Principal Component Analysis for a Mix of Stationary and Nonstationary Variables*

James D. Hamilton†

Xinwei Ma‡

Jin Xi§

May 28, 2022

Revised: February 8, 2025

Abstract

This paper develops a procedure for uncovering the common cyclical factors that drive a mix of stationary and nonstationary variables. The method does not require knowing which variables are nonstationary or the nature of the nonstationarity. An application to the FRED-MD macroeconomic dataset demonstrates that the approach offers similar benefits to those of traditional principal component analysis with some added advantages.

Keywords: principal components, nonstationary, economic activity indexes

*An earlier version of this paper was circulated under the title, "Principal Component Analysis for Nonstationary Time Series." We thank Serena Ng, Mark Watson, Ken West, and anonymous referees and editors for helpful comments.

†Email: jhamilton@ucsd.edu

‡Email: x1ma@ucsd.edu

§Email: xijin@amss.ac.cn

1 Introduction

Principal component analysis (PCA) has become a key tool for building dynamic models of vector time series with a large cross-sectional dimension. A typical application first subtracts the sample mean from each variable and divides the demeaned variables by their sample standard deviations. PCA finds linear combinations of the standardized variables that have maximum sample variance. These principal components are then used to build dynamic models for each individual series. For surveys of PCA and its usefulness in economics see Bai and Ng (2008) and Stock and Watson (2016).

One difficulty with PCA is that many of the time series encountered in economics and finance are nonstationary. For a nonstationary variable, the population mean is undefined and the sample standard deviation diverges to infinity as the number of time-series observations gets large. Onatski and Wang (2021) detailed some of the problems that can arise from trying to apply PCA to nonstationary data. The typical solution to this problem is for researchers to examine each series individually by hand to determine the transformation of that series that needs to be made before calculating principal components of the set of variables.

This approach has two shortcomings. First, while for some variables it may be fairly clear what transformation is necessary to achieve stationarity, for others it is far from obvious. For example, interest rates have a strong downward trend since 1980. Should yields be treated as stationary? If nonstationary, should we take their first differences or deviations from a time trend before performing PCA? Many finance applications take principal components of interest rates without any transformation; see for example Piazzesi (2010) or Joslin et al. (2011). McCracken and Ng (2016) used either first differences or spreads between interest rates as the first step before including interest rates in PCA. Crump and Gospodinov (2022) recommended using either bond returns or the first differences of bond returns in place of yields themselves to reduce the persistence of these data. Some authors work with first differences of inflation and the unemployment rate while others leave those variables as is. Many decisions like these have to be made before applying PCA to large data sets. In this paper, we propose an automatic procedure that allows the researcher to avoid these judgment calls.

A second fundamental problem is the appropriateness of the methodology itself. Suppose we somehow overcame the first problem and knew for certain the true nature of the trend in each individual series. Suppose for illustration we knew correctly that the first variable y_{1t} is a stationary $AR(1)$ process with autoregressive coefficient $\rho = 0.99$ while the second variable y_{2t} is a random walk. The currently prescribed procedure would instruct the researcher to use the first variable as is and the second variable in the form of first differences. But while the levels of y_{1t} and y_{2t} exhibit very similar properties, the level of y_{1t} and first difference of y_{2t} are radically different. Should we expect that there is some linear combination of y_{1t} and Δy_{2t} that can summarize the common economic drivers behind the two variables? If differencing is the appropriate transformation for a random walk, it seems we should be using some similar transformation for an $AR(1)$ process whose root is close but not quite equal to unity.

This paper proposes an approach to PCA that solves these problems. We first note as in Hamilton (2018) that it is possible to use OLS to estimate an h-period-ahead forecast of the level of any variable as a linear function of its own lags without knowing the nature of the nonstationarity. Moreover, the errors from these linear forecasts are stationary for a broad class of underlying nonstationary processes. Our proposal is to use PCA to identify common factors behind the forecast errors. Specifically, we estimate an OLS regression of y_{it} on $\{1, y_{i,t-h}, y_{i,t-h-1}, \ldots, y_{i,t-h-p+1}\}$, where $i = 1, \ldots, N$ represents the index of the variable, h is the forecasting horizon, and p is the number of lags used for the forecast. We then calculate principal components of the residuals. This approach solves both of the problems identified above. The procedure is fully automatic and treats every variable in the same way. The transformation of y_{it} is a continuous function of the estimated autoregressive coefficients and implies no discontinuity in the way that persistent stationary series are treated as the largest autoregressive root tends toward unity.

Hamilton (2018) suggested that a two-year-ahead forecast error could be interpreted as the stationary cyclical component of the series. Our empirical application follows this suggestion and uses $h = 24$ for monthly data to uncover the common cyclical factors behind a range of economic and financial indicators. We demonstrate that using $h = 24$ also offers a practical solution to the problem of outliers which has posed a severe challenge to using PCA on macroeconomic observations in 2020.

Bai and Ng (2004) and Barigozzi et al. (2021) developed methods that are suitable when the observed data are an unknown mix of $I(0)$ and $I(1)$ variables. Their approaches start by applying

PCA to first-differences of all of the variables whether stationary or not. Bai and Ng (2004) estimated factors for the original data by accumulating factors estimated from the differenced data. Barigozzi et al. (2021) estimated factors by applying the loadings estimated from differenced data to the detrended levels of the original data. Barigozzi et al. (2021) showed that their procedure can estimate the set of factors that account for common stochastic trends in the original data. Bai and Ng (2004) allowed for some of the common factors to be $I(0)$ and others to be $I(1)$, but their equations (1)-(3) do not allow the possibility that a single common factor is influencing the level of a stationary variable y_{1t} and the growth rate of a nonstationary variable y_{2t}. In contrast to these approaches, the goal of our analysis is to uncover the cyclical components that are driving both the stationary and nonstationary variables.

Our results are also useful for estimation of dynamic factor models. One common approach to estimating dynamic factor models is to apply a transformation that is believed to make the individual series stationary, calculate principal components of the transformed data, and then fit a vector autoregression to the estimated factors (Stock and Watson (2002, 2016); Bai and Ng (2002, 2006)). Alternatively, the factors and their dynamics can be estimated jointly by maximum likelihood or quasi-maximum likelihood (Bańbura and Modugno (2014); Doz et al. (2012)). For MLE or QMLE, researchers again typically transform each observed variable y_{it} individually to achieve stationarity (e.g., Forni et al. (2009); D'Agostino et al. (2016)). Antolin-Diaz et al. (2017) allowed the growth rate of a known subset of the variables to follow a random walk as part of the hypothesized state-space framework.[1] In contrast, our approach allows the researcher to implement the initial transformations without the need for subjective judgments about each individual series.

Section 2 discusses the use of forecasting regressions to isolate a stationary component of a possibly nonstationary time series. Section 3 describes how PCA could be used to uncover the common factors behind the forecast errors if we somehow could know the true process followed by each variable, and uses standard results to establish the consistency of PCA in that hypothetical setting. Section 4 analyzes the case when PCA is applied to the residuals from estimated OLS regressions, and establishes consistency of the method in that more realistic setting. Simulations in Section 5 investigate what happens when we apply our procedure to a variety of possible

[1] Stock and Watson (2016) recommended a transformation to render the original raw data stationary that allows for a slowly moving growth rate or trend.

data-generating processes. We find that using two-year-ahead forecast errors works well when we have 50 years of data and that $h = 1$ performs quite well even in much shorter samples. Section 6 illustrates the promise of our approach in an empirical analysis using the FRED-MD large macroeconomics data set.

2 Isolating a stationary component from a nonstationary series

Let y_{it} denote the level of a possibly nonstationary variable at time t. Collect its p most recent values as of date $t-h$ along with a constant term in a vector $z_{i,t-h} = (1, y_{t-h}, y_{t-h-1}, ..., y_{t-h-p+1})'$. The population linear projection of y_{it} on $z_{i,t-h}$ is given by $\mathbb{P}(y_{it}|z_{i,t-h}) = \alpha_{i0}' z_{i,t-h}$ where the coefficient α_{i0} is defined as the vector that minimizes the expected squared error of a forecast of y_{it} based on a linear function of $z_{i,t-h}$:

$$\alpha_{i0} = \arg\min_{\alpha} E(y_{it} - \alpha' z_{i,t-h})^2. \tag{1}$$

For example, if y_{it} is covariance stationary with $E(z_{i,t-h}z_{i,t-h}')$ nonsingular, the population linear projection coefficient is given by

$$\alpha_{i0} = \left[E(z_{i,t-h}z_{i,t-h}')\right]^{-1} E(z_{i,t-h}y_{it}). \tag{2}$$

If y_{it} is ergodic for second moments, the population parameter α_{i0} can be consistently estimated by an OLS regression of y_{it} on $z_{i,t-h}$:

$$\hat{\alpha}_i = \left[\textstyle\sum_{t=1}^T z_{i,t-h}z_{i,t-h}'\right]^{-1} \left[\textstyle\sum_{t=1}^T z_{i,t-h}y_{it}\right] = \left[T^{-1}\textstyle\sum_{t=1}^T z_{i,t-h}z_{i,t-h}'\right]^{-1} \left[T^{-1}\textstyle\sum_{t=1}^T z_{i,t-h}y_{it}\right] \xrightarrow{p} \alpha_{i0}.$$

Note that there is no assumption in the definition (1) nor in its stationary solution (2) that the process is linear or that it can be characterized by an ARMA representation of known order. There may be better forecasts of y_{it} that could be obtained using a nonlinear function or more lags like $y_{i,t-h-p}, y_{i,t-h-p-1}, ...$ But there is an optimal forecast within the class of linear forecasts that use only p lags. For a covariance-stationary process, the optimal forecast within that class is characterized by (2).

A solution to (1) also exists if y_{it} is nonstationary but its first difference Δy_{it} is covariance stationary. This can be seen from the accounting identity that holds for any time series:

$$y_{it} = y_{i,t-h} + \Delta y_{i,t-h+1} + \Delta y_{i,t-h+2} + \cdots + \Delta y_{it}. \tag{3}$$

For an $I(1)$ process,[2]

$$\mathbb{P}(y_{it}|z_{i,t-h}) = y_{i,t-h} + \mathbb{P}(\Delta y_{i,t-h+1}|z^{\dagger}_{i,t-h}) + \mathbb{P}(\Delta y_{i,t-h+2}|z^{\dagger}_{i,t-h}) + \cdots + \mathbb{P}(\Delta y_{it}|z^{\dagger}_{i,t-h}) \tag{4}$$

for $z_{i,t-h} = (1, y_{t-h}, y_{t-h-1}, ..., y_{t-h-p+1})'$, $z^{\dagger}_{t-h} = (1, \Delta y_{t-h}, \Delta y_{t-h-1}, ..., \Delta y_{t-h-p+2})'$ and $\mathbb{P}(\Delta y_{i,t-s}|z^{\dagger}_{i,t-h}) = \left[E(\Delta y_{i,t-s} z^{\dagger\prime}_{i,t-h})\right] \left[E(z^{\dagger}_{i,t-h} z^{\dagger\prime}_{i,t-h})\right]^{-1} z^{\dagger}_{i,t-h}$. Moreover, the population linear projection coefficient α_{i0} can again be consistently estimated by an OLS regression of the level of y_{it} on $(1, y_{i,t-h}, y_{i,t-h-1}, ..., y_{i,t-h-p+1})'$. The intuition for this is that an OLS levels regression minimizes the sample analog to the population minimization in (1):

$$\hat{\alpha}_i = \arg\min_{\alpha} T^{-1} \textstyle\sum_{t=1}^{T} (y_{it} - \alpha' z_{i,t-h})^2. \tag{5}$$

If the OLS estimate $\hat{\alpha}_i$ picks up the unit root, the value of (5) converges to a finite number as T gets large. If it does not, the value of (5) would diverge to infinity. Thus in large samples, OLS estimation is dominated by the incentive inherent in the objective function of OLS to remove any nonstationary features of the data. More formally, the fitted values from the regression (5) are numerically identical to the fitted values from an OLS regression of y_{it} on $(y_{i,t-h}, z^{\dagger'}_{i,t-h})$. In the latter regression, the OLS coefficient on the lagged level $y_{i,t-h}$ will converge in probability to one whenever the dependent variable is an $I(1)$ process. In fact, the OLS estimate of this parameter is superconsistent, converging at rate T rather than at rate $\sqrt{T}$, and the asymptotic distribution of the other coefficients is identical to what would result if we regressed Δy_{it} on $z^{\dagger}_{i,t-h}$; see Hamilton (2018).

