefore, the results of the t-tests convey the feeling that *IPR* performs slightly better in terms of empirical size when the actual predictor is weakly rather than strongly persistent; cf. Table S.8.

Finally, DGP 8 no longer considers our patterned loading structure, letting the variables load freely on all factors. For the first time, *GR* is highly conservative, usually detecting only one factor when applied to differenced data ΔX; cf. Table S.9. Interestingly, this does not harm the size of any approach and even improves it for those that were previously oversized, as one can see in Figure 5.

However, note the substantial power loss in panels 3 and 4 of Figure 5, which may occur due to predictability signals from weakly persistent factors being masked by non-predictive highly persistent factors that are also absorbed into the single extracted factor. Since applying *GR* to differenced data appears to be too conservative, we redo the analysis using factor numbers

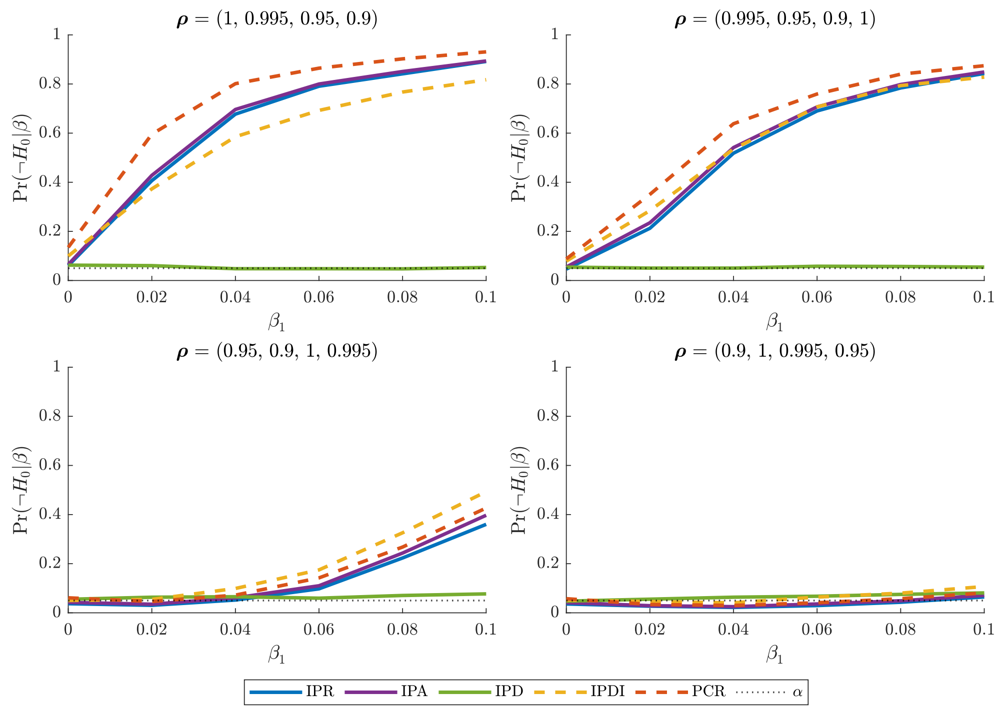

Figure 5: Wald test power curves of jointly testing the extracted factors at 5% significance level. Data generated from DGP 8; $T = 250$, $N = 200$, GR applied to ΔX.

extracted from the original data instead. This yields results similar to the baseline again with the only exception that *IPDI* substantially loses power; see Figure 6 and Table S.9.

In summary, our in-depth simulation study covering a wide variety of setups allows us to conclude that factor-based IVX predictive regression (in combination with suitable bootstrap inference) performs reliably in terms of factor number extraction and inferential as well as forecasting performance so that we may illustrate its empirical applicability, next.

5 Factor-based S&P 500 Equity Premium Prediction

In our empirical application, we are concerned with extracting informative predictors of the S&P 500 equity premium from a data set containing a variety of both financial and macroeconomic variables. The equity premium serves as the dependent variable and follows the standard definition in the literature, *i.e.*

$$\text{eq\_p}_t = \ln(1 + r_{m,t}) - \ln(1 + r_{f,t})$$

where $r_{m,t}$ and $r_{f,t}$ denote the S&P 500 log return and the 3-month treasury bill rate, respectively, with the latter serving as our risk-free rate proxy. Both the return and the t-bill rate data come from an updated version of the data in Welch and Goyal (2008) data. In addition, this dataset provides several presumed predictors of the equity premium that we enrich with more potential

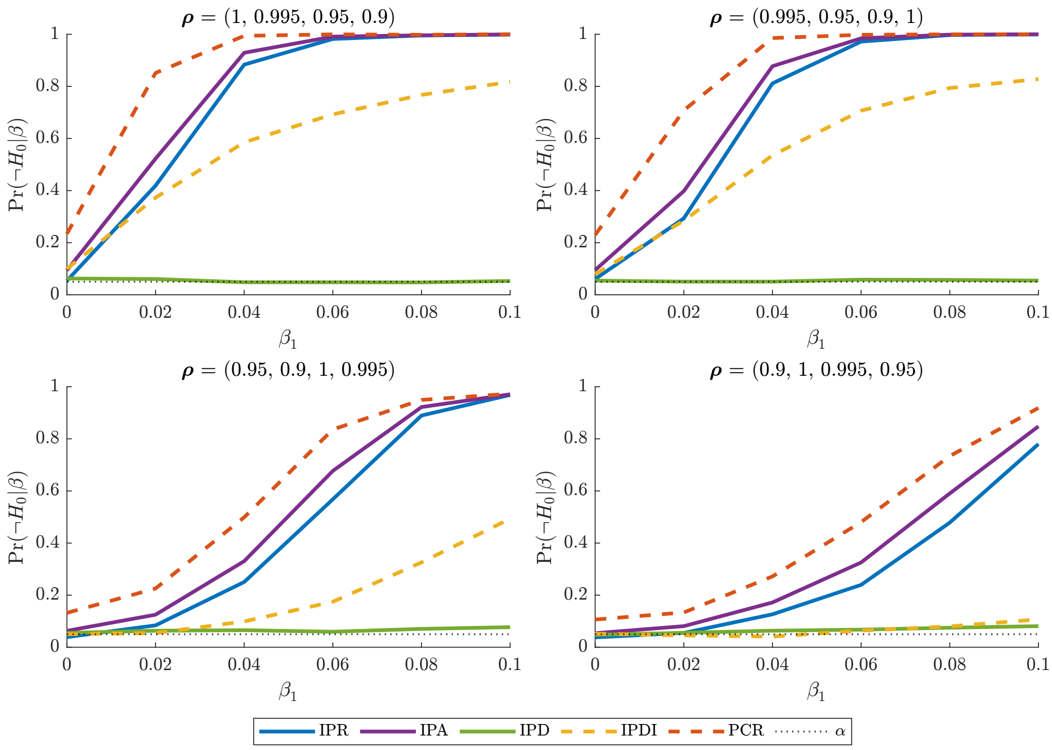

Figure 6: Wald test power curves of jointly testing the extracted factors at 5% significance level. Data generated from DGP 8; $T = 250$, $N = 200$, GR applied to X.

predictors from the dataset of Goyal *et al.* (2024), and the FRED-MD database of McCracken and Ng (2016).[6]

Taken together, our dataset contains observations from January 1970 to December 2019 and $N = 70$ predictors that can be roughly grouped as follows. First, there are financial market-related variables such as dividend-related variables, earnings-related variables, and other measures such as technical indicators like the ones of Neely *et al.* (2014) or stock market volatility items. Interest rates, bond yields, and their corresponding spreads constitute the second group of variables in our dataset. This field is expanded by adding variables that measure the money supply or describe the loan market. In addition, our dataset includes measures of industrial production and personal income, unemployment, and housing, as well as orders to capital goods and price indices; cf. the complete list of variables in Table S.10 in the supplementary material. This table also provides short descriptions and the source of each variable. Of course, this dataset is of rather modest size compared to what is considered to be a large dataset today. Yet, the findings of Xu and Guo (2024) show that even $N = 70$ variables are more than enough to compromise standard IVX inference; which we cannot apply properly anyway due to multicollinearity.

[6]The datasets from Welch and Goyal (2008) and Goyal *et al.* (2024) are available from Amit Goyal's homepage; https://sites.google.com/view/agoyal145. The FRED-MD database is available at https://www.stlouisfed.org/research/economists/mccracken/fred-databases.

However, before extracting and analyzing factors, each variable is individually examined with the results provided in Table S.10. Our first object of interest is the variables' persistence which we asses through AR(1) models fitted to each predictor. Here, we note that many variables have AR(1) coefficients close to 1; for example, see the dividend-price ratio or the book-to-market ratio which yield $\hat{\rho} \approx 0.995$. Thus, some variables are likely highly persistent, which we must account for when testing predictability later.

Secondly, we examine individual-variable equity premium predictability using univariate OLS and IVX regressions. Both non-robust OLS and endogeneity-robust IVX yield a few statistically significant predictors. However, because we ran about 70 individual tests, some findings may simply be a result of multiple testing rather than true predictability and should therefore be interpreted with caution. Nevertheless, many of the significant predictors are interest rates/bond yields (or the corresponding spreads). For this reason, it will be interesting to see whether these individual variables form a common factor later and whether this factor will be a significant predictor of the equity premium, too.

In the final step of our preliminary analysis, we assess the explanatory power of each variable with respect to the equity premium through in-sample [IS] R^2s. As the R^2 values are close to 0 even when the variables are statistically significant, significance does not necessarily imply meaningful explanatory power due to the low individual signal-to-noise ratios.

Taken together, our preliminary analysis suggests that there may be informative predictors in the dataset, but exploiting that information is difficult due to low signal-to-noise ratios and multicollinearity. For this reason, we switch to factor-based models that are supposed to circumvent these issues via dimensional reduction while properly taking into account the fact that predictor persistence is unknown.

5.1 Factor Extraction

The first step is to extract latent factors from the data set. As in the simulation study, the number of factors extracted is chosen according to the *GR* information criterion of Ahn and Horenstein (2013) applied to the data in first differences, knowing that this reduces the risk of incorrectly finding too many factors. Subsequently, four latent factors are extracted by PCA applied to the original data given the factor number estimate from *GR*. As in the simulation study, we do not use first differences to avoid arbitrary transformations that could reduce the factors' explanatory power. Finally, the resulting factors are rotated by an Equamax rotation that is supposed to improve their interpretability.

Upon analyzing the rotated loadings, we find that the variables that load the most (in absolute value) on the first factor are mainly price index variables such as the PPIs for crude and intermediate materials, crude oil, or metals and metal products. Furthermore, unfilled orders for durable goods, business inventories, and loan data are related to $\hat{F}_1$; cf. Table 1. Meanwhile,

$\hat{F}_1$	$\hat{F}_2$	$\hat{F}_3$	$\hat{F}_4$
WPSID62	T5YFFM	ogap	VIXCLSx
PPICMM	T10YFFM	UNRATE	dtoy
OILPRICEx	AAAFFM	e10p	dtoat
WPSID61	BAAFFM	CLAIMSx	svar
AMDMUOx	tms	dy	tchi
WPSFD49207	T1YFFM	dp	ep10
BUSINVx	TB6SMFFM	dfy	avgcor
NONREVSL	TB3SMFFM	PERMIT	rdsp
REALLN	COMPAPFFx	bm	de
M2REAL	FEDFUNDS	baa	lzrt

Table 1: Top-10 variables with the highest loadings (in absolute terms) per extracted factor. Factors extracted from X.

the second factor is mainly driven by the spreads of government and corporate bonds of various maturities, whereas the third factor is a mixture of variables related to the output gap, unemployment, and financial variables. Finally, the fourth factor clearly relates to financial variables that describe the stock market.

Given these results, it is not surprising that the extracted factors are highly persistent, as can be seen from the AR(1) coefficient estimates $\hat{\boldsymbol{\rho}} = (0.999, 0.964, 0.998, 0.935)$ obtained separately for each extracted factor. Hence, one needs an approach such as *IPR* that provides inference results that are robust to predictive regression endogeneity later on.

As a cross-check of the loading-based factor interpretations, we follow the suggestion of Ludvigson and Ng (2011) and regress every observable variable on each of the extracted factors. The resulting R^2 then indicates which factor is correlated with which observable variable. A closer inspection of the resulting R^2 values in Figures S.1 and S.2 in section S.3 of the supplementary appendix supports our previous interpretations, especially with respect to factors 2 and 4 which display the highest R^2 values for bond yield/rate (spread) information and financial market variables, respectively. However, any R^2 should be taken with a grain of salt. Because our dataset contains several I(1) variables and our factors are themselves highly persistent, some of the R^2 values might simply be the result of spurious regressions. Nonetheless, the highest R^2 values are in general in line with our earlier loading-based interpretations.

Since we will compare the performance of *IPR* and *IPA* to that of *IPD* and *IPDI*, a second set of factors is extracted from the data in first differences as the latter two methods utilize them instead. Since these factors are extracted from ΔX instead of X, it is not surprising that the extracted factors differ slightly. First, they are weakly rather than strongly persistent due to the data being in first differences; cf. the estimated AR(1) coefficients $\hat{\boldsymbol{\rho}} = (0.364, 0.246, 0.589, 0.094)$. This then results in the R^2 factor interpretation being much more clear-cut than before, see Figures S.3 and S.4. Furthermore, the bond data factor is now divided into two factors. In

$\hat{F}_1$	$\hat{F}_2$	$\hat{F}_3$	$\hat{F}_4$
GS1	T5YFFM	WPSID61	e10p
GS5	T10YFFM	WPSFD49207	dp
TB6MS	AAAFFM	PCEPI	dtoat
GS10	T1YFFM	CPIAUCSL	dtoy
tbl	TB6SMFFM	WPSID62	bm
aaa	BAAFFM	OILPRICEx	ygap
CP3Mx	TB3SMFFM	M2REAL	ep
baa	FEDFUNDS	ep10	VIXCLSx
lty	COMPAPFFx	de	tchi
FEDFUNDS	tms	PPICMM	avgcor

Table 2: Top-10 variables with the highest loadings (in absolute terms) per extracted factor. Factors extracted from ΔX.

contrast to our earlier results, one comprises the original rate and yield data while the other one summarizes spread information; cf. Table 2 and Figures S.3–S.4. Despite being differenced beforehand, the price index data and the financial market data constitute the two remaining factors.

Finally, note that the factor structure and the factors' interpretation remain unchanged when we cumulate the factors obtained from differenced data to use them in *IPDI*. However, they become strongly persistent, again, as indicated by the resulting estimates of the AR coefficient $\hat{\boldsymbol{\rho}} = (0.985, 0.961, 0.992, 0.967)$. Hence, *IPDI* definitely requires IVX regression under the hood, while *IPD* does not necessarily require this due to the factors' reduced degree of persistence.

5.2 Performance of Factor-based Predictions

Our predictability exercise starts by testing the extracted factors to determine whether they statistically significantly predict the equity premium. This is done for *IPR* (setting $B = 1000$), *IPA*, *IPD*, *IPDI*, and *PCR* using the respective sets of four extracted factors. Afterwards, Wald tests investigate the joint null hypothesis $H_0 : \beta_1 = \ldots = \beta_4 = 0$, finding that *IPD* is the only method suggesting that the extracted factors jointly predict the equity premium at the 5 % level. In addition, *PCR* indicates predictability at the 10 % level at least; cf. Table 3 summarizing the $p^{\mathcal{W}}_{\hat{F}}$-values.

Turning to individual factor tests, there are more predictability findings, with *IPR*, *IPA*, and *PCR* all finding the bond spread factor to predict the equity premium. *PCR* even finds predictability at the 1 % level, whereas the other two do so at the 5 % level. However, the slightly elevated p-value of *IPR* is not a surprise, since it is the most conservative method. Hence, we would expect it to be less susceptible to predictive signals; the same being true for the Wald test p-values where *IPA* is marginally closer to a 10 % rejection than *IPR*. Furthermore, the t-test

	IPR	*IPA*	*IPD*	*IPDI*	*PCR*
$p_{\hat{F}}^{\mathcal{W}}$	0.1380	0.1185	0.0235	0.1671	0.0610
$p_{\hat{F}_1}^{t}$	0.4640	0.5547	0.0068	0.0692	0.3448
$p_{\hat{F}_2}^{t}$	0.0160	0.0132	0.2555	0.4571	0.0099
$p_{\hat{F}_3}^{t}$	0.2840	0.2620	0.0976	0.6478	0.1275
$p_{\hat{F}_4}^{t}$	0.3580	0.3236	0.5129	0.2078	0.3935

Table 3: p-values of Wald tests (H_0 in equation (9)) and t-tests (H_0 in equation (10)) for the extracted predictors.

results give a hint as to why the *IPD* and *PCR* Wald tests rejected their null considering the strong rejections some of the t-tests display in Table 3. For the treasury rate/bond yield factor, the *IPD* t-test signals predictability at the 1 % level with the *PCR*-based test of the spread factor doing so, too. In addition, *IPD* finds that the price data factor contains predictive information (at least at the 10 % level), which, however, is not corroborated by any other method.

The weakest predictability signals are identified by *IPDI* with the only rejection being the bond rate/yield factor t-test that rejects the null at 10 %. This is in contrast to our earlier finding from the simulation study that this method is less robust than *IPR* and *IPA*. However, by internally processing the data, *IPDI* may have obscured the predictability signal, which might have caused a weaker rejection.

Because such transformations may adversely affect forecast performance, we examine the forecast accuracy both in-sample and pseudo-out-of-sample. Here, we compare our methods' ability to forecast the equity premium with an expanding-window historical average, our benchmark forecast, as is standard in the literature. Summing over t instead of over Monte Carlo samples as in equation (13), we quantify in-sample performance differences via

$$R_{IS}^2 = 1 - \frac{\sum_{t=1}^{T-1} \left(y_{t+1} - \widehat{y}_{t+1}\right)^2}{\sum_{t=1}^{T-1} \left(y_{t+1} - \overline{y}_t\right)^2},$$

where $\overline{y}_t = \frac{1}{t}\sum_{j=1}^{t} y_j$ is the expanding-window sample average and $\widehat{y}_{t+1}$ is the predictive regression forecast of y_{t+1}, both computed at time t.

As in-sample forecasts are rather vaguely relevant in practice, we lastly examine forecast performance in a pseudo-out-of-sample fashion, extracting predictors from rolling windows (window width of $T^* = 300$ observations, *i.e.* 25 years of data), estimating the predictive regression model afterwards, and finally forecasting the next, pseudo-OOS observation. Thus, the prediction period begins in January 1995 and lasts until December 2019. By choosing rolling windows of this length, we ensure that there are enough observations to reliably estimate the number of factors, the factors themselves, and the predictive regression coefficients. The resulting equity premium forecasts are then evaluated using the R_{OOS}^2 measure from equation (13), with forecast errors

	IPR	*IPD*	*IPDI*	*PCR*
R^2_{IS}	0.0435	0.0353	0.0419	0.0435
$R^2_{IS,rec}$	0.0915	0.0990	0.0649	0.0923
$R^2_{IS,exp}$	0.0228	0.0093	0.0320	0.0226
R^2_{OOS}	-0.0277	0.0049	-0.0449	-0.0242
$R^2_{OOS,rec}$	0.0148	-0.0551	-0.0326	0.0203
$R^2_{OOS,exp}$	-0.0419	0.0251	-0.0491	-0.0391

Table 4: 1-step-ahead predictability R^2_{FMSE} values relative to expanding window historical averages. Rolling window width of $T = 300$.

averaged over t as for R^2_{IS}. To further assess forecast performance, we also compute R^2_{FMSE} items where the forecast errors are pre-multiplied by an indicator $\mathcal{I}^d_t$, $d = rec, exp$, to include only forecast errors from periods of economic expansions (exp) or contractions (rec), respectively, *i.e.*

$$R^2_{OOS,d} = 1 - \frac{\sum_{t=T^*}^{T-1} \mathcal{I}^d_t \left(y_{t+1} - \widehat{y}_{t+1}\right)^2}{\sum_{t=T^*}^{T-1} \mathcal{I}^d_t \left(y_{t+1} - \overline{y}_t\right)^2}, \quad d = rec,\ exp.$$

Using the NBER Recession Indicator available from the Federal Reserve Bank of St. Louis to indicate economic expansions and contractions,[7] this business cycle $R^2_{(\cdot)}$ is supposed to give an idea of whether equity premium predictability differs along the business cycle; see, among others, Li and Qin (2026, p. 15 f.) for similar considerations.

Compared to our earlier finding that observable variables yield rather uninformative equity premium forecasts, even in-sample, our factor-based predictions outperform the historic average considerably throughout all four approaches; see Table 4. It should be noted that these improvements are mainly caused by improved forecasts during recessions. In line with observations of Dai *et al.* (2025), the recession $R^2_{IS,rec}$ values are substantially higher than their expansion counterparts, which may suggest episodic predictability. Finally, note that *IPD* forecasts also outperform the benchmark despite our earlier simulation results. However, it does so the least because it performs particularly poorly in times of economic expansion, which outweighs its competitive performance in contraction periods.

Finally, an analogous pattern is observed in our pseudo-out-of-sample experiment. However, regression-based forecasts outperform the benchmark even less; in most cases, only during economic contractions. In particular, *IPR* and *PCR* have their strengths in predicting the equity premium in times of business cycle downswings, whereas *IPD* performs better in times of upswings.

Thus, our results support earlier findings that predictability signals are rather weak and difficult to exploit, cf. Welch and Goyal (2008), *inter alia*, for a seminal discussion, even when using generated predictors instead of many individual variables. One potential explanation for

[7] https://fred.stlouisfed.org/series/USREC

this mediocre out-of-sample forecast accuracy arises from the performance difference between periods of economic expansion and contraction. Instead of being predictable throughout the entire sample, the equity premium might just be episodically predictable by a specific predictor; cf. Li and Qin (2026) also finding forecast performance differences along the business cycle. As predictability seems to emerge only from time to time and vanish quickly, exploiting it becomes particularly difficult and would require monitoring the predictability; see, *inter alia*, Demetrescu *et al.* (2022) and Harvey *et al.* (2021) for in-sample and out-of-sample approaches, respectively.

Alternatively, our factor-based approach may have extracted factors that best explain the predictor data rather than those that best predict the equity premium. Instead of treating factor extraction as a separate preliminary step, the dimensional reduction could be designed from the outset to serve both inference and the subsequent forecasting task. For example, this can be achieved by using scaled principal components or partial least squares, which incorporate the forecasting purpose when extracting latent dynamics; cf. Zhang and Xie (2026), among others, for a supervised principal component regression approach. However, predictive regression endogeneity must still be properly addressed when using such methods with data of unknown persistence. Otherwise, the component extraction step might even amplify the endogeneity issue.

In summary, we are nevertheless confident that the predictability findings of *IPR* are not spurious thanks to the method's robustness to predictive regression endogeneity and heteroskedasticity and that there are periods where the equity premium is indeed predictable using a suitable set of variables.

6 Conclusion

In our study of a factor-based extension to the IVX framework of Kostakis *et al.* (2015), we made three contributions. First, we took a closer look at how latent factors can be extracted when the factors are either weakly or strongly persistent. In doing so, we corroborate the result of Onatski and Wang (2025) finding that a fixed number of factors can be extracted consistently from observable data, since the factor structure dominates the impairing effect of strong persistence as long as the factors are "strong". Second, our theoretical results, supported by an extensive simulation study, indicate that factor-based IVX inference performs reliably; in particular, when combined with the adaption of Demetrescu *et al.* (2023)'s residual wild bootstrap that we propose to employ in finite samples. Finally, we have been able to extract factors from financial and macroeconomic data that had not been preprocessed. Afterwards, we have identified the extracted factor that pools interest rate spread (or interest rate) information as a potentially exploitable predictor of the equity premium.

We conclude with a brief discussion of future research directions arising from our findings. First, it should be noted that the extracted factors were chosen *ex ante*, without reference to the subsequent prediction exercise. Thus, our factor-based approach might have extracted factors that best explain the predictor data rather than those that are best suited to predict the equity premium. Instead of treating factor extraction as a separate step, one could tailor it to the forecast task while accounting for predictive regression endogeneity, especially with data of unknown persistence, since component extraction may otherwise worsen the impact of endogeneity on inference. Finally, many observable variables do not necessarily result from a factor structure, although there are many reasons to suspect that in economic data. In such cases, factor extraction may not yield informative factors or fail to improve forecast accuracy. Instead, shrinkage methods such as Ridge or LASSO could be used for dimensional reduction, focusing on a few true predictors within a larger set of potential ones. However, the potentially adverse effect of predictive regression endogeneity should also be taken into account by any candidate shrinkage approach.

References

Ahn, S. C., & Horenstein, A. R. (2013). Eigenvalue Ratio Test for the Number of Factors. *Econometrica*, *81*(3), 1203–1227.

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, *71*(1), 135–171.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, *122*(1), 137–183.

Bai, J., & Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, *70*(1), 191–221.

Bai, J., & Ng, S. (2004). A PANIC Attack on Unit Roots and Cointegration. *Econometrica*, *72*(4), 1127–1177.

Banerjee, A., Marcellino, M., & Masten, I. (2017). Structural FECM: Cointegration in large-scale structural FAVAR models. *Journal of Applied Econometrics*, *32*(6), 1069–1086.

Beckmann, J., Kerkemeier, M., & Kruse-Becher, R. (2025). Regime-specific exchange rate predictability. *Journal of Economic Dynamics and Control*, *176*, 105095.

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

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, *18*(2), 373–399.

Campbell, J. Y., & Shiller, R. J. (1988a). Stock Prices, Earnings, and Expected Dividends. *The Journal of Finance*, *43*(3), 661–676.

Campbell, J. Y., & Shiller, R. J. (1988b). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *The Review of Financial Studies*, *1*(3), 195–228.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *The Review of Financial Studies*, *21*(4), 1509–1531.

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

Cavanagh, C. L., Elliott, G., & Stock, J. H. (1995). Inference in Models with Nearly Integrated Regressors. *Econometric Theory*, *11*(5), 1131–1147.

Chamberlain, G., & Rothschild, M. (1983). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica*, *51*(5), 1281–1304.

Clark, T. E. (2009). *Is the Great Moderation Over? An Empirical Analysis* (Economic Review No. 94 (Q IV), 5–42). Federal Reserve Bank of Kansas City.

Corona, F., Poncela, P., & Ruiz, E. (2017). Determining the number of factors after stationary univariate transformations. *Empirical Economics*, *53*, 351–372.

Dai, Z., Jiang, F., Kang, J., & Xue, B. (2025). Stock return predictability in the frequency domain. *International Journal of Forecasting*, *41*(3), 1126–1147.

Demetrescu, M., Georgiev, I., Rodrigues, P. M. M., & Taylor, A. M. R. (2022). Testing for episodic predictability in stock returns. *Journal of Econometrics*, *227*(1), 85–113.

Demetrescu, M., Georgiev, I., Rodrigues, P. M. M., & Taylor, A. M. R. (2023). Extensions to IVX methods of inference for return predictability. *Journal of Econometrics*, *237*(2, Part C), 105271.

Demetrescu, M., & Hillmann, B. (2022). Nonlinear Predictability of Stock Returns? Parametric Versus Nonparametric Inference in Predictive Regressions. *Journal of Business & Economic Statistics*, *40*(1), 382–397.

Demetrescu, M., Schmidt, F., & Taylor, A. M. R. (2026). *Real-Time Monitoring for Stock Return Predictability in Nonstationary Volatility Environments* (TRR 391 Working Paper No. 8). TRR 391.

Elliott, G., Müller, U. K., & Watson, M. W. (2015). Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis. *Econometrica*, *83*(2), 771–811.

Fama, E. F. (1981). Stock Returns, Real Activity, Inflation, and Money. *The American Economic Review*, *71*(4), 545–565.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, *22*(1), 3–25.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, *25*(1), 23–49.

Folch-Fortuny, A., Arteaga, F., & Ferrer, A. (2015). PCA model building with missing data: New proposals and a comparative study. *Chemometrics and Intelligent Laboratory Systems*, *146*, 77–88.

Gonçalves, S., & Perron, B. (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics*, *182*(1), 156–173.

