Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics and Operations Research

Determining the number of scaled PCA factors - Do the existing methods hold up?

Benhaiem, S.G. (572172)

Supervisor:	Opschoor, PA
Second assessor:	Name of your second assessor
Date final version:	2nd July 2023

Abstract

This paper serves as an extension of Huang et al. (2022), who introduce a development of standard Principal Component Analysis (PCA), called scaled PCA (sPCA). Unlike the traditional method, sPCA incorporates the predictive slopes of each predictor before applying PCA. The authors perform various Monte Carlo simulations for out-of-sample forecasting ability in the context of sPCA. Interestingly, the authors focus less on determining the number of sPCA factors (r), referring to Bai & Ng (2002) and Ahn & Horenstein (2013) for existing methods. This dependency on established methodologies prompts the question whether these estimators for the number of PCA factors remain consistent under sPCA. Our study offers valuable insights into the practical use of existing estimators within the context of sPCA. We find that in a large sample and strong factor scenario, most estimators exhibit consistent performance under sPCA. In the weak factor setting, both sPCA and PCA estimators encounter difficulties. In the context of in-sample prediction, we find that sPCA forecasts using the estimators beats the PCA counterpart in most settings. No specific estimator is identified to excel exclusively under the intended method and not under sPCA.

1 Introduction

Principal Component Analysis (PCA) is a common statistical technique used for data dimensionality reduction. PCA can effectively capture a major part of the variation by linearly transforming the data to a new Cartesian coordinate space. Bai (2003) has developed inferential theory for large dimensional factor models, and Stock & Watson (2002b) have popularised the method for macroeconomic forecasting applications. From the original idea proposed by Pearson (1901), the method has also been revised, such as the "Principal Curves" by Hastie & Stuetzle (1989), which uses explicit manifold construction, or the Sparse PCA by Zou et al. (2006), which finds linear combinations of a subset (instead) of all variables.

Huang et al. (2022) propose another development to the unsupervised method, called scaled principal component analysis (sPCA). Instead of maximising the common variation of the predictors, sPCA first scales the predictors with their respective predictive slope, and then applies PCA. The idea is that the scaling puts extra (less) weight on predictors that are (ir)relevant to the target, before the standard PCA is applied. To test the forecasting and predictive ability of the unsupervised learning method, the authors perform both simulation and empirical studies. The results show an improvement over the traditional method, in both the in- and out-of-sample predictive ability. The authors, however, are less concerned with the choice of the number of (s)PCA factors (r), which is a practical issue often discussed in factor analysis (Hastie et al., 2009). Instead, Huang et al. (2022) refer to Bai & Ng (2002) and Ahn & Horenstein (2013) for further information, stating that we can rely on existing methods for determining the number of factors. This raises the question whether the existing estimators for r are also consistent under sPCA. Furthermore, it would be interesting to study the forecasting performance of sPCA jointly with the estimators for r. In particular, we look at the existing estimators from Bai & Ng (2002) and Ahn & Horenstein (2013), as well as the much-used Kaiser heuristic from Kaiser (1960). As in Huang et al. (2022), we consider settings where factors are either weak or strong, but also partially target-(ir)relevant.

To judge the performance of existing estimators under sPCA, we consult previous literature. Bai & Ng (2002) simply averages Monte Carlo estimates of the number of factors (under various DGPs) to get an idea of the accuracy of their proposed estimators. Ahn & Horenstein (2013) use the root mean square error of the Monte Carlo estimates to compare the efficiency of their proposed estimators with the estimators from Bai & Ng (2002). The authors mainly find that the estimators from Bai & Ng (2002) are not robust to the pre-specified maximum number of PCA factors ($kmax$). Ahn & Horenstein (2013) show their revised estimators for the number of factors are robust to $kmax$. Finally, Huang et al. (2022) use the median of the mean squared errors of their (s)PCA forecasts to compare the predictive ability of forecasts constructed using sPCA and PCA factors. We also decide to test for superior predictive ability (SPA) using the multi-horizon SPA test from Quaedvlieg (2021), which jointly compares forecasts at different horizons, to get a more complete idea of the out-of-sample forecasting ability of (s)PCA factors. The difference with the forecasting setup from Huang et al. (2022) is that we construct point forecasts using $\hat{r}$ (s)PCA factors instead of r, where $\hat{r}$ is an estimated number of (s)PCA factors to include.

In the end, we conduct two distinct Monte Carlo experiments. The first tries to find out whether existing methods to determine the number of PCA factors work reasonably well under sPCA, or whether new methods need to be developed. The second tries to mimic a practical application (i.e. when the true number of (relevant) factors is not known) of sPCA, to find out whether the existing estimators work reasonably well in junction with a forecasting exercise. For both experiments, we compare the results with the PCA counterpart.

We find that in larger samples, and when factors are strong, the existing estimators perform as advertised for both PCA and sPCA, even when we include irrelevant factors. When the factors are weak, however, the estimators become inaccurate and unusable. The Kaiser heuristic performs poorly in larger samples by overestimating, while performing moderate and sometimes best in smaller samples. As for the results of the second experiment, we mostly find a superior in-sample performance from sPCA components (in junction with the estimators for the number of factors) to predict the target, in almost all DGPs we investigate. We do not find a specific estimator that is uniformly better at determining the number of factors for the forecasting exercises and also not one that exclusively works under PCA. As for the in-sample (s)PCA forecasting, we find that under all estimators the sPCA forecasts outperform the PCA forecasts. In the weak factor case, the effect is amplified. This leads us to believe that scaled PCA is better than PCA at gathering variable information, when we use estimators designed to determine the number of PCA factors.

The rest of the paper is divided as follows. Section 2 provides most of the methodology of the paper, starting with the general framework sPCA, a brief explanation of the estimators and heuristics we use, and the methods we employ to determine their performance. Section 3 presents the Monte Carlo experiment specifications and explains the results we obtain. For completeness, we conduct a small empirical study using the same macro variables as Huang et al. (2022), in Section 4. We end with the conclusion, in Section 5. The Appendix can be found after the reference list.

2 Methodology

Within this section, we discuss the general framework for sPCA, the existing methods of determining the number of PCA factors, and the evaluation procedures we use. Section 2.1 describes the setup of sPCA framework and explains its steps. Section 2.2 provides us with the list of existing estimators and heuristics for determining the number of PCA factors. Lastly, Section 2.3 explains the exact metrics and testing procedures we use for both our simulation and empirical results.

2.1 Scaled principal component framework

We set up a similar framework as in Huang et al. (2022). Let $X_{i,t}$ be the (observed) data for the ith cross-section unit at time t, for $i = 1, ..., N$, and $t = 1, ..., T$. We then construct the following factor model as such:

$$X_{i,t} = \mu_i + \lambda_i^{'} F_t + e_{i,t} = \mu_i + \phi_i^{'} g_t + \psi_i^{'} h_t + e_{i,t} \quad , \tag{2.1}$$

$$y_{t+h} = \alpha + \beta^{'} g_t + \epsilon_{t+h} \quad , \tag{2.2}$$

where $F_t = (g_t^{'}, h_t^{'})^{'}$ is the vector of common factors, of which g_t are the $(r_1 \times 1)$ relevant factors (linearly associated with target y_{t+h}), and h_t are the $((r - r_1) \times 1)$ irrelevant factors. Finally, $\lambda_i = (\phi_i^{'}, \psi_i^{'})^{'}$ is the vector of factor loadings for each predictor $i = 1, ..., N$, and $e_{i,t}$ is the idiosyncratic error term. α and β act as the constant and slope parameters and determine how the relevant factors interact with the target.[1] We can set $r_1 = r$ so that the factor model only contains relevant factors. A natural way to estimate the latent factors f_t is PCA. Bai (2003) develops extensive asymptotic theory about the principal components estimator. The author shows that the estimated components are asymptotically normal, with convergence rate equal to $\min(\sqrt{N}, \sqrt{T})$. Huang et al. (2022) go into the caveats of standard PCA in (out-of-sample) forecasting applications, indicating it has the disadvantage of not being able to distinguish between target-relevant and -irrelevant factors. The authors propose a new technique, where the predictors are first scaled by their (linear) predictive slope according to the target, before applying PCA. The hope is that the scaled PCA forecast beats the PCA forecast by assigning more (less) weight to predictors that are (ir)relevant to the target.

Scaled PCA can be divided into two steps. First, we form the scaled predictors, $(\hat{\gamma}_1 X_{1,t}, ..., \hat{\gamma}_N X_{N,t})$, where $\hat{\gamma}_i$ is the estimate as a result of performing OLS on the following regression:

$$y_{t+h} = \nu_i + \gamma_i X_{i,t} + u_{i,t+h} \quad \forall i = 1, ..., N \ . \tag{2.3}$$

Next, we can apply PCA to $(\hat{\gamma}_1 X_{1,t}, ..., \hat{\gamma}_N X_{N,t})$ and predict the target using first r factors. Huang et al. (2022) go into more detail, such as the prior that each predictor i should be demeaned before performing the second step. Forecasts, $\hat{y}_{t+h}^{\text{sPCA}}$, can then be constructed by regressing y_{t+h} on a constant term and the first r_1 sPCA factor estimates from $\hat{F} = [(\hat{g}_1, ..., \hat{g}_T)', (\hat{h}_1, ..., \hat{h}_T)']$, which is denoted by $\hat{f}_t^{\text{sPCA}}$. The authors perform Monte Carlo simulations under various DGPs, showing promising results. Scaled PCA forecasts beat PCA in most instances, but especially when the factors are weak.

Huang et al. (2022) always assume r_1 to be known, which can give a meaningful advantage to the sPCA forecasts (though no guarantees). This leads us to the search for estimators of r or r_1. For standard PCA, rules-of-thumb and estimators have been developed to estimate the number of factors to include. For sPCA, Huang et al. (2022) has a short paragraph about determining the number of sPCA factors, saying one should refer to Bai & Ng (2002) and Ahn & Horenstein (2013) for estimators of r, which we denote by $\hat{r}$. Naturally, it begs the question whether the estimators designed for PCA work as well under sPCA, and whether sPCA forecasts constructed

[1] In Huang et al. (2022), the simulation study uses a two-factor latent factor model with one target relevant factor. The authors construct the target by simply setting α to zero and β to one, such that the h-step-ahead target becomes the factor plus some noise. If we increase the number of target relevant factors, it becomes unclear how we should set β. Stock & Watson (2002a) use $\beta = 1$ for all target-relevant factors. We do the same.

using the first $\hat{r}$ sPCA (instead of r) factor estimates still outperform PCA forecasts. The latter is especially interesting, as it can show whether the simulation setup in Huang et al. (2022) is too beneficial to sPCA[2], while the former attempts to question the general validity of the estimators under sPCA. The results can show if the existing estimators are "good enough" under sPCA, or whether more-tailored estimators are required.

Bai & Ng (2002) indeed explore estimators for r based on model section criteria. The authors explain why standard information criteria are inadequate in a (latent) factor framework and construct adapted estimators. Using Monte Carlo simulation, the proposed estimators show promising results for determining the number of PCA factors to include. We can simply take the same idea and perform it with scaled PCA factors. Ahn & Horenstein (2013) address the issues of the Bai & Ng (2002) estimators, as they are highly sensitive to the choice of the maximum number of factors ($kmax$). The authors then propose two estimators, which are robust to the choice of $kmax$. Again, we can extend the simulation studies from Ahn & Horenstein (2013) by also looking at the performance with sPCA factors. In Huang et al. (2022), factors are allowed to either be strong or weak. The authors show that when the factors are weak, under mild conditions for N and T, the sPCA forecasts outperform the PCA counterpart.

The goal is to perform the Monte Carlo experiments in various scenarios, such as when factors are weak and partially target-relevant, in smaller and larger samples. As a result, we get a further understanding of how and when existing methods to determine the number of PCA factors are also usable for sPCA. We resort to in-sample forecasting, to extend the DGP landscape as much as possible, but we perform a small scale out-of-sample forecasting exercise for completeness. The in-sample forecasting results are less informative, and instead tell us how reliably (s)PCA picks up on variable information. This also makes it more difficult to compare, as the results in Huang et al. (2022) are mainly based on out-of-sample forecasting.

2.2 Heuristics and Estimators

In this section, we describe the selection of heuristics and estimators for r used in the remainder of the paper.

Most of the proposed estimators use the idea of selecting the best factor model (in terms of number of factors). Naturally, we turn to model selection criteria who penalise overfitting and reward high fit. We start with Bai & Ng (2002), who propose a set of alternatives to the popular AIC and BIC. Consider the following model selection criteria:

$$PC(k) = V(k, \hat{F}^k) + kg(N,T) \ , \tag{2.4}$$

$$IC(k) = \ln(V(k, \hat{F}^k)) + kg(N,T) \ , \tag{2.5}$$

where k represents the number of factors, $V(k, \hat{F}^k) = N^{-1}\sum_{i=1}^{N} \hat{\sigma}_i^2$ (where $\hat{\sigma}_i^2 = T^{-1}\hat{e}_i'\hat{e}_i$), and where the penalty function, $g(N,T)$, is chosen such that $PC(k)$ and $IC(k)$ can consistently estimate r. Bai & Ng (2002) construct three specifications of $g(N,T)$ per criteria function, which are represented in the Appendix. In our simulation study, we will use $IC_{p1}(k)$, as this is one of the preferred estimators in Bai & Ng (2002), even in smaller samples (i.e. $min\{N,T\} < 60$). The estimator for r is then

[2]Note that Huang et al. (2022) also construct sPCA forecasts using $(r+1)$ and $(r-1)$ factors to get an idea of this issue.

constructed by taking the arguments of the minima of $PC(k)$ or $IC(k)$. Formally, the estimators for r, denoted by $\hat{k}_{PC}$ ($\hat{k}_{IC}$) for criterion $PC(k)$ ($IC(k)$), look as such:

$$\hat{k}_{PC} = \underset{1 \leq k \leq kmax}{\arg\min} \; PC(k) \; , \tag{2.6}$$

$$\hat{k}_{IC} = \underset{1 \leq k \leq kmax}{\arg\min} \; IC(k) \; , \tag{2.7}$$

where $kmax$ is the maximum number of factors.[3]

Next, we turn to Ahn & Horenstein (2013), who discuss the issues in Bai & Ng (2002), and propose two revised criteria. In particular, the authors use the ratio of two adjacent eigenvalues of $XX'(NT)^{-1}$ for the first criterion function, denoted by $ER(k)$. The second criterion contains the growth rates of residual variance as one fewer factor is used in both the numerator and denominator. Formally, the estimators look as such:

$$ER(k) = \frac{\tilde{\mu}_{NT,k}}{\tilde{\mu}_{NT,k+1}} \; , \tag{2.8}$$

$$GR(k) = \frac{\ln[H(k-1)/H(k)]}{\ln[H(k)/H(k+1)]} \; , \tag{2.9}$$

where $\tilde{\mu}_{NT,k} = \psi_k[X'X(NT)^{-1}]$ (with $\psi_k(A)$ denoting the kth largest eigenvalue of a positive semidefinite matrix A), and $H(k) = \sum_{j=k+1}^{min\{N,T\}} \tilde{\mu}_{NT,j}$. In contrast to $\hat{k}_{PC}$ and $\hat{k}_{IC}$, we have to find the maximisers of $ER(k)$ and $GR(k)$. Denoted by $\hat{k}_{ER}$ and $\hat{k}_{GR}$, we get:

$$\hat{k}_{ER} = \underset{1 \leq k \leq kmax}{\arg\max} \; ER(k) \; , \tag{2.10}$$

$$\hat{k}_{GR} = \underset{1 \leq k \leq kmax}{\arg\max} \; GR(k) \; . \tag{2.11}$$

Ahn & Horenstein (2013) prove consistency of these estimators under reasonable assumptions.