Hamilton (2018) showed that these results generalize to any nonstationary process that is co-

[2]Note that the space spanned by $z_{i,t-h}$ is the same as the space spanned by $(z^{\dagger}_{i,t-h}, y_{i,t-h})$ which is bigger than the space spanned by $z^{\dagger}_{i,t-h}$. However, the coefficient on $y_{i,t-h}$ in the projection $\mathbb{P}(\Delta y_{it}|z^{\dagger}_{i,t-h}, y_{i,t-h})$ is zero, because multiplying the $I(1)$ variable $y_{i,t-h}$ by any nonzero value would produce a forecast of y_{it} that has infinite MSE. For this reason $\mathbb{P}(\Delta y_{it}|z_{i,t-h}) = \mathbb{P}(\Delta y_{it}|z^{\dagger}_{i,t-h})$.

variance stationary around a deterministic polynomial in time t provided that the order of the polynomial is no greater than the number of lags p, or if the process is $I(d)$ with the dth difference of y_{it} covariance stationary for some d no greater than p. One does not need to know the order of the polynomial or the value of d in order to consistently estimate the population linear projection coefficient α_{i0} using a regression of the level of y_{it} on a constant and its p most recent levels as of date $t-h$.

Hamilton (2018) also showed that the residual from the population linear projection, $c_{it} = y_{it} - \alpha_{i0}' z_{i,t-h}$, is covariance stationary for any of the above nonstationary processes. For example, for an $I(1)$ process, we see from (3) and (4) that

$$\begin{aligned} c_{it} &= y_{it} - \mathbb{P}(y_{it}|z_{i,t-h}) \\ &= \left[\Delta y_{i,t-h+1} - \mathbb{P}(\Delta y_{i,t-h+1}|z_{i,t-h}^{\dagger})\right] + \left[\Delta y_{i,t-h+2} - \mathbb{P}(\Delta y_{i,t-h+2}|z_{i,t-h}^{\dagger})\right] + \\ &\quad \cdots + \left[\Delta y_{it} - \mathbb{P}(\Delta y_{it}|z_{i,t-h}^{\dagger})\right] \end{aligned}$$

is covariance stationary for any finite h. It is thus possible to isolate a stationary component of a large range of nonstationary processes from the residuals of an OLS regression of y_{it} on $(1, y_{i,t-h}, y_{i,t-h-1}, ..., y_{i,t-h-p+1})'$ without knowing whether the variable y_{it} is stationary or nonstationary.[3]

It is instructive to compare this approach with other common ways of thinking about the trend in a nonstationary series. If the ith observed variable is a deterministic function of time plus a zero-mean stationary process, $y_{it} = \delta_i^{\text{det}}(t) + c_{it}^{\text{det}}$, we could describe the deterministic time trend as the

[3] Specifically, if either: (i) y_{it} is stationary around a deterministic polynomial function of time of order $d_i \leq p$ satisfying

$$T^{-1/2} \sum_{s=1}^{[Tr]} (y_{it} - \delta_{i0} - \delta_{i1} t - \delta_{i2} t^2 - \cdots - \delta_i t^{d_i}) \Longrightarrow \omega_i W_i(r)$$

where $[Tr]$ denotes the largest integer no greater than Tr, $W_i(r)$ denotes standard Brownian motion, and $\Rightarrow$ denotes weak convergence of associated probability measures; or alternatively if (ii) d_i differences of y_{it} are stationary for some $d_i \leq p$ satisfying

$$T^{-1/2} \sum_{s=1}^{[Tr]} (\Delta^{d_i} y_{it} - \mu_i) \Longrightarrow \omega_i W_i(r);$$

then Hamilton (2018) showed that c_{it} is stationary and that the estimated coefficient $\hat{\alpha}_i$ from an OLS regression gives a consistent estimate of the population parameter α_{i0}.

limit of a forecast that would have been made in the arbitrarily distant past:

$$\delta_i^{\det}(t) = \lim_{h\to\infty}\lim_{p\to\infty} E(y_{it}|y_{i,t-h}, y_{i,t-h-1}, ..., y_{i,t-h-p+1}).$$

By contrast, if the first difference of the ith variable is a zero-mean stationary process, Beveridge and Nelson (1981) suggested that we think of the trend as the forecast of the variable in the arbitrarily distant future. They decomposed $y_{it} = \delta_{it}^{BN} + c_{it}^{BN}$ where

$$\delta_{it}^{BN} = \lim_{h\to\infty}\lim_{p\to\infty} E(y_{i,t+h}|y_{it}, y_{i,t-1}, ..., y_{i,t-p+1}).$$

While these concepts of trend have some appeal, they have the significant practical drawback that both are based on the properties of forecasts at infinite horizons. They thus depend on conjectures of what happens at infinity, conjectures that are impossible to verify on the basis of a finite sample of observations.

By contrast, forecasts at a two-year horizon are something that can be reasonably investigated without auxiliary assumptions in samples of typical size. If we choose h to correspond to a two-year horizon, Hamilton (2018) argued that the decomposition

$$y_{it} = \mathbb{P}(y_{it}|1, y_{i,t-h}, y_{i,t-h-1}, ..., y_{i,t-h-p+1}) + c_{it} \tag{6}$$

can be viewed as a way to implement traditional approaches to trend-cycle decomposition that is practical and robust. Moreover, the primary reason we would go wrong in making a two-year-ahead forecast of most economic time series is due to unforeseen cyclical changes. For example, the value of y_{it} will be significantly below $\mathbb{P}(y_{it}|z_{i,t-h})$ if the economy goes into a recession after $t-h$ and significantly above the forecast if recovery from a downturn is more robust than expected. For this reason, Hamilton (2018) proposed to use h corresponding to a two-year horizon as a way to isolate the stationary cyclical component of the ith variable, for example, $h = 24$ for monthly data.

Another important benefit of setting $h = 24$ is that it offers a solution to the huge practical problem of dealing with outliers. To see why this is the case, suppose for illustration that the ith variable is characterized by a random walk: $y_{it} = y_{i,t-1} + \varepsilon_{it}$. For a monthly random walk, the residual from the two-year population linear projection is given by $c_{it} = \sum_{s=0}^{23} \varepsilon_{i,t-s}$. From the Central Limit Theorem, c_{it} exhibits much less kurtosis than the one-period-ahead forecast error ε_{it}. We find in our empirical application in Section 6 that this feature is extremely helpful in dealing with outliers such as those seen during the COVID-19 pandemic in 2020.

Notwithstanding, these benefits do not come without a cost. The larger the value of h, the more serial correlation there will be in c_{it}, and the more prone PCA can be in small samples to the spurious factor problem identified by Onatski and Wang (2021). Our simulations in Section 5 suggest that if we have 50 years of data, using two-year-ahead forecast errors works pretty well. For shorter samples, the spurious factor problem can be more serious, and our recommendation for those applications is to look for common factors in the one-year-ahead ($h = 12$) or one-month-ahead ($h = 1$) forecasts. In the latter case, additional correction for outliers is likely needed.

To summarize, our procedure is to estimate the same regression for every variable y_{it}, regardless of whether we think it is stationary and without making any conjecture about the nature of any nonstationarity. To allow for persistent seasonal components in y_{it}, we recommend choosing p to be the number of observations in a year. Our procedure estimates the following regression by OLS for every variable,

$$y_{it} = k_i + \alpha_{i1} y_{i,t-h} + \alpha_{i2} y_{i,t-h-1} + \cdots + \alpha_{ip} y_{i,t-h-p+1} + c_{it}, \tag{7}$$

with $h = 8$ and $p = 4$ for quarterly data and $h = 24$ and $p = 12$ for monthly data. We will refer to the residual from the estimated regression $\hat{c}_{it}$ as the OLS residual for variable y_{it} and the residual from the population linear projection c_{it} as the true cyclical component. The value $\hat{c}_{it}$ is a consistent estimate of c_{it}, and the true value c_{it} is stationary as long as any nonstationarity in y_{it} is characterized by either a polynomial time trend of order d_i or an $I(d_i)$ process with $d_i \leq p$. Our procedure is to perform PCA on the regression residuals $\{\hat{c}_{1t}, ..., \hat{c}_{Nt}\}$.

One practical decision is whether a nonlinear transformation of the raw data is necessary for $\Delta^{d_i} y_{it}$ to be stationary for some d_i. If taking first differences of the log is the correct way to produce a stationary series, then taking the change in the level would not produce a stationary series. We recommend using logs for variables like output or prices which are usually described in terms of growth rates. For such variables we use the log of the level of the variable, $y_{it} = \log Y_{it}$ as the variable in the regression (7). For variables like interest rates or the unemployment rate that are already quoted in percentage terms, we use the raw data $y_{it} = Y_{it}$ in the regression.

The true cyclical component c_{it} has mean zero and is stationary for a wide range of processes. However, the population value of α_{i0} is not known but must be estimated by regression. In Section 3 we characterize our assumptions about the factor structure that we hypothesize describes the true values of c_{it}, and use standard results to establish that these population factors could be consistently estimated if the true values of c_{it} were observed without error. Section 4 considers the case when we do not know the value of d_i for each series, do not know whether it is stationary or characterized by a deterministic time trend or an $I(d_i)$ process, and the c_{it} are not observed. In that section we analyze the consequences of performing PCA on the estimated OLS residuals $\hat{c}_{it}$.

3 Principal component analysis when the true cyclical component is observed

In the previous section we defined the true cyclical component c_{it} to be the residual from a population linear projection of y_{it} on $(1, y_{i,t-h}, y_{i,t-h-1}, ..., y_{i,t-h-p+1})'$, and noted that c_{it} is stationary for a broad class of possible processes. In this section we provide sufficient conditions under which the true cyclical components for a collection of N different variables would have a factor structure that could be consistently estimated using PCA if we observed the true value of c_{it} for each variable. The set-up and results in this section closely follow Stock and Watson (2002).

3.1 Assumed factor structure of the true cyclical components

Collect the true cyclical components for the N different series at time t in an $(N \times 1)$ vector $C_t = (c_{1t}, ..., c_{Nt})'$. We postulate that these are characterized by a factor structure of the form

$$\underset{(N\times 1)}{C_t} = \underset{(N\times r)}{\Lambda} \underset{(r\times 1)}{F_t} + \underset{(N\times 1)}{e_t}. \tag{8}$$

The number of latent factors r is much less than the number of variables N, but the r factors are assumed to account for most of the variance of C_t in a sense made formal below. Since the factors are unobserved, $C_t = \Lambda H^{-1} H F_t + e_t$ would imply the identical observable model as (8). Thus some normalizations are necessary in order to talk about consistently estimating the jth factor f_{jt}. In empirical estimation, practitioners typically resolve this ambiguity by estimating the jth column of Λ by the eigenvector associated with the jth largest eigenvalue of $T^{-1}\sum_{t=1}^{T} C_t C_t'$. Note that such a procedure implies a normalization in which the columns of Λ are orthogonal to each other and the elements of F_t are uncorrelated with each other and ordered by the size of their variance. We follow Stock and Watson (2002) in how to characterize these conventions as the cross-section dimension N and time-series dimension T get large.[4]

Assumption 1 (factor structure)**.**

(i) $(\Lambda'\Lambda/N) \rightarrow I_r$.

(ii) $E[F_t F_t'] = \Omega_{FF}$*, where* Ω_{FF} *is a diagonal matrix with* $\omega_{ii} > \omega_{jj} > 0$ *for* $i < j$.

(iii) $|\lambda_{ij}| \leq \bar{\lambda} < \infty$.

(iv) $T^{-1}\sum_t F_t F_t' \xrightarrow{p} \Omega_{FF}$.

In addition to implementing the property that eigenvectors of a symmetric matrix are orthogonal, Assumption 1(*i*) requires that each factor makes a nonnegligible contribution to the average variance of c_{it} across i. That is, if we were to imagine adding more variables (increasing N) with $\lambda_{ij} = 0$ for all i greater than some fixed N_0, then Assumption 1(*i*) could not hold. Likewise 1(*ii*) and 1(*iv*) require that each factor continues to matter as the number of time-series observations T grows. These conditions are consistent with serial dependence of the factors, but rely on the fact that C_t is stationary.