Gonzalo, J., & Pitarakis, J.-Y. (2021). Spurious relationships in high-dimensional systems with strong or mild persistence. *International Journal of Forecasting*, *37*(4), 1480–1497.

Goyal, A., Welch, I., & Zafirov, A. (2024). A Comprehensive 2022 Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, *37*(11), 3490–3557.

Groen, J. J. J., & Kapetanios, G. (2013). Model Selection Criteria for Factor-Augmented Regressions. *Oxford Bulletin of Economics and Statistics*, *75*(1), 37–63.

Hansen, B. E. (1992). Convergence to Stochastic Integrals for Dependent Heterogeneous Processes. *Econometric Theory*, *8*(4), 489–500.

Harvey, D. I., Leybourne, S. J., Sollis, R., & Taylor, A. M. R. (2021). Real-time detection of regimes of predictability in the US equity premium. *Journal of Applied Econometrics*, *36*(1), 45–70.

Jansson, M., & Moreira, M. J. (2006). Optimal Inference in Regression Models with Nearly Integrated Regressors. *Econometrica*, *74*(3), 681–714.

Kapetanios, G. (2010). A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets. *Journal of Business & Economic Statistics*, *28*(3), 397–409.

Keim, D. B., & Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics*, *17*(2), 357–390.

Kostakis, A., Magdalinos, T., & Stamatogiannis, M. P. (2015). Robust Econometric Inference for Stock Return Predictability. *The Review of Financial Studies*, *28*(5), 1506–1553.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411.

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

Li, Z., & Qin, W. (2026). Enhancing market return forecasts with an incident-based ESG indicator. *International Journal of Forecasting*, *(In Press, Corrected Proof)*.

Ludvigson, S. C., & Ng, S. (2011). A Factor Analysis of Bond Risk Premia. In A. Ullah & D. E. A. Giles (Eds.), *Handbook of Empirical Economics and Finance* (1st ed., pp. 313–371). Chapman and Hall/CRC.

Magdalinos, T., & Phillips, P. C. B. (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory*, *25*(2), 482–526.

McCracken, M. W., & Ng, S. (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

Morico, A., & Stauskas, O. (2025). *Robust Tests for Factor-Augmented Regressions with an Application to the novel EA-MD-QD Dataset* (arXiv preprint No. 2504.08455v2). arXiv.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the Equity Risk Premium: The Role of Technical Indicators. *Management Science*, *60*(7), 1772–1791.

Nelson, C. R., & Kim, M. J. (1993). Predictable Stock Returns: The Role of Small Sample Bias. *The Journal of Finance*, *48*(2), 641–661.

Onatski, A. (2010). Determining the Number of Factors From Empirical Distribution of Eigenvalues. *The Review of Economics and Statistics*, *92*(4), 1004–1016.

Onatski, A., & Wang, C. (2025). Spurious factors in data with local-to-unit roots. *Econometric Theory*, *41*(4), 779–816.

Phillips, P. C. B., & Magdalinos, T. (2009). *Econometric Inference in the Vicinity of Unity* (CoFie Working Paper No. 7). Singapore Management University.

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

Stock, J. H., & Watson, M. W. (1998). *Diffusion Indexes* (NBER Working Paper No. 6702). National Bureau of Economic Research. Cambridge, Mass.

Stock, J. H., & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, *44*(2), 293–335.

Stock, J. H., & Watson, M. W. (2002a). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

Stock, J. H., & Watson, M. W. (2002b). Has the Business Cycle Changed and Why? *NBER Macroeconomics Annual*, *17*, 159–218.

van der Vaart, A. W. (1998). *Asymptotic Statistics* [8th printing, 2007]. Cambridge University Press.

Welch, I., & Goyal, A. (2008). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, *21*(4), 1455–1508.

Xu, K.-L., & Guo, J. (2024). A New Test for Multiple Predictive Regression. *Journal of Financial Econometrics*, *22*(1), 119–156.

Zhang, H., He, Q., Jacobsen, B., & Jiang, F. (2020). Forecasting stock returns with model uncertainty and parameter instability. *Journal of Applied Econometrics*, *35*(5), 629–644.

Zhang, T., & Xie, H. (2026). Stock Return Forecasting: A Supervised PCA With Selecting and Scaling. *Journal of Forecasting*, *45*(2), 547–562.

Simulation Results for DGP 1, T=250

DGP 1 ($T=250,\ N=200$)

ρ	β	GR	$\Pr(\neg H_0\|\beta)$ IPR	IPA	IPD	IPDI	PCR	R^2_{OOS} IPR	IPD	IPDI	PCR
	0.00	3.99	0.0767	0.1555	0.0772	0.2979	0.3501	-0.009	-0.003	-0.018	-0.018
1.000	0.02	3.99	0.4472	0.6345	0.0831	0.7611	0.9099	0.015	-0.006	0.017	0.026
1.000	0.04	3.99	0.8972	0.9519	0.0667	0.9448	0.9990	0.109	-0.002	0.063	0.128
1.000	0.06	3.99	0.9829	0.9920	0.0803	0.9799	1.0000	0.213	-0.008	0.126	0.247
1.000	0.08	4.00	0.9965	0.9985	0.0825	0.9915	1.0000	0.346	-0.012	0.169	0.358
	0.10	3.99	0.9985	0.9990	0.0757	0.9935	1.0000	0.449	-0.010	0.221	0.455
	0.00	4.00	0.0696	0.1381	0.0721	0.2508	0.2958	-0.020	-0.010	-0.023	-0.031
0.995	0.02	3.99	0.3305	0.4972	0.0751	0.6525	0.7892	-0.008	-0.002	-0.010	-0.010
0.995	0.04	3.99	0.8301	0.9158	0.0697	0.9368	0.9945	0.035	-0.008	0.008	0.052
0.995	0.06	3.99	0.9780	0.9915	0.0942	0.9900	1.0000	0.133	-0.009	0.078	0.133
0.995	0.08	3.99	0.9980	0.9990	0.1002	0.9975	1.0000	0.222	-0.005	0.142	0.225
	0.10	3.99	0.9995	1.0000	0.0926	0.9975	1.0000	0.328	-0.007	0.221	0.335
	0.00	4.00	0.0350	0.0730	0.0815	0.1000	0.1295	-0.008	-0.011	-0.008	-0.015
0.950	0.02	4.00	0.0985	0.1546	0.0885	0.2106	0.2371	-0.015	-0.009	-0.007	-0.018
0.950	0.04	4.00	0.2570	0.3730	0.0935	0.4545	0.5190	-0.008	-0.016	-0.004	-0.005
0.950	0.06	4.00	0.5868	0.7294	0.0945	0.7809	0.8679	-0.000	-0.009	0.008	0.006
0.950	0.08	4.00	0.9219	0.9730	0.1011	0.9720	0.9950	0.038	-0.004	0.027	0.044
	0.10	4.00	0.9955	0.9980	0.1256	0.9975	1.0000	0.065	-0.009	0.043	0.065
	0.00	4.00	0.0445	0.0750	0.0730	0.0960	0.1070	-0.013	-0.007	-0.004	-0.016
0.900	0.02	4.00	0.0640	0.0965	0.0780	0.1360	0.1535	-0.001	-0.011	-0.005	-0.011
0.900	0.04	4.00	0.1356	0.1981	0.0930	0.2551	0.2856	-0.012	-0.005	-0.004	-0.016
0.900	0.06	4.00	0.2815	0.3840	0.1115	0.4475	0.5095	-0.002	-0.011	0.001	-0.001
0.900	0.08	4.00	0.5290	0.6590	0.1095	0.7155	0.7895	0.020	-0.006	0.013	0.024
	0.10	4.00	0.8220	0.9000	0.1510	0.9235	0.9620	0.046	-0.012	0.028	0.048

DGP 1 ($T=250,\ N=400$)

ρ	β	GR	$\Pr(\neg H_0\|\beta)$ IPR	IPA	IPD	IPDI	PCR	R^2_{OOS} IPR	IPD	IPDI	PCR
	0.00	4.00	0.0775	0.1595	0.0775	0.3060	0.3675	-0.008	-0.004	-0.022	-0.024
1.000	0.02	4.00	0.4535	0.6355	0.0835	0.7525	0.9185	0.004	-0.000	-0.007	0.023
1.000	0.04	4.00	0.9050	0.9535	0.0810	0.9485	0.9990	0.074	-0.007	0.036	0.119
1.000	0.06	4.00	0.9850	0.9910	0.0840	0.9805	1.0000	0.225	-0.004	0.120	0.240
1.000	0.08	4.00	0.9970	0.9990	0.0910	0.9880	1.0000	0.349	-0.014	0.192	0.366
	0.10	4.00	0.9990	1.0000	0.0795	0.9940	1.0000	0.447	-0.007	0.242	0.455
	0.00	4.00	0.0720	0.1470	0.0750	0.2450	0.3115	-0.010	-0.004	-0.023	-0.028
0.995	0.02	4.00	0.3005	0.4815	0.0745	0.6595	0.7900	-0.015	-0.008	-0.009	-0.011
0.995	0.04	4.00	0.8260	0.9195	0.0880	0.9335	0.9930	0.029	-0.007	0.014	0.052
0.995	0.06	4.00	0.9705	0.9895	0.0870	0.9845	1.0000	0.101	-0.008	0.083	0.132
0.995	0.08	4.00	0.9975	0.9990	0.0965	0.9965	1.0000	0.221	0.003	0.139	0.222
	0.10	4.00	0.9995	0.9995	0.0975	0.9990	1.0000	0.314	-0.010	0.210	0.317
	0.00	4.00	0.0420	0.0790	0.0920	0.1165	0.1270	-0.003	-0.006	-0.004	-0.011
0.950	0.02	4.00	0.0950	0.1650	0.0950	0.2220	0.2380	-0.010	-0.010	-0.011	-0.023
0.950	0.04	4.00	0.2430	0.3585	0.0905	0.4555	0.5175	-0.005	-0.015	-0.001	-0.002
0.950	0.06	4.00	0.6010	0.7675	0.1140	0.8110	0.9000	-0.012	-0.007	0.012	-0.003
0.950	0.08	4.00	0.9275	0.9775	0.1220	0.9790	0.9965	0.046	-0.002	0.032	0.049
	0.10	4.00	0.9935	0.9980	0.1195	0.9980	1.0000	0.052	-0.007	0.039	0.053
	0.00	4.00	0.0390	0.0690	0.0785	0.0920	0.0975	0.001	-0.006	-0.002	-0.010
0.900	0.02	4.00	0.0670	0.1050	0.0825	0.1420	0.1555	-0.013	-0.004	-0.003	-0.023
0.900	0.04	4.00	0.1345	0.2030	0.0885	0.2475	0.2885	-0.005	-0.006	-0.007	-0.023
0.900	0.06	4.00	0.2920	0.3865	0.0955	0.4570	0.5145	-0.009	-0.005	-0.004	-0.007
0.900	0.08	4.00	0.5580	0.6785	0.1315	0.7245	0.8000	-0.017	-0.009	0.002	-0.010
	0.10	4.00	0.8290	0.9075	0.1360	0.9255	0.9745	0.031	-0.001	0.024	0.035

Table 5: Average factor numbers (GR), Wald test rejection frequencies ($\Pr(\neg H_0|\beta)$) at 5% significance level, and forecast R^2_{OOS} values. Data from DGP 1 ($T=250,\ N=200$) and DGP 1 ($T=250,\ N=400$).

Table 6: Factor-specific rejection frequencies of two-sided t-tests at 5% significance level. Data from DGP 1 ($T = 250,\ N = 200$) and DGP 1 ($T = 250,\ N = 400$).

ρ	$\hat{F}_i$	β	DGP 1 ($T = 250,\ N = 200$)					DGP 1 ($T = 250,\ N = 400$)				
			IPR	IPA	IPD	IPDI	PCR	IPR	IPA	IPD	IPDI	PCR
$\begin{pmatrix}1.000\\1.000\\1.000\\1.000\end{pmatrix}$	$\hat{F}_1$	0.00	0.0812	0.2613	0.0657	0.4724	0.5306	0.0880	0.2670	0.0675	0.4945	0.5360
		0.02	0.3275	0.6715	0.0746	0.8353	0.9299	0.3235	0.6740	0.0815	0.8465	0.9455
		0.04	0.6755	0.9338	0.0868	0.9378	0.9990	0.6670	0.9300	0.0820	0.9425	0.9995
		0.06	0.8680	0.9885	0.0968	0.9553	1.0000	0.8685	0.9880	0.1010	0.9605	1.0000
		0.08	0.9575	0.9980	0.0995	0.9715	1.0000	0.9460	0.9970	0.1075	0.9665	1.0000
		0.10	0.9855	1.0000	0.1038	0.9699	1.0000	0.9840	0.9995	0.1180	0.9735	1.0000
	$\hat{F}_2$	0.00	0.0421	0.0617	0.0612	0.1048	0.1479	0.0455	0.0665	0.0615	0.1150	0.1560
		0.02	0.0511	0.0686	0.0541	0.1302	0.1452	0.0575	0.0810	0.0560	0.1365	0.1755
		0.04	0.0622	0.0782	0.0612	0.1946	0.1745	0.0555	0.0760	0.0445	0.1775	0.1560
		0.06	0.0647	0.0838	0.0477	0.2393	0.1696	0.0585	0.0755	0.0500	0.2570	0.1655
		0.08	0.0685	0.0885	0.0515	0.2881	0.1801	0.0505	0.0640	0.0460	0.2645	0.1610
		0.10	0.0586	0.0812	0.0456	0.3378	0.1789	0.0580	0.0840	0.0545	0.3195	0.1760
	$\hat{F}_3$	0.00	0.0542	0.0687	0.0562	0.1078	0.1494	0.0485	0.0695	0.0635	0.1050	0.1540
		0.02	0.0541	0.0696	0.0566	0.1307	0.1437	0.0615	0.0845	0.0555	0.1385	0.1760
		0.04	0.0547	0.0732	0.0502	0.1780	0.1680	0.0535	0.0725	0.0545	0.1840	0.1685
		0.06	0.0662	0.0888	0.0647	0.2529	0.1731	0.0455	0.0635	0.0500	0.2260	0.1540
		0.08	0.0640	0.0930	0.0455	0.2921	0.1811	0.0665	0.0845	0.0610	0.2960	0.1665
		0.10	0.0677	0.0942	0.0416	0.3449	0.1865	0.0550	0.0780	0.0475	0.3320	0.1610
	$\hat{F}_4$	0.00	0.0481	0.0612	0.0547	0.1063	0.1489	0.0465	0.0640	0.0445	0.1170	0.1455
		0.02	0.0596	0.0701	0.0566	0.1402	0.1547	0.0470	0.0675	0.0590	0.1285	0.1640
		0.04	0.0577	0.0757	0.0476	0.1805	0.1510	0.0500	0.0695	0.0515	0.1690	0.1475
		0.06	0.0507	0.0758	0.0552	0.2434	0.1565	0.0575	0.0740	0.0520	0.2375	0.1580
		0.08	0.0725	0.0890	0.0550	0.3012	0.1881	0.0675	0.0945	0.0500	0.2950	0.1965
		0.10	0.0607	0.0802	0.0426	0.3223	0.1840	0.0650	0.0915	0.0475	0.3220	0.1850
$\begin{pmatrix}0.995\\0.995\\0.995\\0.995\end{pmatrix}$	$\hat{F}_1$	0.00	0.0821	0.2257	0.0706	0.4129	0.4620	0.0740	0.2385	0.0670	0.4105	0.4770
		0.02	0.3135	0.6490	0.0726	0.8062	0.8983	0.2730	0.6290	0.0670	0.8000	0.8850
		0.04	0.6521	0.9258	0.0882	0.9619	0.9965	0.6590	0.9115	0.0860	0.9510	0.9975
		0.06	0.8883	0.9865	0.1122	0.9900	0.9995	0.8715	0.9860	0.1120	0.9850	1.0000
		0.08	0.9639	0.9980	0.1097	0.9920	1.0000	0.9605	0.9965	0.1105	0.9940	1.0000
		0.10	0.9870	0.9980	0.1342	0.9915	1.0000	0.9855	0.9995	0.1255	0.9975	1.0000
	$\hat{F}_2$	0.00	0.0501	0.0706	0.0531	0.1051	0.1396	0.0490	0.0665	0.0600	0.1055	0.1370
		0.02	0.0576	0.0761	0.0561	0.1122	0.1452	0.0620	0.0865	0.0505	0.1290	0.1630
		0.04	0.0386	0.0541	0.0566	0.1138	0.1383	0.0575	0.0810	0.0640	0.1325	0.1535
		0.06	0.0611	0.0842	0.0531	0.1683	0.1623	0.0555	0.0865	0.0570	0.1605	0.1580
		0.08	0.0671	0.0922	0.0466	0.1934	0.1819	0.0580	0.0805	0.0580	0.2025	0.1590
		0.10	0.0676	0.0896	0.0466	0.2158	0.1748	0.0680	0.0925	0.0545	0.2235	0.1735
	$\hat{F}_3$	0.00	0.0631	0.0806	0.0576	0.1191	0.1416	0.0525	0.0780	0.0580	0.1125	0.1445
		0.02	0.0556	0.0741	0.0511	0.1182	0.1467	0.0495	0.0740	0.0540	0.1220	0.1655
		0.04	0.0546	0.0757	0.0476	0.1368	0.1479	0.0560	0.0775	0.0585	0.1430	0.1540
		0.06	0.0686	0.0877	0.0526	0.1723	0.1688	0.0575	0.0775	0.0515	0.1475	0.1545
		0.08	0.0741	0.1022	0.0571	0.2019	0.1824	0.0635	0.0850	0.0560	0.2180	0.1670
		0.10	0.0581	0.0861	0.0551	0.2273	0.1723	0.0670	0.0955	0.0535	0.2225	0.1775
	$\hat{F}_4$	0.00	0.0511	0.0696	0.0601	0.1091	0.1411	0.0505	0.0750	0.0535	0.0990	0.1480
		0.02	0.0476	0.0741	0.0471	0.1232	0.1402	0.0595	0.0835	0.0595	0.1170	0.1620
		0.04	0.0531	0.0717	0.0511	0.1343	0.1388	0.0625	0.0855	0.0565	0.1330	0.1625
		0.06	0.0581	0.0807	0.0591	0.1748	0.1678	0.0645	0.0910	0.0600	0.1775	0.1705
		0.08	0.0531	0.0772	0.0526	0.1939	0.1623	0.0630	0.0855	0.0420	0.1880	0.1605
		0.10	0.0711	0.1077	0.0491	0.2298	0.1878	0.0660	0.0885	0.0655	0.2295	0.1695

Continued on the next page

Table 6 (continued)

ρ	$\hat{F}_i$	β	DGP 1 ($T = 250,\ N = 200$)					DGP 1 ($T = 250,\ N = 400$)				
			IPR	IPA	IPD	IPDI	PCR	IPR	IPA	IPD	IPDI	PCR
$\begin{pmatrix}0.950\\0.950\\0.950\\0.950\end{pmatrix}$	$\hat{F}_1$	0.00	0.0470	0.1140	0.0655	0.1690	0.1710	0.0495	0.1180	0.0685	0.1770	0.1700
		0.02	0.1441	0.3192	0.0760	0.4017	0.4302	0.1410	0.3345	0.0830	0.4095	0.4485
		0.04	0.3660	0.6900	0.1075	0.7480	0.7985	0.3805	0.6665	0.0890	0.7340	0.7860
		0.06	0.7254	0.9425	0.1156	0.9540	0.9820	0.7390	0.9555	0.1325	0.9605	0.9855
		0.08	0.9515	0.9965	0.1301	0.9965	0.9995	0.9520	0.9970	0.1490	0.9970	0.9990
		0.10	0.9970	1.0000	0.1727	1.0000	1.0000	0.9925	0.9985	0.1630	1.0000	1.0000
	$\hat{F}_2$	0.00	0.0455	0.0665	0.0540	0.0715	0.0940	0.0510	0.0690	0.0665	0.0730	0.0860
		0.02	0.0490	0.0675	0.0655	0.0755	0.0920	0.0545	0.0710	0.0635	0.0810	0.0950
		0.04	0.0620	0.0815	0.0585	0.0860	0.1105	0.0430	0.0675	0.0540	0.0800	0.1005
		0.06	0.0595	0.0825	0.0585	0.0895	0.1101	0.0495	0.0760	0.0525	0.0870	0.1000
		0.08	0.0516	0.0731	0.0450	0.0861	0.0976	0.0570	0.0785	0.0635	0.0885	0.1035
		0.10	0.0576	0.0761	0.0636	0.1041	0.1016	0.0580	0.0790	0.0610	0.1115	0.1035
	$\hat{F}_3$	0.00	0.0500	0.0685	0.0640	0.0800	0.0885	0.0480	0.0680	0.0620	0.0705	0.0890
		0.02	0.0590	0.0705	0.0515	0.0835	0.0945	0.0460	0.0615	0.0530	0.0685	0.0850
		0.04	0.0560	0.0725	0.0610	0.0920	0.1005	0.0510	0.0690	0.0575	0.0770	0.0915
		0.06	0.0630	0.0910	0.0540	0.0915	0.1156	0.0520	0.0715	0.0605	0.0810	0.0935
		0.08	0.0586	0.0856	0.0566	0.0931	0.1036	0.0530	0.0695	0.0540	0.0875	0.0920
		0.10	0.0596	0.0781	0.0561	0.1101	0.1046	0.0590	0.0830	0.0575	0.1055	0.1045
	$\hat{F}_4$	0.00	0.0450	0.0615	0.0545	0.0765	0.0840	0.0500	0.0700	0.0570	0.0740	0.0935
		0.02	0.0510	0.0690	0.0575	0.0800	0.0865	0.0570	0.0740	0.0560	0.0915	0.0975
		0.04	0.0550	0.0735	0.0555	0.0860	0.1065	0.0465	0.0605	0.0640	0.0665	0.0860
		0.06	0.0520	0.0785	0.0480	0.0880	0.1006	0.0580	0.0800	0.0620	0.0900	0.1035
		0.08	0.0521	0.0701	0.0571	0.0886	0.0981	0.0520	0.0765	0.0495	0.0890	0.0930
		0.10	0.0661	0.0861	0.0621	0.1076	0.1141	0.0620	0.0830	0.0580	0.1010	0.1135
$\begin{pmatrix}0.900\\0.900\\0.900\\0.900\end{pmatrix}$	$\hat{F}_1$	0.00	0.0475	0.0770	0.0690	0.1190	0.1140	0.0445	0.0740	0.0715	0.1175	0.1115
		0.02	0.0840	0.1735	0.0735	0.2415	0.2565	0.0940	0.1875	0.0780	0.2655	0.2730
		0.04	0.2426	0.4422	0.0945	0.5103	0.5423	0.2160	0.4015	0.0910	0.4890	0.5125
		0.06	0.4475	0.6855	0.1120	0.7590	0.7955	0.4680	0.7060	0.1295	0.7610	0.7985
		0.08	0.7405	0.9185	0.1285	0.9380	0.9640	0.7210	0.9150	0.1605	0.9400	0.9595
		0.10	0.9270	0.9945	0.1830	0.9920	0.9995	0.9340	0.9900	0.1720	0.9935	0.9980
	$\hat{F}_2$	0.00	0.0465	0.0570	0.0475	0.0690	0.0760	0.0570	0.0660	0.0690	0.0700	0.0830
		0.02	0.0545	0.0700	0.0615	0.0720	0.0885	0.0570	0.0675	0.0645	0.0715	0.0890
		0.04	0.0450	0.0580	0.0595	0.0700	0.0755	0.0395	0.0550	0.0490	0.0610	0.0740
		0.06	0.0480	0.0565	0.0635	0.0645	0.0755	0.0430	0.0575	0.0565	0.0675	0.0785
		0.08	0.0460	0.0585	0.0625	0.0715	0.0755	0.0425	0.0560	0.0655	0.0615	0.0770
		0.10	0.0525	0.0675	0.0605	0.0775	0.0875	0.0550	0.0670	0.0545	0.0845	0.0825
	$\hat{F}_3$	0.00	0.0620	0.0740	0.0560	0.0835	0.0890	0.0435	0.0555	0.0675	0.0580	0.0720
		0.02	0.0385	0.0515	0.0600	0.0565	0.0695	0.0435	0.0530	0.0500	0.0645	0.0710
		0.04	0.0475	0.0570	0.0525	0.0670	0.0825	0.0425	0.0515	0.0600	0.0645	0.0660
		0.06	0.0490	0.0635	0.0705	0.0670	0.0830	0.0415	0.0530	0.0660	0.0610	0.0675
		0.08	0.0475	0.0615	0.0585	0.0690	0.0780	0.0430	0.0550	0.0535	0.0670	0.0740
		0.10	0.0545	0.0710	0.0630	0.0760	0.0875	0.0515	0.0685	0.0610	0.0800	0.0885
	$\hat{F}_4$	0.00	0.0460	0.0560	0.0545	0.0645	0.0690	0.0465	0.0560	0.0565	0.0635	0.0785
		0.02	0.0365	0.0475	0.0570	0.0580	0.0665	0.0430	0.0595	0.0640	0.0635	0.0775
		0.04	0.0425	0.0555	0.0595	0.0630	0.0730	0.0545	0.0680	0.0565	0.0770	0.0950
		0.06	0.0530	0.0700	0.0580	0.0750	0.0820	0.0540	0.0680	0.0510	0.0665	0.0875
		0.08	0.0520	0.0635	0.0590	0.0745	0.0810	0.0470	0.0645	0.0595	0.0780	0.0815
		0.10	0.0560	0.0685	0.0585	0.0855	0.0910	0.0525	0.0680	0.0630	0.0775	0.0925

Supplementary Appendix to "Factor-based IVX Predictive Regression"*

Fabian Schmidt[a]

[a] Department of Statistics, TU Dortmund University

June 9, 2026

Abstract

This supplementary appendix contains three sections. Section S.1 details proofs of the technical results in the main paper. Section S.2 provides additional Monte Carlo results referred to in section 4. Finally, section S.3 contains additional material related to the empirical application reported in section 5 of the main paper.

*The author is grateful for the computing time provided by the Institute for Statistics & Econometrics, Kiel University, and the Linux HPC cluster at TU Dortmund University (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as project 271512359. In addition, financial support from the German Research Foundation (DFG) within TRR 391: Spatio-temporal Statistics for the Transition of Energy and Transport (520388526) is gratefully acknowledged. Address correspondence to: Fabian Schmidt, Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany. Email: `f.schmidt@statistik.tu-dortmund.de`.

S.1 Mathematical Proofs

Before we turn our main result, we collect some auxiliary results for later reference. First, let C be a generic positive constant. Moreover, note that $\varrho \in (0,1)$ by definition. Hence,

$$\sum_{j=0}^{t} \varrho^{kj} = \frac{1-\varrho^{k(t+1)}}{1-\varrho^{k}} \leq CT^{\eta}$$

for any positive t and integer $k > 0$. Thirdly, the following Lemma carries over from Bai and Ng (2002),

Lemma S.1 *Given Assumptions 4.2 and 5, for each t and all T and N*

$$\mathrm{E}\left[\left\|\mathbf{e}_t'\boldsymbol{\Lambda}/\sqrt{N}\right\|^2\right] = \mathrm{E}\left[\left\|N^{-1/2}\sum_{i=1}^{N} e_{it}\boldsymbol{\lambda}_i\right\|^2\right] \leq C < \infty.$$

Proof of Lemma S.1 For the proof, see Bai and Ng (2002, Lemma 1 (ii)). □

First, we prove that our factor estimates are uniformly consistent regardless of the degree of persistence.