As for heuristics, we can look at the Kaiser-criterion, proposed in Kaiser (1960). To determine the number of factors, you simply take the number of factors with eigenvalues greater than 1. The intuition for the eigenvalue-greater-than-one rule is that the average of the eigenvalues is 1 (see proof in the Appendix). The components with eigenvalues less than 1 capture less of the variance than the average component would. Note that the intuition assumes we perform PCA using the correlation matrix, which we do. Braeken & Van Assen (2017) show both theoretical and simulation results, revealing the inaccuracies of the Kaiser-criterion, though it would be interesting to see if similar outcomes show up under scaled PCA.

2.3 Determining the performance of estimators and heuristics

This section describes the functions we use to measure the performance of each estimator, and outlines the techniques we use to quantify the difference in forecasting ability between sPCA and PCA, while keeping the multiple inference issue in mind.

[3]Bai & Ng (2002) only use $kmax = 8$ in their simulation studies.

Let $nrep$ be the number of replications in our Monte Carlo simulation. For the first simulation, we only look at the estimates under both sPCA and PCA. We take the mean over all replications to get an idea of how an estimator would behave (on average), which is what Bai & Ng (2002) use. The second measure (used by Ahn & Horenstein (2013)) is the root mean square error (RMSE), which can be estimated as follows:

$$\text{RMSE}(\hat{k}_1, ..., \hat{k}_{nrep}|\ r) = \sqrt{\sum_{d=1}^{nrep} (\hat{k}_d - r)^2}\ , \tag{2.12}$$

where $\hat{k}_d$ denotes the estimate for r of the dth replication.

The second simulation setup involves a predicted target, $\hat{y}_{t+h}$. As in Huang et al. (2022), we can use the median of the mean square forecast error (MSFE). Using an expanding window from $t = \lfloor \frac{3}{4}T \rfloor, ..., T-1$, we get the forecasts $\{\hat{y}_{\lfloor \frac{3}{4}T \rfloor +1}, ..., \hat{y}_T\}$, and MSFE is constructed as follows[4]:

$$\text{MSFE}(\{\hat{y}_{\lfloor \frac{3}{4}T \rfloor +1}, ..., \hat{y}_T\}) = \sum_{t=\lfloor \frac{3}{4}T \rfloor +1}^{T} (\hat{y}_t - y_t)^2\ . \tag{2.13}$$

The benefit of using the MSFE is that it incorporates both the bias and variance of the (estimated) forecast error (Wackerly, 2008).

For the empirical study, we can test for superior forecast series while keeping in mind the multiple inferences problem. Quaedvlieg (2021) proposes a multi-horizon superior predictive ability (SPA) test. The test is based on an adaptation of the popular Diebold-Mariano test of equal predictive ability in Diebold & Mariano (1995), and falls explicitly into the framework of Hansen (2005) and Giacomini & White (2006). Instead of individually assessing forecasts at various horizons, which can lead to incoherent results, the approach focuses on the combined evaluation of forecasts from different models across a set of different horizons. The author sets up two tests: one that enables us to evaluate superior performance at every forecasting horizon (called uniform SPA), and the average SPA test, which allows inferior performance at certain horizons. As we need a superior forecasting performance at all horizons, the uniform SPA test is more stringent. We opt for the average SPA, however, as this seems more appropriate for our application. Quaedvlieg (2021) demonstrates, through simulation studies, that the described tests exhibit appropriate size and high power.

We construct a point forecast $\hat{y}^h_{t,m}$ at multiple horizons $h \in H$, where $H = \{1, 3, 6, 12\}$ is a set of forecasting horizons.[5] As we compare across three model forecasts, $m \in M = \{\text{sPCA}, \text{PCA}, \text{AR(1)}\}$, we get at each time t $\text{card}(m) \times \text{card}(H) = 12$ point forecasts. Then, for each model m forecast, we construct our losses $L^h_{t,m}(\hat{y}_{t,m}; y_t)$ per horizon using the squared error. For each horizon h, we can take the difference in losses for two different model forecasts. The loss differential between $m_1 \in M$ and $m_2 \in M \backslash \{m_1\}$ can be defined as follows:

$$d^h_{t;m_1,m_2} := L_{t,m_1}(\hat{y}_{t,m_1}; y_t) - L_{t,m_2}(\hat{y}_{t,m_2}; y_t)\ . \tag{2.14}$$

[4]The MSFE is of course not limited to an expanding window from $t = \lfloor \frac{3}{4}T \rfloor, ..., T-1$.

[5]Note that in Quaedvlieg (2021) the idea is to look at the quality of the full path of sequential h-step-ahead forecasts. Instead, we look at a practical set of horizons (i.e. one-month, quarter-year, half-year ahead, and one year), which is not in succession.

The (horizon) weighted average loss differential for model forecasts m_1 and m_2 can be defined as follows:

$$\mu_{m_1,m_2}^{(\mathrm{Avg})} = \sum_{h \in H} w_h \mu_{m_1,m_2}^{h}, \tag{2.15}$$

where $\mu_{m_1,m_2}^{h} := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[d_{t;m_1,m_2}^{h}]$ is the expected loss differential, and w_h are the positive weights we assign for each forecast horizon. Note that the w_h need to sum up to one. As we do not favour a particular forecasting horizon, we settle for equal weighting. The null hypothesis is then

$$H_0 : \mu_{m_1,m_2}^{(\mathrm{Avg})} \leq 0, \tag{2.16}$$

with corresponding alternative $\mu_{m_1,m_2}^{(\mathrm{Avg})} > 0$. Under the alternative, the m_1 forecasts outperform m_1 forecasts on average over all horizons $h \in H$. The null can be tested using a simple t-test. Quaedvlieg (2021) made the test such that the compared models can be nested or non-nested, unlike the Diebold-Mariano test in Diebold & Mariano (1995). The forecasting procedure can only be fixed or rolling window, however. We choose for a rolling window to get as close to the forecasting setup from Huang et al. (2022).

As for the multiple testing problem, we employ the multi-horizon SPA test, which essentially already divides the number of inferences by card(H). The trouble is that we still use the test under various DGPs, which are discussed in detail in Section 3. To combat the problem of multiplicity, we control the family-wise error rate (FWER) by employing the Holm-Bonferroni method proposed in Holm (1979). In contrast to the Bonferroni correction, which simply rejects null hypotheses with p-values less than the prior significance level divided by the number of inferences being made ($\frac{\alpha}{g}$), the Holm-Bonferroni method first sorts every p-value (lowest-to-highest) and then rejects the corresponding hypothesis according to a dynamic significance level. In order to reject H_k, the k-th p-value P_k needs to be less than a significance level of $\frac{P_1}{g+1-k}$, where g is the number of inferences you perform. The catch is that once a hypothesis is not rejected, you stop and all subsequent hypotheses are not rejected. The result is a uniformly more powerful test compared to the Bonferroni correction.

3 Monte Carlo simulation

In order to gain insight into the performance of the estimators discussed in Section 2.2 under scaled PCA, we decide to conduct Monte Carlo simulations under various settings. As in Huang et al. (2022), we look at either a strong or weak factor case. For each case, we divide the study into two. First, we look at the average and (root) mean squared error of the estimates of the number of factors under both PCA and scaled PCA, for which results and specifications for the strong (weak) factor setting are discussed in Section 3.1 (3.2). Then, in Section 3.3 (3.4 for weak factor setting), we discuss the second design, where we use the estimates to determine the number of (s)PCA factors to include in the forecasting model. While the first design looks at the general performance of the estimators under scaled PCA, the second looks at the performance in a practical application of scaled PCA. Note that we construct in-sample forecasts in Sections 3.3 and 3.4. An out-of-sample example can be found in the Appendix.

3.1 Performance of existing estimators for the number of PCA factors under sPCA: strong factors

We construct the predictors, X_{it}, using Equation (2.1). As a base setting, we can follow the simulation design in Bai & Ng (2002). The authors draw the $(N \times r)$ loadings, λ_i, and $(r \times T)$ factors, F_t, from $N(0,1)$ variables.[6] The $(N \times T)$ errors, e_{it}, are drawn from $N(0,\theta)$ variables. The authors also set $\theta = r$ such that the idiosyncratic component has the same variance as the common component [WHY?] Bai & Ng (2002) do not construct a target, as standard PCA does not make use of target information. As we compare the performance of the estimators under both PCA and sPCA, however, we require a target. The target is also relevant for the forecasting simulation, described in Section 3.3. The h-step-ahead target, y_{t+h}, can be simulated using Equation (2.2). For the base case, we can set $\beta = 1$ and $\alpha = 0$, such that we take the sum of all r_1 factors (with $r_1 = r$). For more complex DGPs, we can choose $r_1 < r$ factors to be target irrelevant, and set $\beta = 1$ for the target relevant factors. The main interest is to see how the estimators for r behave in cases that include irrelevant factors when using sPCA. One might think the estimates approach r_1 rather than r, as the scaled predictors are indirectly weighted using information from r_1 relevant factors, though, of course, there is no guarantee. PCA would clearly ignore the target setup. Finally, ϵ_{t+h} can be drawn from $N(0,1)$, in a similar fashion as in Huang et al. (2022). As for our hyperparameters, we run the simulation using 100 (1000) replications for the larger (smaller) sample.

Table 1: Average estimates and RMSEs for IC_{p1}, ER, GR, and Kaiser estimators for strong and fully target relevant factors. $kmax$ set at 8

	sPCA				PCA			
r	IC_{p1}	ER	GR	Kaiser	IC_{p1}	ER	GR	Kaiser
Panel A: Strong relevant factors; $(N,T) = (300,250)$; $kmax = 8$; $\theta = r$								
1	**1 (0)**	**1 (0)**	**1 (0)**	1.21 (0.92)	**1 (0)**	**1 (0)**	**1 (0)**	79.21 (78.26)
2	**2 (0)**	**2 (0)**	**2 (0)**	8.71 (8.85)	**2 (0)**	**2 (0)**	**2 (0)**	72.27 (70.33)
3	**3 (0)**	**3 (0)**	**3 (0)**	15.44 (13.82)	**3 (0)**	**3 (0)**	**3 (0)**	69.53 (66.57)
4	**4 (0)**	**4 (0)**	**4 (0)**	19 (16.47)	**4 (0)**	**4 (0)**	**4 (0)**	68.26 (64.3)
5	**5 (0)**	**5 (0)**	**5 (0)**	20.92 (16.83)	**5 (0)**	**5 (0)**	**5 (0)**	67.16 (62.19)
6	**6 (0)**	**6 (0)**	**6 (0)**	24.72 (19.93)	**6 (0)**	**6 (0)**	**6 (0)**	66.98 (61.01)
7	**7 (0)**	6.94 (0.6)	6.94 (0.6)	24.93 (18.96)	**7 (0)**	**7 (0)**	**7 (0)**	66.59 (59.62)
8	**8 (0)**	6.88 (2.8)	7.72 (1.4)	27.59 (20.62)	**8 (0)**	**8 (0)**	**8 (0)**	66.83 (58.86)
Panel B: Strong relevant factors; $(N,T) = (100,20)$; $kmax = 8$; $\theta = r$								
1	1.013 (0.13)	**1 (0)**	**1 (0)**	7.969 (8.36)	**1 (0)**	**1 (0)**	**1 (0)**	18.917 (17.92)
2	**2.031 (0.42)**	1.354 (0.8)	1.541 (0.68)	10.476 (9.5)	**1.991 (0.09)**	1.941 (0.24)	1.963 (0.19)	18.868 (16.87)
3	**3.005 (0.69)**	1.271 (1.85)	1.516 (1.7)	12.056 (9.88)	**2.896 (0.33)**	2.697 (0.67)	2.805 (0.51)	18.869 (15.87)
4	**4.03 (1.42)**	1.173 (2.89)	1.278 (2.82)	12.962 (9.67)	**3.468 (0.87)**	3.097 (1.39)	3.317 (1.16)	18.879 (14.88)
5	**4.898 (1.9)**	1.119 (3.92)	1.168 (3.89)	14.034 (9.63)	**3.757 (1.63)**	3.291 (2.23)	3.569 (2)	18.898 (13.9)
6	**5.322 (2.29)**	1.102 (4.93)	1.153 (4.89)	14.687 (9.22)	**3.586 (2.83)**	3.31 (3.18)	3.574 (2.96)	18.934 (12.94)
7	**5.568 (2.74)**	1.061 (5.95)	1.101 (5.92)	15.125 (8.65)	3.192 (4.18)	3.32 (4.12)	**3.536 (3.94)**	18.938 (11.94)
8	**5.747 (3.37)**	1.07 (6.94)	1.128 (6.9)	15.381 (7.95)	2.471 (5.8)	3.245 (5.16)	**3.43 (5)**	18.953 (10.96)

Notes. Averages of the estimates for IC_{p1}, ER, GR, and Kaiser estimators, with RMSEs in parenthesis for strong and (fully) relevant factors. The averages with the lowest RMSE for sPCA and PCA for each row are in bold. Panel A shows results for a relatively large sample, while Panel B shows results for a relatively small sample. θ is set to r, $kmax$ is fixed at 8, and all factors are equally relevant (i.e. $\beta = 1$). Note that all factors are relevant here (i.e. $r_1 = r$).