Let γ denote an $(N \times 1)$ vector and $\Gamma = \{\gamma : \gamma'\gamma/N = 1\}$. Note that if γ were the jth column of Λ, the scalar $\gamma'\Lambda F_t/N$ would converge to f_{jt} and $(N^2T)^{-1}\sum_{t=1}^{T}\gamma'\Lambda F_t F_t'\Lambda'\gamma \xrightarrow{p} \omega_{jj}$. The assumption that the idiosyncratic elements e_t do not have a factor structure requires that there is no value of γ for which the analogous operation applied to e_t would lead to anything other than zero: $\sup_{\gamma\in\Gamma}(N^2T)^{-1}\sum_{t=1}^{T}\gamma' e_t e_t'\gamma \xrightarrow{p} 0$. Stock and Watson (2002) used the following assumptions to guarantee the absence of a factor structure in e_t.

Assumption 2 (moments of the errors)**.**

[4]See Bai and Ng (2013) and Stock and Watson (2016) for discussion of alternative normalizations.

(i) $\lim_{N\to\infty} \sup_t \sum_{s=-\infty}^{\infty} |E[e_t' e_{t+s}/N]| < \infty.$

(ii) $\lim_{N\to\infty} \sup_t N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} |E[e_{it}e_{jt}]| < \infty$, *where* e_{it} *denotes the ith element of* e_t.

(iii) $\lim_{N\to\infty} \sup_{t,s} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} |cov[e_{is}e_{it}, e_{js}e_{jt}]| < \infty.$

Some might be concerned that we have simply postulated that the true cyclical components are characterized by Assumptions 1 and 2. But something very similar is done in traditional applications that assume conditions like these characterize specified stationary transformations of the original data. Indeed, insofar as the cyclical components have a common primitive definition in terms of h-period-ahead forecast errors, we would argue that these assumptions are easier to defend in our application than in many others.

3.2 Consequences of applying PCA to the true cyclical components

Recall that the $(N \times 1)$ vector of true cyclical components C_t is stationary and has population mean zero. If C_t was observed directly, its estimated sample variance matrix would be $S = T^{-1} \sum_{t=1}^{T} C_t C_t'$ and a linear combination $\gamma' C_t$ for any $(N \times 1)$ vector γ would have sample variance $\gamma' S \gamma$. If C_t was observed, the first estimated principal component (denoted $\tilde{f}_{1t} = N^{-1}\tilde{\lambda}_1' C_t$) would be defined as the linear combination that has maximum sample variance subject to a normalization condition such as $\gamma \in \Gamma = \{\gamma : \gamma'\gamma/N = 1\}$:

$$\tilde{\lambda}_1 = \arg \sup_{\gamma\in\Gamma} \tilde{R}(\gamma) \tag{9}$$

$$\tilde{R}(\gamma) = (N^2 T)^{-1} \gamma' \sum_{t=1}^{T} C_t C_t' \gamma. \tag{10}$$

Note we are normalizing $\tilde{\lambda}_1' \tilde{\lambda}_1 / N = 1$ as we did asymptotically for the columns of Λ in Assumption 1(i). We also divide the sample variance of $\gamma' C_t$ by N^2 in anticipation of the result that $\tilde{R}(\tilde{\lambda}_1)$ will converge to a fixed constant as N and T grow. The solution to (9) is obtained by setting $\tilde{\lambda}_1$ proportional to the eigenvector of $S = T^{-1} \sum_{t=1}^{T} C_t C_t'$ associated with the largest eigenvalue. For example, if we calculated eigenvectors of this matrix using code that normalizes eigenvectors to have unit length and orders eigenvalues by decreasing size, $\tilde{\lambda}_1$ would be $\sqrt{N}$ times the first eigenvector. The largest eigenvalue of S is equal to $T^{-1} \sum_{t=1}^{T} \tilde{f}_{1t}^2$, the sample variance of the first principal component. The jth principal component $N^{-1}\tilde{\lambda}_j' C_t$ is found by maximizing $\tilde{R}(\gamma)$ subject to the constraint that γ is orthogonal to $\tilde{\lambda}_1, ..., \tilde{\lambda}_{j-1}$. The solution for $\tilde{\lambda}_j$ is proportional to the eigenvector of S associated with the jth largest eigenvalue.

Alternatively, if we observed the true factors F_t and loadings Λ, we could calculate the component of the variance of $\gamma' C_t$ that is attributable to the r factors alone:

$$R^*(\gamma) = (N^2T)^{-1}\gamma' \textstyle\sum_{t=1}^{T} \Lambda F_t F_t' \Lambda' \gamma. \tag{11}$$

Stock and Watson (2002) showed that under Assumptions 1 and 2, the maximum value for (10) (which is given by the largest eigenvalue of S) and the supremum of (11) over all $\gamma \in \Gamma$ converge in probability to the same number ω_{11}, which is the population variance of the first factor, and that $\tilde{\lambda}_1' C_t / N$ gives a consistent estimate of f_{1t} up to a sign. If we were to estimate $k > r$ principal components, the first r would consistently estimate f_{jt} up to a sign normalization and the last $k - r$ would asymptotically have zero variance. We restate their results in the following theorem.

Theorem 1. *(Stock and Watson, 2002). Suppose that Assumptions 1 and 2 hold. Let $R^*(\gamma)$ be the function in (11) and let $\tilde{f}_{1t}, ..., \tilde{f}_{kt}$ denote the first k estimated principal components of C_t ($\tilde{f}_{jt} = \tilde{\lambda}_j' C_t / N$) with $k \geq r$. Let $\underset{(r\times 1)}{\tilde{F}_t} = (\tilde{f}_{1t}, ..., \tilde{f}_{rt})'$ and $\underset{(N\times r)}{\tilde{\Lambda}} = \begin{bmatrix} \tilde{\lambda}_1 & \cdots & \tilde{\lambda}_r \end{bmatrix}$. Then as N and T go to infinity:*

(i) $\sup_{\gamma\in\Gamma} R^*(\gamma) \xrightarrow{p} \omega_{11}$;

(ii) If $\lambda_1^* = \arg\sup_{\gamma\in\Gamma} R^*(\gamma)$ *and* $\lambda_j^* = \arg \sup_{\gamma\in\Gamma, \gamma'\lambda_1^* = \cdots = \gamma'\lambda_{j-1}^* = 0} R^*(\gamma)$, *then* $R^*(\lambda_j^*) \xrightarrow{p} \omega_{jj}$ *for* $j = 1, ..., r$;

(iii) $T^{-1} \sum_{t=1}^{T} \tilde{f}_{jt}^2 = \tilde{R}(\tilde{\lambda}_j) \xrightarrow{p} \omega_{jj}$ *for* $j = 1, ..., r$;

(iv) $T^{-1} \sum_{t=1}^{T} \tilde{f}_{jt}^2 \xrightarrow{p} 0$ *for* $j = r+1, ..., k$;

(v) $\tilde{\Xi}\tilde{\Lambda}'\Lambda / N \xrightarrow{p} I_r$ *where* $\tilde{\Xi}$ *is a diagonal matrix whose row j column j element is* $+1$ *if* $\tilde{\lambda}_j'\lambda_j > 0$ *and* -1 *if* $\tilde{\lambda}_j'\lambda_j < 0$.

(vi) $\tilde{\Xi}\tilde{F}_t - F_t \xrightarrow{p} 0$.

4 Principal component analysis when the cyclical component must be estimated

In this section we assume that we do not observe the true cyclical component c_{it} of series i but have an estimate $\hat{c}_{it} = c_{it} + \hat{v}_{it}$. Let $\hat{C}_t = (\hat{c}_{1t}, ..., \hat{c}_{Nt})'$ and $\hat{V}_t = (\hat{v}_{1t}, ..., \hat{v}_{Nt})'$. We investigate the properties of principal components calculated from the estimated cyclical components:

$$\hat{f}_{jt} = N^{-1}\hat{\lambda}_j'\hat{C}_t \tag{12}$$

$$\hat{\lambda}_j = \arg \sup_{\{\gamma\in\Gamma, \gamma'\hat{\lambda}_1=\cdots=\gamma'\hat{\lambda}_{j-1}=0\}} \hat{R}(\gamma)$$

$$\hat{R}(\gamma) = (N^2T)^{-1}\gamma' \textstyle\sum_{t=1}^T \hat{C}_t\hat{C}_t'\gamma.$$

We first note high-level sufficient conditions under which PCA applied to the estimated cyclical components $\hat{C}_t$ gives consistent estimates of the true factors F_t. Let $\hat{v}_{it} = \hat{c}_{it} - c_{it}$ denote the difference between the estimated and true cyclical component of series i at date t. The conditions require N to grow with T so as to ensure that $N^{-1}\sum_{i=1}^N \hat{v}_{it}^2$ and $T^{-1}\sum_{t=1}^T \hat{v}_{it}^2$ are $o_p(1)$.

Assumption 3 (high-level conditions on $\hat{v}_{it}$)**.** *The cross-section dimension N grows with the number of time series observations T according to a function $N(T)$ such that for all $\delta, \varepsilon > 0$ there exists a $T_0(\delta,\varepsilon)$ such that for all $T > T_0(\delta,\varepsilon)$:*

$$Prob\left\{\frac{1}{N(T)}\sum_{i=1}^{N(T)} \hat{v}_{it}^2 > \delta\right\} < \varepsilon \tag{i}$$

$$Prob\left\{\max_{1\le i\le N} \frac{1}{T}\sum_{t=1}^{T} \hat{v}_{it}^2 > \delta\right\} < \varepsilon. \tag{ii}$$

The following result establishes that if the error in estimating the cyclical component satisfies Assumption 3, then the results $(\hat{f}_{jt}, \hat{\lambda}_j, \hat{R}(\gamma))$ of applying PCA to the estimated cyclical components $\hat{C}_t$ give consistent estimates of the magnitudes that characterize the true cyclical components C_t.

Theorem 2. *Suppose that C_t and e_t in equation (8) satisfy Assumptions 1 and 2. Let $\hat{C}_t = C_t + \hat{V}_t$ for $\hat{V}_t = (\hat{v}_{1t}, ..., \hat{v}_{Nt})'$ where $\hat{v}_{it}$ satisfy Assumption 3. Let $\hat{f}_{1t}, ..., \hat{f}_{kt}$ denote the first k estimated principal* *components of $\hat{C}_t$ ($\hat{f}_{jt} = \hat{\lambda}_j'\hat{C}_t/N$) with $k \geq r$ and let $\underset{(r\times 1)}{\hat{F}_t} = (\hat{f}_{1t}, ..., \hat{f}_{rt})'$ and $\underset{(N\times r)}{\hat{\Lambda}} = \begin{bmatrix} \hat{\lambda}_1 & \cdots & \hat{\lambda}_r \end{bmatrix}$. Then*

(i) $T^{-1}\sum_{t=1}^{T}\hat{f}_{jt}^2 = \hat{R}(\hat{\lambda}_j) \xrightarrow{p} \omega_{jj}$ *for* $j = 1, ..., r$;

(ii) $T^{-1}\sum_{t=1}^{T}\hat{f}_{jt}^2 \xrightarrow{p} 0$ *for* $j = r+1, ..., k$;

(iii) $\hat{\Xi}\hat{\Lambda}'\Lambda/N \xrightarrow{p} I_r$ *where $\hat{\Xi}$ is a diagonal matrix whose row j column j element is $+1$ if $\hat{\lambda}_j'\lambda_j > 0$ and -1 if $\hat{\lambda}_j'\lambda_j < 0$;*

(iv) $\hat{\Xi}\hat{F}_t - F_t \xrightarrow{p} 0$.