Proof of Lemma 1 We follow the proof of Bai (2004). First, we define the normalization $\tilde{\mathbf{F}}^{k\prime}\tilde{\mathbf{F}}^{k}/T^{\theta} = \mathbf{I}_k$ where $\theta = 1$ for stationary factors and $\theta = 2$ for local-to-unit-root/I(1) factors, allowing us to conclude that $T^{-\theta}\sum_{t=1}^{T}\|\tilde{\mathbf{F}}_t\|^2 = O_p(1)$ for all T. Next, we define $\hat{\mathbf{F}}^k = N^{-1}\mathbf{X}\tilde{\boldsymbol{\Lambda}}^k$, $\tilde{\boldsymbol{\Lambda}}^k = T^{-\theta}\mathbf{X}'\tilde{\mathbf{F}}^k$, as well as the matrix $\mathbf{S}^{k\prime} = \left(\tilde{\mathbf{F}}^{k\prime}\mathbf{F}/T^{\theta}\right)(\boldsymbol{\Lambda}'\boldsymbol{\Lambda}/N)$, $\theta \in \{1,2\}$, whose dependence on N and T will be suppressed for notational simplicity, subsequently. For the latter, note that $\|\mathbf{S}^k\| = O_p(1)$ since $\|\mathbf{S}^k\| \leq \|\tilde{\mathbf{F}}^{k\prime}\tilde{\mathbf{F}}^k/T^{\theta}\|^{1/2}\|\mathbf{F}'\mathbf{F}/T^{\theta}\|^{1/2}\|\boldsymbol{\Lambda}'\boldsymbol{\Lambda}/N\|$ because all individual matrix norms are stochastically bounded given Assumptions 1–6.

Given the above definitions, we obtain

$$\begin{aligned}
\hat{\mathbf{F}}_t^k - \mathbf{S}^{k\prime}\mathbf{F}_t &= T^{-2}\sum_{s=1}^{T}\tilde{\mathbf{F}}_s^k\gamma_N(s,t) + T^{-2}\sum_{s=1}^{T}\tilde{\mathbf{F}}_s^k\zeta_{st} + T^{-2}\sum_{s=1}^{T}\tilde{\mathbf{F}}_s^k\vartheta_{st} + T^{-2}\sum_{s=1}^{T}\tilde{\mathbf{F}}_s^k\xi_{st},\\
\zeta_{st} &= \frac{\mathbf{e}_s'\mathbf{e}_t}{N} - \gamma_N(s,t),\\
\vartheta_{st} &= \mathbf{F}_s'\boldsymbol{\Lambda}'\mathbf{e}_t/N,\\
\xi_{st} &= \mathbf{F}_t'\boldsymbol{\Lambda}'\mathbf{e}_s/N = \vartheta_{ts}
\end{aligned}$$

for near-integrated factors, which is the same as the expression Bai (2004) obtained for the special case of I(1) factors, *i.e.* $c_i = 0$ for $i = 1, \ldots, r$.

The norm of this decomposition then satisfies $\left\|\hat{\mathbf{F}}_t^k - \mathbf{S}^{k\prime}\mathbf{F}_t\right\|^2 \leq 4(a_t + b_t + c_t + d_t)$ where

$$a_t = T^{-4}\left\|\sum_{s=1}^{T}\tilde{\mathbf{F}}_s^k\gamma_N(s,t)\right\|^2, \qquad b_t = T^{-4}\left\|\sum_{s=1}^{T}\tilde{\mathbf{F}}_s^k\zeta_{st}\right\|^2$$
$$c_t = T^{-4}\left\|\sum_{s=1}^{T}\tilde{\mathbf{F}}_s^k\vartheta_{st}\right\|^2, \qquad d_t = T^{-4}\left\|\sum_{s=1}^{T}\tilde{\mathbf{F}}_s^k\xi_{st}\right\|^2$$

because $(x+y+z+u)^2 \leq 4(x^2+y^2+z^2+u^2)$. Hence, it suffices to examine the four components a_t, b_t, c_t, and d_t above.

Starting with a_t, we note that $\tilde{\mathbf{F}}_s^k$ and $\gamma_N(s,t)$ remain unchanged to their definition in Bai (2004, Proposition 1). Thus,

$$\begin{aligned} a_t &= T^{-4}\left\|\sum_{s=1}^{T}\tilde{\mathbf{F}}_s^k\gamma_N(s,t)\right\|^2 \leq \left(T^{-2}\sum_{s=1}^{T}\left\|\tilde{\mathbf{F}}_s^k\right\|^2\right)\left(T^{-2}\sum_{s=1}^{T}\gamma_N^2(s,t)\right) \\ &= O_p\left(T^{-2}\right). \end{aligned}$$

Now, recall that an absolutely summable sequence (x_m) is also square-summable because absolute summability, *i.e.* $\sum_{m=1}^{\infty}|x_m| < \infty$, implies $M := \sup_m |x_m| < \infty$. Hence, $|x_m|^2 \leq M|x_m|$ and therefore $\sum_{n=1}^{\infty} x_m^2 \leq M\sum_{n=1}^{\infty}|x_m| < \infty$. Consequently, thanks to $\sum_{s=1}^{T}|\gamma_N(s,t)| \leq C < \infty$ for all t and all T and N, and C being independent of t by Assumption 5.2, the a_t result thus also applies to $\max_{1\leq t\leq T} a_t$. Hence, a_t satisfies the same bound as in Bai (2004, Proposition 1).

Turning to b_t, we note that ζ_{st} did not change either. Now, proceeding along the lines of Bai and Ng (2002, Proof of Theorem 1) results in

$$\begin{aligned} b_t &= T^{-4}\left\|\sum_{s=1}^{T}\tilde{\mathbf{F}}_s^k\zeta_{st}\right\|^2 = T^{-4}\sum_{s=1}^{T}\sum_{u=1}^{T}\tilde{\mathbf{F}}_s^{k\prime}\tilde{\mathbf{F}}_u^k\zeta_{st}\zeta_{ut} \\ &\leq \left(T^{-4}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\tilde{\mathbf{F}}_s^{k\prime}\tilde{\mathbf{F}}_u^k\right)^2\right)^{1/2}\cdot\left(T^{-4}\sum_{s=1}^{T}\sum_{u=1}^{T}(\zeta_{st}\zeta_{ut})^2\right)^{1/2} \\ &= \left(T^{-2}\sum_{s=1}^{T}\left\|\tilde{\mathbf{F}}_s^k\right\|^2\right)\cdot\left(T^{-2}\sum_{s=1}^{T}\|\zeta_{st}\|^2\right). \end{aligned}$$

The first part is $O_p(1)$ uniformly in t by definition as it was for a_t. Next, note that $\|\zeta_{st}\|^2 = \zeta_{st}^2$ so that

$$\mathrm{E}\left[\zeta_{st}^2\right] = \mathrm{E}\left[\left|\mathbf{e}_s'\mathbf{e}_t/N - \gamma_N(s,t)\right|^2\right] \leq N^{-1}\,\mathrm{E}\left[\left|N^{-1/2}\sum_{i=1}^{N}(e_{is}e_{it} - \mathrm{E}\left[e_{is}e_{it}\right])\right|^2\right] \leq N^{-1}C$$

for all (t, s) and all T, N by Assumption 5.5 and the Lyapunov inequality. Thus, $b_t = O_p\left(1/(TN)\right)$ and trivially $O_p\left(T/N\right)$ for all (t, s), T, and N and, since C is again independent of t, this also applies to $\max_{1 \leq t \leq T} b_t$.

For c_t, we have

$$
\begin{aligned}
c_t = T^{-4} \left\| \sum_{s=1}^{T} \tilde{\mathbf{F}}_s^k \vartheta_{st} \right\|^2 &= T^{-4} \left\| \sum_{s=1}^{T} \tilde{\mathbf{F}}_s^k \mathbf{F}_s' \mathbf{\Lambda}' \mathbf{e}_t / N \right\|^2 \\
&\leq N^{-2} \left\| \mathbf{e}_t' \mathbf{\Lambda} \right\|^2 \left(T^{-2} \sum_{s=1}^{T} \left\| \tilde{\mathbf{F}}_s^k \right\|^2 \right) \left(T^{-2} \sum_{s=1}^{T} \left\| \mathbf{F}_s \right\|^2 \right) \\
&= N^{-1} \left\| \frac{\mathbf{e}_t' \mathbf{\Lambda}}{\sqrt{N}} \right\|^2 O_p(1)
\end{aligned}
$$

where $\left\| \mathbf{e}_t' \mathbf{\Lambda} / \sqrt{N} \right\|^2 = O_p(1)$ for each t and all T, N by Lemma S.1 and, thereby, also for $\max_{1 \leq t \leq T} \left\| \mathbf{e}_t' \mathbf{\Lambda} / \sqrt{N} \right\|^2$ because C in Lemma S.1 is independent of t. Since $\left(T^{-2} \sum_{s=1}^{T} \left\| \tilde{\mathbf{F}}_s^k \right\|^2 \right) = O_p(1)$ as usual and, in addition, the result in equation (4) and Prokhorov's theorem, see van der Vaart (1998, p. 8), yield $\left(T^{-2} \sum_{s=1}^{T} \left\| \mathbf{F}_s \right\|^2 \right) = O_p(1)$ which is independent of t, $c_t = O_p\left(N^{-1}\right)$ uniformly in t, too.

Finally, because $d_t = O_p\left(N^{-1}\right)$ follows by similar arguments, we omit the respective details.

The convergence rates for a_t, b_t, c_t, and d_t can thus be matched to those in Bai (2004, Proposition 1) so that $\left\| \hat{\mathbf{F}}_t - \mathbf{S}^{k\prime} \mathbf{F}_t \right\|^2 = O_p\left(T^{-2}\right) + O_p\left(T/N\right)$ for all t and, because the relevant sums in $a_t, \ldots, d_t$ are uniformly bounded in t by Assumption 5 and Lemma S.1, these rates consequently apply to $\max_{1 \leq t \leq T} \left\| \hat{\mathbf{F}}_t - \mathbf{S}^{k\prime} \mathbf{F}_t \right\|^2$ as well, which completes the proof of Lemma 1.2.

In case of weakly persistent predictors, the aforementioned arguments yield the result from Bai and Ng (2002, p. 198): $\max_{1 \leq t \leq T} \left\| \hat{\mathbf{F}}_t - \mathbf{S}^{k\prime} \mathbf{F}_t \right\| = O_p\left(1/\sqrt{T}\right) + O_p\left(1/\sqrt{N}\right)$.

However, one can obtain an even sharper convergence rate than this one. First, note that $\max_t |F_{it}| = O_p\left(T^{1/4}\right)$ according to Demetrescu *et al.* (2023, Suppl. Appx. p. S.8) for our weakly persistent factors. Because the setup of Bai (2003, p. 147) then applies given our assumptions, so does his result

$$
\max_{1 \leq t \leq T} \left\| \hat{\mathbf{F}}_t - \mathbf{S}^{k\prime} \mathbf{F}_t \right\| = O_p\left(T^{-1/2} \left(\min\left(\sqrt{T}, \sqrt{N} \right) \right)^{-1} \right) + O_p\left(\alpha_T T^{-1} \right) + O_p\left(\sqrt{T/N} \right),
$$

where $\max_t \left\| \mathbf{F}_t \right\| = O_p\left(\alpha_T\right)$, and we obtain the result given in Lemma 1.1 for $N \geq C\sqrt{T}$. □

Next, we provide a similar uniform consistency argument with respect to the IVX instruments generated from the estimated factors.

Lemma S.2 *Let the assumptions of Lemma 1 hold and let $T/\sqrt{N} \to 0$ as $N, T \to \infty$. Furthermore, let r be fixed. Given the matrix $\mathbf{S}^k$ defined in the proof of Lemma 1, it holds that*

$$\max_{1\le t\le T} \left\| \hat{\mathbf{z}}_t - \mathbf{S}^{k\prime} \mathbf{z}_t \right\| = o_p(1)$$

under both weak (Assumption 1.1) and strong (Assumption 1.2) persistence with the convergence rates carrying over from Lemma 1.

Proof of Lemma S.2 We generate the instruments according to (6) where the filter is applied row-wise to the factor estimate vector. Thus, we obtain $\hat{\mathbf{z}}_t = \sum_{j=0}^{t-1} \varrho^j \Delta \hat{\mathbf{F}}_{t-j}$ where Δ is again interpreted row-wise. Suitably rewriting this expression results in

$$\hat{\mathbf{z}}_t = \hat{\mathbf{F}}_t - \varrho^{t-2} \hat{\mathbf{F}}_1 + (\varrho - 1) \sum_{j=0}^{t-3} \varrho^j \hat{\mathbf{F}}_{t-2-j}, \tag{S.1}$$

see Demetrescu and Hillmann (2022, Suppl. Appx. p. xvii). Next, expanding $\hat{\mathbf{F}}_t$ in equation (S.1) by $\pm \mathbf{S}^{k\prime} \mathbf{F}_t$ yields

$$\begin{aligned} \hat{\mathbf{z}}_t = \mathbf{S}^{k\prime}\mathbf{F}_t + \left(\hat{\mathbf{F}}_t - \mathbf{S}^{k\prime}\mathbf{F}_t\right) - \varrho^{t-2}\left[\mathbf{S}^{k\prime}\mathbf{F}_1 + \left(\hat{\mathbf{F}}_1 - \mathbf{S}^{k\prime}\mathbf{F}_1\right)\right] \\ - \frac{1}{T^\eta}\sum_{j=0}^{t-3}\varrho^j\left[\mathbf{S}^{k\prime}\mathbf{F}_{t-2-j} + \left(\hat{\mathbf{F}}_{t-2-j} - \mathbf{S}^{k\prime}\mathbf{F}_{t-2-j}\right)\right], \end{aligned}$$

allowing us to split the r.h.s. into the rotated true instruments $\mathbf{S}^{k\prime}\mathbf{z}_t$ plus some estimation error. Then, the following holds for the estimation error:

$$\left\|\hat{\mathbf{z}}_t - \mathbf{S}^{k\prime}\mathbf{z}_t\right\| \le \left\|\hat{\mathbf{F}}_t - \mathbf{S}^{k\prime}\mathbf{F}_t\right\| + \varrho^{t-2}\left\|\hat{\mathbf{F}}_1 - \mathbf{S}^{k\prime}\mathbf{F}_1\right\| + \frac{1}{T^\eta}\sum_{j=0}^{t-3}\varrho^j\left\|\hat{\mathbf{F}}_{t-2-j} - \mathbf{S}^{k\prime}\mathbf{F}_{t-2-j}\right\|.$$

Now, note that $\varrho^{t-2} = (1 - 1/T^\eta)^{t-2} \le 1$ so that

$$\max_{1\le t\le T}\left\|\hat{\mathbf{z}}_t - \mathbf{S}^{k\prime}\mathbf{z}_t\right\| \le \max_{1\le t\le T}\left\|\hat{\mathbf{F}}_t - \mathbf{S}^{k\prime}\mathbf{F}_t\right\| + \max_{1\le t\le T}\left\|\hat{\mathbf{F}}_t - \mathbf{S}^{k\prime}\mathbf{F}_t\right\| + \max_{1\le t\le T}\left\|\hat{\mathbf{F}}_t - \mathbf{S}^{k\prime}\mathbf{F}_t\right\|\frac{1}{T^\eta}\sum_{j=0}^{t-3}\varrho^j$$

given the results in Lemma 1 and $\sum_{j=0}^{t}\varrho^j = O(T^\eta)$. Thus, the convergence rates carry over from the factor estimates and

$$\max_{1\le t\le T}\left\|\hat{\mathbf{z}}_t - \mathbf{S}^{k\prime}\mathbf{z}_t\right\| = \begin{cases} O_p\left(T^{-3/4}\right) + O_p\left(\sqrt{T/N}\right) & \text{given Assumption 1.1} \\ O_p\left(T^{-1}\right) + O_p\left(\sqrt{T/N}\right) & \text{given Assumption 1.2} \end{cases}$$

which completes the proof. □

Ultimately, our test statistics will not need the factor estimates or their instruments on their own. Therefore, Lemmata S.3 and S.4 collect intermediate results with respect to relevant summation expressions that contain estimated factors and instruments. Starting with the weakly persistent case, the following holds:

Lemma S.3 *Let Assumptions 1.1, 2–6 hold. Moreover, let $T/\sqrt{N} \to 0$ as $N, T \to \infty$. Then*

1. $\left\| \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} (y_t - \bar{y}) - \mathbf{S}^{k\prime} \sum_{t=2}^{T} \mathbf{z}_{t-1} u_t \right\| = o_p\left(\sqrt{T}\right)$ *given* H_0.
2. $\left\| \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} \hat{\mathbf{z}}_{t-1}^{\prime} (y_t - \bar{y})^2 - \mathbf{S}^{k\prime} \left(\sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{z}_{t-1}^{\prime} u_t^2 \right) \mathbf{S}^{k} \right\| = o_p(T)$ *given* H_0,
3. $\left\| \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} \hat{\mathbf{F}}_{t-1}^{\prime} - \mathbf{S}^{k\prime} \left(\sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{F}_{t-1}^{\prime} \right) \mathbf{S}^{k} \right\| = o_p(T)$,
4. $\left\| \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} y_t - \mathbf{S}^{k\prime} \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{F}_{t-1}^{\prime} \boldsymbol{\beta} \right\| = o_p(T)$.

Proof of Lemma S.3

1. Expanding $\hat{\mathbf{z}}_{t-1} (y_t - \bar{y})$ by $\pm \mathbf{S}^{k\prime} \mathbf{z}_{t-1}$ and plugging in $\bar{y} = T^{-1} \sum_{t=2}^{T} y_t$ yields

$$\hat{\mathbf{z}}_{t-1} (y_t - \bar{y}) = \left(\mathbf{S}^{k\prime} \mathbf{z}_{t-1} + \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k\prime} \mathbf{z}_{t-1} \right) \right) \left(u_t - T^{-1} \sum_{t=2}^{T} u_t \right)$$

under the null. We then get

$$\begin{aligned} \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} (y_t - \bar{y}) - \mathbf{S}^{k\prime} \mathbf{z}_{t-1} u_t \right) = & - \sum_{t=2}^{T} \mathbf{S}^{k\prime} \mathbf{z}_{t-1} \bar{u} + \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k\prime} \mathbf{z}_{t-1} \right) u_t \\ & - \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k\prime} \mathbf{z}_{t-1} \right) \bar{u} \end{aligned} \tag{S.2}$$

where $\bar{u} = T^{-1} \sum_{t=2}^{T} u_t$ is $O_p\left(T^{-1/2}\right) = o_p(1)$ thanks to u_t being an md sequence with uniformly L_4-bounded increments. Examining the norm of (S.2), since the factors are identified only up to rotation, $\left\| \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} (y_t - \bar{y}) - \mathbf{S}^{k\prime} \sum_{t=2}^{T} \mathbf{z}_{t-1} u_t \right\| \leq (I + II + III)$ where

$$\begin{aligned} I &= \left\| \mathbf{S}^{k\prime} \bar{u} \sum_{t=2}^{T} \mathbf{z}_{t-1} \right\|, \\ II &= \left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k\prime} \mathbf{z}_{t-1} \right) u_t \right\|, \\ III &= \left\| \bar{u} \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k\prime} \mathbf{z}_{t-1} \right) \right\|. \end{aligned}$$

Starting with I, we have

$$I \leq \left\| \mathbf{S}^{k'} \right\| |\bar{u}| \left\| \sum_{t=1}^{T} \mathbf{z}_{t-1} \right\| \leq O_p\left(T^{-1/2}\right) \left\| \sum_{t=1}^{T} \mathbf{z}_{t-1} \right\|$$

since $\left\| \mathbf{S}^{k'} \right\| = O_p(1)$ as usual. Now, we carry over the univariate finding that z_t equals F_t plus some rest term (under weak persistence) that can be controlled in relevant sums and vanishes as $T \to \infty$ (Demetrescu *et al.*, 2023, Proof of Proposition 1.1, Suppl. Appx. p. S.8). This then yields $T^{-1/2} \sum_{t=2}^{T} \mathbf{z}_{t-1} = T^{-1/2} \sum_{t=2}^{T} \mathbf{F}_{t-1} + o_p(1)$ where $T^{-1/2} \sum_{t=2}^{T} \mathbf{F}_{t-1} = O_p(1)$; cf. Demetrescu *et al.* (2023, Suppl. Appx. p. S.8) among others. Thus, $I = o_p\left(\sqrt{T}\right)$ under the null.

Turning to II, recall that $\max_{1 \leq t \leq T} \left\| \hat{\mathbf{F}}_t - \mathbf{S}^{k'} \mathbf{F}_t \right\| = O_p\left(T^{-3/4}\right) + O_p\left(\sqrt{T/N}\right)$ given our assumptions. Since this rate carries over to $\max_{1 \leq t \leq T} \left\| \hat{\mathbf{z}}_t - \mathbf{S}^{k'} \mathbf{z}_t \right\|$ by the arguments in the proof of Lemma S.2, one obtains

$$\begin{aligned} II &\leq \sum_{t=1}^{T} \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| |u_t| \leq \max_{1 \leq t \leq T} \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| \sum_{t=1}^{T} |u_t| \\ &= \left(O_p\left(T^{-3/4}\right) + O_p\left(\sqrt{T/N}\right) \right) O_p(T) \\ &= O_p\left(T^{1/4}\right) + O_p\left(T^{3/2}/N^{1/2}\right) \end{aligned}$$

due to the moment properties of u_t. Hence, II is $o_p\left(T^{1/2}\right)$ as long as $T/\sqrt{N} \to 0$.

Lastly, it holds for III holds that

$$\begin{aligned} \left\| \bar{u} \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right) \right\| &\leq O_p\left(T^{-1/2}\right) \cdot T \cdot \max_{1 \leq t \leq T} \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| \\ &= O_p\left(T^{-1/2}\right) o_p(T) \end{aligned}$$

using the result of Lemma S.2 and the requirement from II that N needs to got to infinity sufficiently faster than T. Thus, $III = o_p\left(T^{1/2}\right)$ as well which completes the proof. □

2. Again, we expand $\hat{\mathbf{z}}_{t-1}$ by $\pm\mathbf{S}^{k'}\mathbf{z}_{t-1}$ and replace $y_t, \bar{y}$ by $u_t, \bar{u}$ since we continue to work under the null. As a result, we obtain

$$\begin{aligned}\sum\nolimits_{t=2}^{T}\hat{\mathbf{z}}_{t-1}\hat{\mathbf{z}}_{t-1}'\left(y_t-\bar{y}_t\right)^2 = \mathbf{S}^{k'}\left(\sum\nolimits_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'u_t^2\right)\mathbf{S}^k - 2\bar{u}\mathbf{S}^{k'}\left(\sum\nolimits_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'u_t\right)\mathbf{S}^k \\ + \bar{u}^2\mathbf{S}^{k'}\left(\sum\nolimits_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)\mathbf{S}^k \\ + \mathbf{S}^{k'}\sum\nolimits_{t=2}^{T}\mathbf{z}_{t-1}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)'u_t^2 - 2\bar{u}\mathbf{S}^{k'}\sum\nolimits_{t=2}^{T}\mathbf{z}_{t-1}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)'u_t \\ + \bar{u}^2\mathbf{S}^{k'}\sum\nolimits_{t=2}^{T}\mathbf{z}_{t-1}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)' \\ + \sum\nolimits_{t=2}^{T}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)\mathbf{z}_{t-1}'u_t^2\mathbf{S}^k - 2\bar{u}\sum\nolimits_{t=2}^{T}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)\mathbf{z}_{t-1}'u_t\mathbf{S}^k \\ + \bar{u}^2\sum\nolimits_{t=2}^{T}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)\mathbf{z}_{t-1}'\mathbf{S}^k \\ + \sum\nolimits_{t=2}^{T}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)'u_t^2 \\ - 2\bar{u}\sum\nolimits_{t=2}^{T}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)'u_t \\ + \bar{u}^2\sum\nolimits_{t=2}^{T}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)'.\end{aligned} \tag{S.3}$$

Now, moving $\mathbf{S}^{k'}\left(\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'u_t^2\right)\mathbf{S}^k$ to the l.h.s. results in

$$\left\|\sum_{t=2}^{T}\hat{\mathbf{z}}_{t-1}\hat{\mathbf{z}}_{t-1}'\left(y_t-\bar{y}_t\right)^2 - \mathbf{S}^{k'}\left(\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'u_t^2\right)\mathbf{S}^k\right\| \leq \widetilde{I} + \widetilde{II} + \ldots + \widetilde{XI}$$

whose individual components are examined next.

Starting with $\widetilde{I}$, note that

$$\begin{aligned}\widetilde{I} = \left\|\bar{u}\mathbf{S}^{k'}\left(\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'u_t\right)\mathbf{S}^k\right\| &\leq |\bar{u}|\left\|\mathbf{S}^{k'}\right\|\left\|\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'u_t\right\|\left\|\mathbf{S}^{k'}\right\| \\ &\leq O_p\left(T^{-1/2}\right)\sum_{t=2}^{T}\left\|\mathbf{z}_{t-1}\mathbf{z}_{t-1}'u_t\right\| \leq O_p\left(T^{-1/2}\right)\sum_{t=2}^{T}\left\|\mathbf{z}_{t-1}\right\|\left\|\mathbf{z}_{t-1}u_t\right\| \\ &\leq O_p\left(T^{-1/2}\right)\sqrt{\left(\sum_{t=2}^{T}\left\|\mathbf{z}_{t-1}\right\|^2\right)\left(\sum_{s=2}^{T}\left\|\mathbf{z}_{s-1}u_s\right\|^2\right)}\end{aligned}$$

thanks to the Cauchy-Schwarz inequality. Then, $\sum_{t=2}^{T}\left\|\mathbf{z}_{t-1}\right\|^2 = \mathrm{tr}\left(\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)$ where $\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}' = O_p(T)$ analogously to the univariate finding of Demetrescu *et al.* (2023, Suppl. Appx. p. S.8). Likewise, $\sum_{t=2}^{T}\left\|\mathbf{z}_{t-1}u_t\right\|^2 = \mathrm{tr}\left(\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'u_t^2\right) = O_p(T)$ since

$\sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}' u_t^2$ is, carrying over the univariate result of Demetrescu *et al.* (2023, Suppl. Appx. p. S.10), too. Therefore, $\widetilde{I} = o_p(T)$.

Continuing with $\widetilde{II}$, we find

$$\widetilde{II} = \left\| \bar{u}^2 \mathbf{S}^{k'} \left(\sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}' \right) \mathbf{S}^{k} \right\| \leq \bar{u}^2 \left\| \mathbf{S}^{k} \right\| \left\| \sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}' \right\| \left\| \mathbf{S}^{k} \right\|.$$

Since it still holds that $\bar{u} = O_p\left(T^{-1/2}\right)$, $\left\| \mathbf{S}^{k'} \right\| = O_p(1)$, and $\sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}' = O_p(T)$, it follows that $\widetilde{II} = o_p(T)$.

Now, focusing on $\widetilde{III}$, observe that

$$\widetilde{III} = \left\| \mathbf{S}^{k'} \sum_{t=2}^{T} \mathbf{z}_{t-1} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right)' u_t^2 \right\| \leq O_p(1) \sum_{t=2}^{T} \left\| \mathbf{z}_{t-1} \right\| \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| u_t^2$$

which is smaller equal than

$$O_p(1) \max_t \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| \sum_{t=2}^{T} \left\| \mathbf{z}_{t-1} \right\| u_t^2 \leq o_p(1) \sqrt{ \left(\sum_{t=2}^{T} \left\| \mathbf{z}_{t-1} \right\|^2 \right) \left(\sum_{s=2}^{T} u_s^4 \right) }.$$

For the expressions on the latter inequality's r.h.s. it now holds that $\sum_{t=2}^{T} \left\| \mathbf{z}_{t-1} \right\|^2 = O_p(T)$ as before and $\sum_{t=2}^{T} u_t^4 = O_p(T)$ by the L_4-boundedness of u_t. Hence, $\widetilde{III} = o_p(T)$. Because the Frobenius norm is self-adjoint, this result directly applies to $\widetilde{VI}$ as well.

Continuing with $\widetilde{IV}$, it turns out that

$$\begin{aligned} \widetilde{IV} &= \left\| \bar{u} \mathbf{S}^{k'} \sum_{t=2}^{T} \mathbf{z}_{t-1} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right)' u_t \right\| \leq O_p\left(T^{-1/2}\right) \sum_{t=2}^{T} \left\| \mathbf{z}_{t-1} u_t \right\| \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| \\ &\leq O_p\left(T^{-1/2}\right) \max_t \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| \sum_{t=2}^{T} \left\| \mathbf{z}_{t-1} u_t \right\|. \end{aligned}$$

Thus, $\widetilde{IV} = o_p(T)$ since $\sum_{t=2}^{T} \left\| \mathbf{z}_{t-1} u_t \right\| \leq \sqrt{T \sum_{t=2}^{T} \left\| \mathbf{z}_{t-1} u_t \right\|^2} = \sqrt{T \operatorname{tr} \left(\sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}' u_t^2 \right)}$ with the same arguments applying to $\widetilde{VII} = o_p(T)$ due to the properties of the Frobenius norm.