Table 1 shows the base case Monte Carlo results for $(N,T) \in \{(300,250),(100,20)\}$ and $kmax = 8$.[7] Looking at Panel A, we see that the Kaiser rule under sPCA is not far from 1 for $r = 1$,

[6] Note that this means that Bai & Ng (2002) only look at (target) relevant factors for the base setting, as $r_1 = r$

[7] Please note that the Kaiser heuristic is not bounded by $kmax$, but rather by $\min(N,T)$

but that the average estimates quickly explode and overestimate for $r > 1$. Under PCA, the rule is completely unusable, which coincides with the results from Braeken & Van Assen (2017). The estimators (i.e. IC_{p1}, ER, and GR) seem to perform much better than the heuristic, with the slight exception of an underestimation by ER and GR under sPCA for 7 and 8 factors. When looking at smaller samples, in Panel B, interesting results show up. Under sPCA, the average estimates for IC_{p1} perform the best for most values of r. The average estimates of ER and GR, under sPCA, never exceed 1.6, making them great estimators for low values of r, but unusable for higher values of r. The PCA counterpart uniformly performs better for ER and GR. What is interesting to note is that, for $r = 3$, the average estimate for IC_{p1} under sPCA is closer to the true number of factors than the PCA parallel, while having a higher RMSE (0.69 against 0.33). We can find similar results for $r \in \{4, 5\}$, which indicates the imprecision of IC_{p1}. The Kaiser rule is again inaccurate under both sPCA and PCA. In general, we find that in the strong and target relevant case, the estimators work well in larger samples and break down in smaller samples, with some exceptions for IC_{p1} under sPCA, while the heuristic is mostly unusable as an accurate estimator for the number of (s)PCA factors.

We also look at the base case with $kmax$ set at 16, for which the table can be found in the Appendix. The main result is that in smaller samples, the average IC_{p1} estimates blow up to 16 (coincidentally where $kmax$ is set at) for both sPCA and PCA. The remaining estimates are unaffected. In the larger sample, however, all estimates are similar to what we find in Table 1.

We can now move away from the base case and look at average estimates under more complex DGPs, one of which is the inclusion of irrelevant factors. As PCA does not make use of target information, the inclusion of irrelevant factors should not change the PCA simulation results from Table 1 (ceteris paribus). Hence, the main interest is to look at how sPCA reacts to irrelevant factors.

Table 2: Average estimates and RMSEs for IC_{p1}, ER, GR, and Kaiser estimators for strong and partially target (ir)relevant factors.

		sPCA: $(N,T) = (300, 250)$				sPCA: $(N,T) = (100, 20)$			
r	r_1	IC_{p1}	ER	GR	Kaiser	IC_{p1}	ER	GR	Kaiser
Panel A: Strong irrelevant factors; $kmax = 8$; $\theta = r$									
2	1	**2 (0)**	**2 (0)**	**2 (0)**	**2 (0)**	**2.095 (0.63)**	1.349 (0.81)	1.512 (0.7)	6.12 (5.55)
3	1	**3 (0)**	**3 (0)**	**3 (0)**	**3 (0)**	**3.339 (1.34)**	1.302 (1.83)	1.499 (1.71)	5.14 (3.53)
3	2	**3 (0)**	**3 (0)**	**3 (0)**	4.38 (2.54)	**3.113 (0.98)**	1.232 (1.87)	1.439 (1.74)	9.271 (7.36)
4	1	**4 (0)**	**4 (0)**	**4 (0)**	**4 (0)**	**4.381 (1.73)**	1.173 (2.89)	1.327 (2.8)	5.223 (3.01)
4	2	**4 (0)**	**4 (0)**	**4 (0)**	4.2 (0.72)	**4.204 (1.53)**	1.146 (2.91)	1.267 (2.84)	8.592 (5.82)
4	3	**4 (0)**	**4 (0)**	**4 (0)**	8.96 (6.75)	**4.079 (1.4)**	1.133 (2.91)	1.245 (2.85)	11.13 (8.04)
5	1	**5 (0)**	**5 (0)**	**5 (0)**	4.99 (0.1)	**5.296 (1.97)**	1.176 (3.88)	1.288 (3.81)	5.015 (2.56)
5	2	**5 (0)**	**5 (0)**	**5 (0)**	5.01 (0.1)	**5.158 (1.97)**	1.134 (3.91)	1.214 (3.86)	8.187 (4.61)
5	3	**5 (0)**	**5 (0)**	**5 (0)**	6.9 (3.09)	**4.986 (1.84)**	1.098 (3.93)	1.19 (3.87)	10.548 (6.61)
5	4	**5 (0)**	**5 (0)**	**5 (0)**	13.31 (9.51)	**4.95 (1.88)**	1.113 (3.92)	1.183 (3.88)	12.474 (8.28)
6	1	**6 (0)**	**6 (0)**	**6 (0)**	5.74 (0.63)	**5.933 (2.12)**	1.184 (4.87)	1.278 (4.81)	4.962 (2.59)
6	2	**6 (0)**	**6 (0)**	**6 (0)**	**6 (0)**	**5.727 (2.1)**	1.116 (4.91)	1.206 (4.86)	7.761 (3.57)
6	3	**6 (0)**	**6 (0)**	**6 (0)**	6.49 (1.18)	**5.62 (2.14)**	1.133 (4.91)	1.194 (4.87)	10.187 (5.47)
6	4	**6 (0)**	**6 (0)**	**6 (0)**	9.94 (5.02)	**5.501 (2.24)**	1.121 (4.91)	1.182 (4.87)	12.131 (7.09)
6	5	**6 (0)**	**6 (0)**	**6 (0)**	16.93 (12.25)	**5.605 (2.14)**	1.111 (4.92)	1.19 (4.87)	13.501 (8.26)

Notes. Average over estimates for IC_{p1}, ER, GR, and Kaiser estimators, with RMSEs in parenthesis for strong and partially irrelevant factors. The target is constructed using r_1 factors with $\beta = 1$. θ is set to r and $kmax = 8$. The left (right) column displays results for the larger (smaller) sample $(N,T) = (300, 250)$ ($(N,T) = (100, 20)$))

Table 2 shows the average estimates and RMSE (based on true value r) for combinations of

$r \in \{2,3,4,5,6\}$ and $r_1 < r$. The PCA average estimates and RMSE are not included in the tables. When looking at the results with the smaller sample, in the right column, we find a few interesting results. The sPCA average estimates do not seem to be very sensitive to the number of relevant factors r_1, and only slightly decrease with higher r_1 (for the same r). Only the Kaiser heuristic gets bigger as r_1 gets closer to r. The question now becomes whether we want the estimates to be close to either r_1 or r. Estimates closer to the true number of target relevant should yield to forecasts with a lower MSFE, as we then approach the DGP of the target.[8] In terms of performance of estimators, there is no clear answer. However, as we want to keep consistency in the showcase of our results, we compute the RMSE by comparing the estimate to the true number of (s)PCA factors, r. For the smaller sample, the RMSEs do not tell us much, as the average sPCA estimates (similar to Table 1) severely underestimate. In particular, on average, the ER and GR never estimate higher than 1.6 (similar to Table 1). The IC_{p1} performs relatively well considering the additional complexity of the target construction. Looking at the left column, it becomes apparent that the estimators perform as advertised in larger samples. This helps us believe that, asymptotically, the estimators are robust to the number of irrelevant factors when using sPCA and tend to estimate r. The Kaiser rule is, however, sensitive to the number of irrelevant factors. The interesting thing is that, in the larger sample, the heuristic is accurate when $r_1 \in \{1,2\}$, but positively biased when r_1 approaches r.

Both when we do and do not include target irrelevant factors, we mainly find that the estimators IC_{p1}, ER, and GR also perform as advertised under sPCA in the larger sample strong factor case. The estimators break down in our smaller sample results, as they were not particularly designed for smaller samples. The heuristic performs poorly all throughout, with some irregular exceptions.

3.2 Performance of existing estimators for the number of PCA factors under sPCA: weak factors

We also perform the same Monte Carlo simulation studies with weak factors, as Huang et al. (2022) show us that the sPCA forecasts then outperform the PCA forecasts. Before looking at forecasts, however, we can look at the general performance of the estimators for r when we have weak factors. As in Huang et al. (2022), we have for each factor $n < N$ randomly chosen loadings that are drawn a uniform distribution with support [0, 1]. Note that the draws are independent for each factor and for each (chosen) loading. The other $(N - n)$ loadings are set to zero. As before, we construct our predictor $X_{i,t}$ using Equation (2.1). We now have weak factors in the sense that only n out of N predictors load on our r factors. We choose $n \in \{10, \frac{N}{2}\}$ to get a setting where factors are very weak, and a setting half of the predictors load on our factors.[9] Again, we look at both a smaller and larger sample with $(N,T) \in \{(300,250),(100,20)\}$.[10] We conduct the Monte Carlo with $kmax = 8$, $\theta = r$ and for 100 (1000) replications for Panel A (B).

[8]There is of course no guarantee, but forecasts which use all factors will have a higher variance than forecasts which only use r_1 target relevant factors. As the MSFE takes into account the (absolute) bias and variance of our forecast, we expect that the forecasts constructed using r_1 target relevant factors to have a lower MSFE. More on this in Section 3.3

[9]When $N = 100$, we choose $n \in \{10, 50\}$, which are the extremes in Huang et al. (2022)

[10]$\frac{N}{2}$ is now 50 in the small sample and 150 in the large sample.

Table 3: Average estimates and RMSEs for IC_{p1}, ER, GR, and Kaiser estimators for weak and fully target-relevant factors.

		sPCA				PCA			
r	n	IC_{p1}	ER	GR	Kaiser	IC_{p1}	ER	GR	Kaiser
Panel A: Weak relevant factors; $(N,T) = (300, 250)$; $kmax = 8$; $\theta = r$									
1	10	7.26 (6.35)	**1 (0)**	**1 (0)**	0.4 (0.77)	**1 (0)**	1.66 (1.48)	1.66 (1.48)	113.87 (112.87)
1	150	**1 (0)**	**1 (0)**	**1 (0)**	**1 (0)**	**1 (0)**	**1 (0)**	**1 (0)**	104.97 (103.98)
2	10	7.55 (5.63)	**1.58 (0.81)**	**1.59 (0.81)**	0.09 (1.93)	**1 (1)**	2.07 (1.66)	2.07 (1.66)	113.79 (111.79)
2	150	**2 (0)**	1.44 (0.75)	1.7 (0.55)	3.04 (2.34)	**2 (0)**	1.99 (0.1)	1.99 (0.1)	104.23 (102.24)
3	10	6.42 (3.76)	**1.55 (1.69)**	**1.56 (1.69)**	0.04 (2.97)	**1 (2)**	2.35 (2.02)	2.35 (2.02)	113.75 (110.75)
3	150	**3 (0)**	1 (2)	1 (2)	8.87 (7.45)	2.38 (0.85)	2.62 (0.87)	**2.7 (0.77)**	104.21 (101.22
4	10	**5.14 (2.22)**	1.5 (2.63)	1.57 (2.62)	0.03 (3.97)	1 (3)	**2.14 (2.44)**	**2.14 (2.44)**	113.86 (109.86)
4	150	**3.73 (0.52)**	1 (3)	1 (3)	16.69 (14.2)	**1.24 (2.79)**	1.12 (2.94)	1.21 (2.89)	104.5 (100.5)
5	10	**3.73 (2.06)**	1.52 (3.58)	1.55 (3.56)	0.02 (4.98)	1 (4)	2.74 (3.04)	**2.78 (3.01)**	113.82 (108.82)
5	150	**2.84 (2.26)**	1 (4)	1 (4)	25.95 (22.26)	**1 (4)**	**1 (4)**	**1 (4)**	104.69 (99.7)
6	10	**3.04 (3.39)**	1.59 (4.55)	1.6 (4.54)	0.07 (5.94)	1 (5)	2.65 (3.92)	**2.7 (3.87)**	114.07 (108.07)
6	150	**1.58 (4.46)**	1 (5)	1 (5)	33.11 (28.22)	**1 (5)**	**1 (5)**	**1 (5)**	104.81 (98.81)
7	10	**2.67 (4.57)**	1.59 (5.48)	1.73 (5.37)	0.08 (6.93)	1 (6)	2.58 (4.84)	**2.64 (4.81)**	113.95 (106.95)
7	150	**1.11 (5.9)**	1 (6)	1 (6)	38.83 (33.27)	**1 (6)**	**1 (6)**	**1 (6)**	104.96 (97.97)
8	10	**2.37 (5.77)**	1.39 (6.64)	1.55 (6.53)	0.08 (7.92)	1 (7)	**2.38 (5.93)**	**2.38 (5.93)**	113.9 (105.9)
8	150	**1.01 (6.99)**	1 (7)	1 (7)	48.17 (41.48)	**1 (7)**	**1 (7)**	**1 (7)**	105.19 (97.19)
Panel B: Weak relevant factors; $(N,T) = (100, 20)$; $kmax = 8$; $\theta = r$									
1	10	1.099 (0.51)	**1.036 (0.3)**	1.043 (0.31)	3.426 (3.38)	**1 (0)**	3.298 (3.25)	2.701 (2.53)	19 (18)
1	50	1.027 (0.17)	**1.002 (0.06)**	1.004 (0.08)	5.633 (5.81)	**1 (0)**	1.008 (0.12)	1.008 (0.12)	19 (18)
2	10	**1.069 (1.01)**	1.066 (1.03)	1.073 (1.02)	6.265 (5.3)	**1 (1)**	3.498 (2.78)	2.754 (2.05)	19 (17)
2	50	**1.103 (0.97)**	1.005 (1)	1.008 (1)	9.71 (8.66)	1.002 (1)	1.314 (1.06)	**1.312 (0.99)**	19 (17)
3	10	**1.05 (1.97)**	1.056 (1.98)	**1.065 (1.97)**	8.534 (6.4)	**1 (2)**	3.545 (2.43)	**2.87 (2)**	19 (16)
3	50	**1.041 (1.97)**	1.002 (2)	1.006 (2)	11.959 (9.82)	1 (2)	1.473 (1.92)	**1.419 (1.87)**	19 (16)
4	10	1.068 (2.97)	1.081 (2.96)	**1.088 (2.95)**	10.327 (7.06)	1 (3)	3.355 (2.42)	**2.705 (2.31)**	19 (15)
4	50	**1.037 (2.98)**	1.001 (3)	1.003 (3)	13.924 (10.55)	1 (3)	1.518 (2.79)	**1.474 (2.75)**	19 (15)
5	10	1.063 (3.95)	1.103 (3.95)	**1.104 (3.94)**	11.508 (7.18)	1 (4)	**3.609 (2.79)**	2.871 (2.93)	19 (14)
5	50	**1.015 (3.99)**	1 (4)	1.001 (4)	15.073 (10.59)	1 (4)	1.597 (3.67)	**1.527 (3.66)**	19 (14)
6	10	**1.079 (4.94)**	1.067 (4.95)	**1.084 (4.94)**	12.844 (7.48)	1 (5)	**3.382 (3.5)**	2.734 (3.78)	19 (13)
6	50	**1.004 (5)**	**1.003 (5)**	**1.003 (5)**	16.073 (10.48)	1 (5)	**1.641 (4.6)**	1.526 (4.63)	19 (13)
7	10	1.067 (5.95)	1.102 (5.93)	**1.107 (5.92)**	13.688 (7.3)	1 (6)	**3.569 (4.17)**	2.804 (4.61)	19 (12)
7	50	**1.013 (5.99)**	1.001 (6)	1.001 (6)	16.871 (10.18)	1 (6)	**1.559 (5.6)**	1.451 (5.65)	19 (12)
8	10	1.054 (6.96)	**1.065 (6.95)**	**1.067 (6.95)**	14.579 (7.13)	1 (7)	**3.439 (5.12)**	2.783 (5.56)	19 (11)
8	50	**1.012 (6.99)**	1.001 (7)	1.002 (7)	17.192 (9.47)	1 (7)	**1.601 (6.55)**	1.472 (6.62)	19 (11)