Before presenting formal sufficient conditions for verifying Assumption 3, we first discuss the intuition for why we might expect it to hold. For $z_{i,t-h} = (1, y_{i,t-h}, y_{i,t-h-1}, ..., y_{i,t-h-p+1})'$, the true cyclical component c_{it} is the residual from a population linear projection of y_{it} on $z_{i,t-h}$ and $\hat{c}_{it}$ is the residual from the corresponding estimated regression:

$$c_{it} = y_{it} - \alpha_{i0}'z_{i,t-h}$$

$$\hat{c}_{it} = y_{it} - \hat{\alpha}_i'z_{i,t-h}$$

$$\hat{v}_{it} = \hat{c}_{it} - c_{it} = (\alpha_{i0} - \hat{\alpha}_i)'z_{i,t-h}$$

$$(\alpha_{i0} - \hat{\alpha}_i) = -\left(\sum_{t=1}^{T} z_{i,t-h}z_{i,t-h}'\right)^{-1}\left(\sum_{t=1}^{T} z_{i,t-h}c_{it}\right) \tag{13}$$

$$\hat{v}_{it}^2 = (\alpha_{i0} - \hat{\alpha}_i)'z_{i,t-h}z_{i,t-h}'(\alpha_{i0} - \hat{\alpha}_i) \tag{14}$$

$$\sum_{t=1}^{T}\hat{v}_{it}^2 = (\alpha_{i0} - \hat{\alpha}_i)'\left(\sum_{t=1}^{T} z_{i,t-h}z_{i,t-h}'\right)(\alpha_{i0} - \hat{\alpha}_i). \tag{15}$$

Expression (15) will be recognized as the OLS Wald statistic for testing the true null hypothesis $H_0 : \alpha_i = \alpha_{i0}$ multiplied by $\hat{\sigma}_i^2 = (T-k)^{-1}\sum_{t=1}^{T}\hat{c}_{it}^2$, the average squared regression residual. As we now sketch, this statistic would be expected to be $O_p(1)$ for a wide class of stationary and nonstationary processes. This would lead us to expect that the individual terms $\hat{v}_{it}^2$ should be $o_p(1)$, with some uniformity conditions across i then guaranteeing Assumption 3(i). Likewise if we divide (15) by T we should again get an $o_p(1)$ random variable, as required by Assumption 3(ii).

We now explore the intuition for why we would typically expect (15) to be $O_p(1)$. Note that the expression can be written

$$\begin{aligned}
\textstyle\sum_{t=1}^T \hat{v}_{it}^2 &= \left(\textstyle\sum_{t=1}^T c_{it} z_{i,t-h}'\right)\left(\textstyle\sum_{t=1}^T z_{i,t-h} z_{i,t-h}'\right)^{-1}\left(\textstyle\sum_{t=1}^T z_{i,t-h} c_{it}\right) \\
&= \left(\textstyle\sum_{t=1}^T c_{it} z_{i,t-h}' \Upsilon_{iT}^{-1}\right)\left(\Upsilon_{iT}^{-1}\textstyle\sum_{t=1}^T z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1}\right)^{-1}\left(\Upsilon_{iT}^{-1}\textstyle\sum_{t=1}^T z_{i,t-h} c_{it}\right) \qquad (16)
\end{aligned}$$

where Υ_{iT}^{-1} could be any sequence of nonsingular matrices. Note that the researcher does not need to know the value of Υ_{iT}. This matrix is just a device to calculate the asymptotic properties of the left side of (16) (which does not depend on Υ_{iT}) under various possible forms of nonstationarity for z_{it}.

If y_{it} were stationary, we would use $\Upsilon_{iT}^{-1} = T^{-1/2} I_{p+1}$ to find the asymptotic distribution of $\sum_{t=1}^T \hat{v}_{it}^2$. In this case, $z_{i,t-h} c_{it}$ is a stationary random variable which by the definition of the true cyclical component c_{it} has population mean zero. In this case, a central limit theorem could be used to establish that $\Upsilon_{iT}^{-1} \sum_{t=1}^T z_{i,t-h} c_{it} = T^{-1/2} \sum_{t=1}^T z_{i,t-h} c_{it}$ converges to a Normal distribution. Likewise in the stationary case,

$$\Upsilon_{iT}^{-1} \textstyle\sum_{t=1}^T z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1} = T^{-1} \textstyle\sum_{t=1}^T z_{i,t-h} z_{i,t-h}' \xrightarrow{p} E(z_{i,t-h} z_{i,t-h}').$$

For this reason, $\sum_{t=1}^T \hat{v}_{it}^2$ would typically be $O_p(1)$ when y_{it} is stationary.

Consider next the case when $d_i = 1$ and Δy_{it} (denoted below by u_{it}) is a zero-mean stationary process. Taking a regression with $p = 3$ lags for illustration, an OLS regression of y_{it} on $z_{i,t-h} = (1, y_{i,t-h}, y_{i,t-h-1}, y_{i,t-h-2})'$ has the identical fitted values as an OLS regression of y_{it} on $R_i^{-1} z_{i,t-h} = (u_{i,t-h}, u_{i,t-h-1}, 1, y_{i,t-h})'$. In this case to calculate the asymptotic distribution of (16) we would take

$$\Upsilon_{iT}^{-1} = \begin{bmatrix} T^{-1/2} I_3 & 0 \\ 0 & T^{-1} \end{bmatrix} R_i^{-1}$$

$$\Upsilon_{iT}^{-1} \sum_{t=1}^{T} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1}$$

$$= \begin{bmatrix} T^{-1}\sum_{t=1}^{T} u_{i,t-h}^2 & T^{-1}\sum_{t=1}^{T} u_{i,t-h}u_{i,t-h-1} & T^{-1}\sum_{t=1}^{T} u_{i,t-h} & T^{-3/2}\sum_{t=1}^{T} u_{i,t-h}y_{i,t-h} \\ T^{-1}\sum_{t=1}^{T} u_{i,t-h-1}u_{i,t-h} & T^{-1}\sum_{t=1}^{T} u_{i,t-h-1}^2 & T^{-1}\sum_{t=1}^{T} u_{i,t-h-1} & T^{-3/2}\sum_{t=1}^{T} u_{i,t-h-1}y_{i,t-h} \\ T^{-1}\sum_{t=1}^{T} u_{i,t-h} & T^{-1}\sum_{t=1}^{T} u_{i,t-h-1} & 1 & T^{-3/2}\sum_{t=1}^{T} y_{i,t-h} \\ T^{-3/2}\sum_{t=1}^{T} y_{i,t-h}u_{i,t-h} & T^{-3/2}\sum_{t=1}^{T} y_{i,t-h}u_{i,t-h-1} & T^{-3/2}\sum_{t=1}^{T} y_{i,t-h} & T^{-2}\sum_{t=1}^{T} y_{i,t-h}^2 \end{bmatrix}.$$

Under standard unit-root asymptotics (e.g., Hamilton (1994, p. 506)), we expect $\Upsilon_{iT}^{-1} \sum_{t=1}^{T} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1} \xrightarrow{d} Q_i$ with

$$Q_i = \begin{bmatrix} \gamma_{i0} & \gamma_{i1} & 0 & 0 \\ \gamma_{i1} & \gamma_{i0} & 0 & 0 \\ 0 & 0 & 1 & \omega_i \int_0^1 W_i(r)dr \\ 0 & 0 & \omega_i \int_0^1 W_i(r)dr & \omega_i^2 \int_0^1 \left[W_i(r)\right]^2 dr \end{bmatrix} \tag{17}$$

for $\gamma_{ij} = E(u_{it}u_{i,t-j})$, $\omega_i^2 = \sum_{j=-\infty}^{\infty} \gamma_{ij}$, and $W_i(r)$ standard Brownian motion. Similar tools can be used to establish that $\Upsilon_{iT}^{-1} \sum_{t=1}^{T} z_{i,t-h} c_{it} \xrightarrow{d} q_i$ and $\sum_{t=1}^{T} \hat{v}_{it}^2 \xrightarrow{d} q_i' Q_i^{-1} q_i \sim O_p(1)$.

If instead $E(\Delta y_{it}) = \mu_i \neq 0$, the level y_{it} would be dominated asymptotically by a deterministic time trend $\mu_i t$.[5] In this case (and for general $p \geq 2$) we would use $R_i^{-1} z_{i,t-h} = (\Delta y_{i,t-h} - \mu_i, \Delta y_{i,t-h-1} - \mu_i, ..., \Delta y_{i,t-h+2} - \mu_i, 1, y_{i,t-h})'$ and

$$\Upsilon_{iT}^{-1} = \begin{bmatrix} T^{-1/2} I_p & 0 \\ 0 & T^{-3/2} \end{bmatrix} R_i^{-1}$$

to establish that $\sum_{t=1}^{T} \hat{v}_{it}^2 \sim O_p(1)$. General cases for $d_i \leq p$ are examined in Hamilton (2018). Note that we do not need to know the value of d_i or Υ_{iT} to estimate any magnitudes – in every case we are talking about a regression of y_{it} on the lagged levels $z_{i,t-h} = (1, y_{i,t-h}, y_{i,t-h-1}, ..., y_{i,t-h-p+1})'$. Instead, Υ_{iT} is just a device to analyze the properties if the levels regression is applied to a variety of different stationary and nonstationary processes.

We are now in a position to state formally sufficient conditions under which Assumption 3

[5] That is, for any nonnegative integer ν,

$$T^{-(\nu+1)} \sum_{t=1}^{T} y_{it}^{\nu} - T^{-(\nu+1)} \sum_{t=1}^{T} (\mu_i t)^{\nu} \xrightarrow{p} 0$$

as $T \to \infty$ and $T^{-(\nu+1)} \sum_{t=1}^{T} (\mu_i t)^{\nu} \to \mu_i^{\nu}/(\nu+1)$.

would be satisfied, for which we let $\|X\|$ denote the Frobenius/Euclidean norm (the square root of the sums of squares of the elements of a vector or matrix X).

Assumption 4 (sufficient conditions for Assumption 3)**.** *For each i, there exist some nonsingular matrix Υ_{iT} and a possibly random matrix Q_i, such that*

(i) For some constant $\mathfrak{c}_1$,

$$\max_{1\leq i\leq N}\max_{1\leq t\leq T} E(\|\sqrt{T}\Upsilon_{iT}^{-1}z_{i,t-h}\|^2)\leq \mathfrak{c}_1 \quad \textit{and} \quad \max_{1\leq i\leq N} E(\|\sum_{t=1}^{T}\Upsilon_{iT}^{-1}z_{i,t-h}c_{it}\|^2)\leq \mathfrak{c}_1.$$

(ii) For some constants $\mathfrak{c}_2$, $\mathfrak{c}_3$, $\mathfrak{c}_4$, and $\mathfrak{c}_5>0$,

$$\max_{1\leq i\leq N} Prob\Big(\varrho_{\min}(Q_i)<\delta\Big)\leq \mathfrak{c}_2\exp\Big\{-\mathfrak{c}_3\Big(\frac{1}{\delta}\Big)^{1/\mathfrak{c}_4}\Big\} \quad \textit{for all } \delta\leq \mathfrak{c}_5,$$

where $\varrho_{\min}$ denotes the minimum eigenvalue. In addition, $N\log^{2\mathfrak{c}_4}(N)/T\to 0$.

(iii) For some sequence $\mathfrak{c}_6(T,N)=o(\log^{-\mathfrak{c}_4}(N))$,

$$Prob\Big(\max_{1\leq i\leq N}\Big\|\sum_{t=1}^{T}\Upsilon_{iT}^{-1}z_{i,t-h}z_{i,t-h}'\Upsilon_{iT}^{-1}-Q_i\Big\|>\mathfrak{c}_6(T,N)\Big)\to 0.$$

Condition 4(i) is quite mild, as it only requires finite moments on the transformed regressors and the cyclical components after suitable normalization. For example, when Δy_{it} is a zero-mean stationary process, the first inequality requires Δy_{it} and $y_{it}/\sqrt{T}$ to have finite variance. Conditions 4(ii–iii) are key to our theoretical developments. Specifically, they posit that the matrix that gets inverted in (16) can be uniformly approximated by a possibly random limit Q_i, and that the minimum eigenvalue of the limiting matrix exhibits an exponential-type tail at zero.

Theorem 3. *Assumption 4 implies Assumption 3.*

Next we relate Assumption 4 to conditions for general linear processes.

Assumption 5 (linear process)**.** *There exist positive constants $\mathfrak{c}_7$–$\mathfrak{c}_{12}$, such that for each $i\in\{1,2,\ldots,N\}$ there exist a μ_i and an integer d_i for which $u_{it}=(1-L)^{d_i}y_{it}-\mu_i=\sum_{l=0}^{\infty}\psi_{il}\eta_{i,t-l}$, and*

(i) $\max_{1\leq i\leq N}|\psi_{il}|\leq \mathfrak{c}_7 l^{-\mathfrak{c}_8}$ with $\mathfrak{c}_8>2$.

(ii) $\min_{1\le i\le N} |\sum_{l=0}^{\infty} \psi_{il}| \ge \mathfrak{c}_9 > 0$.

(iii) η_{it} *is an iid process with zero mean, unit variance, and* $\max_{1\le i\le N} E(\eta_{it}^4) \le \mathfrak{c}_{10}$.

(iv) For each i, either $\mu_i = 0$ *or* $|\mu_i| \ge \mathfrak{c}_{11} > 0$.