For $\widetilde{V}$ and $\widetilde{VIII}$,

$$
\begin{aligned}
\widetilde{V} &\leq \left\| \bar{u}^2 \mathbf{S}^{k'} \sum_{t=2}^{T} \mathbf{z}_{t-1} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right)' \right\| \leq \left| \bar{u}^2 \right| \left\| \mathbf{S}^{k'} \right\| \left\| \sum_{t=2}^{T} \mathbf{z}_{t-1} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right) \right\| \\
&\leq \left| \bar{u}^2 \right| \left\| \mathbf{S}^{k'} \right\| \max_t \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| \sum_{t=2}^{T} \left\| \mathbf{z}_{t-1} \right\| \\
&= o_p(T)
\end{aligned}
$$

using already known arguments.

Afterwards, examining $\widetilde{IX}$ shows that

$$
\begin{aligned}
\widetilde{IX} &= \left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right) \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right)' u_t^2 \right\| \\
&\leq \sum_{t=2}^{T} \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| u_t^2 \\
&\leq \left(\max_t \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| \right)^2 \sum_{t=2}^{T} u_t^2,
\end{aligned}
$$

i.e. $\widetilde{IX} = o_p(T)$ because $T^{-1} \sum_{t=2}^{T} u_t^2 = O_p(1)$, cf. Demetrescu *et al.* (2023, Lemma 3).

Continuing with $\widetilde{X}$, it turns out that

$$
\begin{aligned}
\widetilde{X} &= \left\| \bar{u} \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right) \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right)' u_t \right\| \\
&\leq \left| \bar{u} \right| \left(\max_t \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| \right)^2 \sum_{t=2}^{T} \left| u_t \right|
\end{aligned}
$$

so that $\widetilde{X} = o_p(T)$ by aforementioned findings, too.

Finally,

$$
\begin{aligned}
\widetilde{XI} &= \left\| \bar{u}^2 \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right) \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right)' \right\| \\
&\leq \bar{u}^2 \cdot T \cdot \left(\max_t \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| \right)^2 \\
&= o_p(T)
\end{aligned}
$$

which completes the proof. □

3. Expanding $\sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1}\hat{\mathbf{F}}'_{t-1}$ by $\pm\mathbf{S}^{k'}\mathbf{z}_t$ and $\pm\mathbf{S}^{k'}\mathbf{F}_t$ results in

$$\begin{aligned}\left\| \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1}\hat{\mathbf{F}}'_{t-1} - \mathbf{S}^{k'} \left(\sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{F}'_{t-1} \right) \mathbf{S}^{k} \right\| \\ \leq \left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'}\mathbf{z}_{t-1} \right) \mathbf{F}'_{t-1} \right\| + \left\| \sum_{t=2}^{T} \mathbf{S}^{k'}\mathbf{z}_{t-1} \left(\hat{\mathbf{F}}_{t-1} - \mathbf{S}^{k'}\mathbf{F}_{t-1} \right)' \right\| \\ + \left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'}\mathbf{z}_{t-1} \right) \left(\hat{\mathbf{F}}_{t-1} - \mathbf{S}^{k'}\mathbf{F}_{t-1} \right)' \right\|.\end{aligned} \tag{S.4}$$

Now, arguments analog to the ones in the proofs of Lemmata S.3.1 and S.3.2 yield

$$\left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'}\mathbf{z}_{t-1} \right) \mathbf{F}'_{t-1} \right\| \leq \max_t \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'}\mathbf{z}_{t-1} \right\| \sum_{t=2}^{T} \left\| \mathbf{F}_{t-1} \right\| = o_p(1) \sum_{t=2}^{T} \left\| \mathbf{F}_{t-1} \right\|$$

where $\sum_{t=2}^{T} \left\| \mathbf{F}_{t-1} \right\| = O_p(T)$ can be inferred from the Cauchy-Schwarz inequality and multivariate counterparts to the univariate results of Demetrescu *et al.* (2023, Suppl. Appx. p. S.8).

Moreover, we obtain

$$\begin{aligned}\left\| \sum_{t=2}^{T} \mathbf{S}^{k'}\mathbf{z}_{t-1} \left(\hat{\mathbf{F}}_{t-1} - \mathbf{S}^{k'}\mathbf{F}_{t-1} \right)' \right\| \leq \max_t \left\| \hat{\mathbf{F}}_{t-1} - \mathbf{S}^{k'}\mathbf{F}_{t-1} \right\| \left\| \mathbf{S}^{k'} \right\| \sum_{t=2}^{T} \left\| \mathbf{z}_{t-1} \right\| \\ \leq o_p(1) \sum_{t=2}^{T} \left\| \mathbf{z}_{t-1} \right\|\end{aligned}$$

where the last part is $O_p(T)$ by the Cauchy-Schwarz inequality, again; cf. Demetrescu *et al.* (2023, Suppl. Appx. p. S.8) for univariate results extending to the multivariate case.

Finally,

$$\begin{aligned}\left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'}\mathbf{z}_{t-1} \right) \left(\hat{\mathbf{F}}_{t-1} - \mathbf{S}^{k'}\mathbf{F}_{t-1} \right)' \right\| \\ \leq \sum_{t=2}^{T} \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'}\mathbf{z}_{t-1} \right\| \left\| \hat{\mathbf{F}}_{t-1} - \mathbf{S}^{k'}\mathbf{F}_{t-1} \right\| \\ = o_p(T)\end{aligned}$$

thanks to the uniform consistency results in Lemmata 1 and S.2 for $T/\sqrt{N} \to 0$, thereby completing the proof. □

4. Our examination of $\sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} y_t$ is again built around an expansion, this time using $\pm \mathbf{S}^{k'} \mathbf{z}_t$ and $y_t = \mathbf{F}'_{t-1}\boldsymbol{\beta} + u_t$. Consequently, we obtain

$$\begin{aligned} \left\| \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} y_t - \mathbf{S}^{k'} \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{F}'_{t-1} \boldsymbol{\beta} \right\| &\leq \left\| \mathbf{S}^{k'} \sum_{t=2}^{T} \mathbf{z}_{t-1} u_t \right\| \\ &+ \left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right) \mathbf{F}'_{t-1} \boldsymbol{\beta} \right\| + \left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right) u_t \right\| . \end{aligned} \tag{S.5}$$

An examination of the first summand yields

$$\left\| \mathbf{S}^{k'} \sum_{t=2}^{T} \mathbf{z}_{t-1} u_t \right\| \leq O_p(1) \left\| \sum_{t=2}^{T} \mathbf{z}_{t-1} u_t \right\| = o_p(T)$$

since $\sum_{t=2}^{T} \mathbf{z}_{t-1} u_t = O_p\left(\sqrt{T}\right)$ given Assumption 1.1 after carrying over the univariate result of Demetrescu *et al.* (2023, Suppl. Appx. p. S.8) to our multivariate setup.

The second summand then satisfies

$$\left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right) \mathbf{F}'_{t-1} \boldsymbol{\beta} \right\| \leq \max_t \left\| \hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right\| \|\boldsymbol{\beta}\| \sum_{t=2}^{T} \left\| \mathbf{F}_{t-1} \right\| .$$

Since $\sum_{t=2}^{T} \left\| \mathbf{F}_{t-1} \right\| = O_p(T)$, see above, and b_i is finite so that $\|\boldsymbol{\beta}\| = O\left(T^{-1/2}\right)$ by the definition in Assumption 2, the second summand is $o_p\left(T^{1/2}\right)$.

Finally, $\left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right) u_t \right\| = o_p(T)$ as before which completes the proof. □

Meanwhile, the following holds under strong persistence.

Lemma S.4 *Let Assumptions 1.2, 2–6 hold and let $T/\sqrt{N} \to 0$ as $N, T \to \infty$, then*

1. $\left\| \sum_{t=1}^{T} \hat{\mathbf{z}}_{t-1} (y_t - \bar{y}) - \mathbf{S}^{k'} \sum_{t=1}^{T} \mathbf{z}_{t-1} u_t \right\| = o_p\left(T^{1/2+\eta/2}\right)$ *given* H_0.
2. $\left\| \sum_{t=1}^{T} \hat{\mathbf{z}}_{t-1} \hat{\mathbf{z}}'_{t-1} (y_t - \bar{y})^2 - \mathbf{S}^{k'} \left(\sum_{t=1}^{T} \mathbf{z}_{t-1} \mathbf{z}'_{t-1} u_t^2 \right) \mathbf{S}^{k} \right\| = o_p\left(T^{1+\eta}\right)$ *given* H_0.
3. $\left\| \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} \hat{\mathbf{F}}'_{t-1} - \mathbf{S}^{k'} \left(\sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{F}'_{t-1} \right) \mathbf{S}^{k} \right\| = o_p\left(T^{1+\eta}\right)$ *for* $\eta \geq 1/2$.
4. $\left\| \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} y_t - \mathbf{S}^{k'} \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{F}'_{t-1} \boldsymbol{\beta} \right\| = o_p\left(T^{1+\eta}\right)$.

Proof of Lemma S.4

1. Like in Lemma S.3.1, we start from the expansion in (S.2). For local-to-unit roots factors, *i.e.* under Assumption 1.2, this results in

$$I \leq \left\|\mathbf{S}^{k'}\right\| \left\|\bar{u}\sum_{t=1}^{T}\mathbf{z}_{t-1}\right\| \leq O_p\left(T^{-1/2}\right)\cdot\left\|\sum_{t=1}^{T}\mathbf{z}_{t-1}\right\|$$

since $\left\|\mathbf{S}^{k'}\right\| = O_p(1)$ still holds. Next, carrying over the univariate result of Demetrescu *et al.* (2023, Lemma 5(a)) to the multivariate case provides $\sum_{t=1}^{T}\mathbf{z}_{t-1} = O_p\left(T^{1/2+\eta}\right)$. Therefore,

$$I = O_p\left(T^{-1/2}\right)O_p\left(T^{1/2+\eta}\right) = O_p\left(T^{\eta}\right) = o_p\left(T^{1/2+\eta/2}\right).$$

For II, recall that $\max_{1\leq t\leq T}\left\|\hat{\mathbf{F}}_t - \mathbf{S}^{k'}\mathbf{F}_t\right\| = O_p\left(T^{-1}\right) + O_p\left(\sqrt{T/N}\right)$ with this rate carrying over to $\max_{1\leq t\leq T}\left\|\hat{\mathbf{z}}_t - \mathbf{S}^{k'}\mathbf{z}_t\right\|$ by the arguments in the proof of Lemma S.2.

$$\begin{aligned} II &\leq \sum_{t=1}^{T}\left\|\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'}\mathbf{z}_{t-1}\right\| |u_t| \leq \max_{1\leq t\leq T}\left\|\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'}\mathbf{z}_{t-1}\right\|\sum_{t=1}^{T}|u_t| \\ &= \left(O_p\left(T^{-1}\right) + O_p\left(\sqrt{T/N}\right)\right)O_p(T) \\ &= O_p(1) + O_p\left(T^{3/2}/N^{1/2}\right). \end{aligned}$$

Thus, II is $o_p\left(T^{1/2+\eta/2}\right)$ as long as N grows sufficiently fast, *i.e.* as long as $T^{1-\eta/2}/\sqrt{N} \to 0$ (which it has to anyway due to the factors' unknown degree of persistence and the weakly persistent alternative requiring $T/\sqrt{N} \to 0$).

Finally,

$$\begin{aligned} III &= \left\|\sum_{t=1}^{T}\left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'}\mathbf{z}_{t-1}\right)\bar{u}\right\| \leq \max_{1\leq t\leq T}\left\|\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'}\mathbf{z}_{t-1}\right\|\sum_{t=1}^{T}|\bar{u}| \\ &= o_p\left(T^{1/2+\eta/2}\right) \end{aligned}$$

for N going faster to infinity than T which completes the proof. □

2. Under the assumption of local-to-unity factors, starting from the expansion in (S.3) allows us to proceed analogously to the weakly persistent case in Lemma S.3.2 and we thus leave out the proof to avoid repitition. □

3. An examination of the first summand in (S.4), cf. the proof of Lemma S.3.3, under strong persistence yields

$$\left\|\sum_{t=2}^{T}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)\mathbf{F}_{t-1}'\right\| \leq \max_t \left\|\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right\| \sum_{t=2}^{T}\left\|\mathbf{F}_{t-1}\right\|$$
$$= o_p(1)\sum_{t=2}^{T}\left\|\mathbf{F}_{t-1}\right\|$$

where $\sum_{t=2}^{T}\left\|\mathbf{F}_{t-1}\right\| = O_p\left(T^{3/2}\right)$ by the Cauchy-Schwarz inequality and the result in equation (4). Thus, the first part turns out to be $o_p\left(T^{1+\eta}\right)$ as long as $\eta \geq 0.5$.

For the second summand, it holds that

$$\left\|\sum_{t=2}^{T}\mathbf{S}^{k'}\mathbf{z}_{t-1}\left(\hat{\mathbf{F}}_{t-1}-\mathbf{S}^{k'}\mathbf{F}_{t-1}\right)'\right\| \leq \max_t \left\|\hat{\mathbf{F}}_{t-1}-\mathbf{S}^{k'}\mathbf{F}_{t-1}\right\| \left\|\mathbf{S}^{k'}\right\| \sum_{t=2}^{T}\left\|\mathbf{z}_{t-1}\right\|$$
$$\leq o_p(1) O_p\left(T^{1+\eta/2}\right)$$

by the uniform consistency of the factor estimates for $T/N \to 0$. Thus, the second summand is $o_p\left(T^{1+\eta}\right)$, too.

Lastly, the third summand still satisfies

$$\left\|\sum_{t=2}^{T}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)\left(\hat{\mathbf{F}}_{t-1}-\mathbf{S}^{k'}\mathbf{F}_{t-1}\right)'\right\| = o_p(T)$$

by the same arguments as in Lemma S.3.3, thereby completing the proof. □

4. Finally, re-investigating the expansion in (S.5), see the proof of Lemma S.3.4, yields

$$\left\|\mathbf{S}^{k'}\sum_{t=2}^{T}\mathbf{z}_{t-1}u_t\right\| \leq O_p(1)\left\|\sum_{t=2}^{T}\mathbf{z}_{t-1}u_t\right\| = o_p\left(T^{1+\eta}\right)$$

because $\sum_{t=2}^{T}\mathbf{z}_{t-1}u_t = O_p\left(T^{1/2+\eta/2}\right)$ given Assumption 1.2 and the multivariate counterpart to Demetrescu *et al.* (2023, Lemma 5(d)).

The convergence rate for the second summand changes as well. Starting from

$$\left\|\sum_{t=2}^{T}\left(\hat{\mathbf{z}}_{t-1}-\mathbf{S}^{k'}\mathbf{z}_{t-1}\right)\mathbf{F}_{t-1}'\boldsymbol{\beta}\right\| \leq o_p(1)\left\|\boldsymbol{\beta}\right\|\sum_{t=2}^{T}\left\|\mathbf{F}_{t-1}\right\|,$$

the Cauchy-Schwarz inequality leads to

$$\sum_{t=2}^{T} \left\| \mathbf{F}_{t-1} \right\| \leq \sqrt{T \sum_{t=2}^{T} \left\| \mathbf{F}_{t-1} \right\|^2}.$$

Then, $\sum_{t=2}^{T} \mathbf{F}_{t-1}\mathbf{F}'_{t-1} = O_p(T^2)$, b_i being finite, and $\boldsymbol{\beta} = \mathbf{b}/T^{1/2+\eta/2}$ by Assumption 2 together yield

$$\|\boldsymbol{\beta}\| \sum_{t=2}^{T} \left\| \mathbf{F}_{t-1} \right\| = O_p\left(T^{-1/2-\eta/2} T^{3/2}\right) = O_p\left(T^{1-\eta/2}\right)$$

from which it is straightforward to conclude that the second summand is $o_p\left(T^{1+\eta}\right)$.

Finally, $\left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'} \mathbf{z}_{t-1} \right) u_t \right\| = o_p(T)$ has been found earlier, which completes the proof. □

Now, proofs of our main results are presented:

Proof of Proposition 1 Thanks to Lemmata S.3 and S.4, we know that

$$\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \hat{\mathbf{z}}_{t-1} \hat{\mathbf{F}}'_{t-1} = \mathbf{S}^{k'} \left(\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \mathbf{z}_{t-1} \mathbf{F}'_{t-1} \right) \mathbf{S}^{k} + o_p(1)$$

and

$$\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \hat{\mathbf{z}}_{t-1} y_t = \mathbf{S}^{k'} \left(\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \mathbf{z}_{t-1} \mathbf{F}'_{t-1} \boldsymbol{\beta} \right) + o_p(1)$$

for $T/N \to 0$ as $N, T \to 0$ and $\eta \geq 0.5$, where $\theta = 1$ ($\theta = 2$) when the factors are weakly (strongly) persistent. The result is then directly obtained from the standard IV estimator and an application of Slutsky's theorem. □

Proof of Proposition 2 We prove the result in two steps; first, we show that our estimated test statistic converges to an expression solely based on rotated versions of their true counterparts, and then, we show that the rotation does not matter to the Wald test statistic.

From Lemmata S.3 and S.4, we know that

$$\frac{1}{T^{1/2+(\theta-1)\eta/2}} \sum_{t=1}^{T} \hat{\mathbf{z}}_{t-1} \hat{u}_t = \mathbf{S}^{k'} \left(\frac{1}{T^{1/2+(\theta-1)\eta/2}} \sum_{t=1}^{T} \mathbf{z}_{t-1} u_t \right) + o_p(1)$$

and

$$\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \hat{\mathbf{z}}_{t-1} \hat{\mathbf{z}}'_{t-1} \hat{u}_t^2 = \mathbf{S}^{k'} \left(\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \mathbf{z}_{t-1} \mathbf{z}'_{t-1} u_t^2 \right) \mathbf{S}^{k} + o_p(1)$$

under the null where $T/\sqrt{N} \to 0$ as $N, T \to 0$ and $\theta \in \{1,2\}$ depends on the degree of persistence as in Proposition 1. Up to the rotation matrix, the convergence limits of the r.h.s. expressions are given by the multivariate counterparts of the results in Demetrescu *et al.* (2023); see their results in Lemma 4 plus the first part of the proof of Proposition 2 for stationary factors and Lemma 5 (d) plus the second part of the proof of Proposition 2 for near-integrated factors, respectively. Plugging these results into the test statistic and applying Slutsky's theorem, we obtain for $\mathbf{R} = \mathbf{I}_k$, *i.e.* under the null hypothesis (9), that

$$\begin{aligned}
\mathcal{W}^{EW} &= \left(\left(\sum_{t=1}^{T} \mathbf{z}_{t-1}' u_t\right) \mathbf{S}^k \mathbf{R}'\right) \cdot \left(\mathbf{R}\mathbf{S}^{k'} \left(\sum_{t=1}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}' u_t^2\right) \mathbf{S}^k \mathbf{R}'\right)^{-1} \\
&\quad \cdot \left(\mathbf{R}\mathbf{S}^{k'} \left(\sum_{t=1}^{T} \mathbf{z}_{t-1} u_t\right)\right) + o_p(1) \\
&= \left(\sum_{t=1}^{T} \mathbf{z}_{t-1}' u_t\right) \mathbf{S}^k \left(\mathbf{S}^k\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}' u_t^2\right)^{-1} \left(\mathbf{S}^{k'}\right)^{-1} \mathbf{S}^{k'} \\
&\quad \cdot \left(\sum_{t=1}^{T} \mathbf{z}_{t-1} u_t\right) + o_p(1) \\
&= \left(\sum_{t=1}^{T} \mathbf{z}_{t-1}' u_t\right) \left(\sum_{t=1}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}' u_t^2\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{z}_{t-1} u_t\right) + o_p(1).
\end{aligned}$$

The test statistic's asymptotic χ^2-distribution can then be found in a straightforward, yet tedious, way extending the univariate arguments and results of Demetrescu *et al.* (2023); see Remarks 9, 18, 22 of Demetrescu *et al.* □

Proof of Corollary 1 Plugging in $y_t = \mathbf{F}_t'\boldsymbol{\beta} + u_t$ into $\hat{\boldsymbol{\beta}}_F$ yields

$$\hat{\boldsymbol{\beta}}_F = \left(\sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1}\hat{\mathbf{F}}_{t-1}'\right)^{-1} \left(\sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1}\mathbf{F}_{t-1}'\boldsymbol{\beta} + \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} u_t\right)$$

Now, let $\theta = 1$ ($\theta = 2$) if the predictors are weakly (strongly) persistent. We then know from Lemmas S.3.3 and S.4.3 that

$$\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \hat{\mathbf{z}}_{t-1}\hat{\mathbf{F}}_{t-1}' = \mathbf{S}^{k'} \left(\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \mathbf{z}_{t-1}\mathbf{F}_{t-1}'\right) \mathbf{S}^k + o_p(1)$$

for $\eta \geq 0.5$ and $T/N \to 0$ as $T, N \to \infty$. Similarly,

$$\begin{aligned}
\left\| \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1}\mathbf{F}_{t-1}'\boldsymbol{\beta} - \sum_{t=2}^{T} \mathbf{S}^{k'}\mathbf{z}_{t-1}\mathbf{F}_{t-1}'\boldsymbol{\beta} \right\| &= \left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k'}\mathbf{z}_{t-1}\right) \mathbf{F}_{t-1}'\boldsymbol{\beta} \right\| \\
&= o_p\left(T^{1/2+(\theta-1)(1/2-\eta/2)}\right),
\end{aligned}$$

see the second summands in the proofs of Lemmas S.3.4 and S.4.4, which is thus $o_p(T^{1/2+(\theta-1)\eta/2})$ if $\eta \geq 0.5$. Finally, recall

$$\left\| \sum_{t=2}^{T} \hat{\mathbf{z}}_{t-1} u_t - \sum_{t=2}^{T} \mathbf{S}^{k\prime} \mathbf{z}_{t-1} u_t \right\| = \left\| \sum_{t=2}^{T} \left(\hat{\mathbf{z}}_{t-1} - \mathbf{S}^{k\prime} \mathbf{z}_{t-1} \right) u_t \right\|$$
$$= o_p \left(T^{1/2+(\theta-1)\eta/2} \right)$$

if $T/\sqrt{N} \to 0$; see the second summands in the proofs of Lemmas S.3.1 and S.4.1.

Consequently,

$$T^{1/2+(\theta-1)\eta/2} \hat{\boldsymbol{\beta}}_F = \left(\mathbf{S}^{k\prime} \left(\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \mathbf{z}_{t-1} \mathbf{F}_{t-1}' \right) \mathbf{S}^k \right)^{-1} \cdot \left(\mathbf{S}^{k\prime} \frac{1}{T^{1/2+(\theta-1)\eta/2}} \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{F}_{t-1}' \boldsymbol{\beta} + \mathbf{S}^{k\prime} \frac{1}{T^{1/2+(\theta-1)\eta/2}} \sum_{t=2}^{T} \mathbf{z}_{t-1} u_t \right) + o_p(1)$$

for $\eta \geq 0.5$ and $T/\sqrt{N} \to 0$ as $T, N \to \infty$. This can be simplified to

$$T^{1/2+(\theta-1)\eta/2} \left(\hat{\boldsymbol{\beta}}_F - (\mathbf{S}^k)^{-1} \boldsymbol{\beta} \right) = \left(\mathbf{S}^k \right)^{-1} \left(\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \mathbf{z}_{t-1} \mathbf{F}_{t-1}' \right)^{-1} \cdot \left(\frac{1}{T^{1/2+(\theta-1)\eta/2}} \sum_{t=2}^{T} \mathbf{z}_{t-1} u_t \right) + o_p(1).$$

Lastly, note that

$$\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \hat{\mathbf{z}}_{t-1} \hat{\mathbf{z}}_{t-1}' \hat{u}_t^2 = \mathbf{S}^{k\prime} \left(\frac{1}{T^{1+(\theta-1)\eta}} \sum_{t=1}^{T} \mathbf{z}_{t-1} \mathbf{z}_{t-1}' u_t^2 \right) \mathbf{S}^k + o_p(1)$$

by the convergence results concerning residual estimation given in the proof of Proposition 1 (with the univariate results carrying over) in Demetrescu *et al.* (2023, Suppl. Appx. pp. S.8 ff.) and intermediate results in the proofs of Lemmas S.3.2 and S.4.2.