Notes. Average over estimates for IC_{p1}, ER, GR, and Kaiser estimators, with RMSEs in parenthesis for weak and (fully) relevant factors. θ is set to r and $kmax = 8$. Panel A displays results for the larger sample $(N,T) = (300, 250)$, while Panel B displays the results for the smaller sample $(N,T) = (100, 20)$.

Table 3 shows the average estimates and RMSE for the estimators and heuristic for fully weak and target-relevant (with $\beta = 1$) factors. Looking at Panel A, we see that the IC_{p1} estimator from Bai & Ng (2002) performs counter-intuitively for $n = 10$ under sPCA, as the averages decline when r grows. For $n = 150$, we see accurate performance for $r \in \{1, 2, 3, 4\}$, but also a strange decline for $r > 4$. This performance adds up, as the estimator was not designed to work with weak factor structures. Under PCA, the estimator stays stuck at 1 for most values of r. The ER and GR from Ahn & Horenstein (2013) perform similarly in the sense that their averages stay both under 2, for sPCA, and under 3 for PCA. This behaviour is unexpected, as Ahn & Horenstein (2013) show that their estimators perform well in a weak factor setting. The authors did only look at a three-factor static model, where they varied the number of weak factors. Under PCA, at $r = 3$, we indeed see that the average estimates are not too far off, for both the smaller and larger sample. The accurate performance cannot be seen for other values of r or when we look at sPCA. As for the heuristic, the Kaiser rule is completely unusable in larger samples. For PCA, it predicts at around 110, and for sPCA the rule predicts irregularly (predicting close to 0 for $n = 10$). One would think that a larger n gives the estimators more chance to be accurate, but this is not the case for most values of r. Panel B shows the same results but for the smaller sample. The IC_{p1} estimator stays stuck at 1 for both PCA and sPCA now, which differs from the results from Panel B in Table 1. This time, the Kaiser rule predicts similar to Table 1 under sPCA, but stays stuck at 19 under PCA. The estimators ER and GR also perform poorly in smaller samples, predicting no more than 1.2 under sPCA and no more than 3.7 under PCA. One thing we see in both panels is that the average estimates for ER and GR are pulled to 1 when $n = \frac{N}{2}$.

We also perform Monte Carlo simulations for the weak but target-irrelevant setting which we have included in the Appendix. In general, though, none of the estimators seem to perform well under a weak and relevant factor structure.

3.3 In-sample forecasting using $\hat{r}$ (s)PCA factors: strong factors

The second design closely follows from the framework in Huang et al. (2022), who use Monte Carlo in order to compare the out-of-sample forecasting ability of factor models under scaled PCA and PCA. This time, we use the estimated number of (s)PCA factors to predict the target, instead of known r.[11] We also choose in-sample forecasting, which gives a better idea of how reliably (s)PCA pick up on variable information. The DGP of the h-step-ahead target, y_{t+h}, becomes important here. Using the first $\hat{r}$ (s)PCA factors (according to an estimator described in Section 2.2), we try to predict the h-step-ahead target using a window rule. Huang et al. (2022) only show the use an expanding window for 1-step-ahead. Looking at higher horizons (i.e. $h > 1$) would be meaningful only if our target is time-dependent, which requires a dynamic latent factor model (i.e. time-dependent factors). As the estimators from Bai & Ng (2002) and Ahn & Horenstein (2013) are designed for a static approximate factor model, we settle for a time-independent target and $h = 1$.[12]

As for the window rule, we use a rolling window in order to stay consistent with our empirical study.[13] At each prediction we estimate the parameters using a fixed window size of $\lfloor \frac{3}{4}T \rfloor$ (approx. 75% of the observations). The next prediction requires a new identification of the parameters using an updated information set which is shifted by one in the direction of T. We keep predicting until we reach T. The result is a forecast series from t equals $(\lfloor \frac{3}{4}T \rfloor + 1)$ to $(T + 1)$. As described in Section 2.3, we can use the median of the MSFEs (using Equation (2.13)) to evaluate the forecast series. Note that not only the performance of the estimators for r is important but also the quality of the estimated (s)PCA factors when comparing the median MSFE of sPCA and PCA. Within sPCA or PCA, however, the forecasting performance only depends on the quality of the estimators, which means we can form expectations using the associated results from Section 3.1 (3.2) for the strong (weak) factor case. It is important to note that we confine the Kaiser estimates further for the forecasting exercise, as we have cases where it predicts higher than the window size minus 1 and also lower than 1. Specifically, if in one replication the Kaiser estimate is 0 we replace it to be 1 and if the Kaiser estimate is higher than window size minus 1 we replace it to be the window size minus 1. The latter bound is set due to identification problems of the parameters for the rolling window forecasts.[14] The estimators IC_{p1}, ER, and GR do not have this issue as there are bounded

[11]Huang et al. (2022) use a static two-factor model to predict their target using 1, 2, and 3 (s)PCA factors. We use a static r-factor model to predict a target using an estimated number of factors.

[12]The empirical study looks at time-dependent targets such as the monthly inflation and industrial production growth. Predicting $(h + 1)$-step-ahead will be different than h-step-ahead.

[13]As a rolling or fixed window is required for the multi-horizon SPA test from Quaedvlieg (2021). Note that we stray from Huang et al. (2022) who use an expanding window.

[14]In particular, we need more observations than parameters for OLS estimation (Hastie et al., 2009). As we include a constant term and the (s)PCA factors, we bound the number of factors by the window size minus 1. We could use an elastic net regularization (Zou & Hastie, 2005) in order to remove the need for bounds, but we keep things simple and settle for OLS estimation

by 1 and $kmax$.

As for the DGPs, we can use the same ones described in Section 3.1. For the base case, all factor are target relevant which means that we can simply take the (s)PCA factors corresponding to the $\hat{r}$ largest eigenvalues. For the case where we include target irrelevant factors, taking (s)PCA factors corresponding to the r_1 largest eigenvalues gives us no guarantee that we only pick the target relevant factors (which are "relevant" for prediction). Hence, it becomes interesting to see how forecasts constructed using (s)PCA factors corresponding to the $\hat{r}$ largest eigenvalues in a setting that includes target irrelevant factors perform for both sPCA and PCA. As for our hyperparameters, we run the simulation using 100 replications for the larger sample and 1000 for the smaller sample for computational reasons. We set $kmax$ to 8 and θ to r.

Table 4: Median of MSFEs using IC_{p1}, ER, GR, and Kaiser estimators, for sPCA and PCA forecasts with strong and fully relevant factors.

	sPCA					PCA				
r	r	IC_{p1}	ER	GR	Kaiser	r	IC_{p1}	ER	GR	Kaiser
Panel A: Strong relevant factors; $(N,T) = (300, 250)$; $kmax = 8$; $\theta = r$										
1	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**1**	**1**	**1**	**1**	1.77
2	**1.04**	**1.04**	**1.04**	**1.04**	1.06	**1.06**	**1.06**	**1.06**	**1.06**	1.69
3	**1.03**	**1.03**	**1.03**	**1.03**	1.09	**1.04**	**1.04**	**1.04**	**1.04**	1.67
4	**1.01**	**1.01**	**1.01**	**1.01**	1.13	**1.02**	**1.02**	**1.02**	**1.02**	1.58
5	**1.09**	**1.09**	**1.09**	**1.09**	1.19	**1.11**	**1.11**	**1.11**	**1.11**	1.62
6	**1.16**	**1.16**	**1.16**	**1.16**	1.28	**1.17**	**1.17**	**1.17**	**1.17**	1.76
7	**1.24**	**1.24**	**1.24**	**1.24**	1.35	**1.22**	**1.22**	**1.22**	**1.22**	1.91
8	**1.3**	**1.3**	1.31	1.31	1.39	**1.26**	**1.26**	**1.26**	**1.26**	1.88
Panel B: Strong relevant factors; $(N,T) = (100, 20)$; $kmax = 8$; $\theta = r$										
1	0.86	**0.85**	0.86	0.86	2.21	**1.04**	**1.04**	**1.04**	**1.04**	97.61
2	**0.93**	**0.93**	**0.93**	**0.93**	6.74	**1.19**	**1.19**	1.2	1.2	110.3
3	1.08	1.09	**1**	**1**	16.5	**1.46**	1.47	1.52	1.52	102.02
4	1.23	1.28	**1.15**	**1.15**	26.17	**1.76**	1.87	1.92	1.92	114.35
5	1.58	1.62	**1.32**	**1.32**	31.42	**2.24**	2.55	2.62	2.62	126.98
6	2.03	2.04	**1.57**	**1.57**	45.13	**2.8**	3.44	3.53	3.53	146.37
7	2.91	2.53	**1.83**	**1.83**	73.76	**4.03**	4.75	4.63	4.63	178.35
8	3.85	2.97	**2.15**	**2.15**	82.78	**5.33**	6.22	5.7	5.7	201.21

Notes. Median of MSFEs for r, IC_{p1}, ER, GR, and Kaiser estimators, for rolling window forecasts. Strong and fully relevant factors with θ is set to r and β to one.

The results for the base case (relevant factors) are found in Table 4. Note that we also include the median MSFE for when we use the true number of (s)PCA factors r to predict the target. Bold numbers denote the lowest median MSFE for both sPCA and PCA for each row. Looking at Panel A, we see that at almost all values of r the median MSFE for estimators IC_{p1}, ER, and GR is equal to the median MSFE using the true number of (s)PCA factors r - both for sPCA and PCA forecasts. This can be linked to the estimators' accurate performance from Table 1. What is interesting is that the Kaiser rule estimates inaccurately in large samples (Table 1), but that only the PCA forecasts are heavily detrimented. This can be justified by the greater bias of the Kaiser rule under PCA. Panel B shows rather troubling results for the sPCA forecasts. At most values of r, the forecasts using $\hat{k}_{ER}$ or $\hat{k}_{GR}$ sPCA factors perform better than forecasts using r factors. It is possible that this is due to a small prediction window (5 observations) and not enough replications, leading to inaccurate Monte Carlo results. What we can see, however, is that forecasts using the Kaiser rule perform the worst - both for sPCA and PCA factors. On the PCA side, we can see more natural results. In Table 1 Panel B, the estimators under PCA performed well for lower values of r but poorly for higher values of r. This is somewhat reflected in Table 4 where the median MSFEs using estimators IC_{p1}, ER, and GR are close to the median MSFE using the true number of factors, for low values of r, and further apart for higher values of r. In general, we see that for the strong and target-relevant factor case the Kaiser estimates are unusable for small sample (s)PCA forecasting, while estimates from IC_{p1}, ER, and GR prove to be competent in larger samples and worthy in smaller samples.