(v) Let $\tilde{Q}_i$ *be the covariance matrix of* $(u_{it}, u_{i,t-1}, \ldots, u_{i,t-p+d_i+1})$. *Then* $\varrho_{\min}(\tilde{Q}_i) \ge \mathfrak{c}_{12} > 0$.

Conditions 5(i) and 5(ii) would be satisfied if u_{it} is any zero-mean stationary $AR(g)$ process for some g,

$$u_{it} = \phi_{i1}u_{i,t-1} + \cdots + \phi_{ig}u_{i,t-g} + \eta_{it} \quad \text{with } \|z\| > 1 \text{ for any } z \text{ satisfying } \left|1 - \phi_{i1}z - \cdots - \phi_{ig}z^g\right| = 0.$$

Alternatively, 5(i) and 5(ii) would also be satisfied for any $MA(g)$ process that has a representation that is not arbitrarily close to being noninvertible,

$$u_{it} = \eta_{it} + \theta_{i1}\eta_{i,t-1} + \cdots + \theta_{ig}\eta_{i,t-g} \quad \text{with } \left|1 + \theta_{i1} + \cdots + \theta_{ig}\right| > \mathfrak{c}_9.$$

Condition 5 (ii) defines the necessary degree of differencing d_i. For example, if $y_{it} = \psi_i(L)\eta_{it}$ with $\psi_{i\ell}$ satisfying $\psi_i(1) > \mathfrak{c}_9$, then the $MA(\infty)$ coefficients $\tilde{\psi}_{i\ell}$ of the first difference $(1-L)y_{it} = (1-L)\psi_i(L)\eta_{it} = \tilde{\psi}_i(L)\eta_{it}$ would not satisfy (ii) because $\tilde{\psi}_i(1) = (1-1)\psi_i(1) = 0$.

When $\mu_i \neq 0$, y_{it} is dominated asymptotically by $(\mu_i t)^{d_i}$ whenever $d_i > 0$ as in footnote 5. Condition 5(iv) allows us to use this fact in establishing a rate of convergence when $\mu_i \neq 0$.

Theorem 4. *Suppose that* $p \ge 1$ *and Assumption 5 is satisfied for every* $i \in \{1, 2, \ldots, N\}$ *for some* $d_i \in \{0, 1\}$, *and that the true cyclical components* c_{it} *satisfy Assumptions 1 and 2. If* $N \log^{12}(N)/T \to 0$, *then results (i)–(iv) of Theorem 2 hold.*

While our proof of Theorem 4 only covers the case when $d_i \le 1$, we conjecture that a similar result could be obtained for any $d_i \le p$. Extending the proof to this case requires verifying the three conditions in Assumption 4. Assumption 4(i) requires only finite moments of the regressors and the cyclical components after suitable transformation and rescaling by the appropriate Υ_{iT}. This is true for processes satisfying Assumption 5 with $d_i = 2$, although the calculations demonstrating this would be more tedious than the $d_i = 1$ case we review. It is also possible to establish 4(iii) for $d_i \geq 2$. The key step in our proof (result (A-21) in Appendix A) builds on the coupling result of Komlós et al. (1976), which provides a uniform (in both $1 \leq i \leq N$ and $1 \leq t \leq T$) strong approximation of a random walk by sums of independent Gaussian variables. This result can also be applied to provide a uniform (in $1 \leq i \leq N$) approximation of the Q_{iT} matrix using n-fold integrals of Brownian motion.

The main technical challenge in extending Theorem 4 to the case when $d_i \geq 2$ is verifying Assumption 4(ii). When $d_i \geq 2$ and $\mu_i = 0$,the limiting random matrix Q_i involves n-fold integrals of Brownian motion. Based on the literature of small ball probabilities for n-fold integrals of Brownian motion (Chen and Li, 2003; Gao et al., 2003), we conjecture that 4(ii) holds with $\mathfrak{c}_4 = 2\max_{1\leq i\leq N} d_i$, though we were unable to locate a concrete demonstration of this in the literature. Thus, although our current Theorem 4 applies only to stationary and $I(1)$ processes, we expect that it extends to series with any order of integration, with the only change that the exponent on the log of N may differ.

Finally, we note that no additional proof is necessary to extend Theorem 4 to the case when $2 \leq d_i \leq p$ if the number of series for which $d_i \geq 2$ does not increase with N or T.

5 Results from simulations

In this section we report results from applying our method in a variety of different settings. In these simulations we take the number of cross-section variables to be $N = 100$ and vary the number of time-series observations T from 100 to 1000. For each generated sample, we calculate the $(N \times N)$ correlation matrices of: (1) the raw data; (2) the OLS regression residuals for that data set; and (3) the true cyclical components for that data set implied by the particular process that was used to generate that sample. For each of these correlation matrices, let $\hat{\xi}_j$ denote the jth largest eigenvalue of the correlation matrix. The fraction of the variance of the sample explained by the jth principal component is

$$R_j^2 = \hat{\xi}_j / N. \tag{18}$$

We also calculated the number of factors that would be selected for that sample based on the IC_{p2} criterion of Bai and Ng (2002) recommended by Stock and Watson (2016, p. 436),

$$r^* = \arg\left\{\min_{r\in\{0,1,\ldots,r_0\}} \log\left(1 - \frac{\sum_{j=0}^{r} \hat{\xi}_j}{N}\right) + r\left(\frac{N+T}{NT}\right)\min\{N,T\}\right\}, \tag{19}$$

with $\hat{\xi}_0$ defined to be 0. We took $r_0 = 10$ and for each of the different cases generated 100 different samples.

5.1 Mix of unrelated stationary and nonstationary variables

For our first example, half the variables are random walks and the other half are white noise,

$$y_{it} = \begin{cases} y_{it-1} + \varepsilon_{it} = \varepsilon_{it} + \varepsilon_{i,t-1} + \cdots + \varepsilon_{i1} & \text{for } i = 1,2,...,N/2 \\ \varepsilon_{it} & \text{for } i = (N/2)+1,...,N \end{cases}$$

for $t = 1, ..., T$. The innovations $\varepsilon_{it} \sim N(0,1)$ are independent across all i and t. Thus each of the N variables is completely independent of the others and there is no factor structure in the true data-generating process.

The first two columns of Table 1 report the results from calculating principal components of the raw data. In a sample of $T = 100$ observations, the first three principal components seem to account for 38% of the variance of the full set of $N = 100$ variables. This result is entirely spurious, and illustrates the cautions raised by Onatski and Wang (2021) about using PCA when some of the variables are nonstationary. The criterion (19) would always lead us incorrectly to conclude that there is more than one factor in a sample of size $T = 100$. This problem in the apparent number of factors gets even worse when the sample size increases. The latter is the expected result, since (19) is in the class of criteria for which Onatski and Wang (2021, p. 602) demonstrated that the number of factors selected diverges as N and T go to infinity.

The next two columns of Table 1 report what the results would be if we somehow knew the true cyclical component of each variable. For this case, we applied PCA to a sample of $T - h$

Table 1: Mixture of independent random walks and white noise

	Raw data		c_t (h=24)		$\hat{c}_t$ (h=24)		$\hat{c}_t$ (h=12)		$\hat{c}_t$ (h=8)		$\hat{c}_t$ (h=1)	
j	R^2	r^*	R^2	r^*	R^2	r^*	R^2	r^*	R^2	r^*	R^2	r^*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$T = 100$												
0	—	0	—	0	—	1	—	13	—	96	—	100
1	22.6	0	16.0	0	11.9	26	10.3	48	8.4	4	4.0	0
2	9.9	67	12.0	6	9.4	54	8.5	35	7.1	0	3.8	0
3	5.6	33	8.3	94	7.3	19	6.9	4	6.1	0	3.6	0
$T = 200$												
0	—	0	—	0	—	0	—	78	—	100	—	100
1	22.7	0	9.4	0	10.2	0	6.6	22	5.2	0	2.9	0
2	9.4	13	7.8	3	8.2	8	5.7	0	4.6	0	2.7	0
3	5.3	87	6.6	97	6.5	92	4.9	0	4.2	0	2.6	0
$T = 400$												
0	—	0	—	23	—	3	—	100	—	100	—	100
1	22.5	0	5.9	51	6.5	34	4.3	0	3.5	0	2.2	0
2	9.4	0	5.2	22	5.5	47	3.8	0	3.2	0	2.1	0
3	5.1	100	4.5	4	4.7	16	3.5	0	3.0	0	2.1	0
$T = 600$												
0	—	0	—	99	—	79	—	100	—	100	—	100
1	22.4	0	4.6	1	5.0	21	3.4	0	2.9	0	1.9	0
2	9.3	0	4.1	0	4.4	0	3.1	0	2.7	0	1.9	0
3	5.2	100	3.7	0	3.9	0	2.9	0	2.5	0	1.8	0
$T = 800$												
0	—	0	—	100	—	100	—	100	—	100	—	100
1	22.7	0	4.0	0	4.2	0	3.0	0	2.6	0	1.8	0
2	9.3	0	3.6	0	3.7	0	2.7	0	2.4	0	1.7	0
3	5.0	100	3.3	0	3.4	0	2.6	0	2.3	0	1.7	0
$T = 1000$												
0	—	0	—	100	—	100	—	100	—	100	—	100
1	22.6	0	3.5	0	3.7	0	2.7	0	2.4	0	1.7	0
2	9.3	1	3.2	0	3.3	0	2.5	0	2.2	0	1.7	0
3	5.1	99	3.0	0	3.0	0	2.4	0	2.1	0	1.6	0

Notes to Table 1. R^2 indicates the percentage of total variance accounted for by the jth principal component for $j = 1, 2$ or 3. r^* indicates the percentage of samples for which the criterion (19) selects the number of factors to be $j = 0, 1, 2,$ or ≥ 3. In every case, the true number of factors is $r = 0$ and the cross-section dimension is $N = 100$.

observations for which the ith observed variable for $t = h+1, h+2, ..., T$ is given by

$$c_{it} = \begin{cases} y_{it} - y_{i,t-h} = \varepsilon_{it} + \varepsilon_{i,t-1} + \cdots + \varepsilon_{i,t-h+1} & \text{for } i = 1, 2, ..., N/2 \\ \varepsilon_{it} & \text{for } i = (N/2)+1, ..., N \end{cases}. \tag{20}$$

Note that the cyclical components in (20) can be serially correlated, but this autocorrelation vanishes for observations separated by more than h periods. The cyclical components c_{it} thus satisfy by construction the conditions under which Bai and Ng (2002) demonstrated that (19) would asymptotically select the correct number of factors. We find in our simulations that (19) does indeed correctly conclude there is no factor structure for these data sets provided the number of time-series observations is 600 or larger. For smaller T it is less reliable. The reason is that there is a small-sample version of the Onatski and Wang (2021) spurious factor problem that arises from the serial correlation of some of the variables that is induced by the definition of the true cyclical component in equation (20). If T is large enough, this problem goes away, but for smaller T it can make a difference.

Columns (5) and (6) examine the case where the analysis is based on the residuals from running an OLS regression on the raw data Y_{it} for all variables $i = 1, ..., N$ without making any judgments about which variables are stationary and which are not. For large samples, the results are similar to those that we would obtain if we somehow knew the exact correct transformation to use for every variable.

The small-sample problem in columns (3)-(6) results from the serial correlation that is a consequence defining the cyclical component to be the error from a 24-period-ahead forecast. Columns (7) and (8) report results if we instead were to look for common factors in the 12-period-ahead forecast errors. This typically would reach the correct conclusion even in a sample of only $T = 200$ observations. Columns (9) and (10) consider 8-period-ahead forecast errors, such as our suggested cyclical calculation would use for quarterly data. The results indicate that if we have more than 50 years of data ($T = 600$ for monthly data or $T = 200$ for quarterly data), conducting PCA on the two-year-ahead OLS forecast residuals ($h = 24$ for monthly data or $h = 8$ for quarterly data) is reasonably reliable. With less than 50 years of data, some researchers might want to use a smaller value for h or place less reliance on (19) as a criterion for selecting the number of factors.

The last two columns of Table 1 examine looking for common factors in the one-period-ahead forecast errors. In these simulations, this reaches the correct conclusion 100% of the time that there are zero factors in these data sets even for a sample of $T = 100$ observations. Thus our proposed method appears to be quite reliable if the interest is in identifying common factors behind one-period-ahead forecast errors. However, one-period-ahead forecast errors are more sensitive to outliers. This is an important consideration, as will be demonstrated in our analysis of actual data in Section 6.