Thanks to $\frac{1}{T^{1/2+(\theta-1)\eta/2}} \sum_{t=1}^{T} \mathbf{z}_{t-1} u_t$ converging, irrespective of the type of persistence, to a vector of Brownian motions — multivariate analogue of the results in Lemmas 4 and 5(d) of Demetrescu *et al.* (2023, Suppl. Appx.); additionally, see the proof of Proposition 2 of Demetrescu *et al.* for the respective (univariate) quadratic variation results — a linear combination resulting from pre-multiplying a row of $\left(\mathbf{S}^k\right)^{-1}$ is itself, regardless of the kind of persistence, (mixed) Gaussian asymptotically under the null in (10), *i.e.* $\left[(\mathbf{S}^k)^{-1}\boldsymbol{\beta}\right]_i = 0$, thereby completing the proof. □

S.2 Additional Simulation Results

ρ	β	DGP 1 ($T=500,\ N=200$)										DGP 1 ($T=500,\ N=400$)									
		GR	$\Pr(\neg H_0\|\beta)$					R^2_{OOS}				GR	$\Pr(\neg H_0\|\beta)$					R^2_{OOS}			
			IPR	IPA	IPD	IPDI	PCR	IPR	IPD	IPDI	PCR		IPR	IPA	IPD	IPDI	PCR	IPR	IPD	IPDI	PCR
(1.000, 1.000, 1.000, 1.000)	0.00	4.00	0.0800	0.1415	0.0695	0.2550	0.3305	-0.003	0.000	-0.009	-0.005	4.00	0.0910	0.1430	0.0685	0.2645	0.3635	-0.001	-0.001	-0.004	-0.005
	0.02	4.00	0.9045	0.9550	0.0690	0.9475	0.9995	0.052	-0.004	0.026	0.067	4.00	0.8965	0.9395	0.0645	0.9400	0.9985	0.038	-0.005	0.036	0.063
	0.04	4.00	0.9970	0.9985	0.0780	0.9910	1.0000	0.193	0.001	0.122	0.208	4.00	0.9975	0.9990	0.0685	0.9940	1.0000	0.176	-0.004	0.114	0.185
	0.06	4.00	0.9995	1.0000	0.0625	0.9945	1.0000	0.404	-0.001	0.218	0.408	4.00	0.9995	1.0000	0.0760	0.9970	1.0000	0.396	-0.004	0.199	0.397
	0.08	4.00	1.0000	1.0000	0.0640	0.9990	1.0000	0.531	-0.002	0.294	0.531	4.00	1.0000	1.0000	0.0775	0.9975	1.0000	0.521	0.003	0.282	0.522
	0.10	4.00	1.0000	1.0000	0.0730	0.9995	1.0000	0.645	-0.000	0.357	0.646	4.00	1.0000	1.0000	0.0735	0.9980	1.0000	0.623	-0.002	0.335	0.622
(0.995, 0.995, 0.995, 0.995)	0.00	4.00	0.0700	0.1225	0.0690	0.1925	0.2445	-0.007	-0.005	-0.004	-0.013	4.00	0.0680	0.1040	0.0675	0.1690	0.2265	-0.003	-0.002	-0.006	-0.007
	0.02	4.00	0.7920	0.8780	0.0610	0.9085	0.9845	0.017	-0.005	0.013	0.022	4.00	0.7640	0.8700	0.0680	0.9075	0.9850	0.012	-0.002	0.004	0.014
	0.04	4.00	0.9965	0.9985	0.0600	0.9980	1.0000	0.115	-0.003	0.089	0.117	4.00	0.9925	0.9950	0.0795	0.9970	1.0000	0.092	-0.004	0.070	0.092
	0.06	4.00	0.9995	1.0000	0.0770	1.0000	1.0000	0.175	-0.003	0.133	0.172	4.00	1.0000	1.0000	0.0715	1.0000	1.0000	0.202	0.001	0.150	0.206
	0.08	4.00	1.0000	1.0000	0.0715	1.0000	1.0000	0.288	0.001	0.213	0.287	4.00	1.0000	1.0000	0.0920	1.0000	1.0000	0.330	-0.002	0.244	0.331
	0.10	4.00	1.0000	1.0000	0.0925	0.9995	1.0000	0.412	-0.005	0.305	0.412	4.00	1.0000	1.0000	0.0885	1.0000	1.0000	0.408	-0.000	0.304	0.408
(0.950, 0.950, 0.950, 0.950)	0.00	4.00	0.0385	0.0590	0.0510	0.0650	0.0845	-0.003	-0.000	-0.001	-0.006	4.00	0.0545	0.0750	0.0640	0.0820	0.0915	-0.003	0.002	-0.002	-0.004
	0.02	4.00	0.1495	0.1995	0.0600	0.2315	0.2620	-0.011	-0.001	-0.003	-0.013	4.00	0.1465	0.1885	0.0655	0.2245	0.2695	-0.005	0.000	-0.004	-0.012
	0.04	4.00	0.5870	0.6715	0.0775	0.6890	0.7780	-0.006	-0.001	-0.005	-0.004	4.00	0.5670	0.6550	0.0835	0.6745	0.7755	0.005	-0.007	0.006	0.010
	0.06	4.00	0.9795	0.9910	0.0815	0.9815	0.9990	0.023	-0.009	0.016	0.023	4.00	0.9855	0.9910	0.1045	0.9875	0.9980	0.032	-0.004	0.026	0.033
	0.08	4.00	1.0000	1.0000	0.1045	1.0000	1.0000	0.048	-0.001	0.035	0.047	4.00	1.0000	1.0000	0.1260	0.9995	1.0000	0.071	0.000	0.039	0.071
	0.10	4.00	1.0000	1.0000	0.1355	1.0000	1.0000	0.088	-0.004	0.059	0.087	4.00	1.0000	1.0000	0.1500	1.0000	1.0000	0.070	0.003	0.057	0.069
(0.900, 0.900, 0.900, 0.900)	0.00	4.00	0.0420	0.0555	0.0660	0.0655	0.0750	-0.003	-0.006	-0.001	-0.004	4.00	0.0460	0.0565	0.0670	0.0635	0.0750	-0.003	-0.004	-0.001	-0.004
	0.02	4.00	0.0915	0.1035	0.0595	0.1235	0.1430	-0.009	0.001	-0.006	-0.011	4.00	0.0830	0.1030	0.0720	0.1310	0.1560	-0.008	-0.008	-0.002	-0.009
	0.04	4.00	0.2855	0.3345	0.0880	0.3645	0.4250	0.002	-0.000	-0.000	0.000	4.00	0.2565	0.3130	0.0825	0.3510	0.4040	0.008	-0.004	0.007	0.011
	0.06	4.00	0.6595	0.7100	0.0945	0.7265	0.8140	0.005	-0.006	0.004	0.006	4.00	0.6605	0.7165	0.1140	0.7370	0.8130	0.016	-0.007	0.011	0.014
	0.08	4.00	0.9425	0.9660	0.1215	0.9620	0.9870	0.034	-0.007	0.019	0.034	4.00	0.9520	0.9690	0.1270	0.9735	0.9915	0.022	-0.003	0.016	0.023
	0.10	4.00	0.9980	0.9995	0.1570	0.9990	1.0000	0.050	0.001	0.032	0.049	4.00	0.9995	1.0000	0.1670	1.0000	1.0000	0.029	-0.009	0.022	0.029

Table S.1: Average factor numbers (GR), Wald test rejection frequencies ($\Pr(\neg H_0|\beta)$) at 5% significance level, and forecast R^2_{OOS} values. Data from DGP 1 ($T=500,\ N=200$) and DGP 1 ($T=500,\ N=400$).

Table S.2: Factor-specific rejection frequencies of two-sided t-tests at 5% significance level. Data from DGP 1 ($T = 500,\ N = 200$) and DGP 1 ($T = 500,\ N = 400$).

ρ	$\hat{F}_i$	β	DGP 1 ($T = 500,\ N = 200$)					DGP 1 ($T = 500,\ N = 400$)				
			IPR	IPA	IPD	IPDI	PCR	IPR	IPA	IPD	IPDI	PCR
$\begin{pmatrix}1.000\\1.000\\1.000\\1.000\end{pmatrix}$	$\hat{F}_1$	0.00	0.0870	0.2375	0.0550	0.4715	0.5425	0.0845	0.2535	0.0615	0.4730	0.5510
		0.02	0.6665	0.9115	0.0665	0.9380	0.9975	0.6425	0.9085	0.0745	0.9435	0.9990
		0.04	0.9360	0.9925	0.0745	0.9675	1.0000	0.9370	0.9955	0.0685	0.9725	1.0000
		0.06	0.9915	1.0000	0.0845	0.9695	1.0000	0.9905	1.0000	0.0945	0.9720	1.0000
		0.08	0.9980	1.0000	0.1005	0.9780	1.0000	0.9985	1.0000	0.0995	0.9820	1.0000
		0.10	1.0000	1.0000	0.1040	0.9795	1.0000	1.0000	1.0000	0.1045	0.9805	1.0000
	$\hat{F}_2$	0.00	0.0460	0.0600	0.0545	0.0985	0.1595	0.0500	0.0615	0.0600	0.1085	0.1580
		0.02	0.0510	0.0660	0.0535	0.1770	0.1520	0.0500	0.0615	0.0455	0.1745	0.1580
		0.04	0.0615	0.0805	0.0545	0.2730	0.1790	0.0645	0.0810	0.0500	0.2845	0.1760
		0.06	0.0715	0.0880	0.0385	0.3730	0.1845	0.0685	0.0855	0.0545	0.3680	0.1950
		0.08	0.0870	0.1115	0.0470	0.4425	0.2235	0.0895	0.1090	0.0530	0.4385	0.2285
		0.10	0.0990	0.1295	0.0395	0.5170	0.2605	0.0950	0.1245	0.0425	0.5025	0.2580
	$\hat{F}_3$	0.00	0.0535	0.0665	0.0605	0.1100	0.1515	0.0490	0.0625	0.0565	0.1060	0.1540
		0.02	0.0465	0.0650	0.0540	0.1775	0.1580	0.0535	0.0725	0.0485	0.1745	0.1590
		0.04	0.0500	0.0670	0.0610	0.2920	0.1660	0.0560	0.0770	0.0570	0.2885	0.1825
		0.06	0.0625	0.0780	0.0550	0.3790	0.1840	0.0695	0.0935	0.0475	0.3795	0.2035
		0.08	0.0800	0.1005	0.0445	0.4505	0.2090	0.0880	0.1105	0.0500	0.4470	0.2175
		0.10	0.1025	0.1245	0.0415	0.5020	0.2500	0.1120	0.1500	0.0390	0.5145	0.2860
	$\hat{F}_4$	0.00	0.0425	0.0595	0.0595	0.1025	0.1470	0.0510	0.0645	0.0510	0.1070	0.1520
		0.02	0.0505	0.0660	0.0555	0.1710	0.1470	0.0530	0.0735	0.0525	0.1855	0.1660
		0.04	0.0520	0.0700	0.0555	0.2865	0.1690	0.0695	0.0840	0.0420	0.2920	0.1775
		0.06	0.0665	0.0795	0.0510	0.3955	0.1890	0.0730	0.0930	0.0490	0.3715	0.1930
		0.08	0.0715	0.0925	0.0490	0.4435	0.2050	0.0825	0.1045	0.0500	0.4315	0.2080
		0.10	0.0900	0.1140	0.0370	0.5015	0.2500	0.0975	0.1220	0.0415	0.4795	0.2480
$\begin{pmatrix}0.995\\0.995\\0.995\\0.995\end{pmatrix}$	$\hat{F}_1$	0.00	0.0765	0.2090	0.0620	0.3325	0.3925	0.0715	0.2025	0.0610	0.3325	0.3980
		0.02	0.6725	0.9195	0.0685	0.9535	0.9920	0.6560	0.9175	0.0740	0.9490	0.9895
		0.04	0.9750	0.9975	0.0895	0.9975	1.0000	0.9715	0.9960	0.0890	0.9975	1.0000
		0.06	0.9980	1.0000	0.1075	0.9995	1.0000	0.9960	1.0000	0.1055	0.9995	1.0000
		0.08	1.0000	1.0000	0.0970	0.9985	1.0000	1.0000	1.0000	0.1315	0.9990	1.0000
		0.10	1.0000	1.0000	0.1370	0.9995	1.0000	1.0000	1.0000	0.1415	1.0000	1.0000
	$\hat{F}_2$	0.00	0.0555	0.0790	0.0490	0.1045	0.1435	0.0515	0.0805	0.0535	0.1005	0.1405
		0.02	0.0595	0.0825	0.0470	0.1070	0.1540	0.0590	0.0775	0.0560	0.1125	0.1515
		0.04	0.0635	0.0945	0.0380	0.1565	0.1595	0.0525	0.0780	0.0595	0.1505	0.1520
		0.06	0.0700	0.0950	0.0490	0.2040	0.1700	0.0835	0.1070	0.0490	0.2095	0.1820
		0.08	0.0840	0.1195	0.0470	0.2500	0.1950	0.0800	0.1095	0.0440	0.2620	0.1815
		0.10	0.1050	0.1405	0.0460	0.3165	0.2100	0.1040	0.1350	0.0475	0.3190	0.2205
	$\hat{F}_3$	0.00	0.0545	0.0720	0.0490	0.0940	0.1380	0.0525	0.0680	0.0520	0.0925	0.1275
		0.02	0.0570	0.0790	0.0505	0.1050	0.1490	0.0555	0.0735	0.0480	0.1120	0.1415
		0.04	0.0690	0.0925	0.0425	0.1465	0.1540	0.0495	0.0725	0.0550	0.1535	0.1405
		0.06	0.0660	0.0870	0.0465	0.1985	0.1595	0.0805	0.1090	0.0445	0.2260	0.1845
		0.08	0.0840	0.1075	0.0470	0.2540	0.1900	0.0745	0.1115	0.0520	0.2670	0.1870
		0.10	0.1025	0.1415	0.0395	0.3045	0.2215	0.0910	0.1220	0.0435	0.3120	0.2050
	$\hat{F}_4$	0.00	0.0605	0.0760	0.0505	0.1055	0.1260	0.0505	0.0705	0.0545	0.1015	0.1385
		0.02	0.0670	0.0845	0.0570	0.1095	0.1525	0.0580	0.0805	0.0505	0.1060	0.1375
		0.04	0.0585	0.0800	0.0520	0.1475	0.1430	0.0535	0.0755	0.0455	0.1405	0.1475
		0.06	0.0910	0.1150	0.0465	0.2045	0.1940	0.0735	0.1050	0.0595	0.2220	0.1680
		0.08	0.0840	0.1205	0.0510	0.2710	0.2060	0.1005	0.1345	0.0485	0.2875	0.2095
		0.10	0.1065	0.1435	0.0445	0.3065	0.2340	0.1050	0.1455	0.0450	0.3300	0.2215

Continued on the next page

Table S.2 (continued)

ρ	$\hat{F}_i$	β	DGP 1 ($T = 500,\ N = 200$) IPR	IPA	IPD	IPDI	PCR	DGP 1 ($T = 500,\ N = 400$) IPR	IPA	IPD	IPDI	PCR
(0.950, 0.950, 0.950, 0.950)′	$\hat{F}_1$	0.00	0.0470	0.0855	0.0540	0.1060	0.1080	0.0475	0.0875	0.0645	0.1040	0.1150
		0.02	0.2445	0.4220	0.0670	0.4760	0.5265	0.2445	0.4270	0.0765	0.4775	0.5260
		0.04	0.7680	0.9240	0.1035	0.9335	0.9570	0.7595	0.9170	0.1045	0.9225	0.9570
		0.06	0.9965	1.0000	0.1375	0.9990	1.0000	0.9945	1.0000	0.1230	0.9975	1.0000
		0.08	1.0000	1.0000	0.1660	1.0000	1.0000	1.0000	1.0000	0.1885	1.0000	1.0000
		0.10	1.0000	1.0000	0.2320	1.0000	1.0000	1.0000	1.0000	0.2270	1.0000	1.0000
	$\hat{F}_2$	0.00	0.0395	0.0495	0.0550	0.0510	0.0610	0.0455	0.0525	0.0600	0.0605	0.0670
		0.02	0.0430	0.0605	0.0535	0.0600	0.0720	0.0490	0.0625	0.0535	0.0640	0.0810
		0.04	0.0585	0.0690	0.0540	0.0820	0.0875	0.0500	0.0630	0.0460	0.0650	0.0720
		0.06	0.0555	0.0685	0.0430	0.0735	0.0825	0.0560	0.0695	0.0570	0.0715	0.0845
		0.08	0.0620	0.0725	0.0505	0.0875	0.0930	0.0605	0.0740	0.0460	0.0930	0.0865
		0.10	0.0710	0.0865	0.0520	0.1015	0.0985	0.0700	0.0830	0.0545	0.1080	0.1045
	$\hat{F}_3$	0.00	0.0435	0.0540	0.0545	0.0580	0.0675	0.0475	0.0605	0.0455	0.0620	0.0735
		0.02	0.0470	0.0600	0.0555	0.0625	0.0705	0.0460	0.0560	0.0465	0.0585	0.0695
		0.04	0.0515	0.0610	0.0645	0.0650	0.0765	0.0470	0.0640	0.0545	0.0650	0.0740
		0.06	0.0670	0.0830	0.0495	0.0820	0.0965	0.0570	0.0725	0.0490	0.0885	0.0840
		0.08	0.0550	0.0675	0.0525	0.0795	0.0785	0.0600	0.0705	0.0570	0.0780	0.0870
		0.10	0.0625	0.0750	0.0530	0.1100	0.0895	0.0660	0.0850	0.0600	0.1100	0.0970
	$\hat{F}_4$	0.00	0.0505	0.0695	0.0485	0.0700	0.0860	0.0535	0.0655	0.0450	0.0685	0.0760
		0.02	0.0470	0.0560	0.0395	0.0650	0.0680	0.0500	0.0615	0.0580	0.0575	0.0715
		0.04	0.0570	0.0680	0.0510	0.0680	0.0810	0.0500	0.0655	0.0620	0.0650	0.0765
		0.06	0.0605	0.0720	0.0600	0.0715	0.0835	0.0580	0.0710	0.0600	0.0780	0.0925
		0.08	0.0620	0.0755	0.0585	0.0900	0.0905	0.0650	0.0810	0.0605	0.1080	0.0945
		0.10	0.0680	0.0855	0.0450	0.1125	0.1045	0.0695	0.0845	0.0520	0.1005	0.1020
(0.900, 0.900, 0.900, 0.900)′	$\hat{F}_1$	0.00	0.0460	0.0625	0.0595	0.0815	0.0880	0.0500	0.0645	0.0700	0.0790	0.0880
		0.02	0.1475	0.2340	0.0675	0.2750	0.2945	0.1340	0.2245	0.0740	0.2730	0.2945
		0.04	0.4630	0.6280	0.1115	0.6710	0.7195	0.4330	0.6100	0.0995	0.6525	0.6970
		0.06	0.8510	0.9395	0.1355	0.9375	0.9660	0.8530	0.9390	0.1485	0.9440	0.9685
		0.08	0.9880	0.9985	0.1860	0.9975	1.0000	0.9910	0.9990	0.1925	0.9995	1.0000
		0.10	1.0000	1.0000	0.2465	1.0000	1.0000	1.0000	1.0000	0.2485	1.0000	1.0000
	$\hat{F}_2$	0.00	0.0430	0.0475	0.0580	0.0525	0.0630	0.0420	0.0495	0.0555	0.0530	0.0630
		0.02	0.0455	0.0455	0.0485	0.0475	0.0565	0.0455	0.0485	0.0505	0.0505	0.0570
		0.04	0.0585	0.0630	0.0555	0.0610	0.0740	0.0465	0.0520	0.0515	0.0550	0.0635
		0.06	0.0530	0.0555	0.0485	0.0645	0.0695	0.0510	0.0530	0.0615	0.0715	0.0680
		0.08	0.0540	0.0615	0.0620	0.0660	0.0740	0.0435	0.0490	0.0590	0.0610	0.0665
		0.10	0.0595	0.0640	0.0620	0.0760	0.0805	0.0625	0.0720	0.0525	0.0840	0.0830
	$\hat{F}_3$	0.00	0.0480	0.0505	0.0535	0.0555	0.0615	0.0450	0.0505	0.0595	0.0585	0.0625
		0.02	0.0410	0.0455	0.0555	0.0450	0.0585	0.0475	0.0605	0.0495	0.0630	0.0675
		0.04	0.0500	0.0565	0.0545	0.0545	0.0680	0.0485	0.0575	0.0540	0.0655	0.0675
		0.06	0.0505	0.0570	0.0530	0.0600	0.0675	0.0505	0.0545	0.0610	0.0595	0.0685
		0.08	0.0475	0.0535	0.0515	0.0705	0.0695	0.0520	0.0630	0.0525	0.0740	0.0730
		0.10	0.0610	0.0700	0.0505	0.0780	0.0870	0.0545	0.0635	0.0515	0.0825	0.0795
	$\hat{F}_4$	0.00	0.0500	0.0555	0.0535	0.0555	0.0645	0.0540	0.0595	0.0595	0.0645	0.0715
		0.02	0.0440	0.0510	0.0495	0.0540	0.0650	0.0435	0.0530	0.0590	0.0530	0.0645
		0.04	0.0445	0.0480	0.0500	0.0535	0.0620	0.0485	0.0565	0.0505	0.0625	0.0710
		0.06	0.0460	0.0560	0.0525	0.0630	0.0620	0.0510	0.0560	0.0580	0.0575	0.0700
		0.08	0.0510	0.0580	0.0490	0.0700	0.0695	0.0540	0.0610	0.0620	0.0785	0.0735
		0.10	0.0525	0.0620	0.0620	0.0735	0.0775	0.0530	0.0640	0.0585	0.0750	0.0775

ρ	β	DGP 2 GR	DGP 2 $\Pr(\neg H_0\|\beta)$ IPR	IPA	IPD	IPDI	PCR	DGP 2 R^2_{OOS} IPR	IPD	IPDI	PCR
$(1.000, 0.995, 0.950, 0.900)'$	0.00	4.00	0.0536	0.1061	0.0746	0.2237	0.2503	-0.005	-0.001	-0.011	-0.015
	0.02	4.00	0.3575	0.5195	0.0795	0.6740	0.8640	0.023	-0.004	0.007	0.034
	0.04	4.00	0.8664	0.9295	0.0635	0.9210	0.9970	0.113	-0.000	0.056	0.131
	0.06	4.00	0.9780	0.9905	0.0780	0.9650	1.0000	0.245	-0.011	0.099	0.257
	0.08	4.00	0.9955	0.9975	0.0680	0.9810	1.0000	0.360	-0.008	0.142	0.363
	0.10	4.00	0.9975	0.9990	0.0645	0.9900	1.0000	0.458	-0.008	0.187	0.462
$(0.995, 0.950, 0.900, 1.000)'$	0.00	4.00	0.0595	0.1031	0.0720	0.1866	0.2381	-0.016	-0.010	-0.023	-0.034
	0.02	4.00	0.2705	0.4160	0.0695	0.5710	0.7340	-0.005	-0.002	-0.011	-0.015
	0.04	4.00	0.7909	0.8834	0.0710	0.9115	0.9920	0.032	-0.009	0.001	0.056
	0.06	4.00	0.9720	0.9880	0.0896	0.9845	1.0000	0.134	-0.008	0.077	0.145
	0.08	4.00	0.9985	0.9995	0.0905	0.9965	1.0000	0.221	-0.004	0.131	0.222
	0.10	4.00	0.9990	0.9995	0.0760	0.9945	1.0000	0.343	-0.007	0.214	0.345
$(0.950, 0.900, 1.000, 0.995)'$	0.00	4.00	0.0375	0.0685	0.0810	0.1070	0.1370	-0.007	-0.011	-0.010	-0.018
	0.02	4.00	0.0900	0.1446	0.0895	0.2146	0.2551	-0.018	-0.008	-0.018	-0.027
	0.04	4.00	0.2611	0.3592	0.0945	0.4567	0.5378	-0.004	-0.013	-0.006	-0.003
	0.06	4.00	0.6021	0.7307	0.0926	0.8083	0.8874	0.003	-0.010	0.009	0.012
	0.08	4.00	0.9329	0.9780	0.1011	0.9775	0.9965	0.051	-0.004	0.030	0.050
	0.10	4.00	0.9995	1.0000	0.1301	0.9990	1.0000	0.055	-0.010	0.049	0.054
$(0.900, 1.000, 0.995, 0.950)'$	0.00	4.00	0.0370	0.0595	0.0710	0.1001	0.1126	-0.010	-0.007	-0.003	-0.015
	0.02	4.00	0.0635	0.0905	0.0840	0.1475	0.1560	-0.009	-0.012	-0.008	-0.020
	0.04	3.99	0.1482	0.2083	0.0856	0.2819	0.3005	-0.011	-0.006	-0.010	-0.019
	0.06	4.00	0.2841	0.3752	0.1091	0.4682	0.5318	0.002	-0.010	0.004	-0.005
	0.08	4.00	0.5463	0.6563	0.1061	0.7344	0.7974	0.007	-0.005	0.012	0.012
	0.10	4.00	0.8348	0.9024	0.1431	0.9309	0.9615	0.029	-0.010	0.016	0.036

ρ	β	DGP 3 GR	DGP 3 $\Pr(\neg H_0\|\beta)$ IPR	IPA	IPD	IPDI	PCR	DGP 3 R^2_{OOS} IPR	IPD	IPDI	PCR
$(1.000, 0.995, 0.950, 0.900)'$	0.00	4.00	0.0580	0.1195	0.0805	0.2245	0.2195	-0.008	-0.009	-0.007	-0.008
	0.02	4.00	0.2330	0.3945	0.0905	0.6220	0.7380	0.012	-0.010	0.011	0.019
	0.04	4.00	0.7320	0.8760	0.0820	0.9060	0.9800	0.079	-0.003	0.072	0.109
	0.06	4.00	0.9465	0.9845	0.0810	0.9700	0.9995	0.226	-0.010	0.128	0.229
	0.08	4.00	0.9925	0.9985	0.0745	0.9905	1.0000	0.330	-0.011	0.170	0.336
	0.10	4.00	0.9990	0.9995	0.0820	0.9950	1.0000	0.424	-0.008	0.216	0.425
$(0.995, 0.950, 0.900, 1.000)'$	0.00	4.00	0.0605	0.1235	0.0770	0.1925	0.2300	-0.017	-0.016	-0.011	-0.030
	0.02	4.00	0.2180	0.3765	0.0805	0.5325	0.6070	-0.011	-0.005	-0.002	-0.013
	0.04	4.00	0.6450	0.8240	0.0780	0.8855	0.9600	0.027	-0.015	0.020	0.058
	0.06	4.00	0.9485	0.9835	0.0900	0.9760	0.9990	0.137	-0.009	0.086	0.143
	0.08	4.00	0.9955	0.9990	0.0895	0.9970	1.0000	0.217	-0.014	0.138	0.219
	0.10	4.00	0.9995	1.0000	0.0850	0.9970	1.0000	0.333	-0.008	0.203	0.334
$(0.950, 0.900, 1.000, 0.995)'$	0.00	4.00	0.0485	0.0995	0.0810	0.1345	0.1545	-0.012	-0.013	-0.007	-0.018
	0.02	4.00	0.1050	0.1730	0.0945	0.2325	0.2475	-0.019	-0.015	-0.011	-0.031
	0.04	4.00	0.2525	0.3825	0.1045	0.4265	0.4785	-0.011	-0.012	0.001	-0.007
	0.06	4.00	0.4870	0.6605	0.1005	0.7095	0.7725	0.002	-0.008	0.016	0.015
	0.08	4.00	0.8050	0.9105	0.0955	0.9215	0.9730	0.026	-0.014	0.030	0.038
	0.10	4.00	0.9695	0.9945	0.1080	0.9830	0.9995	0.056	-0.008	0.054	0.054
$(0.900, 1.000, 0.995, 0.950)'$	0.00	4.00	0.0450	0.0790	0.0755	0.1085	0.1150	-0.013	-0.009	0.001	-0.021
	0.02	4.00	0.0715	0.1220	0.0860	0.1620	0.1750	-0.016	-0.010	-0.008	-0.027
	0.04	4.00	0.1290	0.2030	0.0865	0.2500	0.2830	-0.014	-0.007	-0.003	-0.021
	0.06	4.00	0.2585	0.3660	0.0985	0.4030	0.4515	-0.014	-0.016	0.004	-0.017
	0.08	4.00	0.4340	0.5855	0.1000	0.6185	0.6895	0.002	-0.017	0.011	0.009
	0.10	4.00	0.6970	0.8295	0.1300	0.8350	0.8900	0.015	-0.011	0.019	0.030

Table S.3: Average factor numbers (GR), Wald test rejection frequencies ($\Pr(\neg H_0|\beta)$) at 5% significance level, and forecast R^2_{OOS} values. Data from DGP 2 and DGP 3.

Table S.4: Factor-specific rejection frequencies of two-sided t-tests at 5% significance level. Data from DGP 2 and DGP 3.