Table 5: Median of MSFEs using IC_{p1}, ER, GR, and Kaiser estimators, for sPCA and PCA forecasts with strong and partially (ir)relevant factors.

		sPCA					PCA				
r	r_1	r (r_1)	IC_{p1}	ER	GR	Kaiser	r (r_1)	IC_{p1}	ER	GR	Kaiser
Panel A: Strong irrelevant factors; $(N,T) = (300,250)$; $kmax = 8$; $\theta = r$											
2	1	**0.99** (**0.99**)	**0.99**	**0.99**	**0.99**	**0.99**	**1** (1.49)	**1**	**1**	**1**	1.68
3	1	1.05 (**1.04**)	1.05	1.05	1.05	1.05	**1.08** (1.73)	**1.08**	**1.08**	**1.08**	1.65
3	2	**1.01** (**1.01**)	**1.01**	**1.01**	**1.01**	1.02	**1.02** (1.41)	**1.02**	**1.02**	**1.02**	1.63
4	1	**0.96** (**0.96**)	**0.96**	**0.96**	**0.96**	**0.96**	**0.99** (1.81)	**0.99**	**0.99**	**0.99**	1.58
4	2	**1.06** (1.07)	**1.06**	**1.06**	**1.06**	**1.06**	**1.08** (2.03)	**1.08**	**1.08**	**1.08**	1.61
4	3	**1.03** (1.06)	**1.03**	**1.03**	**1.03**	1.06	**1.07** (1.51)	**1.07**	**1.07**	**1.07**	1.64
5	1	**1.05** (**1.05**)	**1.05**	**1.05**	**1.05**	**1.05**	**1.08** (1.8)	**1.08**	**1.08**	**1.08**	1.63
5	2	**0.99** (1.04)	**0.99**	**0.99**	**0.99**	**0.99**	**1.02** (2.18)	**1.02**	**1.02**	**1.02**	1.6
5	3	**1.04** (1.11)	**1.04**	**1.04**	**1.04**	1.05	**1.06** (2.25)	**1.06**	**1.06**	**1.06**	1.6
5	4	**1.09** (1.11)	**1.09**	**1.09**	**1.09**	1.12	**1.1** (1.44)	**1.1**	**1.1**	**1.1**	1.72
6	1	**1.02** (1.04)	**1.02**	**1.02**	**1.02**	**1.02**	**1.05** (1.78)	**1.05**	**1.05**	**1.05**	1.58
6	2	**1.03** (1.06)	**1.03**	**1.03**	**1.03**	**1.03**	**1.06** (2.28)	**1.06**	**1.06**	**1.06**	1.64
6	3	**1.09** (1.16)	**1.09**	**1.09**	**1.09**	**1.09**	**1.11** (2.55)	**1.11**	**1.11**	**1.11**	1.68
6	4	**1.1** (1.15)	**1.1**	**1.1**	**1.1**	1.14	**1.1** (2.08)	**1.1**	**1.1**	**1.1**	1.71
6	5	**1.12** (1.16)	**1.12**	**1.12**	**1.12**	1.18	**1.14** (1.41)	**1.14**	**1.14**	**1.14**	1.7
Panel B: Strong irrelevant factors; $(N,T) = (100,20)$; $kmax = 8$; $\theta = r$											
2	1	0.88 (**0.82**)	0.87	0.86	0.86	1.3	**1.15** (1.4)	**1.15**	**1.15**	**1.15**	121.71
3	1	0.93 (**0.82**)	0.96	0.84	0.84	1.22	1.32 (1.7)	**1.3**	1.32	1.32	112.54
3	2	0.95 (**0.92**)	0.97	0.88	0.88	3.12	**1.34** (1.64)	**1.34**	1.37	1.37	93.73
4	1	1 (**0.81**)	1.07	0.84	0.84	1.17	1.52 (1.71)	**1.5**	**1.5**	**1.5**	101.53
4	2	1.1 (0.95)	1.16	**0.92**	**0.92**	2.59	**1.56** (1.92)	1.58	1.62	1.62	86.23
4	3	1.12 (1.08)	1.16	**1**	**1**	6.75	**1.54** (1.85)	1.65	1.65	1.65	126.52
5	1	1.1 (**0.79**)	1.25	0.81	0.81	1.13	1.73 (1.81)	**1.62**	**1.62**	**1.62**	109.66
5	2	1.27 (0.99)	1.33	**0.93**	**0.93**	2.22	**1.83** (2.14)	1.93	1.99	1.99	105.23
5	3	1.41 (1.18)	1.42	**1.08**	**1.08**	5.31	**2** (2.32)	2.09	2.12	2.12	108.84
5	4	1.47 (1.36)	1.54	**1.23**	**1.23**	18.46	**2.06** (2.18)	2.22	2.33	2.33	137.42
6	1	1.32 (**0.79**)	1.34	0.81	0.81	1.13	2.02 (1.85)	**1.77**	1.78	1.78	104.87
6	2	1.44 (1)	1.44	**0.93**	**0.93**	2.05	2.22 (2.35)	**2.1**	**2.1**	**2.1**	106.23
6	3	1.68 (1.22)	1.63	**1.09**	**1.09**	4.42	**2.45** (2.56)	2.49	2.54	2.54	118.48
6	4	1.75 (1.45)	1.73	**1.23**	**1.23**	14.05	**2.49** (2.63)	2.78	2.83	2.83	123.9
6	5	1.94 (1.72)	1.89	**1.37**	**1.37**	32.58	**2.68** (**2.68**)	3.08	3.21	3.21	134.71

Notes. Median of MSFEs for r, r_1, IC_{p1}, ER, GR, and Kaiser estimators, for rolling window forecasts. Strong and partially (ir)relevant factors with θ is set to r and β to one.

Table 5 shows median MSFEs in the strong but partially (ir)relevant case. Note that we include the median MSFEs of forecasts constructed using both r and r_1 (s)PCA factors. The important thing here is that there is no guarantee that we pick the target relevant factors when choosing the (s)PCA factors corresponding to the r_1 largest eigenvalues, which means that the forecasts using r_1 (s)PCA factors are not assured to have lowest median MSFE. Choosing r (s)PCA factors (corresponding to the r largest eigenvalues) makes sure we pick irrelevant factors which can inflate the variance of our forecasting error and lead to a higher (median) MSFE. From Panel A, however, we can see that choosing r factors almost always results in the best forecasts. The median MSFEs using the estimators IC_{p1}, ER, and GR are similar to those from r - which aligns with Table 2 as the average estimates indicate that IC_{p1}, ER, and GR successfully predicts r (rather than r_1 for sPCA). Under PCA, the large sample predictions using the Kaiser rule have inferior forecasting performance, though often better than forecasts using r_1 PCA factors. Under sPCA, the difference in forecasts using the Kaiser rule or r_1 sPCA factors is more muted. The former result is surprising as the Kaiser average estimates are close to 53 (Table 2). Panel B also shows interesting results. Under sPCA, it is either forecasts using r_1 or ER and GR that have the lowest median MSFE. In particular, we can see that the forecasts using r_1 outperform the rest when $r_1 = 1$, though the difference is not very large. Under PCA, forecasts are always inferior to the sPCA counterpart - which is also present in Table 4. This aligns with the results found in Huang et al. (2022).

In general, we see that in the strong factor case the estimators are quite competent for determining the number of factors for in-sample predictions. The heuristic works well in larger samples and under sPCA, but not so much under PCA. We also see a uniform superior performance from sPCA forecasts. One thing that remains unclear is whether r or r_1 (s)PCA factors are more efficient at predicting the one-step-ahead target, though there seems to be a slight preference for r (and estimators that predict close to r).

3.4 In-sample forecasting using $\hat{r}$ (s)PCA factors: weak factors

Next, we look at the forecasting performance when using the estimators to determine the number of (s)PCA factors to include for weak factors. Again, we have for each factor $n < N$ randomly chosen non-zero loadings that are drawn a uniform distribution with support [0, 1]. The remaining $(N - n)$ loadings are set to zero. We look at cases with $n \in \{10, \frac{N}{2}\}$.

Table 6: Median of MSFEs using IC_{p1}, ER, GR, and Kaiser estimators, for sPCA and PCA forecasts with weak and fully relevant factors.

	sPCA					PCA				
r	r	IC_{p1}	ER	GR	Kaiser	r	IC_{p1}	ER	GR	Kaiser
Panel A: Weak relevant factors; $(N, T) = (300, 250)$; $kmax = 8$; $\theta = r$; $n = 10$										
1	**1.23**	**1.23**	**1.23**	**1.23**	**1.23**	1.52	1.52	**1.51**	**1.51**	3.22
2	1.76	**1.69**	1.85	1.85	1.95	**2.59**	2.66	2.61	2.61	4.79
3	2.49	**2.35**	2.75	2.75	2.91	**3.77**	3.85	3.81	3.81	6.61
4	**3.22**	**3.22**	3.71	3.71	3.78	**4.86**	4.91	4.75	4.75	8.79
5	**3.98**	4.05	4.64	4.64	4.75	**5.86**	5.91	5.88	5.88	10.98
6	**4.77**	5.01	5.51	5.51	5.74	**7.04**	7.09	7.14	7.14	13.84
7	**5.21**	5.82	6.03	6.03	6.17	**7.66**	7.75	7.77	7.77	15.56
8	**6.04**	6.73	6.98	6.98	7.23	9.12	**9.06**	9.1	9.1	17.99
Panel B: Weak relevant factors; $(N, T) = (100, 20)$; $kmax = 8$; $\theta = r$; $n = 10$										
1	**0.58**	**0.58**	**0.58**	**0.58**	0.68	**1.85**	**1.85**	1.99	1.99	143.44
2	0.9	**0.85**	**0.85**	**0.85**	1.6	3.06	**2.93**	3.17	3.17	227.23
3	1.29	**1.1**	**1.1**	**1.1**	2.99	4.43	**3.94**	4.42	4.42	303.26
4	1.76	**1.41**	**1.41**	**1.41**	6.08	5.65	**4.6**	5.29	5.29	355.42
5	2.42	**1.7**	**1.7**	**1.7**	15.92	8	**5.83**	6.65	6.65	422.75
6	3.15	**2.01**	**2.01**	**2.01**	44.72	10.81	**7.07**	8.32	8.32	546.81
7	4.57	2.29	**2.28**	**2.28**	88.61	14.47	**8.11**	8.87	8.87	646.26
8	6.02	**2.75**	2.79	2.79	150.18	19.04	**8.95**	10.24	10.24	709.33

Notes. Median of MSFEs for r, r_1, IC_{p1}, ER, GR, and Kaiser estimators, for rolling window forecasts. Weak and relevant factors with θ is set to r and β to one.

Table 6 shows us the median MSFEs over 100 (1000) replications for the larger (smaller) sample for the weak and fully relevant factor case for $n = 10$. The table continues for $n = \frac{N}{2}$. Compared to the strong factor case we see (very) weak factors inferior forecasts. The increase in median MSFE also grows much faster with r compared to Table 4. As for the performance of the estimators and heuristic in junction with in-sample forecasting for weak factors, we see that the IC_{p1} performs relatively better in the smaller sample (compared to the strong factor counterpart), that the Kaiser rule now also becomes unusable for larger samples, but that r (s)PCA factors are still the best number to construct the one-step-ahead forecast. Revisiting Table 3 reveals some interesting things. In the smaller sample under PCA, the IC_{p1} kept predicting 1 which now translates to the best forecasts (within PCA). This means that constructing forecasts using a lower number of PCA factors is beneficial - we also see this when $n = \frac{N}{2}$, more on this later. Also under sPCA in the smaller sample we see that the estimators that predicted near 1 (i.e. smaller than r for most values of r) now perform the best for determining the number of sPCA factors for forecasting (IC_{p1}, ER, and GR). We also find a case where the estimators predicted not near 1 but still lower than r, such as estimators ER and GR under PCA (small sample), which indeed turns into worse forecasts than IC_{p1} (which predicted on average 1) but still better than the forecasts constructed using the true number of PCA factors. For the larger sample we do not find this, but rather median MSFEs that are very close to each other for r and estimators IC_{p1}, ER, and GR. Similar to the strong factor case, we find superior forecasts from sPCA for both the smaller and larger sample.

Table 6: *(Continued)* Median of MSFEs using IC_{p1}, ER, GR, and Kaiser estimators, for sPCA and PCA forecasts with weak and fully relevant factors.

	sPCA					PCA				
r	r	IC_{p1}	ER	GR	Kaiser	r	IC_{p1}	ER	GR	Kaiser
Panel A: Weak relevant factors; $(N,T) = (300, 250)$; $kmax = 8$; $\theta = r$; $n = 150$										
1	**1.02**	**1.02**	**1.02**	**1.02**	**1.02**	**1.04**	**1.04**	**1.04**	**1.04**	2.31
2	**1.02**	**1.02**	**1.02**	**1.02**	1.03	**1.07**	**1.07**	**1.07**	**1.07**	2.37
3	1.07	1.07	**1.06**	**1.06**	1.11	**1.1**	1.11	**1.1**	**1.1**	2.51
4	1.06	1.06	**1.02**	**1.02**	1.15	1.11	**1.1**	**1.1**	**1.1**	2.56
5	**1.19**	**1.19**	**1.19**	**1.19**	1.3	1.25	**1.24**	**1.24**	**1.24**	2.68
6	**1.26**	**1.26**	1.27	1.27	1.45	1.32	**1.31**	**1.31**	**1.31**	2.88
7	**1.32**	1.34	1.34	1.34	1.6	1.39	**1.36**	**1.36**	**1.36**	3.08
8	**1.39**	1.41	1.41	1.41	1.76	1.43	**1.42**	**1.42**	**1.42**	3.09
Panel B: Weak relevant factors; $(N,T) = (100, 20)$; $kmax = 8$; $\theta = r$; $n = 50$										
1	**0.59**	**0.59**	**0.59**	**0.59**	0.88	**1.17**	**1.17**	**1.17**	**1.17**	117.37
2	**0.68**	**0.68**	**0.68**	**0.68**	0.99	**1.15**	**1.15**	**1.15**	**1.15**	110.98
3	0.74	0.7	**0.69**	**0.69**	2.83	1.48	**1.42**	1.43	1.43	125.98
4	1.23	**0.87**	0.88	0.88	38.78	2.4	2.1	**2.06**	**2.06**	166
5	1.58	**1.01**	**1.01**	**1.01**	66.67	3.17	2.28	**2.25**	**2.25**	171.97
6	1.93	**1.08**	**1.08**	**1.08**	89.5	3.73	2.67	**2.61**	**2.61**	179.43
7	2.56	**1.23**	**1.23**	**1.23**	126.3	4.91	2.97	**2.95**	**2.95**	209.81
8	3.5	**1.38**	**1.38**	**1.38**	117.01	6.27	**3.33**	3.36	3.36	205.7

Notes. Median of MSFEs for r, r_1, IC_{p1}, ER, GR, and Kaiser estimators, for rolling window forecasts. Weak and relevant factors with θ is set to r and β to one.

For $n = \frac{N}{2}$, we find different results. As more predictors are loaded on our r factors, we expect an improvement in forecasting performance. Both PCA and sPCA forecast seem to improve over the case where $n = 10$. Comparing to the forecasting performance when the small sample factors are strong, in Table 4, we find that the both the sPCA and PCA forecasts constructed using the estimates improve while the (s)PCA forecasts using r do not. This might seem counter intuitive, as in Table 1 and 3 we see that in the small sample the estimators are competent for strong factors, but completely inaccurate in the weak factor case (predicting close to 1 when $n = 50$). This aligns with the results we get from when $n = 10$. In the smaller sample weak factor case, using 1 factor to construct forecasts seems to be the superior choice. In Huang et al. (2022), in the weak factor case, overpredicting the number of (s)PCA factors to include in the target construction does not seem to be detrimental, while underpredicting is. This is opposite to what we find. In the larger sample we do not find an improvement in forecasting performance in the weak factor case. What we do find is that although the estimators are not accurate in the large sample weak factor case (from Table 3), the forecasts using the estimators are comparable (or even better) to the forecasts constructed using r, similar to the strong factor case. In the strong factor case, however, the estimators are very accurate in the large sample. This weak factor result is interesting as it shows that underpredicting the number of (s)PCA we choose in not detrimental for in-sample forecasting ability. Finally, we can look at the forecasting ability in junction with the estimators and heuristic in a weak but partially target-(ir)relevant factors, for which the results are discussed and shown in the Appendix.