5.2 Mix of unrelated stationary variables with differing persistence

In our second example, for $i = 1, 2, ..., N/2$ the variables are generated by a stationary but persistent AR(1) process:

$$y_{i1} \sim N(0, 1/(1-\rho^2))$$

$$y_{it} = \rho y_{i,t-1} + \varepsilon_{it} \quad \text{for } t = 2, 3, ..., T.$$

The remaining $N/2$ variables are white noise ($y_{it} = \varepsilon_{it}$ for $i = (N/2) + 1, ..., N$). Our example uses $\rho = 0.99$, so all the variables are stationary but half of them are highly persistent. The innovations $\varepsilon_{it} \sim N(0, 1)$ are independent across all i and t, so there is no factor structure in the true data-generating processes.

Columns (1) and (2) of Table 2 report the results from applying PCA to the raw data. Note that for this example, the raw data themselves satisfy the Bai and Ng (2002) conditions for asymptotic validity of PCA. Nevertheless, even in a sample of size $T = 1000$, the first principal component alone appears to explain a third of the data, and the criterion in (19) would always conclude incorrectly that there is at least one factor. This is a small-sample manifestation of the Onatski and Wang (2021) spurious factor phenomenon. In a sufficiently large sample, this problem would go away. But $T = 1000$ is not large enough for persistence characterized by $\rho = 0.99$.

In columns (3)-(4) we apply PCA to the residuals from a 24-period-ahead forecasting regression. Again we estimated the same regression for all variables, whether persistent or not. And again using regression residuals solves the problem pretty reliably in samples larger than $T = 600$. If we look for a factor structure in the one-period-ahead regression residuals as in columns (5)-(6), the problem is solved 100% of the time even in a sample of $T = 100$.

Table 2: Monte Carlo results in other settings

	$\rho = 0.99$						Cointegrated				Mixed			
	Raw data		$\hat{c}_t$ (h=24)		$\hat{c}_t$ (h=1)		Raw data		$\hat{c}_t$ (h=24)		Raw data		$\hat{c}_t$ (h=24)	
j	R^2	r^*	R^2	r^*	R^2	r^*	R^2	r^*	R^2	r^*	R^2	r^*	R^2	r^*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
$T = 100$														
0	—	0	—	0	—	100	—	0	—	0	—	0	—	0
1	43.3	98	16.0	1	4.1	0	46.1	100	40.3	100	27.5	0	23.8	0
2	3.6	2	10.6	99	3.8	0	2.8	0	3.4	0	18.1	0	11.2	1
3	2.8	0	5.5	0	3.6	0	2.5	0	3.1	0	8.5	100	8.7	99
$T = 200$														
0	—	0	—	0	—	100	—	0	—	0	—	0	—	0
1	42.5	82	10.9	0	2.9	0	48.1	100	45.7	100	25.1	0	22.6	0
2	3.3	18	8.7	2	2.7	0	2.1	0	2.3	0	18.0	0	9.8	0
3	2.3	0	6.9	98	2.6	0	2.0	0	2.2	0	8.7	100	7.9	100
$T = 400$														
0	—	0	—	7	—	100	—	0	—	0	—	0	—	0
1	40.5	54	6.3	26	2.2	0	48.8	100	46.5	100	23.8	0	21.2	0
2	3.4	46	5.5	54	2.1	0	1.8	0	1.8	0	17.9	0	6.4	0
3	2.2	0	4.8	13	2.1	0	1.7	0	1.7	0	8.8	100	5.4	100
$T = 600$														
0	—	0	—	96	—	100	—	0	—	0	—	0	—	0
1	38.5	40	4.8	4	1.9	0	49.3	100	47.0	100	23.3	0	20.8	0
2	3.5	57	4.3	0	1.9	0	1.6	0	1.6	0	17.9	0	4.9	4
3	2.3	3	3.9	0	1.8	0	1.5	0	1.6	0	8.9	100	4.3	96
$T = 800$														
0	—	0	—	100	—	100	—	0	—	0	—	0	—	0
1	36.9	32	4.1	0	1.8	0	49.4	100	47.0	100	23.2	0	20.5	33
2	3.4	60	3.7	0	1.7	0	1.5	0	1.5	0	17.7	0	4.1	43
3	2.4	8	3.3	0	1.7	0	1.5	0	1.5	0	9.2	100	3.7	24
$T = 1000$														
0	—	0	—	100	—	100	—	0	—	0	—	0	—	0
1	35.4	33	3.6	0	1.7	0	49.5	100	47.1	100	22.8	0	20.4	86
2	3.5	60	3.3	0	1.7	0	1.5	0	1.5	0	17.6	0	3.7	14
3	2.5	7	3.0	0	1.6	0	1.4	0	1.4	0	9.2	100	3.3	0

Notes to Table 2. R^2 indicates the percentage of total variance accounted for by the jth principal component for $j = 1, 2$ or 3. r^* indicates the percentage of samples for which the criterion (19) selects the number of factors to be $j = 0, 1, 2$, or ≥ 3. In columns (1)-(6), the true number of factors is $r = 0$. In columns (7)-(14), the true number of factors is $r = 1$. In every case, the cross-section dimension is $N = 100$.

5.3 Cointegration

In our next example, the nonstationary variables are cointegrated and there is no factor structure for the stationary variables. The single common factor for the nonstationary variables follows a random walk:

$$F_t = F_{t-1} + v_t \quad t = 1, ..., T; \; F_0 = 0$$

$$y_{it} = F_t + \varepsilon_{it} \quad i = 1, 2, ..., N/2$$

$$y_{it} = \varepsilon_{it} \quad i = (N/2) + 1, ..., N$$

with $v_t \sim N(0,1)$ and $\varepsilon_{it} \sim N(0,1)$ independent for all i and t. Notice that the first $(N/2)$ variables are characterized by

$$y_{it} = F_{t-h} + v_{t-h+1} + v_{t-h+2} + \cdots + v_t + \varepsilon_{it}$$

$$\mathbb{P}(y_{it}|1, y_{i,t-h}, y_{i,t-h-1}, ..., y_{i,t-h-p+1}) \simeq F_{t-h}$$

$$c_{it} \simeq v_{t-h+1} + v_{t-h+2} + \cdots + v_t + \varepsilon_{it}.$$

Thus there is a single factor (namely $v_{t-h+1} + v_{t-h+2} + \cdots + v_t$) that is common to the cyclical component of the first $(N/2)$ variables,[6] and the true number of common factors in the sample of N variables is $r = 1$.

Columns (7) and (8) of Table 2 report the results from applying PCA to the raw data. Note that even though half the variables are nonstationary, PCA always correctly concludes that there is a single common factor in these data. This is an illustration of the well-known result that PCA on levels data can correctly identify cointegrating relations; for more discussion see Harris (1997) and Onatski and Wang (2018).

Columns (9) and (10) report results from applying PCA to the residuals from 24-period-ahead forecasting regressions. Again the same regression is estimated in the same way for stationary and nonstationary observations. And again PCA on the regression residuals results in the correct answer 100% of the time, even if the sample size is as small as $T = 100$.

[6] Another way to express this is that the variables are cointegrated with $(N/2) - 1$ linearly independent cointegrating relations given by $y_{it} - y_{1t} \sim I(0)$ for $i = 2, 3, ..., (N/2)$.

5.4 Stationary factor with a mix of stationary and nonstationary indicators

We next consider data-generating processes in which the common factor is stationary: $F_t = \rho_F F_{t-1} + v_t$ with $\rho_F = 0.8$. The ith observed variable y_{it} is related to F_t with a weight ω_i, $y_{it} = \omega_i F_t + g_{it}$, where the idiosyncratic components g_{it} are a mix of stationary and nonstationary processes: $g_{it} = \rho_i g_{i,t-1} + e_{it}$ where v_t and e_{it} $(i = 1, ..., N)$ are mutually independent $N(0, 1)$. Columns (11)-(14) of Table 2 report results for the following example:

$$\begin{array}{ll} \rho_i = 1, \omega_i = 1 & \text{for } i = 1, ..., N/4 \\ \rho_i = 0.5, \omega_i = 1 & \text{for } i = 1 + N/4, ..., N/2 \\ \rho_i = 1, \omega_i = 0 & \text{for } i = 1 + N/2, ..., 3N/4 \\ \rho_i = 0.5, \omega_i = 0 & \text{for } i = 1 + 3N/4, ..., N \end{array}.$$

Thus for this example half of the N observed variables are nonstationary. The variables are independent of each other apart from their potential common dependence on the $r = 1$-dimensional factor F_t, and this single factor affects some of the observed variables but not others. When PCA is applied directly to the observed data y_{it}, the familiar Onatski and Wang (2021) result is observed in column (11) of Table 2: higher-order factors spuriously appear to explain a large amount of the variance of the data. When PCA is applied instead to the $h = 24$-period-ahead regression forecast residuals $\hat{c}_{it}$, the contribution of higher-order factors $\hat{F}_{jt}$ for $j \geq 1$ is substantially lower, though the selection criterion (19) would still typically incorrectly conclude that $r > 2$ for $T \leq 600$. The correct conclusion $(r = 1)$ would be reached 86% of the time when $T = 1000$. Moreover, we found that the average correlation between the true realization of F_t for a particular simulation and the first estimated factor $\hat{F}_{1t}$ of the OLS regression residuals from that simulation is 0.98 for $T \geq 600$. In other words, the first principal component of the OLS regression residuals accurately uncovers the true single common feature of these data.

Overall, these results confirm the asymptotic theory that applying PCA to OLS regression residuals is a promising approach to handling both nonstationarity and stationary persistence of unknown form provided that the time-series dimension T is reasonably large.

6 Characterizing a large macroeconomic data set

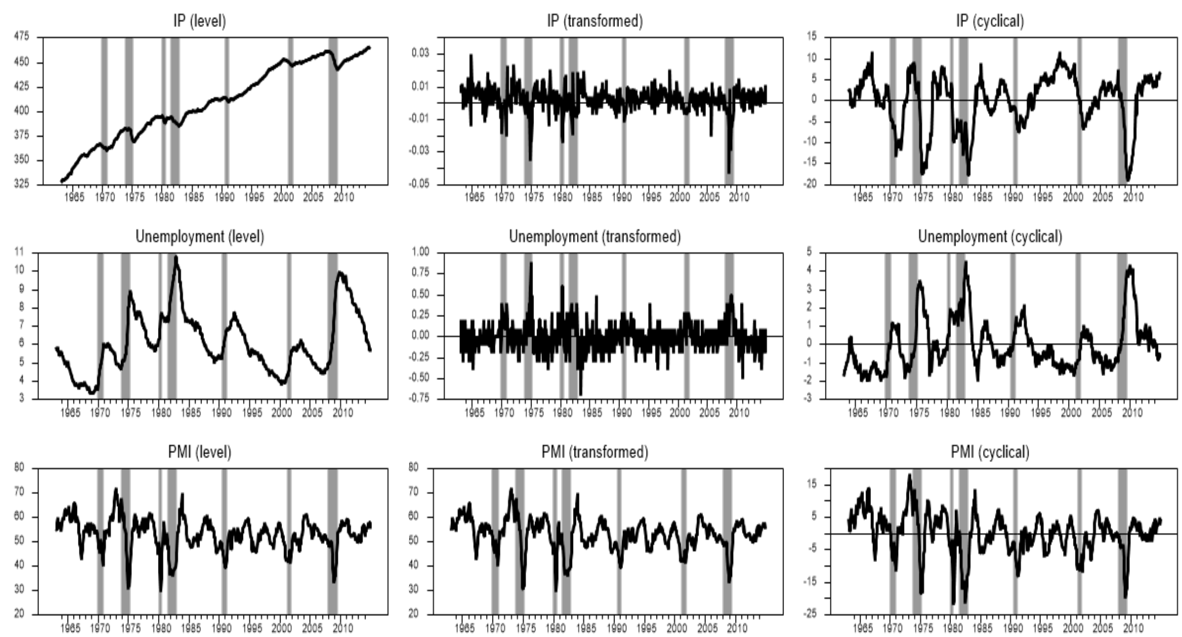
The use of large macroeconomic data sets was pioneered by Stock and Watson (1999), whose goal was to use the information of 168 different macroeconomic variables to produce better forecasts of inflation. They found that the first principal component of macroeconomic variables that are related to the level of real economic activity produced the best inflation forecasts over the period 1959:1 to 1997:9. Their findings led to the development of the Chicago Fed National Activity Index (CFNAI), which is the first principal component of a subset of 85 different measures of economic activity.[7]

McCracken and Ng (2016) reviewed subsequent uses of large macroeconomic sets and developed the FRED-MD database whose 2015:4 vintage covered 134 macroeconomic variables. These variables include monthly measures divided into 8 broad categories: (1) output and income; (2) labor market; (3) housing; (4) consumption, orders, and inventories; (5) money and credit; (6) interest and exchange rates; (7) prices; and (8) stock market. This data set offers benefits of continuity and continuous updating and is the basis for our analysis in this paper.