ρ	$\hat{F}_i$	β	DGP 2 IPR	DGP 2 IPA	DGP 2 IPD	DGP 2 IPDI	DGP 2 PCR	DGP 3 IPR	DGP 3 IPA	DGP 3 IPD	DGP 3 IPDI	DGP 3 PCR
$\begin{pmatrix}1.000\\0.995\\0.950\\0.900\end{pmatrix}$	$\hat{F}_1$	0.00	0.0661	0.1652	0.0661	0.3729	0.3969	0.0610	0.1715	0.0645	0.3360	0.3180
		0.02	0.3500	0.6295	0.0730	0.7835	0.9135	0.2605	0.5560	0.0615	0.7525	0.8040
		0.04	0.7724	0.9350	0.0870	0.9185	0.9990	0.7055	0.9335	0.0690	0.9390	0.9875
		0.06	0.9320	0.9920	0.0970	0.9450	1.0000	0.9160	0.9935	0.0785	0.9780	0.9990
		0.08	0.9780	0.9990	0.0980	0.9650	1.0000	0.9840	0.9995	0.0890	0.9845	1.0000
		0.10	0.9930	0.9990	0.1011	0.9685	1.0000	0.9960	1.0000	0.0955	0.9845	1.0000
	$\hat{F}_2$	0.00	0.0430	0.0611	0.0621	0.1016	0.1351	0.0530	0.0780	0.0630	0.1075	0.1460
		0.02	0.0525	0.0690	0.0540	0.1225	0.1385	0.0705	0.0975	0.0640	0.1070	0.1560
		0.04	0.0545	0.0715	0.0600	0.1911	0.1471	0.0615	0.1020	0.0620	0.1575	0.1540
		0.06	0.0520	0.0735	0.0470	0.2350	0.1385	0.0610	0.0965	0.0510	0.2005	0.1495
		0.08	0.0600	0.0805	0.0520	0.2966	0.1571	0.0680	0.1035	0.0580	0.2585	0.1625
		0.10	0.0630	0.0875	0.0455	0.3492	0.1686	0.0730	0.1055	0.0480	0.3035	0.1685
	$\hat{F}_3$	0.00	0.0435	0.0641	0.0566	0.0791	0.0966	0.0450	0.0710	0.0645	0.0770	0.0870
		0.02	0.0420	0.0565	0.0500	0.0970	0.0920	0.0435	0.0675	0.0590	0.0845	0.0915
		0.04	0.0515	0.0675	0.0470	0.1316	0.0960	0.0500	0.0740	0.0480	0.1100	0.1005
		0.06	0.0515	0.0780	0.0555	0.1915	0.1060	0.0605	0.0855	0.0725	0.1585	0.1015
		0.08	0.0540	0.0760	0.0350	0.2271	0.1121	0.0540	0.0830	0.0485	0.1895	0.1035
		0.10	0.0500	0.0690	0.0295	0.2671	0.1026	0.0485	0.0805	0.0535	0.2185	0.1010
	$\hat{F}_4$	0.00	0.0400	0.0516	0.0501	0.0721	0.0681	0.0470	0.0660	0.0575	0.0680	0.0735
		0.02	0.0440	0.0580	0.0550	0.0850	0.0825	0.0460	0.0710	0.0620	0.0775	0.0850
		0.04	0.0465	0.0615	0.0470	0.1171	0.0800	0.0455	0.0705	0.0515	0.0995	0.0800
		0.06	0.0440	0.0570	0.0465	0.1670	0.0830	0.0525	0.0695	0.0570	0.1340	0.0835
		0.08	0.0415	0.0540	0.0375	0.1911	0.0785	0.0495	0.0765	0.0450	0.1660	0.0905
		0.10	0.0450	0.0610	0.0230	0.2266	0.0870	0.0555	0.0745	0.0390	0.1925	0.0830
$\begin{pmatrix}0.995\\0.950\\0.900\\1.000\end{pmatrix}$	$\hat{F}_1$	0.00	0.0640	0.1691	0.0710	0.3217	0.3837	0.0730	0.1880	0.0625	0.3100	0.3240
		0.02	0.3010	0.5740	0.0710	0.7425	0.8695	0.2730	0.5460	0.0640	0.6945	0.7490
		0.04	0.6873	0.9150	0.0890	0.9530	0.9965	0.6315	0.9205	0.0680	0.9355	0.9775
		0.06	0.9069	0.9865	0.1081	0.9880	1.0000	0.9010	0.9935	0.0870	0.9865	0.9995
		0.08	0.9725	0.9975	0.1101	0.9920	1.0000	0.9815	0.9985	0.0785	0.9950	1.0000
		0.10	0.9880	0.9990	0.1341	0.9915	1.0000	0.9950	1.0000	0.0985	0.9930	1.0000
	$\hat{F}_2$	0.00	0.0480	0.0695	0.0515	0.0895	0.1056	0.0570	0.0835	0.0565	0.0905	0.1100
		0.02	0.0465	0.0685	0.0600	0.0875	0.1090	0.0480	0.0760	0.0650	0.0835	0.0950
		0.04	0.0370	0.0605	0.0565	0.0920	0.0955	0.0490	0.0745	0.0620	0.0900	0.0950
		0.06	0.0551	0.0746	0.0465	0.1286	0.1106	0.0570	0.0850	0.0500	0.1150	0.1135
		0.08	0.0595	0.0875	0.0380	0.1626	0.1196	0.0685	0.0985	0.0490	0.1650	0.1225
		0.10	0.0520	0.0780	0.0400	0.1721	0.1116	0.0690	0.1035	0.0490	0.1655	0.1215
	$\hat{F}_3$	0.00	0.0475	0.0670	0.0530	0.0735	0.0930	0.0515	0.0740	0.0635	0.0740	0.0910
		0.02	0.0410	0.0565	0.0545	0.0735	0.0795	0.0490	0.0655	0.0650	0.0665	0.0810
		0.04	0.0425	0.0565	0.0480	0.0885	0.0830	0.0485	0.0675	0.0525	0.0765	0.0800
		0.06	0.0445	0.0611	0.0480	0.1026	0.0811	0.0530	0.0725	0.0485	0.0920	0.0835
		0.08	0.0420	0.0595	0.0500	0.1336	0.0890	0.0450	0.0695	0.0540	0.1215	0.0900
		0.10	0.0480	0.0690	0.0415	0.1496	0.0895	0.0535	0.0725	0.0475	0.1350	0.0860
	$\hat{F}_4$	0.00	0.0455	0.0595	0.0600	0.0970	0.1576	0.0530	0.0795	0.0570	0.1040	0.1580
		0.02	0.0415	0.0535	0.0465	0.1000	0.1645	0.0535	0.0745	0.0560	0.0920	0.1570
		0.04	0.0500	0.0670	0.0515	0.1326	0.1691	0.0655	0.0895	0.0660	0.1235	0.1745
		0.06	0.0576	0.0786	0.0606	0.1652	0.1842	0.0625	0.0905	0.0685	0.1590	0.1805
		0.08	0.0600	0.0775	0.0530	0.1926	0.1861	0.0640	0.0935	0.0625	0.1675	0.1890
		0.10	0.0665	0.0870	0.0490	0.2386	0.2191	0.0755	0.1105	0.0535	0.2140	0.2055

Continued on the next page

Table S.4 (continued)

ρ	$\hat{F}_i$	β	DGP 2					DGP 3				
			IPR	IPA	IPD	IPDI	PCR	IPR	IPA	IPD	IPDI	PCR
$\begin{pmatrix}0.950\\0.900\\1.000\\0.995\end{pmatrix}$	$\hat{F}_1$	0.00	0.0390	0.1080	0.0650	0.1820	0.1990	0.0525	0.1460	0.0670	0.1775	0.1955
		0.02	0.1416	0.3117	0.0745	0.4362	0.4752	0.1380	0.3390	0.0675	0.4015	0.4445
		0.04	0.3287	0.6423	0.1086	0.7609	0.8344	0.3150	0.6300	0.0850	0.6740	0.7260
		0.06	0.6481	0.9079	0.1151	0.9580	0.9850	0.5750	0.8875	0.0920	0.9000	0.9455
		0.08	0.9019	0.9855	0.1286	0.9960	1.0000	0.8350	0.9815	0.1040	0.9815	0.9935
		0.10	0.9685	0.9960	0.1726	0.9995	1.0000	0.9650	0.9995	0.1185	0.9980	1.0000
	$\hat{F}_2$	0.00	0.0390	0.0510	0.0540	0.0650	0.0845	0.0500	0.0725	0.0600	0.0755	0.0915
		0.02	0.0465	0.0605	0.0615	0.0705	0.0885	0.0595	0.0755	0.0665	0.0850	0.0895
		0.04	0.0470	0.0610	0.0570	0.0770	0.0840	0.0540	0.0715	0.0640	0.0805	0.0855
		0.06	0.0435	0.0631	0.0611	0.0796	0.0916	0.0550	0.0805	0.0555	0.0815	0.1025
		0.08	0.0430	0.0586	0.0470	0.0741	0.0886	0.0525	0.0700	0.0520	0.0815	0.0920
		0.10	0.0490	0.0630	0.0640	0.0820	0.0890	0.0585	0.0800	0.0580	0.0840	0.0955
	$\hat{F}_3$	0.00	0.0380	0.0405	0.0640	0.0680	0.1125	0.0515	0.0585	0.0720	0.0875	0.1100
		0.02	0.0330	0.0440	0.0530	0.0790	0.1131	0.0455	0.0615	0.0620	0.0875	0.1230
		0.04	0.0480	0.0560	0.0570	0.0850	0.1306	0.0515	0.0640	0.0695	0.0915	0.1275
		0.06	0.0485	0.0571	0.0495	0.1011	0.1316	0.0555	0.0780	0.0585	0.1090	0.1430
		0.08	0.0521	0.0556	0.0546	0.0936	0.1441	0.0580	0.0775	0.0625	0.1025	0.1380
		0.10	0.0570	0.0695	0.0515	0.1201	0.1626	0.0685	0.0840	0.0575	0.1255	0.1735
	$\hat{F}_4$	0.00	0.0410	0.0505	0.0585	0.0795	0.1160	0.0535	0.0725	0.0605	0.0900	0.1170
		0.02	0.0330	0.0425	0.0560	0.0750	0.1036	0.0410	0.0635	0.0640	0.0895	0.1200
		0.04	0.0450	0.0520	0.0525	0.0910	0.1226	0.0530	0.0710	0.0620	0.0915	0.1250
		0.06	0.0531	0.0611	0.0511	0.0956	0.1216	0.0550	0.0730	0.0605	0.1060	0.1305
		0.08	0.0536	0.0681	0.0576	0.1006	0.1286	0.0565	0.0825	0.0680	0.1195	0.1410
		0.10	0.0595	0.0660	0.0620	0.1161	0.1391	0.0580	0.0860	0.0660	0.1225	0.1395
$\begin{pmatrix}0.900\\1.000\\0.995\\0.950\end{pmatrix}$	$\hat{F}_1$	0.00	0.0380	0.0910	0.0700	0.1536	0.1471	0.0390	0.1110	0.0660	0.1365	0.1420
		0.02	0.0785	0.1990	0.0730	0.3110	0.3240	0.0845	0.2310	0.0605	0.2760	0.2990
		0.04	0.2218	0.4617	0.0936	0.5704	0.6139	0.2025	0.4560	0.0690	0.4950	0.5410
		0.06	0.3892	0.7109	0.1106	0.8029	0.8529	0.3635	0.6750	0.0805	0.7035	0.7585
		0.08	0.6548	0.9075	0.1291	0.9590	0.9770	0.5770	0.8820	0.0935	0.8970	0.9315
		0.10	0.8849	0.9820	0.1817	0.9970	0.9990	0.8095	0.9830	0.1255	0.9780	0.9940
	$\hat{F}_2$	0.00	0.0390	0.0405	0.0480	0.0760	0.0950	0.0495	0.0530	0.0520	0.0795	0.1040
		0.02	0.0310	0.0355	0.0590	0.0700	0.0985	0.0490	0.0550	0.0605	0.0805	0.1065
		0.04	0.0356	0.0381	0.0531	0.0806	0.1097	0.0515	0.0575	0.0595	0.0875	0.1125
		0.06	0.0405	0.0435	0.0600	0.0685	0.1126	0.0475	0.0565	0.0700	0.0795	0.1060
		0.08	0.0500	0.0560	0.0545	0.0920	0.1266	0.0525	0.0610	0.0590	0.0760	0.1130
		0.10	0.0611	0.0621	0.0621	0.0971	0.1401	0.0660	0.0730	0.0580	0.1040	0.1400
	$\hat{F}_3$	0.00	0.0430	0.0490	0.0585	0.0760	0.1016	0.0515	0.0690	0.0540	0.0870	0.0990
		0.02	0.0395	0.0490	0.0600	0.0725	0.0970	0.0515	0.0605	0.0585	0.0790	0.0980
		0.04	0.0441	0.0531	0.0551	0.0836	0.1047	0.0520	0.0655	0.0555	0.0895	0.1085
		0.06	0.0450	0.0535	0.0705	0.0730	0.1041	0.0495	0.0655	0.0660	0.0885	0.1035
		0.08	0.0515	0.0560	0.0585	0.0870	0.1166	0.0530	0.0635	0.0600	0.0900	0.1115
		0.10	0.0716	0.0746	0.0671	0.1126	0.1366	0.0735	0.0900	0.0705	0.1125	0.1365
	$\hat{F}_4$	0.00	0.0500	0.0605	0.0530	0.0695	0.0945	0.0495	0.0750	0.0590	0.0750	0.0905
		0.02	0.0355	0.0505	0.0555	0.0660	0.0795	0.0445	0.0690	0.0570	0.0665	0.0850
		0.04	0.0371	0.0536	0.0621	0.0751	0.0881	0.0485	0.0705	0.0635	0.0715	0.0830
		0.06	0.0520	0.0620	0.0570	0.0775	0.0970	0.0600	0.0795	0.0570	0.0810	0.0985
		0.08	0.0520	0.0660	0.0585	0.0920	0.0895	0.0580	0.0865	0.0565	0.0860	0.1070
		0.10	0.0541	0.0646	0.0576	0.0866	0.0901	0.0630	0.0850	0.0610	0.0975	0.1040

ρ	β	DGP 4										DGP 5									
		GR	$\Pr(\neg H_0\|\beta)$					R^2_{OOS}				GR	$\Pr(\neg H_0\|\beta)$					R^2_{OOS}			
			IPR	IPA	IPD	IPDI	PCR	IPR	IPD	IPDI	PCR		IPR	IPA	IPD	IPDI	PCR	IPR	IPD	IPDI	PCR
	0.00	4.00	0.0775	0.1685	0.0830	0.2280	0.2930	-0.018	-0.012	-0.023	-0.035	3.66	0.0547	0.1143	0.0810	0.2183	0.2462	-0.003	-0.004	-0.008	-0.018
1.000	0.02	4.00	0.5040	0.6785	0.0810	0.6575	0.8795	0.009	-0.008	-0.015	0.029	3.63	0.3747	0.5493	0.0782	0.6667	0.8490	0.014	-0.003	0.016	0.026
0.995	0.04	4.00	0.8505	0.9250	0.0765	0.8805	0.9980	0.132	-0.009	0.018	0.168	3.58	0.8403	0.9101	0.0748	0.9079	0.9944	0.098	-0.003	0.044	0.120
0.950	0.06	4.00	0.9520	0.9780	0.0710	0.9425	1.0000	0.296	-0.011	0.061	0.329	3.59	0.9666	0.9827	0.0736	0.9671	1.0000	0.222	-0.014	0.096	0.232
0.900	0.08	4.00	0.9875	0.9940	0.0640	0.9725	1.0000	0.416	-0.002	0.084	0.435	3.64	0.9923	0.9984	0.0653	0.9791	1.0000	0.327	-0.006	0.121	0.327
	0.10	4.00	0.9955	0.9975	0.0560	0.9825	1.0000	0.532	-0.005	0.139	0.546	3.61	0.9961	0.9983	0.0714	0.9906	1.0000	0.425	-0.009	0.175	0.429
	0.00	4.00	0.0690	0.1395	0.0870	0.1865	0.2785	-0.025	-0.011	-0.046	-0.057	3.62	0.0591	0.1171	0.0768	0.1994	0.2370	-0.019	-0.007	-0.028	-0.027
0.995	0.02	4.00	0.3005	0.4760	0.0700	0.5190	0.7395	-0.014	-0.005	-0.042	-0.035	3.64	0.2821	0.4358	0.0675	0.5746	0.7272	-0.008	-0.003	-0.011	-0.014
0.950	0.04	4.00	0.7425	0.8740	0.0685	0.8720	0.9950	0.024	-0.012	-0.051	0.032	3.59	0.7767	0.8764	0.0774	0.9042	0.9889	0.027	-0.005	-0.005	0.040
0.900	0.06	4.00	0.9420	0.9760	0.0900	0.9720	1.0000	0.093	-0.011	0.034	0.128	3.58	0.9575	0.9793	0.0911	0.9726	0.9983	0.112	-0.008	0.062	0.123
1.000	0.08	4.00	0.9855	0.9910	0.0845	0.9905	1.0000	0.194	-0.000	0.101	0.218	3.66	0.9951	0.9984	0.0913	0.9923	1.0000	0.211	-0.005	0.129	0.220
	0.10	4.00	0.9950	0.9980	0.0740	0.9910	1.0000	0.342	-0.006	0.207	0.355	3.61	0.9978	0.9994	0.0858	0.9939	1.0000	0.323	-0.006	0.202	0.327
	0.00	4.00	0.0420	0.0850	0.0865	0.1085	0.1505	-0.020	-0.013	-0.015	-0.039	3.63	0.0419	0.0728	0.0766	0.1092	0.1400	-0.011	-0.003	-0.013	-0.026
0.950	0.02	4.00	0.0850	0.1505	0.0820	0.2040	0.2535	-0.015	-0.011	-0.037	-0.043	3.66	0.0956	0.1415	0.0918	0.2212	0.2447	-0.015	-0.007	-0.019	-0.026
0.900	0.04	4.00	0.1905	0.3290	0.1000	0.3925	0.4915	-0.019	-0.015	-0.029	-0.023	3.61	0.2499	0.3490	0.0903	0.4371	0.4936	-0.003	-0.011	0.001	0.002
1.000	0.06	4.00	0.4450	0.6325	0.0890	0.6870	0.8040	-0.035	-0.016	-0.021	-0.020	3.67	0.5499	0.6921	0.0877	0.7515	0.8272	-0.006	-0.011	0.002	-0.001
0.995	0.08	4.00	0.8305	0.9355	0.0890	0.9495	0.9825	0.033	-0.008	0.017	0.038	3.64	0.8818	0.9434	0.0940	0.9593	0.9791	0.029	-0.006	0.022	0.040
	0.10	4.00	0.9755	0.9925	0.1095	0.9980	0.9995	0.025	-0.010	0.022	0.025	3.65	0.9808	0.9940	0.1124	0.9923	0.9984	0.038	-0.004	0.038	0.042
	0.00	4.00	0.0515	0.0920	0.0780	0.1125	0.1475	-0.014	-0.009	-0.013	-0.035	3.59	0.0441	0.0753	0.0786	0.1188	0.1216	-0.015	-0.007	-0.003	-0.015
0.900	0.02	4.00	0.0565	0.1010	0.0845	0.1450	0.1785	-0.016	-0.010	-0.030	-0.029	3.67	0.0632	0.0937	0.0877	0.1607	0.1607	-0.017	-0.006	-0.012	-0.021
1.000	0.04	4.00	0.1225	0.1905	0.0935	0.2410	0.3070	-0.025	-0.010	-0.028	-0.043	3.62	0.1397	0.1999	0.0845	0.2634	0.2943	-0.012	-0.001	-0.003	-0.018
0.995	0.06	4.00	0.2070	0.3160	0.0985	0.3785	0.4785	-0.007	-0.009	-0.012	-0.011	3.63	0.2438	0.3440	0.1112	0.4458	0.4854	0.001	-0.009	0.000	-0.003
0.950	0.08	4.00	0.4200	0.5615	0.1065	0.6210	0.7060	-0.001	-0.009	-0.003	-0.001	3.65	0.4844	0.6056	0.0960	0.6840	0.7444	0.002	-0.006	0.008	0.013
	0.10	4.00	0.6855	0.8105	0.1245	0.8395	0.9060	0.010	-0.010	-0.007	0.021	3.61	0.7421	0.8362	0.1411	0.8727	0.9181	0.011	-0.008	0.011	0.024

Table S.5: Average factor numbers (GR), Wald test rejection frequencies ($\Pr(\neg H_0|\beta)$) at 5% significance level, and forecast R^2_{OOS} values. Data from DGP 4 and DGP 5.

Table S.6: Factor-specific rejection frequencies of two-sided t-tests at 5% significance level. Data from DGP 4 and DGP 5.

ρ	$\hat{F}_i$	β	DGP 4					DGP 5				
			IPR	IPA	IPD	IPDI	PCR	IPR	IPA	IPD	IPDI	PCR
$\begin{pmatrix}1.000\\0.995\\0.950\\0.900\end{pmatrix}$	$\hat{F}_1$	0.00	0.0820	0.2015	0.0585	0.3655	0.4515	0.0678	0.1668	0.0722	0.3632	0.3802
		0.02	0.3790	0.6230	0.0585	0.7025	0.9195	0.3537	0.6116	0.0722	0.7554	0.8975
		0.04	0.6720	0.8840	0.0635	0.8555	0.9950	0.7404	0.9263	0.0832	0.9135	0.9955
		0.06	0.8645	0.9700	0.0575	0.8900	1.0000	0.9164	0.9861	0.0953	0.9409	0.9994
		0.08	0.9415	0.9875	0.0590	0.9260	1.0000	0.9742	0.9973	0.0988	0.9555	1.0000
		0.10	0.9695	0.9970	0.0670	0.9240	1.0000	0.9895	0.9983	0.1035	0.9718	1.0000
	$\hat{F}_2$	0.00	0.0370	0.0565	0.0730	0.1095	0.1235	0.0427	0.0618	0.0629	0.1089	0.1373
		0.02	0.0415	0.0690	0.0520	0.1430	0.1285	0.0457	0.0705	0.0573	0.1118	0.1383
		0.04	0.0440	0.0665	0.0620	0.2175	0.1430	0.0503	0.0754	0.0586	0.1904	0.1446
		0.06	0.0445	0.0665	0.0530	0.2790	0.1440	0.0485	0.0741	0.0491	0.2375	0.1505
		0.08	0.0440	0.0705	0.0560	0.3160	0.1470	0.0615	0.0835	0.0505	0.3009	0.1538
		0.10	0.0485	0.0715	0.0595	0.3645	0.1530	0.0570	0.0842	0.0437	0.3505	0.1645
	$\hat{F}_3$	0.00	0.0390	0.0580	0.0575	0.0905	0.0995	0.0481	0.0635	0.0536	0.0810	0.0990
		0.02	0.0370	0.0615	0.0585	0.1165	0.1020	0.0430	0.0612	0.0496	0.0981	0.0953
		0.04	0.0395	0.0590	0.0530	0.1555	0.0970	0.0503	0.0681	0.0491	0.1329	0.1005
		0.06	0.0445	0.0685	0.0540	0.1970	0.1045	0.0585	0.0825	0.0613	0.2074	0.1109
		0.08	0.0455	0.0685	0.0405	0.2365	0.1055	0.0566	0.0807	0.0379	0.2378	0.1109
		0.10	0.0460	0.0690	0.0405	0.2975	0.1080	0.0554	0.0770	0.0338	0.2724	0.1096
	$\hat{F}_4$	0.00	0.0325	0.0495	0.0495	0.0730	0.0725	0.0405	0.0558	0.0574	0.0722	0.0711
		0.02	0.0340	0.0555	0.0585	0.1030	0.0830	0.0413	0.0545	0.0529	0.0799	0.0716
		0.04	0.0350	0.0530	0.0490	0.1285	0.0855	0.0441	0.0592	0.0463	0.1122	0.0743
		0.06	0.0330	0.0515	0.0540	0.1660	0.0745	0.0379	0.0552	0.0502	0.1661	0.0814
		0.08	0.0350	0.0545	0.0500	0.1985	0.0775	0.0500	0.0643	0.0373	0.2026	0.0873
		0.10	0.0430	0.0655	0.0280	0.2360	0.0975	0.0404	0.0587	0.0238	0.2375	0.0770
$\begin{pmatrix}0.995\\0.950\\0.900\\1.000\end{pmatrix}$	$\hat{F}_1$	0.00	0.0720	0.1760	0.0735	0.2975	0.4215	0.0773	0.1652	0.0713	0.3144	0.3602
		0.02	0.2705	0.5265	0.0585	0.6555	0.8575	0.3150	0.5626	0.0615	0.7130	0.8436
		0.04	0.5585	0.8280	0.0675	0.9010	0.9950	0.6748	0.8998	0.0846	0.9443	0.9944
		0.06	0.8260	0.9575	0.0685	0.9635	1.0000	0.8977	0.9832	0.1045	0.9816	0.9994
		0.08	0.9170	0.9835	0.0740	0.9815	1.0000	0.9661	0.9962	0.1017	0.9891	1.0000
		0.10	0.9530	0.9935	0.0855	0.9810	1.0000	0.9878	0.9994	0.1152	0.9884	1.0000
	$\hat{F}_2$	0.00	0.0435	0.0645	0.0530	0.0980	0.1065	0.0453	0.0663	0.0558	0.0928	0.1050
		0.02	0.0405	0.0590	0.0615	0.0925	0.1080	0.0467	0.0681	0.0609	0.0840	0.1054
		0.04	0.0350	0.0605	0.0545	0.1070	0.1020	0.0412	0.0635	0.0512	0.0980	0.1013
		0.06	0.0405	0.0655	0.0510	0.1265	0.1090	0.0537	0.0788	0.0497	0.1269	0.1118
		0.08	0.0475	0.0725	0.0495	0.1460	0.1110	0.0618	0.0875	0.0410	0.1636	0.1247
		0.10	0.0390	0.0675	0.0425	0.1675	0.1215	0.0565	0.0880	0.0432	0.1822	0.1202
	$\hat{F}_3$	0.00	0.0400	0.0600	0.0530	0.0805	0.0980	0.0536	0.0691	0.0564	0.0762	0.0989
		0.02	0.0420	0.0580	0.0580	0.0815	0.0885	0.0401	0.0593	0.0445	0.0796	0.0801
		0.04	0.0325	0.0515	0.0500	0.0960	0.0800	0.0412	0.0568	0.0484	0.0874	0.0768
		0.06	0.0445	0.0640	0.0490	0.1120	0.0965	0.0442	0.0548	0.0470	0.1084	0.0811
		0.08	0.0385	0.0530	0.0515	0.1260	0.0860	0.0470	0.0629	0.0438	0.1395	0.0935
		0.10	0.0415	0.0580	0.0505	0.1445	0.0900	0.0532	0.0764	0.0343	0.1611	0.1013
	$\hat{F}_4$	0.00	0.0465	0.0650	0.0545	0.1050	0.1640	0.0497	0.0641	0.0630	0.1110	0.1630
		0.02	0.0460	0.0670	0.0560	0.1155	0.1735	0.0494	0.0642	0.0538	0.0988	0.1553
		0.04	0.0425	0.0665	0.0485	0.1425	0.1620	0.0462	0.0585	0.0596	0.1353	0.1698
		0.06	0.0525	0.0800	0.0625	0.1820	0.1725	0.0643	0.0788	0.0676	0.1638	0.1778
		0.08	0.0605	0.0820	0.0575	0.1885	0.1845	0.0667	0.0897	0.0503	0.1969	0.1975
		0.10	0.0625	0.0885	0.0510	0.2400	0.2080	0.0714	0.0875	0.0504	0.2481	0.2198

Continued on the next page

Table S.6 (continued)