For all the simulation results, we never find a specific DGP where the estimators and heuristic solely works with PCA factors and not with sPCA factors, not in the estimators' accuracy nor in the ability to determine the number of factors for in-sample forecasting. In the forecasting ability we mostly find dominance from sPCA factors, even if the estimators for the number of factors are inaccurate. We also find that the number of target-irrelevant factors in the DGP do not matter much for the estimators' accuracy and ability to determine the number of factors for in-sample forecasting.

To answer the research question(s) we can say that many of the existing methods for determining the number of PCA factors work in a similar way under sPCA. We mostly find that strong factor large samples are useful for both accuracy and forecasting ability of these estimators, but that the estimators become inaccurate in the weak setting, though still usable for determining the number of (s)PCA factors in in-sample forecasting. Only the Kaiser rule behaves differently under sPCA and PCA where there seems to be a higher bias under PCA. The heuristic is still mostly inaccurate, even in base case settings.

4 Empirical example

For the empirical example, we use the same predictors and target from Huang et al. (2022). The authors take 123 monthly macro variables from the FRED-MD spanning from January 1960 to December 2019. Then, they construct one-month-ahead forecasts for the U.S. inflation, industrial production growth, change in the unemployment rate, and the VIX based on r sPCA factors. We only try to predict the U.S. inflation and industrial production growth, but we construct our forecasts using $\hat{r}$ sPCA factors. In contrast to Huang et al. (2022), we use a rolling window (with window size of approx. 75%) and multiple horizons (i.e. $h \in \{1, 3, 6, 12\}$). Then, we use the multi-horizon SPA test from Quaedvlieg (2021) to determine whether sPCA forecasts in junction with an estimator (i.e. IC_{p1}, ER, GR, Kaiser) outperform the PCA counterpart over multiple horizons. As the compared models may be nested, we perform the test across multiple forecast series. In particular, we test for superior predictive ability in between PCA and sPCA forecasts for each estimator or heuristic. This gives us 4 tests. We also test, within either sPCA or PCA, for SPA in between (s)PCA forecasts using an estimator against an AR(1) model - in the same fashion as Huang et al. (2022). This means we perform 12 inferences. As we try to predict both the U.S. inflation and industrial production growth, the total number of inferences climb to 24. At a 5% prior significance level and using the Holm-Bonferroni method, we reject the first hypothesis (according to the lowest p-value) using a corrected significance level of $\frac{0.05}{24} \approx 0.00417$. If the first is rejected, the next hypothesis is rejected using significance level of $\frac{0.05}{23} \approx 0.004545$. Assuming all hypothesis are rejected, the kth hypothesis uses significance level $\frac{0.05}{24+1-k}$ with $k = \{1, 2, ..., 24\}$.

Table 7: p-values for multi-horizon SPA test

	Industrial production				Inflation			
	IC_{p1}	ER	GR	Kaiser	IC_{p1}	ER	GR	Kaiser
Panel A: SPA of sPCA against PCA forecasts								
sPCA vs. PCA	>0.99	>0.99	>0.99	0.002*	0.015	0.037	0.04	0.001*
Panel B: SPA of (s)PCA against AR(1) forecasts								
sPCA vs. AR(1)	0.80	0.64	0.64	0.84	0.005	0.023	0.033	0.005
PCA vs. AR(1)	0.029	0.026	0.029	>0.99	0.04	0.032	0.028	>0.99

Notes. Reported values are the p-values of the multi-horizon SPA test for horizons 1, 3, 6, and 12 using squared error losses; we place an asterisk for significance according to the Holm-Bonferroni method. Panel A shows the results for when we compare sPCA forecasts constructed using a factor-augmented regression with $\hat{r}$ estimated sPCA factors with the PCA counterpart. Panel B shows the results for when we compare sPCA forecasts constructed using a factor-augmented regression with $\hat{r}$ estimated sPCA factors with forecasts constructed using an AR(1) model. The forecasts are constructed using a rolling window with window size at 75%. Left column shows results for the industrial production growth as the target, and the right column shows results for the change in inflation as a target.

Table 7 shows p-values for the multi-horizon SPA test for multiple comparisons. Looking at Panel A, we can see that only the forecasts using the Kaiser rule to determine the number of (s)PCA factors are different for PCA and sPCA, so much so that we find significant superior predictive ability from the sPCA forecasts for both the inflation and industrial production. For the industrial production predictions using the remaining estimators IC_{p1}, ER, and GR, there is no evidence to call for SPA, and we can even suppose that the PCA forecasts are superior to the sPCA counterpart.[15] When we try to predict inflation using the estimators, we again do not find any significant SPA, though serious evidence can be found for the sPCA forecasts constructed using IC_{p1}, ER, or GR factors. When looking at Panel B, we compare each (s)PCA forecast series (across horizons) against the AR(1) forecasts. We find plenty of evidence for PCA SPA, most notably for the estimators IC_{p1}, ER and GR. It also becomes evident that the PCA forecasts using the Kaiser rule are inferior to both sPCA and AR(1) forecasts. From Table 7 we can also see that estimators ER and GR behave similarly in the sense that the PCA forecasts seem to be more competent then the sPCA counterpart. In the end, however, it is only sPCA forecasts using the Kaiser rule that display significant SPA for both targets. From our simulation study this was also quite clear in terms of superior performance for average estimates and in-sample median MSFE, so it is interesting to also see it for an empirical out-of-sample scenario.

[15] a p-value close to one indicates that the average (across horizons) loss differences are likely negative which means that the PCA forecasts outperform the sPCA counterpart.

5 Conclusion

The main intention of this paper is to find whether existing estimators to determine the number of PCA factors still work under sPCA in their indented environments. We look at estimators from Bai & Ng (2002) and Ahn & Horenstein (2013), as these were cited by Huang et al. (2022), and a heuristic proposed in Kaiser (1960). To determine the performance of the estimators we look at both desirable properties of estimators and application in in-sample forecasting. We run Monte Carlo experiments under various factor setups and target constructions to come up with an idea of the performance of the existing estimators. We find that most of the estimators perform as advertised under sPCA in a large sample and strong factor setting. In the weak factor setting, the estimators struggle both under sPCA and PCA. When we construct in-sample forecasts using the number of (s)PCA factor determined by the estimates, we see that sPCA predictive ability is superior to the PCA counterpart. The gap in predictive ability grows when factors become weaker. In general, however, we do not find a specific estimator that works well under PCA and not under sPCA.

The empirical part of our study shows that out-of-sample forecasts are not always better when using sPCA factors. When predicting the monthly U.S. inflation rate across multiple horizons, we can find evidence for superior predictive ability of sPCA forecasts. For the monthly U.S. industrial production growth, it seems that the PCA forecasts are better. However, only for the forecasts constructed using the eigenvalue-greater-than-one (Kaiser) heuristic can we find significant superior predictive ability when comparing the sPCA forecasts with the PCA counterpart.

As for the limitations of this study we can name a few possible amendments. We can look at time-dependent factors in a dynamic latent factor setting to gain a further understanding of the ability of the existing estimators in the context of sPCA forecasting. Looking at out-of-sample forecasting can also tell us more about the accuracy required of the estimators when determining the number (s)PCA factors, which would also be more in line with the setup from Huang et al. (2022). Additionally, Huang et al. (2022) use a factor-augmented prediction model without lags of the target in the simulation study, but with lags in the empirical study. As an amendment, we can use the lags of the target in the simulation study, to see whether an autoregressive component helps or damages forecasts. The opposite can be done for the empirical study, where we look at the difference in forecasts without the autoregressive component.

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. & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.

Braeken, J. & Van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological methods*, *22*(3), 450.

Diebold, F. X. & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, *13*(3).

Giacomini, R. & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, *23*(4), 365–380.

Hastie, T. & Stuetzle, W. (1989). Principal curves. *Journal of the American Statistical Association*, *84*(406), 502–516.

Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Huang, D., Jiang, F., Li, K., Tong, G. & Zhou, G. (2022). Scaled PCA: A new approach to dimension reduction. *Management Science*, *68*(3), 1678–1695.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, *20*(1), 141–151.

Pearson, K. (1901). LIII. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, *2*(11), 559–572.

Quaedvlieg, R. (2021). Multi-horizon forecast comparison. *Journal of Business & Economic Statistics*, *39*(1), 40–53.

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). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.

Wackerly, D. D. (2008). *Mathematical Statistics with Applications 7th ed.*

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301–320.

Zou, H., Hastie, T. & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, *15*(2), 265–286.

6 Appendix

The Appendix is divided as follows. We start with Section A which reports the median Monte Carlo MSFEs in an out-of-sample forecasting setting. Section B gives us the missing and/or full tables from Section 3 with extra explanation. Section C provides a partial replication of the original sPCA paper by Huang et al. (2022). In Section D we can find the exact specifications $g(N, T)$ which where briefly discussed in Section 2. Finally, a short proof for intuition for the eigenvalue-greater-than-one rule can be found in Section E.

A Out-of-sample forecasting

In this section we present the out-of-sample forecasting performance (s)PCA factors in a strong and fully relevant factor case. A rolling window at 75% percent is used for $(N, T) = (300, 250)$.

Table 8: Median of MSFEs using IC_{p1}, ER, GR and Kaiser estimates for out-of-sample sPCA and PCA forecasts

	sPCA					PCA				
r	r	IC_{p1}	ER	GR	Kaiser	r	IC_{p1}	ER	GR	Kaiser
1	**1.99**	**1.99**	**1.99**	**1.99**	2.36	**1.98**	**1.98**	**1.98**	**1.98**	6824.08
2	**3.07**	**3.07**	**3.07**	**3.07**	3.93	**3.04**	**3.04**	**3.04**	**3.04**	12997.89
3	**4.17**	**4.17**	**4.17**	**4.17**	5.26	**4.14**	**4.14**	**4.14**	**4.14**	16114.08
4	**4.9**	**4.91**	**4.91**	**4.91**	6.64	**4.87**	**4.87**	**4.87**	**4.87**	20844.92
5	**5.97**	**5.97**	**5.97**	**5.97**	8.61	**5.89**	**5.89**	**5.89**	**5.89**	17468.09
6	**7.08**	**7.08**	**7.08**	**7.06**	10.46	**7.03**	**7.03**	**7.03**	**7.03**	31385.34
7	**8.04**	**8.04**	**8.04**	**8.04**	12.09	**7.97**	**7.97**	**7.97**	**7.97**	45841.25
8	9.32	9.32	9.32	**9.31**	14.61	**9.12**	**9.12**	**9.12**	**9.12**	50843.36

Notes. Reported are the median of the out-of-sample MSFE in a strong and fully target relevant factor case. The forecasts are constructed using either r (first column) or $\hat{r}$ (s)PCA factors. In bold are the median MSFEs that are lowest for each row for both sPCA and PCA. $kmax$ is set at 8 and $\theta = r$. 50 replications are used.

Table 8 presents median MSFEs of out-of-sample forecasts. For almost all values of r, the forecasts constructed using IC_{p1}, ER, or GR to determine the number of factors to include are not inferior to the forecasts constructed using the true number of factors. This aligns with the accuracy of the estimators in the strong and fully relevant factor case in Table 1. What we do see is that, in contrast with in-sample forecasting in Table 4, the out-of-sample PCA forecasts are superior to their sPCA counterpart. It is only when using the Kaiser rule to determine the number of factors where the sPCA forecasts win - this time by a landslide. Just as in Table 4, this is justified by the greater bias from the Kaiser estimates under PCA.

B Full tables

Table 9: Average estimates and RMSEs for IC_{p1}, ER, GR, and Kaiser estimators for strong and fully target relevant factors. $kmax$ set at 16

	sPCA				PCA			
r	IC_{p1}	ER	GR	Kaiser	IC_{p1}	ER	GR	Kaiser
	Panel A: Strong relevant factors; $(N,T)=(300,250)$; $kmax=16$; $\theta=r$							
1	**1 (0)**	**1 (0)**	**1 (0)**	1.06 (0.45)	**1 (0)**	**1 (0)**	**1 (0)**	79.54 (78.59)
2	**2 (0)**	**2 (0)**	**2 (0)**	8.09 (8.11)	**2 (0)**	**2 (0)**	**2 (0)**	72.21 (70.26)
3	**3 (0)**	**3 (0)**	**3 (0)**	14.31 (12.69)	**3 (0)**	**3 (0)**	**3 (0)**	69.52 (66.57)
4	**4 (0)**	**4 (0)**	**4 (0)**	18.23 (15.41)	**4 (0)**	**4 (0)**	**4 (0)**	68.27 (64.32)
5	**5 (0)**	**5 (0)**	**5 (0)**	21.28 (17.44)	**5 (0)**	**5 (0)**	**5 (0)**	67.6 (62.64)
6	**6 (0)**	5.95 (0.5)	**6 (0)**	24.52 (19.49)	**6 (0)**	**6 (0)**	**6 (0)**	66.9 (60.93)
7	**7 (0)**	6.88 (0.85)	**7 (0)**	25.72 (19.75)	**7 (0)**	**7 (0)**	**7 (0)**	66.55 (59.58)
8	**8 (0)**	7.09 (2.52)	7.79 (1.21)	26.11 (19.32)	**8 (0)**	**8 (0)**	**8 (0)**	66.32 (58.35)
	Panel B: Strong relevant factors; $(N,T)=(100,20)$; $kmax=16$; $\theta=r$							
1	15.955 (14.98)	**1 (0)**	**1 (0)**	8.501 (8.82)	15.31 (14.65)	**1 (0)**	**1 (0)**	18.906 (17.91)
2	16 (14)	1.335 (0.82)	**1.516 (0.7)**	11.036 (9.97)	15.972 (13.99)	1.947 (0.23)	**1.969 (0.18)**	18.861 (16.87)
3	16 (13)	1.201 (1.88)	**1.418 (1.76)**	12.209 (9.97)	16 (13)	2.7 (0.67)	**2.779 (0.57)**	18.857 (15.86)
4	16 (12)	1.172 (2.95)	**1.299 (2.81)**	12.845 (9.55)	16 (12)	3.197 (1.31)	**3.396 (1.09)**	18.884 (14.89)
5	16 (11)	1.125 (3.91)	**1.183 (3.88)**	13.791 (9.46)	16 (11)	3.255 (2.28)	**3.585 (1.98)**	18.898 (13.9)
6	16 (10)	1.063 (4.95)	**1.114 (4.92)**	14.679 (9.2)	16 (10)	3.364 (3.22)	**3.676 (2.89)**	18.917 (12.92)
7	16 (9)	1.114 (5.95)	**1.163 (5.88)**	14.988 (8.56)	16 (9)	3.49 (4.25)	**3.494 (4)**	18.925 (11.93)
8	16 (8)	1.099 (6.95)	**1.118 (6.91)**	15.607 (8.09)	16 (8)	3.714 (5.16)	**3.523 (4.92)**	18.946 (10.95)

Notes. Average over estimates for IC_{p1}, ER, GR, and Kaiser estimators, with RMSEs in parenthesis for strong and (fully) relevant factors. θ is set to r and all factors are equally relevant (i.e. $\beta = 1$). Panel A displays results for the larger sample $(N,T) = (300, 250)$, while Panel B displays the results for the smaller sample $(N,T) = (100, 20)$.