In previous applications of PCA to large macroeconomic data sets, each of the variables needed to be transformed using a detrending method that was selected individually for each series. For details of how this has been done for the CFNAI see Federal Reserve Bank of Chicago (2021) and for FRED-MD see the data appendix to McCracken and Ng (2016). Figure 1 illustrates these transformations for three important macroeconomic indicators. The first column plots the raw data, while the second column plots the data as transformed by McCracken and Ng (2016) in order to ensure stationarity, using the same data set as in their original paper. Everyone agrees that industrial production (row 1) is nonstationary, and all previous researchers have used first differences of the log of industrial production shown in panel (1,2). While there is little doubt that this is a good way to generate a stationary series for this variable, monthly growth rates of industrial production exhibit a lot of high-frequency fluctuations around the dominant cyclical patterns. For the unemployment rate (row 2), it is less clear whether the series should be regarded as stationary. McCracken and Ng (2016) used first differences of unemployment, which behave

[7]The variables used to calculate the Chicago Fed National Activity Index (CFNAI) fall into four broad groups: (1) production and income; (2) employment, unemployment, and hours; (3) personal consumption and housing; and (4) sales, orders, and inventories

Figure 1: Level, transformed value, and cyclical component of industrial production, unemployment, and PMI Composite, 1962:3 to 2014:12



quite differently from the level. The purchasing managers composite index from the Institute of Supply Management (row 3) appears to be stationary, and McCracken and Ng (2016) entered this series directly into PCA without any transformation.

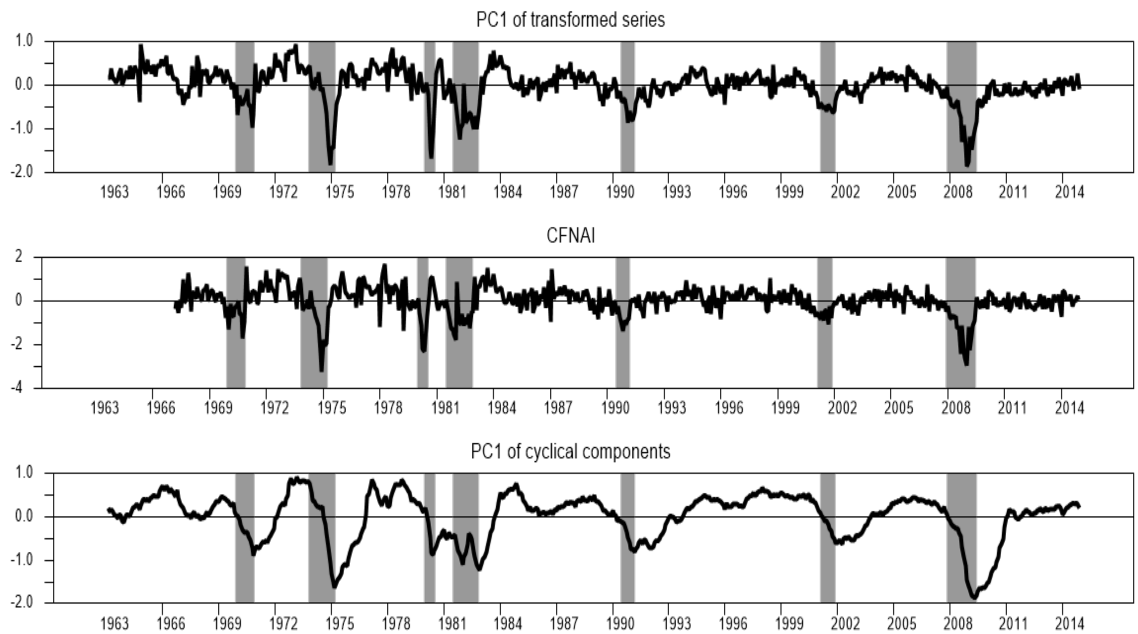
The top panel of Figure 2 plots the first principal component of the transformed series arrived at by McCracken and Ng (2016).[8] This inherits some of the high-frequency fluctuations seen in the (1,2) and (2,2) panels of Figure 1. Indeed, McCracken and Ng (2016) regarded this series as too volatile to reliably identify business cycles and turning points, and instead plotted in their Figure 3 the accumulation of this series. The CFNAI (shown in panel 2 of Figure 2) is very similar to the first principal component of the FRED-MD macro data set.

The third column of Figure 1 plots the cyclical components of industrial production, unemployment, and PMI as estimated by the residuals of the OLS regression (7) with $h = 24$ and $p = 12$.[9] PMI is almost impossible to predict two years in advance, and our cyclical component

[8] We generated this figure using the exact data and code posted at https://research.stlouisfed.org/econ/mccracken/fred-databases/. Note that we have multiplied the series by -1 in order to give it the property that the factor declines in recessions, and that their Figure 3 plots the accumulations ($s_{1t} = \sum_{j=1}^{t} \tilde{f}_{1t}$) whereas our graph shows $\tilde{f}_{1t}$ itself.

[9] For those series that McCracken and Ng (2016) transformed using logs, first differences of logs, or second differences of logs (their transformations 4-6), we simply took the log of the variable before performing the regresssion. Thus

Figure 2: First PC of FRED-MD variables as transformed by McCracken and Ng (2016), the Chicago Fed National Activity Index, and first PC of cyclical components of FRED-MD variables, 1962:3 to 2014:12



is almost identical to the original series. Thus both our method and the traditional approach use this variable essentially as is. There is some but not much predictability of the unemployment rate at the two-year horizon, so for this variable our transformation much more closely resembles the original series than it does the first-difference transformation. For industrial production, our approach takes out the broad trend while retaining the essential cyclical behavior observed in the raw data. The three variables in the third column, unlike those in the second column, all share a common characterization of what is happening over the business cycle. Consistent with a long tradition in business cycle research, when plotted this way PMI appears as a leading indicator, industrial production as a coincident indicator, and unemployment as a coincident or lagging indicator, with all three clearly following the same cycle.

The first principal component of the estimated cyclical components of the variables in the

for example the series plotted in the upper left panel of Figure 1 is 100 times the natural logarithm of the industrial production index. For those series that they used as is, as first differences, or second differences (their transformations 1-3), we simply used the variable as is. They employed a special transformation (7) for nonborrowed reserves. One would have expected to take logs of a variable like this, but the variable took on negative values in 2008. For this series their transformation was $y_{it} = \Delta(x_t/x_{t-1} - 1.0)$ and we used $y_{it} = x_t/x_{t-1}$.

data set is plotted in the bottom panel of Figure 2. Unlike the CFNAI, this provides a very clean summary of historical business cycles. There is another interesting difference between the third panel and the first two. The NBER defines the business-cycle trough (the end point of the shaded regions) as the low point in the *level* of overall economic activity. For example, in the first month of a new expansion, the unemployment rate is still very high, but it has started to come down. Our series in the bottom panel of Figure 2 captures this feature very well, reaching a trough at exactly the point identified by the NBER Business Cycle Dating Committee. By contrast, the low point in the series plotted in the first two panels typically comes more towards the middle of the recession. This is because the *rate of decline* of real output (a common raw input in the variables as usually transformed) starts to ease well before the recession has ended. Both in terms of the cleanness of the series and its timing, we would suggest that our approach offers a better characterization of the state of the U.S. business cycle over this sample period.

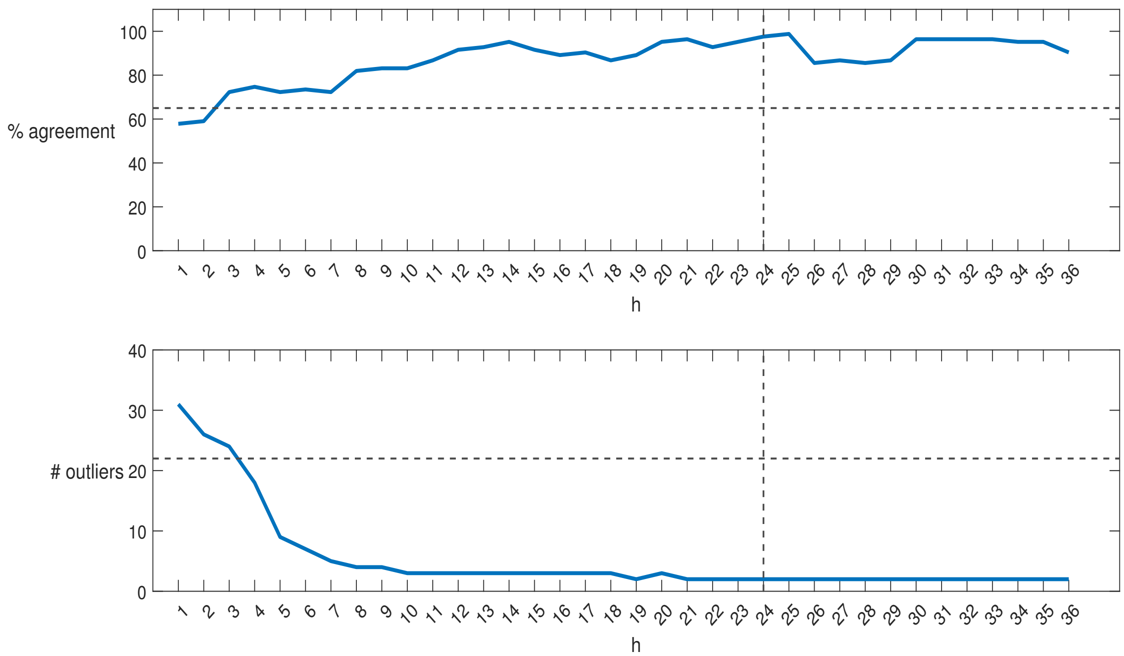
6.1 Choice of h

Our method identifies common components in h-period-ahead forecast errors. Any value of $h \geq 1$ is sufficient to produce a stationary series. Which value of h to use depends on which common components the researcher is interested in.

McCracken and Ng did not propose their method as a tool to identify business-cycle turning points, and neither do we. They nevertheless reported an exercise that we were interested in reproducing. They applied the Bry and Boschan (1971) algorithm for dating business-cycle turning points to the first principal component that resulted from their analysis. They found that the resulting series was in agreement with NBER recession dates only 65% of the time, meaning that 65% of NBER recession periods fell within the peak-to-trough phases of the series. We were able to reproduce this result by adapting the code in Stock and Watson (2014).[10] We found a similar result (67% agreement) when Bry and Boschan is applied to the CFNAI.

[10] As discussed in Stock and Watson (2014, f.n. 2), the first step in applying the Bry and Boschan algorithm is to take a centered moving average of the series for which turning points are to be assigned, which in this case is a moving average of the first principal component. Specifying nma = 12 seems to come closest to reproducing the numbers reported in McCracken and Ng (2016, p. 583). Our numbers, however, were not in all instances identical. This appears to be in part because McCracken and Ng were reporting results for a data set that has some slightly different series from those in 2015-04.csv. In the case of the percent agreement with NBER recession dates, our calculation (65%) exactly reproduces theirs. Other numbers reported here in the text are based on our reproduction using the 2015-04.csv data, which sometimes differ from those reported in McCracken and Ng (2016, p. 583).

Figure 3: Agreement with NBER-dated recessions and number of series exhibiting outliers as a function of forecast horizon h, 1967:2-2014:12



Notes to Figure 3. Horizontal axis: forecast horizon h. Top panel: percent of NBER-dated recession periods that fall within Bry-Boschan-designated downturns. Bottom panel: number of series exhibiting outliers. Horizontal lines denote results for the first principal component found by McCracken and Ng (2016).