ρ	$\hat{F}_i$	β	DGP 4					DGP 5				
			IPR	IPA	IPD	IPDI	PCR	IPR	IPA	IPD	IPDI	PCR
(0.950, 0.900, 1.000, 0.995)	$\hat{F}_1$	0.00	0.0430	0.1045	0.0590	0.1685	0.2060	0.0430	0.1058	0.0673	0.1648	0.1852
		0.02	0.1080	0.2645	0.0515	0.3670	0.4505	0.1327	0.3004	0.0666	0.4140	0.4380
		0.04	0.2520	0.5420	0.0780	0.6575	0.7830	0.3147	0.6022	0.1014	0.7130	0.7762
		0.06	0.4865	0.8030	0.0825	0.9085	0.9640	0.6109	0.8643	0.1177	0.9346	0.9624
		0.08	0.7540	0.9400	0.0905	0.9910	0.9985	0.8675	0.9714	0.1358	0.9874	0.9967
		0.10	0.8875	0.9820	0.1070	0.9985	1.0000	0.9556	0.9923	0.1508	0.9989	0.9989
	$\hat{F}_2$	0.00	0.0340	0.0480	0.0570	0.0720	0.0920	0.0375	0.0518	0.0551	0.0628	0.0827
		0.02	0.0455	0.0580	0.0665	0.0815	0.0965	0.0442	0.0617	0.0677	0.0754	0.0890
		0.04	0.0395	0.0575	0.0675	0.0790	0.0930	0.0421	0.0593	0.0598	0.0759	0.0853
		0.06	0.0405	0.0600	0.0620	0.0790	0.0900	0.0447	0.0676	0.0616	0.0834	0.0937
		0.08	0.0415	0.0560	0.0510	0.0775	0.1005	0.0435	0.0600	0.0468	0.0776	0.0864
		0.10	0.0395	0.0585	0.0720	0.0845	0.0985	0.0488	0.0702	0.0647	0.0834	0.0905
	$\hat{F}_3$	0.00	0.0365	0.0535	0.0560	0.0790	0.1210	0.0369	0.0402	0.0595	0.0777	0.1108
		0.02	0.0405	0.0505	0.0655	0.0895	0.1225	0.0382	0.0448	0.0579	0.0868	0.1109
		0.04	0.0420	0.0595	0.0580	0.0855	0.1370	0.0465	0.0565	0.0548	0.0859	0.1274
		0.06	0.0470	0.0655	0.0490	0.0940	0.1410	0.0414	0.0529	0.0469	0.0992	0.1351
		0.08	0.0480	0.0655	0.0670	0.0905	0.1445	0.0589	0.0726	0.0561	0.1034	0.1524
		0.10	0.0520	0.0645	0.0595	0.1065	0.1565	0.0510	0.0702	0.0510	0.1141	0.1558
	$\hat{F}_4$	0.00	0.0295	0.0445	0.0625	0.0755	0.1110	0.0369	0.0513	0.0540	0.0728	0.1075
		0.02	0.0335	0.0465	0.0600	0.0815	0.1215	0.0360	0.0486	0.0563	0.0754	0.1032
		0.04	0.0385	0.0530	0.0625	0.0840	0.1250	0.0465	0.0548	0.0582	0.0837	0.1175
		0.06	0.0465	0.0615	0.0570	0.0900	0.1270	0.0529	0.0599	0.0512	0.1003	0.1292
		0.08	0.0475	0.0700	0.0610	0.1130	0.1345	0.0534	0.0721	0.0622	0.1139	0.1353
		0.10	0.0595	0.0750	0.0570	0.1160	0.1535	0.0625	0.0741	0.0614	0.1179	0.1465
(0.900, 1.000, 0.995, 0.950)	$\hat{F}_1$	0.00	0.0430	0.0945	0.0710	0.1365	0.1665	0.0402	0.0920	0.0753	0.1489	0.1445
		0.02	0.0610	0.1780	0.0680	0.2625	0.3225	0.0724	0.1830	0.0752	0.3034	0.3034
		0.04	0.1680	0.3955	0.0690	0.4860	0.5665	0.2131	0.4197	0.0906	0.5323	0.5660
		0.06	0.2815	0.5770	0.0875	0.6895	0.7810	0.3605	0.6516	0.1057	0.7496	0.8013
		0.08	0.4975	0.7980	0.1045	0.8810	0.9415	0.6007	0.8612	0.1229	0.9210	0.9440
		0.10	0.7275	0.9325	0.1280	0.9745	0.9945	0.8201	0.9568	0.1621	0.9751	0.9884
	$\hat{F}_2$	0.00	0.0400	0.0495	0.0565	0.0825	0.1035	0.0379	0.0418	0.0446	0.0892	0.0982
		0.02	0.0420	0.0515	0.0625	0.0885	0.1045	0.0403	0.0419	0.0545	0.0746	0.0937
		0.04	0.0375	0.0435	0.0605	0.0730	0.1170	0.0392	0.0442	0.0514	0.0872	0.1154
		0.06	0.0405	0.0475	0.0570	0.0735	0.1170	0.0429	0.0495	0.0649	0.0771	0.1106
		0.08	0.0565	0.0640	0.0575	0.0955	0.1330	0.0483	0.0560	0.0581	0.0960	0.1218
		0.10	0.0620	0.0770	0.0710	0.0945	0.1515	0.0681	0.0664	0.0581	0.0991	0.1467
	$\hat{F}_3$	0.00	0.0335	0.0510	0.0600	0.0710	0.1090	0.0491	0.0563	0.0591	0.0820	0.1065
		0.02	0.0360	0.0470	0.0565	0.0760	0.1020	0.0359	0.0463	0.0588	0.0735	0.0964
		0.04	0.0440	0.0565	0.0630	0.0785	0.1035	0.0420	0.0536	0.0629	0.0784	0.1016
		0.06	0.0405	0.0530	0.0670	0.0720	0.1135	0.0435	0.0517	0.0633	0.0737	0.1040
		0.08	0.0480	0.0575	0.0625	0.0915	0.1230	0.0516	0.0603	0.0576	0.0916	0.1245
		0.10	0.0555	0.0690	0.0645	0.1075	0.1385	0.0742	0.0825	0.0626	0.1085	0.1401
	$\hat{F}_4$	0.00	0.0455	0.0615	0.0550	0.0780	0.1085	0.0502	0.0641	0.0508	0.0736	0.0976
		0.02	0.0385	0.0550	0.0575	0.0735	0.0845	0.0408	0.0550	0.0566	0.0681	0.0839
		0.04	0.0395	0.0585	0.0645	0.0825	0.0990	0.0475	0.0591	0.0635	0.0817	0.0856
		0.06	0.0405	0.0640	0.0605	0.0840	0.1055	0.0567	0.0688	0.0523	0.0848	0.1029
		0.08	0.0470	0.0625	0.0715	0.0895	0.1055	0.0505	0.0647	0.0521	0.0900	0.0965
		0.10	0.0430	0.0615	0.0555	0.0825	0.1005	0.0576	0.0736	0.0598	0.0930	0.1058

ρ	β	DGP 6										DGP 7									
		GR	$\Pr(\neg H_0\vert\beta)$					R^2_{OOS}				GR	$\Pr(\neg H_0\vert\beta)$					R^2_{OOS}			
			IPR	IPA	IPD	IPDI	PCR	IPR	IPD	IPDI	PCR		IPR	IPA	IPD	IPDI	PCR	IPR	IPD	IPDI	PCR
	0.00	4.00	0.0575	0.0935	0.0810	0.1786	0.2236	-0.005	-0.004	-0.014	-0.017	3.22	0.0480	0.0642	0.0718	0.0622	0.0495	-0.004	-0.006	-0.003	-0.005
1.000	0.02	4.00	0.2140	0.3000	0.0795	0.3740	0.5135	0.006	-0.009	0.005	0.018	3.25	0.2777	0.3158	0.0679	0.2443	0.3350	0.012	0.001	0.008	0.011
0.995	0.04	4.00	0.5931	0.6932	0.0646	0.7012	0.8328	0.076	-0.005	0.041	0.099	3.26	0.6941	0.7200	0.0523	0.6077	0.7495	0.092	-0.001	0.032	0.092
0.950	0.06	4.00	0.8404	0.8964	0.0665	0.8819	0.9650	0.187	-0.008	0.085	0.215	3.27	0.8887	0.9028	0.0584	0.8127	0.9290	0.195	-0.008	0.087	0.204
0.900	0.08	4.00	0.9410	0.9575	0.0530	0.9435	0.9890	0.310	-0.000	0.149	0.331	3.26	0.9577	0.9642	0.0489	0.9209	0.9743	0.308	-0.004	0.145	0.312
	0.10	4.00	0.9780	0.9870	0.0430	0.9770	0.9985	0.437	-0.001	0.202	0.438	3.27	0.9818	0.9858	0.0421	0.9382	0.9894	0.397	-0.001	0.153	0.400
	0.00	4.00	0.0460	0.0820	0.0660	0.1230	0.1530	-0.010	-0.008	-0.013	-0.028	3.27	0.0611	0.0763	0.0515	0.0687	0.0606	-0.008	-0.009	-0.005	-0.009
0.995	0.02	4.00	0.1551	0.2186	0.0675	0.2771	0.3482	-0.011	-0.005	-0.009	-0.018	3.28	0.2173	0.2594	0.0583	0.2183	0.2604	-0.003	-0.004	-0.007	-0.006
0.950	0.04	4.00	0.4770	0.5766	0.0611	0.5841	0.7047	0.022	-0.008	0.013	0.032	3.32	0.5746	0.6174	0.0590	0.5630	0.6436	0.032	-0.005	0.014	0.032
0.900	0.06	4.00	0.7799	0.8259	0.0680	0.8229	0.9085	0.103	-0.005	0.054	0.118	3.28	0.8279	0.8516	0.0601	0.7905	0.8647	0.103	-0.005	0.052	0.109
1.000	0.08	4.00	0.9215	0.9470	0.0665	0.9385	0.9805	0.197	-0.012	0.150	0.211	3.27	0.9367	0.9438	0.0638	0.9084	0.9580	0.198	-0.007	0.139	0.202
	0.10	4.00	0.9630	0.9710	0.0620	0.9655	0.9930	0.284	-0.006	0.185	0.285	3.29	0.9727	0.9772	0.0522	0.9453	0.9828	0.269	-0.003	0.181	0.277
	0.00	4.00	0.0430	0.0640	0.0825	0.0930	0.1155	-0.005	-0.012	-0.009	-0.012	3.16	0.0568	0.0774	0.0674	0.0608	0.0719	-0.011	-0.009	-0.002	-0.010
0.950	0.02	3.99	0.0781	0.1157	0.0746	0.1437	0.1733	-0.006	-0.004	-0.014	-0.022	3.19	0.0994	0.1250	0.0778	0.1140	0.1099	-0.010	-0.003	-0.007	-0.013
0.900	0.04	4.00	0.2371	0.3087	0.0850	0.3242	0.3887	-0.009	-0.005	-0.005	-0.016	3.25	0.2652	0.3104	0.0773	0.2903	0.2998	-0.004	0.000	0.000	-0.002
1.000	0.06	4.00	0.4645	0.5420	0.0720	0.5585	0.6260	0.015	-0.005	0.015	0.023	3.25	0.4925	0.5461	0.0736	0.5251	0.5441	0.012	-0.007	0.011	0.014
0.995	0.08	4.00	0.7292	0.7808	0.0866	0.7808	0.8413	0.041	-0.008	0.031	0.049	3.19	0.7333	0.7732	0.0773	0.7384	0.7712	0.024	-0.006	0.020	0.030
	0.10	4.00	0.8679	0.9055	0.0860	0.8964	0.9360	0.051	-0.005	0.060	0.058	3.19	0.8573	0.8775	0.0812	0.8538	0.8734	0.055	-0.001	0.050	0.057
	0.00	4.00	0.0640	0.1150	0.0720	0.2350	0.2595	-0.014	-0.010	-0.014	-0.031	3.33	0.0632	0.0793	0.0657	0.0672	0.0622	-0.014	-0.003	-0.005	-0.011
0.900	0.02	4.00	0.0770	0.1366	0.0790	0.2621	0.2796	-0.015	-0.009	-0.018	-0.028	3.32	0.0676	0.0898	0.0701	0.0888	0.0792	-0.012	-0.002	-0.009	-0.010
1.000	0.04	3.99	0.1632	0.2539	0.0786	0.3701	0.4221	0.001	-0.003	-0.001	-0.013	3.41	0.1540	0.1934	0.0753	0.1717	0.1768	-0.006	-0.006	-0.010	-0.009
0.995	0.06	4.00	0.2943	0.4194	0.0861	0.5140	0.5891	-0.003	-0.011	-0.003	0.000	3.32	0.3019	0.3559	0.0934	0.3352	0.3352	0.003	-0.003	0.003	0.001
0.950	0.08	4.00	0.5005	0.6150	0.0870	0.6970	0.7620	0.004	-0.002	0.003	0.005	3.33	0.4917	0.5497	0.0914	0.5078	0.5351	0.004	-0.002	0.011	0.008
	0.10	4.00	0.6967	0.8013	0.0991	0.8338	0.8804	0.019	-0.014	0.026	0.027	3.31	0.6808	0.7246	0.1027	0.6928	0.7080	0.021	-0.010	0.007	0.020

Table S.7: Average factor numbers (GR), Wald test rejection frequencies ($\Pr(\neg H_0|\beta)$) at 5% significance level, and forecast R^2_{OOS} values. Data from DGP 6 and DGP 7.

Table S.8: Factor-specific rejection frequencies of two-sided t-tests at 5% significance level. Data from DGP 6 and DGP 7.

ρ	$\hat{F}_i$	β	DGP 6					DGP 7				
			IPR	IPA	IPD	IPDI	PCR	IPR	IPA	IPD	IPDI	PCR
$\begin{pmatrix}1.000\\0.995\\0.950\\0.900\end{pmatrix}$	$\hat{F}_1$	0.00	0.0400	0.0525	0.0545	0.1076	0.1446	0.0475	0.0480	0.0597	0.0536	0.0516
		0.02	0.1745	0.1970	0.0535	0.2165	0.3405	0.2560	0.2610	0.0578	0.1956	0.2914
		0.04	0.4860	0.5295	0.0490	0.4835	0.6972	0.6402	0.6524	0.0589	0.5041	0.6885
		0.06	0.7189	0.7584	0.0500	0.6978	0.8724	0.8499	0.8535	0.0569	0.6888	0.8872
		0.08	0.8675	0.8950	0.0615	0.8120	0.9670	0.9420	0.9461	0.0580	0.8221	0.9567
		0.10	0.9355	0.9540	0.0675	0.8764	0.9915	0.9782	0.9807	0.0588	0.8677	0.9853
	$\hat{F}_2$	0.00	0.0680	0.1626	0.0655	0.3247	0.3602	0.0447	0.0496	0.0705	0.0521	0.0453
		0.02	0.0605	0.1485	0.0720	0.3235	0.3480	0.0545	0.0606	0.0509	0.0770	0.0636
		0.04	0.0556	0.1321	0.0531	0.3373	0.3133	0.0660	0.0727	0.0539	0.1672	0.0733
		0.06	0.0570	0.1276	0.0670	0.3687	0.2996	0.0777	0.0813	0.0542	0.2325	0.0801
		0.08	0.0495	0.1165	0.0570	0.3815	0.2815	0.0939	0.0963	0.0457	0.2840	0.0945
		0.10	0.0510	0.0980	0.0500	0.4247	0.2301	0.1053	0.1095	0.0389	0.3363	0.1119
	$\hat{F}_3$	0.00	0.0465	0.0660	0.0560	0.0850	0.1066	0.0582	0.0625	0.0484	0.0600	0.0662
		0.02	0.0500	0.0660	0.0500	0.0980	0.0985	0.0515	0.0570	0.0527	0.0685	0.0558
		0.04	0.0475	0.0686	0.0445	0.1421	0.1041	0.0569	0.0588	0.0369	0.1133	0.0575
		0.06	0.0555	0.0835	0.0505	0.2036	0.1156	0.0723	0.0813	0.0458	0.1705	0.0777
		0.08	0.0540	0.0815	0.0380	0.2300	0.1265	0.0752	0.0806	0.0349	0.2148	0.0824
		0.10	0.0530	0.0750	0.0285	0.2726	0.1086	0.0892	0.0952	0.0293	0.2837	0.1011
	$\hat{F}_4$	0.00	0.0395	0.0485	0.0500	0.0675	0.0750	0.0537	0.0571	0.0504	0.0563	0.0521
		0.02	0.0445	0.0595	0.0515	0.0860	0.0855	0.0517	0.0533	0.0541	0.0656	0.0607
		0.04	0.0470	0.0611	0.0425	0.1141	0.0831	0.0530	0.0554	0.0449	0.1035	0.0554
		0.06	0.0480	0.0680	0.0520	0.1726	0.0990	0.0561	0.0643	0.0496	0.1554	0.0602
		0.08	0.0420	0.0570	0.0330	0.1995	0.0935	0.0524	0.0500	0.0352	0.1933	0.0491
		0.10	0.0495	0.0570	0.0285	0.2366	0.0870	0.0591	0.0543	0.0291	0.2243	0.0591
$\begin{pmatrix}0.995\\0.950\\0.900\\1.000\end{pmatrix}$	$\hat{F}_1$	0.00	0.0370	0.0505	0.0490	0.0875	0.1070	0.0541	0.0551	0.0435	0.0561	0.0566
		0.02	0.1536	0.1721	0.0585	0.1921	0.2656	0.2006	0.2133	0.0522	0.1986	0.2224
		0.04	0.4009	0.4424	0.0606	0.4665	0.6066	0.5449	0.5554	0.0610	0.4849	0.5897
		0.06	0.6583	0.6998	0.0530	0.7074	0.8324	0.7910	0.8021	0.0591	0.7188	0.8289
		0.08	0.8444	0.8724	0.0725	0.8519	0.9480	0.9124	0.9190	0.0643	0.8648	0.9387
		0.10	0.9275	0.9410	0.0740	0.9145	0.9825	0.9595	0.9641	0.0648	0.9094	0.9732
	$\hat{F}_2$	0.00	0.0490	0.1275	0.0675	0.2075	0.2175	0.0435	0.0453	0.0588	0.0484	0.0496
		0.02	0.0485	0.1136	0.0615	0.1911	0.2021	0.0570	0.0576	0.0618	0.0674	0.0594
		0.04	0.0450	0.1041	0.0485	0.1802	0.1777	0.0497	0.0503	0.0509	0.0727	0.0533
		0.06	0.0560	0.1116	0.0560	0.2086	0.1801	0.0792	0.0841	0.0512	0.1060	0.0786
		0.08	0.0400	0.0935	0.0570	0.2121	0.1586	0.0704	0.0735	0.0496	0.1427	0.0765
		0.10	0.0385	0.0895	0.0510	0.2345	0.1485	0.0759	0.0844	0.0449	0.1888	0.0886
	$\hat{F}_3$	0.00	0.0500	0.0585	0.0575	0.0715	0.0915	0.0637	0.0668	0.0613	0.0613	0.0631
		0.02	0.0400	0.0535	0.0520	0.0710	0.0860	0.0496	0.0557	0.0508	0.0576	0.0576
		0.04	0.0435	0.0586	0.0526	0.0901	0.0826	0.0552	0.0600	0.0455	0.0806	0.0576
		0.06	0.0440	0.0630	0.0465	0.1161	0.0910	0.0701	0.0737	0.0475	0.1225	0.0750
		0.08	0.0435	0.0585	0.0495	0.1431	0.0900	0.0863	0.0900	0.0423	0.1427	0.0931
		0.10	0.0455	0.0615	0.0410	0.1565	0.0905	0.1069	0.1081	0.0352	0.1809	0.1154
	$\hat{F}_4$	0.00	0.0355	0.0420	0.0620	0.0700	0.1135	0.0547	0.0593	0.0570	0.0624	0.0578
		0.02	0.0395	0.0470	0.0485	0.0850	0.1251	0.0537	0.0643	0.0477	0.0613	0.0643
		0.04	0.0400	0.0475	0.0485	0.1236	0.1231	0.0561	0.0591	0.0650	0.1012	0.0547
		0.06	0.0455	0.0540	0.0555	0.1651	0.1401	0.0630	0.0661	0.0600	0.1522	0.0661
		0.08	0.0575	0.0640	0.0495	0.1831	0.1541	0.0790	0.0883	0.0558	0.1774	0.0968
		0.10	0.0655	0.0740	0.0445	0.2320	0.1690	0.0738	0.0799	0.0479	0.2291	0.0875

Continued on the next page

Table S.8 (continued)

ρ	$\hat{F}_i$	β	DGP 6 IPR	DGP 6 IPA	DGP 6 IPD	DGP 6 IPDI	DGP 6 PCR	DGP 7 IPR	DGP 7 IPA	DGP 7 IPD	DGP 7 IPDI	DGP 7 PCR
$\begin{pmatrix}0.950\\0.900\\1.000\\0.995\end{pmatrix}$	$\hat{F}_1$	0.00	0.0400	0.0550	0.0600	0.0710	0.0860	0.0563	0.0608	0.0513	0.0533	0.0588
		0.02	0.0946	0.1112	0.0501	0.1152	0.1462	0.1044	0.1109	0.0562	0.1009	0.1099
		0.04	0.2671	0.2911	0.0550	0.3032	0.3587	0.2627	0.2783	0.0618	0.2657	0.2838
		0.06	0.4880	0.5320	0.0705	0.5445	0.6265	0.4765	0.4935	0.0787	0.4689	0.5105
		0.08	0.7232	0.7518	0.0911	0.7508	0.8303	0.7162	0.7288	0.0788	0.6970	0.7414
		0.10	0.8369	0.8634	0.0925	0.8639	0.9130	0.8285	0.8376	0.0852	0.8180	0.8381
	$\hat{F}_2$	0.00	0.0350	0.0910	0.0750	0.1525	0.1555	0.0496	0.0542	0.0595	0.0456	0.0562
		0.02	0.0431	0.0851	0.0641	0.1457	0.1437	0.0601	0.0607	0.0621	0.0555	0.0627
		0.04	0.0430	0.0900	0.0660	0.1431	0.1481	0.0588	0.0614	0.0576	0.0658	0.0601
		0.06	0.0380	0.0680	0.0580	0.1150	0.1230	0.0594	0.0619	0.0639	0.0683	0.0639
		0.08	0.0480	0.0761	0.0536	0.1406	0.1246	0.0748	0.0767	0.0472	0.0852	0.0767
		0.10	0.0480	0.0685	0.0630	0.1286	0.1126	0.0733	0.0798	0.0484	0.0942	0.0812
	$\hat{F}_3$	0.00	0.0340	0.0375	0.0610	0.0560	0.0965	0.0509	0.0522	0.0588	0.0449	0.0549
		0.02	0.0320	0.0346	0.0516	0.0736	0.0936	0.0496	0.0496	0.0542	0.0529	0.0470
		0.04	0.0430	0.0460	0.0605	0.0770	0.1086	0.0702	0.0721	0.0601	0.0721	0.0696
		0.06	0.0425	0.0450	0.0540	0.1005	0.1160	0.0664	0.0715	0.0517	0.0830	0.0722
		0.08	0.0526	0.0531	0.0511	0.1041	0.1261	0.0774	0.0833	0.0498	0.0925	0.0898
		0.10	0.0515	0.0580	0.0505	0.1181	0.1416	0.0772	0.0838	0.0602	0.1139	0.0916
	$\hat{F}_4$	0.00	0.0370	0.0460	0.0550	0.0665	0.0890	0.0509	0.0525	0.0664	0.0509	0.0532
		0.02	0.0330	0.0381	0.0551	0.0671	0.0816	0.0492	0.0523	0.0614	0.0530	0.0477
		0.04	0.0410	0.0480	0.0545	0.0825	0.0990	0.0659	0.0659	0.0570	0.0703	0.0674
		0.06	0.0430	0.0550	0.0500	0.0905	0.1075	0.0502	0.0611	0.0560	0.0699	0.0575
		0.08	0.0526	0.0561	0.0556	0.1036	0.1181	0.0697	0.0757	0.0527	0.1001	0.0682
		0.10	0.0555	0.0580	0.0585	0.1211	0.1236	0.0682	0.0742	0.0569	0.1071	0.0787
$\begin{pmatrix}0.900\\1.000\\0.995\\0.950\end{pmatrix}$	$\hat{F}_1$	0.00	0.0420	0.0550	0.0615	0.0765	0.0860	0.0531	0.0561	0.0606	0.0531	0.0541
		0.02	0.0620	0.0815	0.0640	0.0890	0.1151	0.0716	0.0812	0.0565	0.0777	0.0893
		0.04	0.1858	0.2073	0.0606	0.2238	0.2609	0.2020	0.2192	0.0662	0.2096	0.2192
		0.06	0.3594	0.3859	0.0671	0.3869	0.4805	0.3786	0.3968	0.0777	0.3629	0.4048
		0.08	0.5605	0.6000	0.0895	0.5965	0.6805	0.5810	0.6002	0.0732	0.5568	0.6022
		0.10	0.7292	0.7703	0.0996	0.7593	0.8478	0.7402	0.7533	0.0856	0.7160	0.7538
	$\hat{F}_2$	0.00	0.0730	0.1650	0.0675	0.3780	0.3975	0.0608	0.0658	0.0486	0.0602	0.0670
		0.02	0.0705	0.1421	0.0755	0.3612	0.3517	0.0546	0.0608	0.0583	0.0602	0.0589
		0.04	0.0626	0.1372	0.0681	0.3255	0.3465	0.0597	0.0556	0.0591	0.0585	0.0669
		0.06	0.0465	0.1131	0.0606	0.3113	0.3123	0.0683	0.0750	0.0781	0.0886	0.0750
		0.08	0.0425	0.1045	0.0690	0.2920	0.2680	0.0699	0.0686	0.0717	0.0790	0.0809
		0.10	0.0501	0.0921	0.0726	0.2808	0.2417	0.0714	0.0795	0.0652	0.0962	0.0819
	$\hat{F}_3$	0.00	0.0500	0.0685	0.0560	0.1130	0.1400	0.0596	0.0621	0.0584	0.0633	0.0615
		0.02	0.0470	0.0640	0.0605	0.1091	0.1276	0.0503	0.0491	0.0571	0.0546	0.0491
		0.04	0.0601	0.0831	0.0586	0.1272	0.1527	0.0639	0.0645	0.0568	0.0657	0.0669
		0.06	0.0536	0.0731	0.0666	0.1086	0.1647	0.0670	0.0738	0.0603	0.0732	0.0701
		0.08	0.0575	0.0830	0.0535	0.1275	0.1665	0.0650	0.0729	0.0564	0.0735	0.0668
		0.10	0.0686	0.0911	0.0606	0.1527	0.1747	0.0875	0.0937	0.0571	0.1055	0.0956
	$\hat{F}_4$	0.00	0.0540	0.0720	0.0525	0.0855	0.1015	0.0565	0.0565	0.0501	0.0508	0.0565
		0.02	0.0415	0.0575	0.0540	0.0800	0.0895	0.0458	0.0458	0.0551	0.0465	0.0480
		0.04	0.0441	0.0581	0.0576	0.0861	0.0936	0.0472	0.0512	0.0566	0.0546	0.0506
		0.06	0.0541	0.0766	0.0556	0.1001	0.1096	0.0597	0.0632	0.0533	0.0739	0.0632
		0.08	0.0500	0.0725	0.0550	0.1050	0.1120	0.0644	0.0708	0.0531	0.0722	0.0630
		0.10	0.0556	0.0766	0.0591	0.1081	0.1076	0.0670	0.0741	0.0584	0.0783	0.0755

ρ	β	DGP 8 (GR applied to ΔX)										DGP 8 (GR applied to X)									
		GR	$\Pr(\neg H_0\|\beta)$					R^2_{OOS}				GR	$\Pr(\neg H_0\|\beta)$					R^2_{OOS}			
			IPR	IPA	IPD	IPDI	PCR	IPR	IPD	IPDI	PCR		IPR	IPA	IPD	IPDI	PCR	IPR	IPD	IPDI	PCR
	0.00	1.00	0.0605	0.0650	0.0625	0.0995	0.1351	-0.003	-0.001	-0.003	-0.000	3.89	0.0540	0.0935	0.0625	0.0995	0.2340	-0.010	-0.001	-0.003	-0.021
1.000	0.02	1.00	0.4065	0.4280	0.0605	0.3730	0.5955	0.014	-0.001	0.008	0.021	3.84	0.4190	0.5225	0.0605	0.3730	0.8515	0.033	-0.001	0.008	0.038
0.995	0.04	1.00	0.6773	0.6963	0.0480	0.5853	0.8014	0.083	-0.001	0.028	0.091	3.86	0.8835	0.9290	0.0480	0.5853	0.9940	0.125	-0.001	0.028	0.131
0.950	0.06	1.00	0.7910	0.8000	0.0480	0.6925	0.8645	0.187	-0.002	0.056	0.189	3.83	0.9820	0.9910	0.0480	0.6925	0.9995	0.255	-0.002	0.056	0.259
0.900	0.08	1.00	0.8414	0.8509	0.0475	0.7674	0.9025	0.209	-0.001	0.061	0.215	3.86	0.9955	0.9960	0.0475	0.7674	0.9980	0.359	-0.001	0.061	0.358
	0.10	1.00	0.8919	0.8939	0.0531	0.8173	0.9309	0.293	-0.001	0.112	0.306	3.89	0.9985	0.9995	0.0531	0.8173	1.0000	0.457	-0.001	0.112	0.457
	0.00	1.00	0.0465	0.0550	0.0545	0.0790	0.0880	-0.003	-0.000	-0.001	0.000	3.86	0.0605	0.0950	0.0545	0.0790	0.2295	-0.013	-0.000	-0.001	-0.028
0.995	0.02	1.00	0.2126	0.2351	0.0505	0.2846	0.3507	0.001	-0.001	-0.000	-0.004	3.88	0.2935	0.3980	0.0505	0.2846	0.7080	-0.002	-0.001	-0.000	-0.007
0.950	0.04	1.00	0.5183	0.5413	0.0505	0.5353	0.6383	0.021	0.001	0.008	0.028	3.87	0.8115	0.8775	0.0505	0.5353	0.9855	0.051	0.001	0.008	0.055
0.900	0.06	1.00	0.6903	0.7069	0.0580	0.7069	0.7594	0.052	-0.001	0.035	0.058	3.86	0.9720	0.9845	0.0580	0.7069	0.9980	0.139	-0.001	0.035	0.141
1.000	0.08	1.00	0.7841	0.7981	0.0571	0.7936	0.8402	0.104	-0.002	0.058	0.111	3.85	0.9970	0.9985	0.0571	0.7936	0.9995	0.219	-0.002	0.058	0.219
	0.10	1.00	0.8430	0.8490	0.0545	0.8280	0.8745	0.135	-0.000	0.087	0.143	3.87	0.9995	1.0000	0.0545	0.8280	1.0000	0.336	-0.000	0.087	0.338
	0.00	1.00	0.0370	0.0425	0.0551	0.0501	0.0611	-0.001	-0.001	-0.001	0.002	3.84	0.0390	0.0630	0.0551	0.0501	0.1320	-0.010	-0.001	-0.001	-0.017
0.950	0.02	1.00	0.0305	0.0350	0.0635	0.0545	0.0460	-0.001	-0.004	-0.003	-0.003	3.86	0.0840	0.1250	0.0635	0.0545	0.2255	-0.014	-0.004	-0.003	-0.021
0.900	0.04	1.00	0.0520	0.0610	0.0655	0.0990	0.0720	-0.002	-0.001	-0.002	-0.002	3.84	0.2505	0.3305	0.0655	0.0990	0.4990	-0.008	-0.001	-0.002	0.003
1.000	0.06	1.00	0.0976	0.1102	0.0596	0.1748	0.1422	-0.005	-0.003	-0.000	-0.007	3.86	0.5680	0.6770	0.0596	0.1748	0.8360	0.006	-0.003	-0.000	0.012
0.995	0.08	1.00	0.2232	0.2432	0.0706	0.3258	0.2678	0.002	-0.000	-0.001	0.000	3.83	0.8895	0.9220	0.0706	0.3258	0.9495	0.042	-0.000	-0.001	0.045
	0.10	1.00	0.3602	0.3972	0.0770	0.4932	0.4267	-0.012	-0.002	-0.005	-0.009	3.85	0.9685	0.9710	0.0770	0.4932	0.9730	0.047	-0.002	-0.005	0.047
	0.00	1.00	0.0360	0.0415	0.0475	0.0520	0.0575	0.000	-0.002	-0.002	-0.003	3.86	0.0380	0.0540	0.0475	0.0520	0.1065	-0.012	-0.002	-0.002	-0.014
0.900	0.02	1.00	0.0270	0.0290	0.0555	0.0455	0.0340	-0.004	-0.002	-0.003	-0.002	3.85	0.0550	0.0810	0.0555	0.0455	0.1335	-0.007	-0.002	-0.003	-0.011
1.000	0.04	1.00	0.0220	0.0250	0.0636	0.0410	0.0300	0.001	-0.001	-0.001	-0.003	3.87	0.1265	0.1720	0.0636	0.0410	0.2720	-0.015	-0.001	-0.001	-0.021
0.995	0.06	1.00	0.0295	0.0360	0.0675	0.0640	0.0410	-0.002	-0.003	-0.002	-0.003	3.85	0.2395	0.3255	0.0675	0.0640	0.4805	-0.010	-0.003	-0.002	-0.009
0.950	0.08	1.00	0.0430	0.0490	0.0750	0.0800	0.0575	-0.004	0.005	-0.002	-0.006	3.84	0.4785	0.5905	0.0750	0.0800	0.7340	0.003	0.005	-0.002	0.015
	0.10	1.00	0.0645	0.0710	0.0810	0.1070	0.0810	-0.004	-0.001	-0.003	-0.007	3.88	0.7800	0.8480	0.0810	0.1070	0.9185	0.023	-0.001	-0.003	0.041

Table S.9: Average factor numbers (GR), Wald test rejection frequencies ($\Pr(\neg H_0|\beta)$) at 5% significance level, and forecast R^2_{OOS} values. Data from DGP 8 (GR applied to ΔX) and DGP 8 (GR applied to X).