As in Ahn & Horenstein (2013), we check for the performance of the estimators for different values $kmax$. As already mentioned, the authors find that the estimators described in Bai & Ng (2002) are not robust to the choice of $kmax$. Table 9 shows the average estimates under the exact same specification as in Table 1, except for $kmax$ which is now equal to 16. As expected, the average IC_{p1} estimates show very different results (predicting the level of $kmax$) in smaller samples. The estimates according to the Kaiser rule should not be different (as it does not depend on $kmax$), which indeed does not seem to be the case.[16] The ER and GR estimators perform similar with $kmax = 16$ and $kmax = 8$, which helps confirm the promise of robustness to choices of $kmax$. In the larger sample, the estimators IC_{p1}, ER, and GR perform similar to when we set $kmax$ to 8.

[16]There is a slight difference which is likely due to randomness in the Monte Carlo experiment

Table 10: Average estimates and RMSEs for IC_{p1}, ER, GR, and Kaiser estimators for weak and partially target (ir)relevant factors.

		sPCA: $n = 10$				sPCA: $n = \frac{N}{2}$			
r	r_1	IC_{p1}	ER	GR	Kaiser	IC_{p1}	ER	GR	Kaiser
Panel A: Weak irrelevant factors; $(N,T) = (300, 250)$; $kmax = 8$; $\theta = r$									
2	1	6.59 (4.81)	**1.05 (0.97)**	**1.07 (0.97)**	0.02 (1.98)	**2 (0)**	1 (1)	1 (1)	1.17 (0.91)
3	1	5.93 (3.38)	**1.18 (1.9)**	**1.18 (1.9)**	0 (3)	**2.95 (0.3)**	1 (2)	1 (2)	1 (2)
3	2	6.63 (3.93)	**1.42 (1.68)**	1.59 (1.7)	0.01 (2.99)	**2.91 (0.3)**	1 (2)	1 (2)	2.82 (0.53)
4	1	**5.19 (2.1)**	1.23 (2.84)	1.3 (2.81)	0 (4)	**3.31 (0.99)**	1 (3)	1 (3)	0.95 (3.06)
4	2	**5.86 (2.6)**	1.39 (2.67)	1.56 (2.57)	0 (4)	**2.93 (1.24)**	1 (3)	1 (3)	2.02 (2.12)
4	3	**5.58 (2.28)**	1.71 (2.55)	1.78 (2.51)	0 (4)	**3.27 (0.87)**	1 (3)	1 (3)	5.81 (3.46)
5	1	**4.36 (1.88)**	1.39 (3.71)	1.5 (3.68)	0 (5)	**3.06 (2.22)**	1 (4)	1 (4)	0.86 (4.15)
5	2	**5.29 (1.78)**	1.49 (3.59)	1.64 (3.5)	0 (5)	**2.62 (2.48)**	1 (4)	1 (4)	1.57 (3.53)
5	3	**4.93 (1.69)**	1.4 (3.69)	1.46 (3.66)	0 (5)	2.69 (2.39)	1 (4)	1 (4)	**5 (1.59)**
5	4	**4.39 (1.86)**	1.65 (3.52)	1.75 (3.49)	0 (5)	**2.91 (2.17)**	1 (4)	1 (4)	11.12 (7.8)
6	1	**3.99 (2.62)**	1.35 (4.7)	1.39 (4.67)	0 (6)	**2.79 (3.39)**	1 (5)	1 (5)	0.64 (5.38)
6	2	**3.91 (2.56)**	1.71 (4.44)	1.73 (4.42)	0 (6)	**2.24 (3.84)**	1 (5)	1 (5)	1.06 (4.95)
6	3	**3.79 (2.73)**	1.42 (4.65)	1.55 (4.58)	0 (6)	2.13 (3.94)	1 (5)	1 (5)	**3.86 (2.65)**
6	4	**3.52 (2.96)**	1.53 (4.6)	1.54 (4.6)	0 (6)	**1.92 (4.14)**	1 (5)	1 (5)	9.14 (4.95)
6	5	**3.49 (3.05)**	1.52 (4.57)	1.54 (4.55)	0.03 (5.97)	**1.68 (4.36)**	1 (5)	1 (5)	19.71 (15.35)
Panel B: Weak irrelevant factors; $(N,T) = (100, 20)$; $kmax = 8$; $\theta = r$									
2	1	**1.083 (1.05)**	1.088 (1.07)	**1.093 (1.05)**	3.008 (2.38)	**1.17 (0.99)**	1.038 (1)	1.047 (1)	4.485 (3.94)
3	1	**1.076 (1.98)**	1.1 (1.99)	1.125 (1.99)	2.882 (2.13)	**1.143 (1.92)**	1.044 (1.98)	1.057 (1.97)	4.196 (3.04)
3	2	1.077 (1.97)	1.089 (1.97)	**1.099 (1.96)**	5.729 (4.03)	**1.092 (1.97)**	1.006 (2)	1.013 (1.99)	8.34 (6.57)
4	1	1.09 (2.96)	1.11 (2.95)	**1.123 (2.94)**	**2.842 (2.36)**	1.164 (2.91)	1.059 (2.96)	1.066 (2.95)	**3.79 (2.62)**
4	2	1.085 (2.97)	1.091 (2.96)	**1.107 (2.94)**	5.785 (3.4)	**1.106 (2.93)**	1.011 (2.99)	1.016 (2.99)	7.955 (5.36)
4	3	1.05 (2.97)	1.054 (2.97)	**1.076 (2.95)**	8.188 (5.32)	**1.043 (2.96)**	1.009 (2.99)	1.009 (2.99)	11.402 (8.32)
5	1	1.042 (3.97)	1.097 (3.94)	**1.107 (3.93)**	2.834 (3.07)	1.173 (3.89)	1.075 (3.95)	1.089 (3.94)	**3.612 (2.91)**
5	2	1.09 (3.94)	1.114 (3.93)	1.129 (3.91)	**5.52 (2.98)**	**1.095 (3.93)**	1.032 (3.98)	1.037 (3.98)	7.391 (4.3)
5	3	1.066 (3.96)	1.089 (3.94)	**1.103 (3.93)**	8.273 (4.54)	**1.052 (3.97)**	1.007 (3.99)	1.007 (3.99)	10.965 (7.07)
5	4	**1.096 (3.94)**	1.078 (3.95)	**1.092 (3.94)**	10.202 (6.1)	**1.026 (3.98)**	1.005 (4)	1.006 (4)	13.435 (9.18)
6	1	1.074 (4.94)	1.101 (4.93)	1.121 (4.91)	**2.845 (3.75)**	1.258 (4.83)	1.114 (4.91)	1.127 (4.9)	**3.472 (3.53)**
6	2	1.076 (4.95)	1.096 (4.94)	1.113 (4.92)	**5.668 (2.87)**	1.101 (4.92)	1.041 (4.97)	1.044 (4.97)	**7.152 (3.7)**
6	3	1.093 (4.94)	1.102 (4.92)	1.111 (4.92)	**8.106 (3.83)**	**1.039 (4.97)**	1.009 (4.99)	1.011 (4.99)	10.301 (5.7)
6	4	1.049 (4.96)	1.065 (4.95)	**1.076 (4.94)**	9.996 (5.13)	**1.036 (4.97)**	1.006 (4.99)	1.007 (4.99)	12.928 (7.9)
6	5	1.061 (4.95)	**1.094 (4.94)**	**1.091 (4.94)**	11.394 (6.25)	**1.019 (4.99)**	1 (5)	1.003 (5)	14.877 (9.44)

Notes. Average over estimates for IC_{p1}, ER, GR, and Kaiser estimators, with RMSEs in parenthesis for weak and partially target irrelevant factors. θ is set to r and $kmax = 8$. Panel A displays results for the larger sample $(N,T) = (300, 250)$, while Panel B displays the results for the smaller sample $(N,T) = (100, 20)$. The left column is for $n = 10$ and right column is for $n = \frac{N}{2}$

Table 10 shows results for the same setup but now including irrelevant factors. Again, we leave out Panel B and can be found in the Appendix. As for the strong factor setting, PCA completely ignores the (ir)relevance of a factor which means we do not present it again (Table 3 for PCA results). Instead, we display in the left (right) column the average estimates and RMSE for $n = 10$ ($n = \frac{N}{2}$). We include combinations of $r \in \{2, 3, 4, 5, 6\}$ and $r_1 < r$, and compute the RMSE using r as the true value. Looking at Panel B, we see that the average estimates are close to 1 for most combinations of r and r_1. Only the average Kaiser estimates seem to be sensitive to r_1 (for fixed r), which aligns with the results found in Table 2. The results seem to be similar to the ones for weak but relevant factor sPCA for the smaller sample in Table **??**. For Panel A in Table 10, we also see some resemblance from the target-relevant counterpart. The average Kaiser estimates are pulled to 0 for $n = 10$, and the average ER and GR estimates settle to 1 for $n = \frac{N}{2}$.

Table 11: Median of MSFEs using IC_{p1}, ER, GR, and Kaiser estimators, for sPCA and PCA forecasts with weak and partially (ir)relevant factors.

		sPCA					PCA				
r	r_1	r (r_1)	IC_{p1}	ER	GR	Kaiser	r (r_1)	IC_{p1}	ER	GR	Kaiser
		Panel A: Weak irrelevant factors; $(N, T) = (300, 250)$; $kmax = 8$; $\theta = r$; $n = 10$									
2	1	1.36 (1.36)	**1.35**	**1.35**	**1.35**	1.36	**1.94 (2.05)**	2.05	2	2	3.69
3	1	1.51 (1.5)	**1.46**	1.5	1.5	1.5	**2 (2.06)**	2.06	2.07	2.07	4.08
3	2	**1.96 (2.02)**	1.97	2.14	2.14	2.2	**2.89 (2.94)**	3	2.92	2.92	5.38
4	1	**1.48 (1.56)**	**1.48**	1.56	1.56	1.56	**1.98 (1.98)**	**1.98**	**1.98**	**1.98**	4.33
4	2	2.05 (2.11)	**2.02**	2.2	2.2	2.26	**2.76 (2.84)**	2.88	2.83	2.83	5.75
4	3	2.69 (2.75)	**2.65**	2.93	2.93	3.03	**3.89 (3.98)**	4.05	4.02	4.02	7.17
5	1	**1.5 (1.61)**	1.55	1.61	1.61	1.61	2.06 (2.06)	2.06	**2.05**	**2.05**	4.54
5	2	**2.09 (2.27)**	2.12	2.39	2.39	2.42	**2.91 (2.94)**	3	2.99	2.99	6.19
5	3	**2.69 (2.89)**	**2.69**	3.03	3.03	3.1	3.86 (3.93)	3.97	**3.82**	**3.82**	7.72
5	4	**3.41 (3.47)**	3.47	3.88	3.88	4.01	**5.03 (5.05)**	5.1	5.05	5.05	9.29
6	1	**1.46 (1.61)**	1.53	1.61	1.61	1.61	**1.97 (1.97)**	**1.97**	2	2	4.5
6	2	**2.04 (2.21)**	2.15	2.34	2.34	2.38	2.88 (2.89)	2.89	**2.87**	**2.87**	6.35
6	3	**2.7 (2.9)**	2.84	3.13	3.13	3.17	3.96 (3.94)	3.94	**3.88**	**3.88**	8.05
6	4	**3.33 (3.4)**	3.41	3.82	3.82	3.87	**4.72 (4.69)**	4.77	4.73	4.73	9.3
6	5	**3.98 (4.06)**	4.34	4.47	4.47	4.65	**5.82 (5.72)**	5.98	5.89	5.89	11.71
		Panel B: Weak irrelevant factors; $(N, T) = (100, 20)$; $kmax = 8$; $\theta = r$; $n = 10$									
2	1	**0.65 (0.61)**	**0.61**	**0.61**	**0.61**	0.71	**2.11 (1.98)**	**1.98**	2.28	2.28	160.74
3	1	**0.68 (0.59)**	0.6	**0.59**	**0.59**	0.7	**2.26 (2.05)**	**2.05**	2.3	2.3	144.06
3	2	1.03 (0.95)	**0.88**	**0.88**	**0.88**	1.44	3.34 (3.02)	**2.93**	3.4	3.4	217.9
4	1	**0.78 (0.62)**	**0.62**	**0.62**	**0.62**	0.71	**2.67 (2.07)**	**2.07**	2.38	2.38	183.53
4	2	1.16 (0.95)	**0.88**	**0.88**	**0.88**	1.41	3.58 (3.15)	**2.96**	3.4	3.4	241.3
4	3	1.54 (1.41)	**1.2**	**1.2**	**1.2**	3	5 (4.5)	**4.03**	4.55	4.55	287.37
5	1	**0.88 (0.61)**	**0.61**	**0.61**	**0.61**	0.71	**2.96 (1.95)**	**1.95**	2.29	2.29	147.96
5	2	1.29 (0.99)	**0.9**	**0.9**	**0.9**	1.4	4.34 (3.23)	**3.03**	3.52	3.52	284.22
5	3	1.74 (1.34)	**1.2**	**1.2**	**1.2**	3.18	5.57 (4.58)	**4.01**	4.57	4.57	331.95
5	4	2.02 (1.84)	**1.47**	**1.47**	**1.47**	6.43	6.85 (6.03)	**4.94**	5.74	5.74	387.26
6	1	**1.01 (0.65)**	**0.65**	**0.65**	**0.65**	0.74	**3.48 (2.12)**	**2.12**	2.53	2.53	174.02
6	2	1.44 (1)	0.94	**0.93**	**0.93**	1.37	5.06 (3.31)	**3.07**	3.63	3.63	229.43
6	3	1.96 (1.4)	**1.2**	**1.2**	**1.2**	3.1	6.52 (4.72)	**4.06**	4.8	4.8	318.45
6	4	2.43 (1.97)	**1.5**	**1.5**	**1.5**	6.2	8.01 (6.54)	**5.13**	6.02	6.02	425.45
6	5	2.8 (2.51)	**1.73**	**1.75**	**1.75**	19	9.28 (7.98)	**5.84**	6.54	6.54	437.29

Notes. Median of MSFEs for r, r_1, IC_{p1}, ER, GR, and Kaiser estimators, for rolling window forecasts. Weak and irrelevant factors with θ is set to r and β to one.