S.3 Data Description and Additional Results

Table S.10: List of variables providing a short description of the variables and the data source (GW: Welch and Goyal (2008), https://sites.google.com/view/agoyal145; GW2: Goyal *et al.* (2024), https://sites.google.com/view/agoyal145; FRED: McCracken and Ng (2016), https://www.stlouisfed.org/research/economists/mccracken/fred-databases). In addition, the table summarizes the preliminary results regarding the individual variables' peristence, measured in terms of estimated AR(1) coefficients $\hat{\rho}$, standard predictive regression p-values p_{OLS} and p_{IVX} ($a = 1, \eta = 0.95$) belonging to two-tailed t-tests using asymptotic inference, and standard in-sample R^2 values as a measure of each variable's predictive ability regarding the equity premium. Monthly data from January 1970 to December 2019.

	Description; Data source	$\hat{\rho}$	p_{OLS}	p_{IVX}	R^2
dp	Dividend-price ratio; GW2	0.9947	0.8830	0.7552	0.0000
dy	Dividend yield; GW2	0.9948	0.7923	0.7022	0.0001
ep	Earnings-price ratio; GW2	0.9900	0.9882	0.9501	-0.0000
e10p	Earnings' ten-year moving average divided by price; GW	0.9937	0.8730	0.8389	0.0000
ep10	Earnings divided by prices' ten-year moving average; GW	0.9934	0.9469	0.9357	0.0000
de	Dividend-earnings ratio; GW2	0.9852	0.8571	0.7962	0.0001
svar	Stock market variance; GW	0.4629	0.0056	0.0048	0.0167
VIXCLSx	VIX; FRED	0.8519	0.7615	0.6709	0.0003
skvw	Average stock skewness; GW2	0.0653	0.0346	0.0242	0.0089
tail	Tail risk from cross section; GW2	0.8879	0.0263	0.2092	0.0071
dtoy	Nearness to Dow 52-week high; GW2	0.8865	0.9721	0.7339	-0.0000
dtoat	Nearness to Dow all-time high; GW2	0.9389	0.4327	0.2843	0.0017
ygap	Stock-bond yield gap; GW2	0.9892	0.9454	0.8851	-0.0000
rdsp	Stock return dispersion; GW2	0.6328	0.3011	0.3568	0.0035
tchi	Technical indicators; GW2	0.9158	0.0416	0.1005	0.0103
avgcor	Average correlation of daily stock returns; GW2	0.8828	0.2165	0.4347	0.0037

Continued on the next page

Table S.10 Variable list (continue)

	Description; Data source	$\hat{\rho}$	p_{OLS}	p_{IVX}	R^2
lzrt	9 illiquidity measures; GW2	0.8381	0.0380	0.0555	0.0086
bm	Book-to-market ratio; GW	0.9956	0.6276	0.7644	0.0005
fbmlr	Single factor from book-to-market cross section (log returns); GW2	0.9704	0.1382	0.1502	0.0036
ntis	Net equity expansion; GW	0.9809	0.6607	0.9946	0.0004
FEDFUNDS	Effective Federal Funds Rate; FRED	0.9902	0.0277	0.0095	0.0087
CP3Mx	3-Month AA Financial Commercial Paper Rate; FRED	0.9894	0.0303	0.0086	0.0084
tbl	3-Month Treasury bill rate; GW	0.9910	0.0495	0.0308	0.0065
TB6MS	6-Month Treasury Bill; FRED	0.9922	0.0548	0.0334	0.0063
GS1	1-Year Treasury Rate; FRED	0.9927	0.0601	0.0364	0.0061
GS5	5-Year Treasury Rate; FRED	0.9956	0.0876	0.0584	0.0049
GS10	10-Year Treasury Rate; FRED	0.9968	0.1135	0.0790	0.0041
ltr	Long-term government bond return; GW	0.0332	0.0262	0.0327	0.0105
lty	Long-term government bond yield; GW	0.9962	0.1045	0.0715	0.0042
COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS; FRED	0.7548	0.2107	0.2614	0.0040
TB3SMFFM	3-Month Treasury C Minus FEDFUNDS; FRED	0.8794	0.0132	0.0047	0.0165
TB6SMFFM	6-Month Treasury C Minus FEDFUNDS; FRED	0.8873	0.0087	0.0024	0.0161
T1YFFM	1-Year Treasury C Minus FEDFUNDS; FRED	0.8610	0.0291	0.0286	0.0113
T5YFFM	5-Year Treasury C Minus FEDFUNDS; FRED	0.9376	0.0394	0.0456	0.0090
T10YFFM	10-Year Treasury C Minus FEDFUNDS; FRED	0.9510	0.0305	0.0351	0.0100
tms	Term spread (lty - tbl); GW2	0.9488	0.1494	0.4475	0.0037
aaa	Corporate bond yield; AAA-rated firms; GW	0.9984	0.1133	0.0744	0.0039
baa	Corporate bond yield; BAA-rated firms; GW	0.9990	0.1526	0.1024	0.0031
corpr	Long-term corporate bond returns; GW	0.1005	0.0054	0.0055	0.0257

Continued on the next page

Table S.10 Variable list (continue)

	Description; Data source	$\hat{\rho}$	p_{OLS}	p_{IVX}	R^2
AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS; FRED	0.9635	0.0286	0.0383	0.0100
BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS; FRED	0.9652	0.0384	0.0830	0.0101
dfy	Default yield spread (BAA - AAA); GW	0.9625	0.7636	0.8954	0.0002
dfr	Default return spread (corpr - ltr); GW	-0.0610	0.2029	0.1776	0.0068
INDPRO	Industrial production index; FRED	0.9991	0.3417	0.5152	0.0014
CUMFNS	Capacity Utilization: Manufacturing; FRED	0.9898	0.1394	0.2493	0.0047
ogap	Industrial production output gap; GW2	0.9944	0.2546	0.2988	0.0023
RPI	Real personal income; FRED	1.0015	0.1937	0.1535	0.0026
UNRATE	Civilian unemployment rate; FRED	0.9944	0.0998	0.3571	0.0042
CLAIMSx	Initial Claims; FRED	0.9751	0.3316	0.7160	0.0019
HOUST	Housing Starts: Total New Privately Owned; FRED	0.9644	0.5061	0.4929	0.0007
PERMIT	New Private Housing Permits (SAAR); FRED	0.9798	0.8198	0.7796	0.0001
DPCERA3M086SBEA	Real personal consumption expenditures; FRED	1.0014	0.1889	0.1502	0.0027
AMDMNOx	New Orders for Durable Goods; FRED	0.9932	0.2036	0.2224	0.0025
ndrbl	New orders/shipments to durable goods; GW2	0.7122	0.1267	0.1118	0.0055
ANDENOx	New Orders for Nondefense Capital Goods; FRED	0.9724	0.1731	0.1087	0.0028
AMDMUOx	Unfilled Orders for Durable Goods; FRED	1.0002	0.2187	0.2544	0.0023
BUSINVx	Total Business Inventories; FRED	1.0014	0.1368	0.0580	0.0034
M2REAL	Real M2 Money Stock; FRED	1.0043	0.0929	0.0207	0.0037
BUSLOANS	Commercial and Industrial Loans; FRED	1.0040	0.2185	0.2222	0.0022
REALLN	Real Estate Loans at All Commercial Banks; FRED	1.0024	0.2505	0.2645	0.0021
NONREVSL	Total Nonrevolving Credit; FRED	1.0044	0.1279	0.0536	0.0031

Continued on the next page

Table S.10 Variable list (continue)

	Description; Data source	$\hat{\rho}$	p_{OLS}	p_{IVX}	R^2
WPSFD49207	Producer Price Index: Finished Goods; FRED	0.9995	0.1123	0.0456	0.0040
WPSID61	PPI: Intermediate materials; FRED	0.9988	0.1681	0.2019	0.0031
WPSID62	PPI: Crude materials; FRED	0.9951	0.5645	0.9927	0.0006
OILPRICEx	Crude Oil, spliced WTI and Cushing; FRED	0.9911	0.8105	0.7909	0.0001
wtexas	Oil price changes; GW2	0.1996	0.1287	0.1288	0.0057
PPICMM	PPI: Metals and metal products; FRED	0.9944	0.7386	0.7852	0.0002
CPIAUCSL	Consumer price index: All Items; FRED	1.0000	0.1209	0.0443	0.0039
infl	Inflation; GW	0.6036	0.1095	0.1092	0.0058
PCEPI	Personal Cons. Expend.: Chain Index; FRED	0.9995	0.1091	0.0316	0.0041

Both the 3-Month AA Financial Commercial Paper Rate, CP3Mx, and its corresponding spread contain one missing observation in April 2020, *i.e.* when COVID-19 severly hit the markets in early 2020. We imputed these two observations using the Principal Component-Trimmed Score Regression approach of Folch-Fortuny *et al.* (2015) as follows: First, difference the entire dataset and select the number of factors using *GR*, dropping all observations from April 2020. Afterwards, use the Matlab code of Folch-Fortuny *et al.* (2015) to impute the two missing first differences. Subsequently, use the imputed changes to obtain April 2020 observations for CP3Mx and its spread by adding the imputed first differences to the March 2020 observations. Finally, check whether this operation changed the factor number by applying *GR* to the imputed dataset again; *GR* indicated 4 factors before and after the imputation. Morico and Stauskas (2025, p. 21) proceed in almost the same manner when imputing missings in their partially non-stationary EA-MD-QD dataset.

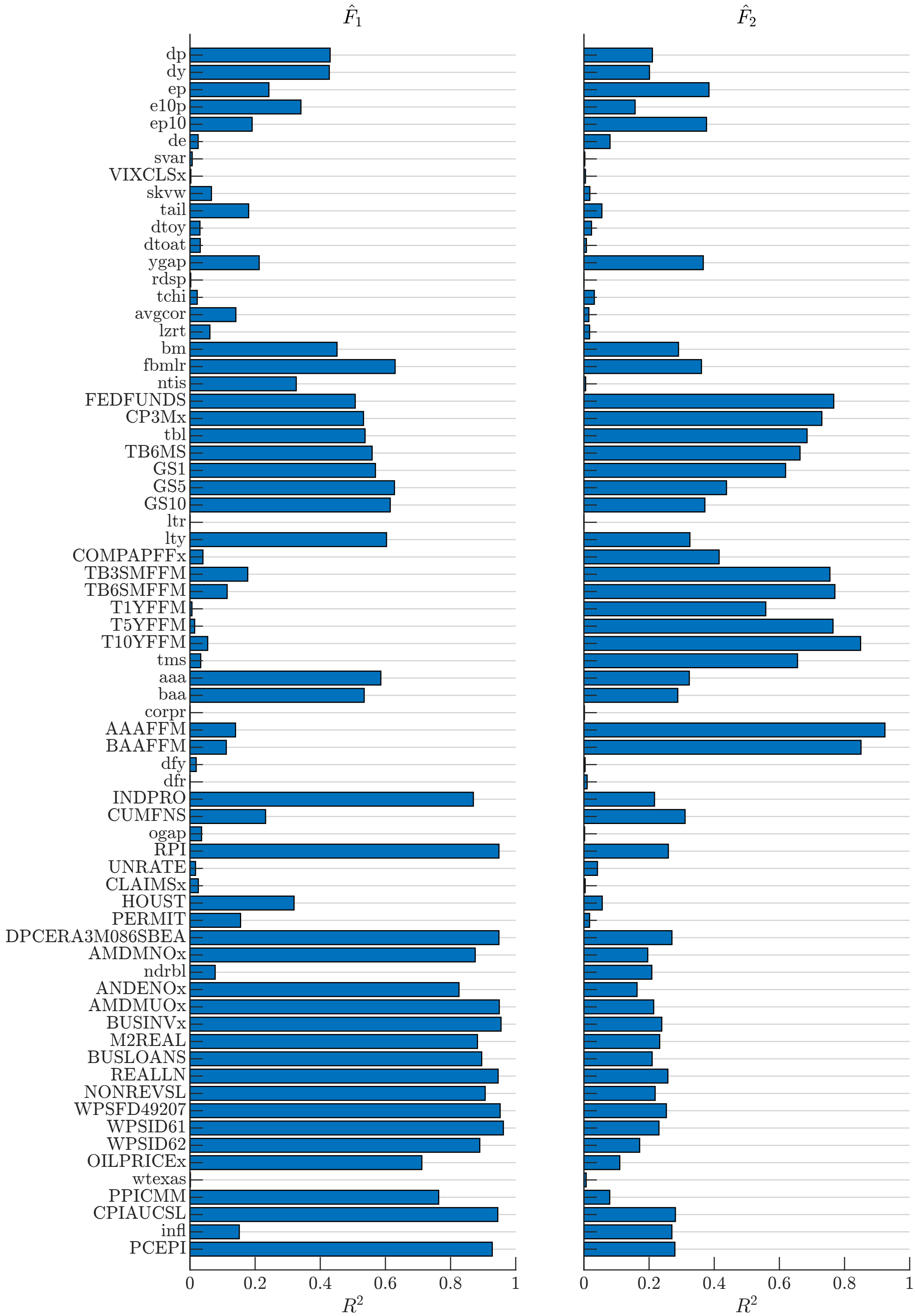

Figure S.1: R^2 of regressing observable variable i on $\hat{F}_1$, $\hat{F}_2$; $\hat{F}_i$ obtained from X.

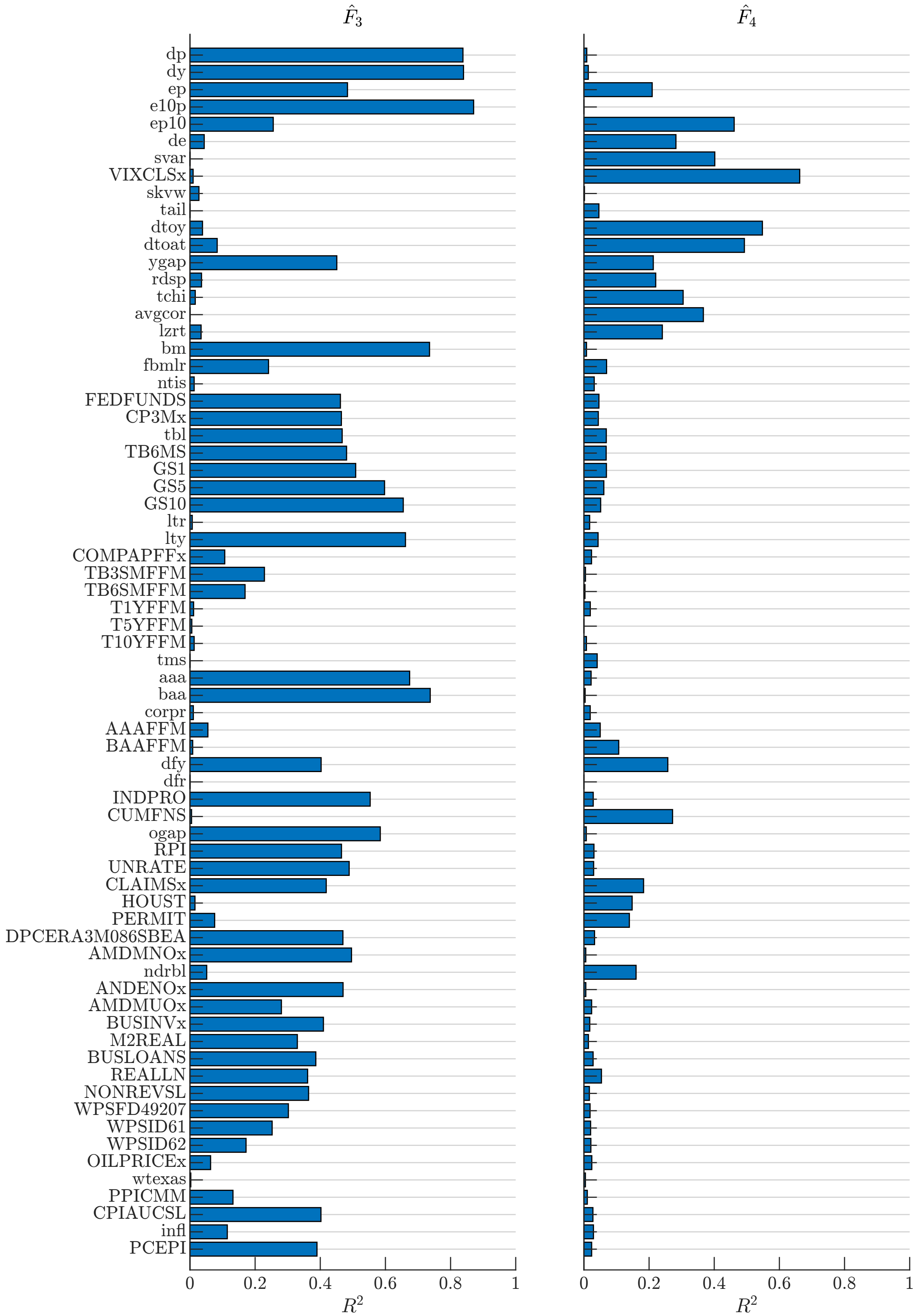

Figure S.2: R^2 of regressing observable variable i on $\hat{F}_3$, $\hat{F}_4$; $\hat{F}_i$ obtained from X.

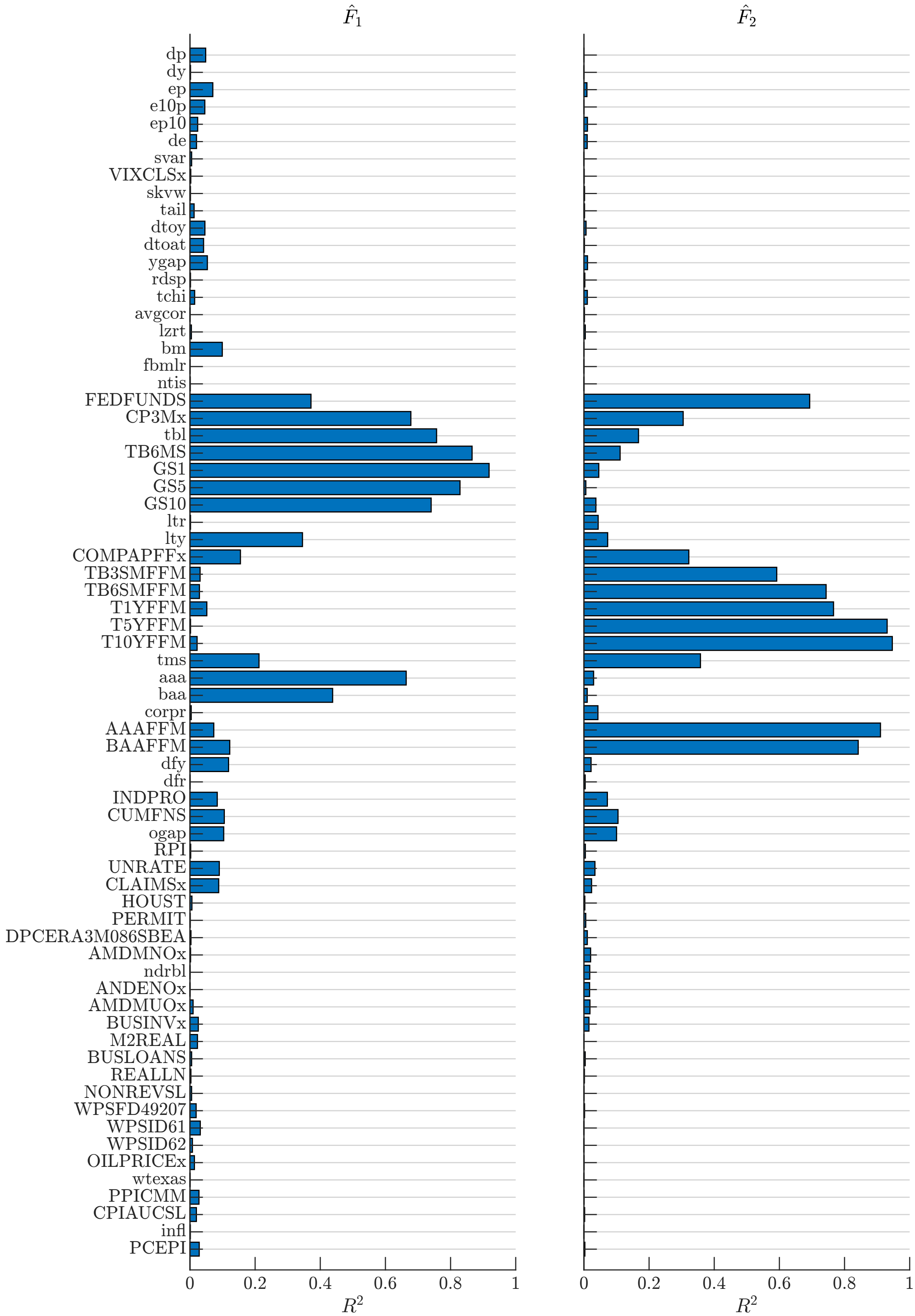

Figure S.3: R^2 of regressing observable variable i on $\hat{F}_1$, $\hat{F}_2$; $\hat{F}_i$ obtained from ΔX.

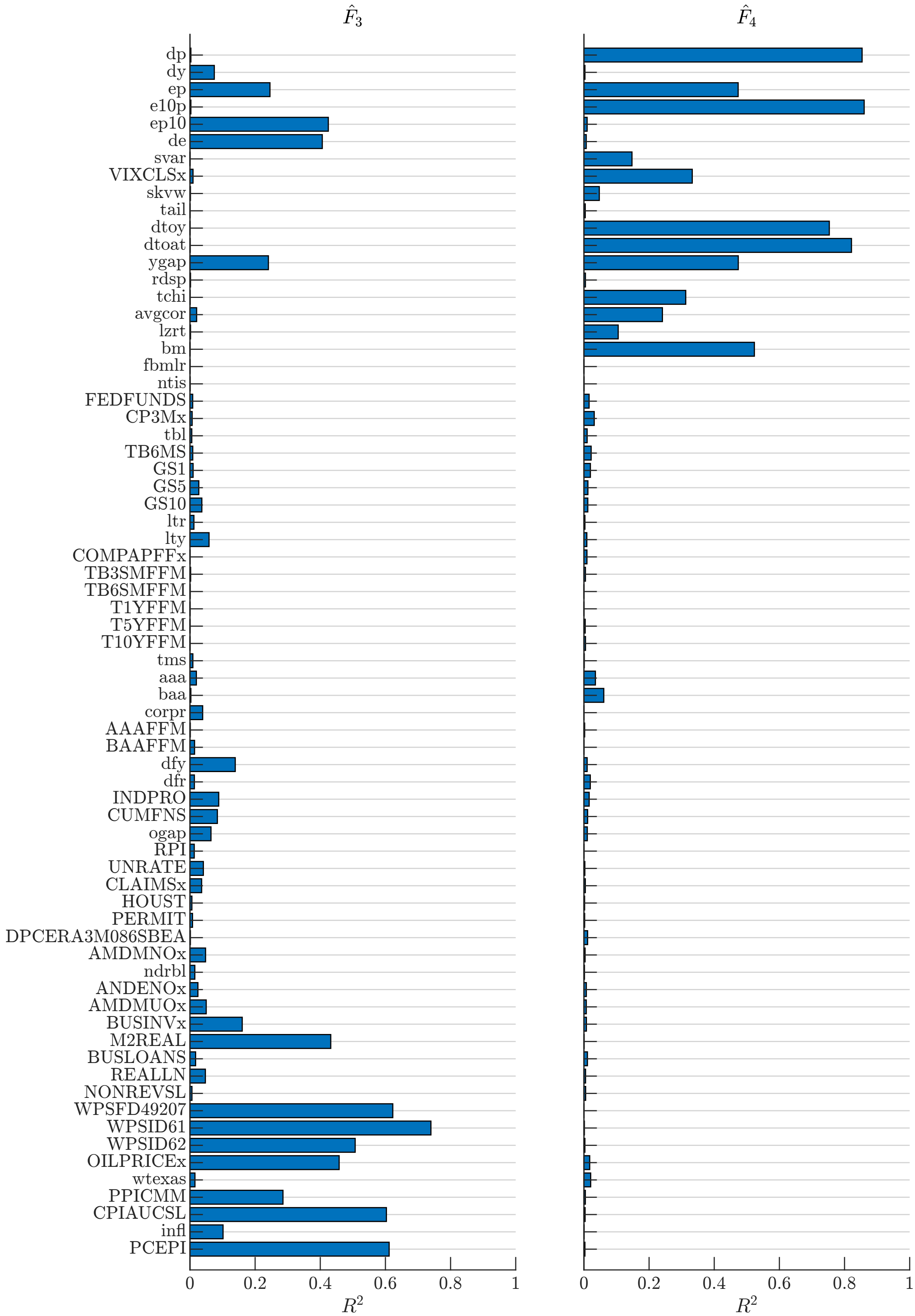

Figure S.4: R^2 of regressing observable variable i on $\hat{F}_3$, $\hat{F}_4$; $\hat{F}_i$ obtained from ΔX.