Table 11 exhibits the median MSFEs for weak and target irrelevant factors for $n = 10$ and $n = \frac{N}{2}$ (below). We can see that the sPCA forecasts beat the PCA counterparts for every estimator, similar to the weak factor case in Table 6. For the smaller sample, using r_1 (or IC_p1) factors is better than r. In the larger sample, using the true number of factors (r) is average the better choice. When $n = \frac{N}{2}$ we can find similar results. The difference in median MSFE is now more muted compared to when $n = 10$. Under the small sample sPCA, it is now the estimators that are the better choice.

Table 11: *Continued:* Median of MSFEs using IC_{p1}, ER, GR, and Kaiser estimators, for sPCA and PCA forecasts with weak and partially (ir)relevant factors.

		sPCA					PCA				
r	r_1	r (r_1)	IC_{p1}	ER	GR	Kaiser	r (r_1)	IC_{p1}	ER	GR	Kaiser
Panel A: Weak irrelevant factors; $(N, T) = (300, 250)$; $kmax = 8$; $\theta = r$; $n = \frac{N}{2}$											
2	1	**1.01 (1.1)**	**1.01**	1.1	1.1	1.09	**1.05 (1.54)**	**1.05**	**1.05**	**1.05**	2.4
3	1	**1.02 (1.12)**	**1.02**	1.12	1.12	1.12	**1.04 (1.66)**	1.17	1.08	1.08	2.41
3	2	**1.13 (1.21)**	1.15	1.21	1.21	1.17	**1.17 (1.43)**	1.28	1.21	1.21	2.52
4	1	**1.14 (1.24)**	1.16	1.24	1.24	1.24	**1.16 (1.77)**	1.75	1.72	1.72	2.62
4	2	**1.21 (1.31)**	1.27	1.32	1.32	1.29	**1.2 (1.71)**	1.94	1.98	1.98	2.69
4	3	**1.16 (1.22)**	1.21	1.24	1.24	1.17	**1.18 (1.4)**	1.77	1.77	1.77	2.59
5	1	**1.21 (1.29)**	**1.21**	1.29	1.29	1.29	**1.19 (1.84)**	1.84	1.84	1.84	2.65
5	2	**1.32 (1.42)**	1.41	1.45	1.45	1.41	**1.27 (1.95)**	2.14	2.14	2.14	2.81
5	3	**1.34 (1.46)**	1.45	1.46	1.46	1.37	**1.31 (1.67)**	2.24	2.24	2.24	2.96
5	4	**1.19 (1.3)**	1.3	1.37	1.37	1.25	**1.21 (1.37)**	1.87	1.87	1.87	2.83
6	1	**1.26 (1.32)**	1.28	1.32	1.32	1.32	**1.26 (1.87)**	1.87	1.87	1.87	2.67
6	2	**1.43 (1.53)**	1.51	1.52	1.52	1.52	**1.34 (2.15)**	2.32	2.32	2.32	2.92
6	3	**1.54 (1.6)**	1.65	1.66	1.66	1.59	**1.42 (2.07)**	2.57	2.57	2.57	3.02
6	4	**1.48 (1.56)**	1.61	1.6	1.6	1.52	**1.43 (1.8)**	2.45	2.45	2.45	3.24
6	5	**1.33 (1.39)**	1.49	1.5	1.5	1.4	**1.34 (1.46)**	1.99	1.99	1.99	2.96
Panel B: Weak irrelevant factors; $(N, T) = (100, 20)$; $kmax = 8$; $\theta = r$; $n = \frac{N}{2}$											
2	1	**0.66 (0.64)**	**0.64**	**0.64**	**0.64**	0.86	1.43 (1.57)	1.57	**1.52**	**1.52**	101.84
3	1	0.74 (0.66)	**0.65**	**0.65**	**0.65**	0.8	**1.78 (1.74)**	**1.74**	**1.74**	**1.74**	113.32
3	2	0.91 (0.85)	**0.78**	0.79	0.79	2.2	**2.1 (1.96)**	2.02	1.99	1.99	141.05
4	1	0.82 (0.62)	0.62	**0.61**	**0.61**	0.82	**2.09 (1.83)**	**1.83**	1.84	1.84	135.09
4	2	1.07 (0.89)	**0.82**	**0.82**	**0.82**	2.04	2.54 (2.31)	2.31	**2.28**	**2.28**	171.68
4	3	1.2 (1.08)	**0.89**	**0.89**	**0.89**	7.17	2.53 (2.33)	2.27	**2.24**	**2.24**	175.87
5	1	**0.92 (0.68)**	**0.68**	**0.68**	**0.68**	0.84	**2.4 (1.83)**	**1.83**	1.86	1.86	128.7
5	2	1.24 (0.94)	**0.87**	**0.87**	**0.87**	1.92	**3.11 (2.44)**	2.47	2.48	2.48	143.81
5	3	1.38 (1.12)	**0.94**	**0.94**	**0.94**	6.08	3.23 (2.85)	**2.74**	2.76	2.76	175.89
5	4	1.55 (1.36)	**0.97**	0.98	0.98	24.57	3.16 (2.86)	2.57	**2.5**	**2.5**	162.35
6	1	**1.07 (0.67)**	0.68	**0.67**	**0.67**	0.81	**2.85 (1.9)**	**1.9**	1.92	1.92	127.99
6	2	1.42 (0.92)	**0.85**	**0.85**	**0.85**	1.73	3.48 (2.5)	**2.42**	2.5	2.5	180.09
6	3	1.67 (1.16)	1.06	**1.05**	**1.05**	4.96	3.89 (3.07)	3.05	**3.03**	**3.03**	197.75
6	4	1.89 (1.44)	**1.11**	**1.11**	**1.11**	18.06	4.26 (3.47)	**3.12**	3.15	3.15	211.16
6	5	2.01 (1.78)	**1.16**	**1.16**	**1.16**	60.61	4.11 (3.72)	3.05	**3.01**	**3.01**	193.64

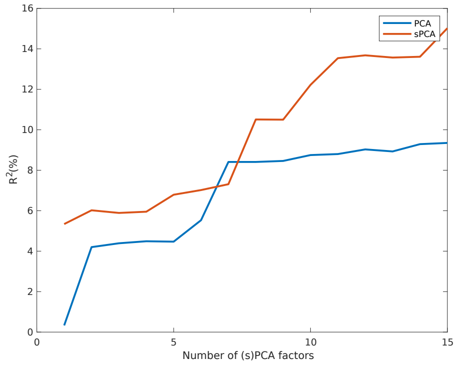
Notes. Median of MSFEs for r, r_1, IC_{p1}, ER, GR, and Kaiser estimators, for rolling window forecasts. Weak and irrelevant factors with θ is set to r and β to one.

C Replication

In this section we demonstrate a partial replication of the results in Huang et al. (2022) using the Matlab code provided by the authors for the empirical part and self-written R code for the simulation study. In particular, we reproduce the both the in-sample and out-of-sample forecasting performance for the PCA and sPCA factors for U.S. inflation and industrial production.

Figure 1: In-sample forecasting performance for PCA and sPCA factors

(a) Panel A: U.S. Inflation



(b) Panel B: Industrial production

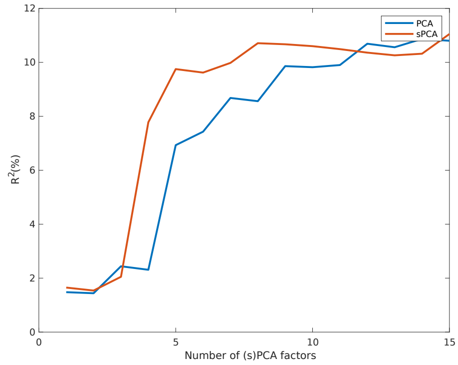
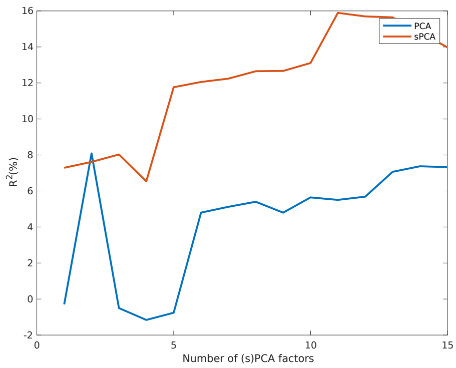


Figure 1 displays the in-sample forecasting performance using the in-sample R^2 (in percentage). In general, these metrics are non-decreasing with the number of (s)PCA factors. We find similar panels in Huang et al. (2022).

Figure 2: Out-of-sample forecasting performance for PCA and sPCA factors

(a) Panel A: U.S. Inflation



(b) Panel B: Industrial production

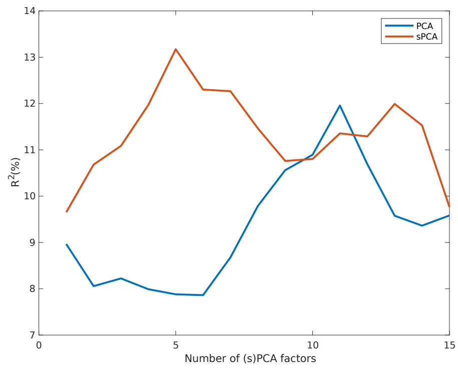


Figure 2 exhibits the out-of-sample forecasting performance using the out-of-sample (OS) R^2. The metric is constructed by comparing the mean square prediction error of the (s)PCA forecasts against AR(1) forecasts. The (s)PCA forecasts are constructed using a factor-augmented prediction model with lagged values of the target. The lags that are included are determined by a model selection criteria (such as the BIC). We omit this and simply use the first lag (essentially combining an AR(1) model with our estimated factors). The resulting graphs are very different to the ones in Huang et al. (2022), except for the first few number of factors. We also show the performance up until 15 factors for consistency with Figure 1.

We also replicate a part of the simulation study. In particular, using only 30 replications, we perform an out-of-sample forecasting exercise for weak factor setting with heterogeneous idiosyncratic errors. The sample size is $(N, T) = (500, 250)$ with expanding window start at 200.

Table 12: The median MSFEs of the sPCA and PCA forecasts with weak factors

	sPCA			PCA		
n	One factor	Two factors	Three factors	One factor	Two factors	Three factors
Panel A: Heterogeneous idiosyncratic error (cross-sectionally); $(N, T) = (500, 250)$						
50	2.028547	2.085022	2.07782	2.033285	2.036143	1.9749
40	2.077837	2.083983	2.06577	2.053383	2.0547	1.966774
30	1.97937	2.020459	2.021038	1.972748	1.928224	1.984553
20	1.866362	1.953081	2.060749	1.840224	2.054716	1.99488
10	2.012405	2.044547	2.142803	1.978784	1.979258	2.042895

In Table 12, we can find the median MSFEs of sPCA and PCA forecasts using one, two, and three factors. We unsuccessfully replicate this part of the study. The exact draws that are used in Huang et al. (2022) probably not correct in our replication. Using time-dependence (unreported) does not change much of the results. The choice that might have mixed up the values is that we perform singular value decomposition and extract the factors and loadings from the eigenvectors, while the authors use the "princomp" package in Matlab.

D Specifications of $g(N, T)$

$$PC_{p1}(k) = V(k, \hat{F}^k) + k\hat{\sigma}^2 \left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right) , \tag{D.1}$$

$$PC_{p2}(k) = V(k, \hat{F}^k) + k\hat{\sigma}^2 \left(\frac{N+T}{NT}\right) \ln C_{NT}^2 , \tag{D.2}$$

$$PC_{p3}(k) = V(k, \hat{F}^k) + k\hat{\sigma}^2 \left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right) , \tag{D.3}$$

$$IC_{p1}(k) = \ln(V(k, \hat{F}^k)) + k \left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right) , \tag{D.4}$$

$$IC_{p2}(k) = \ln(V(k, \hat{F}^k)) + k \left(\frac{N+T}{NT}\right) \ln C_{NT}^2 , \tag{D.5}$$

$$IC_{p3}(k) = \ln(V(k, \hat{F}^k)) + k \left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right) . \tag{D.6}$$

E Proof for intuition of eigenvalue-greater-than-one rule

Let R be the $(N \times N)$ correlation matrix of our predictor data X $(T \times N)$, with diagonal elements ρ_{ii}. We can then compute a matrix V of eigenvectors that diagonalises R in the following way: $V^{-1}RV = D$, where D is the diagonal matrix of eigenvalues (with diagonal elements λ_{ii} being the singular values). The sum of the eigenvalues $\sum_{i=1}^{N} \lambda_{ii}$ is equal to the trace of the correlation matrix R, which is equal to the sum of the diagonal elements of R, ρ_{ii}. As our correlation matrix is standardised (by definition), $\rho_{ii} = 1$ for all values of $i = 1, ..., N$. The result is that $\sum_{i=1}^{N} \lambda_{ii} = \sum_{i=1}^{N} 1 = N$. The mean of λ_{ii} is then equal to 1. The intuition follows from here. As the eigenvalues are on average equal to 1, the Kaiser or eigenvalue-greater-than-one rule chooses the number of factors based the number of eigenvalues that are greater than. This is because the remaining components (with eigenvalues less than one) capture less of the variance than the average component would.