We next were interested to calculate the percent agreement with NBER recessions that results from our method for every value of h between 1 and 36 months. The result is plotted in the top panel of Figure 3. The common component of $h = 1$- or 2-month-ahead forecast errors has less correspondence with NBER recession dates than either the McCracken and Ng series or the CFNAI. We attribute this to the fact that replacing stationary series like the PMI in the bottom row of Figure 1 with a 1-month-ahead forecast error removes what we would normally think of as the business-cycle indicator provided by the series. By contrast, for every $h \geq 3$, our approach has a closer correspondence to NBER recessions than either McCracken-Ng or the CFNAI. The maximal agreement (99%) is obtained using $h = 25$. Choosing $h = 24$ is very similar (98%).

In commenting on this, McCracken and Ng noted that the accumulation of their series $\hat{F}_t^{MN} = \sum_{s=1}^{t} \hat{f}_s^{MN}$ has a much better correspondence with recessions, with 93% agreement in our reproduction of their analysis. Indeed, Figure 3 in their paper plots the accumulated measure $\hat{F}_t^{MN}$

rather than the principal component $\hat{f}_t^{MN}$ itself, reflecting their judgment that $\hat{F}_t^{MN}$ is a better indicator of the state of the business cycle than is the principal component $\hat{f}_t^{MN}$ itself.[11] We found that 100% of the Bry-Boschan dated recessions for the accumulated CFNAI were also characterized as recessions by NBER.

McCracken and Ng noted that agreement of the original $\hat{f}_t^{MN}$ with NBER expansion dates was even weaker. We found that only 51% of the expansion dates based on $\hat{f}_t^{MN}$ were also designated as expansion by NBER and still only 61% agreement using $\hat{F}_t^{MN}$. Our series with $h = 24$ does a little better than either of these with 65% agreement.

Our conclusion is that $h = 24$ months affords a useful characterization of the common business-cycle factors across a broad range of economic indicators. In the next subsection we note another important reason why researchers should be interested in the common components in forecast errors at this horizon.

6.2 Outliers

Previous users of large macro data sets devoted a lot of attention to outliers and implemented procedures to mitigate their influence. Prior to the COVID-19 recession of 2020, the CFNAI discarded observations that were more than six times the interquartile range, as did Stock and Watson (1999) in some of their analysis. McCracken and Ng (2016) discarded observations that were more than ten times the interquartile range. This criterion identifies 79 different observations on 22 different variables as outliers in the 1960:3 to 2014:12 data set.

We noted in Section 2 that as a result of the Central Limit Theorem, we should expect fewer outliers when we calculate principal components of h-period-ahead forecasts for larger values of h. To calculate outliers for our method, we constructed forecast errors from leave-one-out regressions[12] and designated values exceeding ten times the interquartile range as outliers. The number of series exhibiting outliers is plotted as a function of h in the bottom panel of Figure 3. Using $h = 1$ results in 31 series exhibiting outliers, substantially more than in calculating the McCracken and Ng (2016) principal components. Many of the additional outliers come from interest rate

[11]Notwithstanding, McCraken and Ng noted that a drawback of $\hat{F}_t^{MN}$ is that $\hat{F}_t^{MN} = 0$ at $t = 0$ and $t = T$ by construction. Note that our approach completely avoids this problem.

[12]That is, we calculated $\tilde{c}_{it} = y_{it} - \tilde{\alpha}_{i,t}\tilde{z}_{i,t-h}$ with $\tilde{\alpha}_{i,t} = \left(\sum_{s=1,s\neq t}^{T} \tilde{z}_{i,s-h}\tilde{z}'_{i,s-h}\right)^{-1} \left(\sum_{s=1,s\neq t}^{T} \tilde{z}_{i,s-h}y_{is}\right)$ estimated separately for each i and t and then divided $\tilde{c}_{i.t}$ by its observed interquartile range.

spreads. McCracken and Ng (2016) used these as is without filtering, which turns out to be close to our definition of the cyclical component of these variables. By contrast, using $h = 1$-month-ahead forecast errors highlights some of the unusual behavior of interest rates during the Volcker monetary contraction in 1980-81. The number of outliers steadily decreases as h is increased. Only two series exhibit outliers for $h \geq 21$.[13]

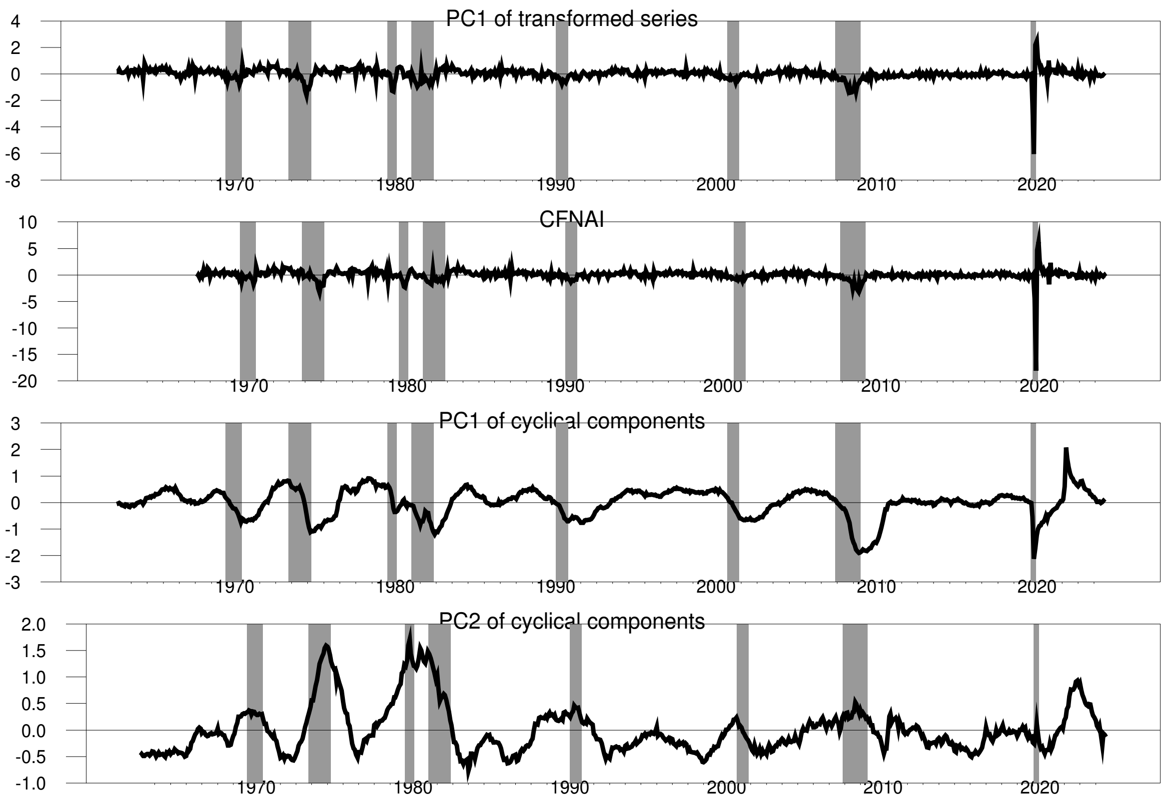
Our recommended procedure is to use two-year-ahead regression residuals and make no corrections for outliers. The series that we have plotted in the bottom panel of Figure 2 is the unadjusted first principal component of the full set of OLS residuals $\hat{c}_{it}$.

Outliers are an even bigger issue when data for 2020 are included. For the 2024:12 vintage of FRED-MD, the McCracken-Ng procedure would identify 40 of the 126 variables as all being outliers in the single month of 2020:4. Despite dropping all of these 40 observations, the first principal component calculated using their algorithm shows an enormous decline in this month. Indeed, in order to include the 2020 observations in the top panel of Figure 4, the scale must be so large that it makes all the previous cyclical fluctuations barely noticeable. The CFNAI modified its procedure for dealing with anomalous observations to handle these observations. Even so the CFNAI still displays an unprecedented drop in 2020, as seen in the second panel.

By contrast, only two variables are identified as outliers for 2020:4 for purposes of our approach, these being new claims for unemployment insurance and the number unemployed for less than 5 weeks. The result of applying our procedure to the FRED-MD database available as of December of 2024 with no corrections for outliers is displayed in the third panel of Figure 4. Note that, unlike the top two panels, our series describes the downturn in 2020 on a comparable scale as earlier recessions, although our series indicates that the speed of the downturn was unprecedented, as was the growth in the first two months of the recovery. Also in contrast to the first two panels, our series indicates (correctly, in our view) that the economy did not fully recover from the COVID-19 shock until September of 2021. The sharp spike up in our series in April of 2022 reflects the fact that most macro variables were substantially higher in April 2022 than one would have predicted based on observations in April 2020. This conclusion is also consistent with the aggressive actions of policy makers in the spring of 2020. Our series further indicates that economic activity remained unusually strong through the fall of 2023.

[13] These are total bank reserves and nonborrowed reserves in the Fed's response to the financial crisis.

Figure 4: First PC of FRED-MD variables as transformed by McCracken and Ng (2016), the Chicago Fed National Activity Index, and first and second PC of cyclical components of FRED-MD variables, 1962:3 to 2024:9



6.3 Missing observations

Another issue with large data sets comes from discontinued, newly added, or missing variables. McCracken and Ng (2016) adapted the Stock and Watson (2002) algorithm for unbalanced panels, though they found in their original data set that the results are essentially identical if one simply drops variables as needed to create a balanced panel. For our application, we have simply calculated principal components of $\hat{c}_{it}$ on a balanced panel, though there is no obstacle to applying the Stock and Watson (2002) algorithm to an unbalanced panel of $\hat{c}_{it}$.[14]

[14] A balanced panel was created from the 126 variables in the 2024:12 dataset by: using only data over 1960:1-2024:9; dropping the Michigan Survey of Consumer Sentiment (UMCSENT), trade-weighted exchange rate (TWEXAFEGSMTH), and new orders for consumer goods (ACOGNO) and nondefense capital goods (ANDENO), which are the same four series dropped by McCracken and Ng to create a balanced panel from the 2015:4 dataset; dropping the VIX (VIXCLS), which was not included in the 2015:4 dataset and whose first value is July 1962; and dropping the financial commercial paper rate (CP3M) and the commercial paper-fed funds spread (COMPAPFF) which were not reported for April 2020. The particular variables used in our analysis of the 2023 vintage dataset are described in Table B1.

6.4 Uses of macroeconomic cyclical factors

A key use of PCA is to summarize the statistical information in a large cross section of indicators; for illustrations see Bernanke et al. (2005), Bai and Ng (2008), Forni et al. (2009), Bai and Ng (2010), and Stock and Watson (2016). The movement in variable i that is captured by the jth factor alone is given by $\hat{\lambda}_{ij}\hat{f}_{jt}$. Since $\hat{c}_{it}$ is normalized to have unit variance, the fraction of the variance of the stationary component of variable i that is explained by the jth macro factor is given by $\hat{\lambda}_{ij}^2 T^{-1}\sum_{t=1}^{T}\hat{f}_{jt}^2$.

Table B1 reports the R^2 explained by the first macroeconomic cyclical factor for each of the 119 variables used in our analysis of the 2023 vintage data set. The first factor alone accounts for almost 2/3 of the variance of typical indicators of real output or income and more than half of the variance of typical indicators of labor-market conditions. The first cyclical factor is far less successful at describing financial indicators and nominal prices. It is interesting that when we add the second cyclical factor, the R^2 for the median price indicator rises to 66%. The first factor thus seems mainly to capture real economic conditions and the second characterizes nominal prices and interest rates.

The fourth panel in Figure 4 plots the second cyclical factor. This by construction is orthogonal to the first, and often continues to fall even as the recovery in real economic activity is beginning. This is consistent with the view that nominal variables may respond sluggishly to business-cycle developments. It describes events beginning in 2022 as a third big U.S. inflation wave, though less dramatic than the big inflations of 1973-74 and 1979-81.

We next explore the use of the cyclical factors in forecasting. Stock and Watson (1999) demonstrated that the first principal component of a large data set of real macroeconomic variables could be very helpful for forecasting inflation. Their finding gave rise to the Chicago Fed National Activity Index (CFNAI), a PCA-based indicator that is still widely used today. We compare the usefulness for forecasting of the CFNAI (denoted $\hat{f}_t^{CF}$), the first or second principal component calculated using the algorithm and data set of McCracken and Ng (2016) (denoted $\hat{f}_t^{MN1}$ and $\hat{f}_t^{MN2}$, respectively), or the first or second principal component of the forecasting residuals (denoted $\hat{f}_t^{HX1}$ or $\hat{f}_t^{HX2}$) calculated from the FRED-MD data set using equation (12). Our approach to comparing different forecasts is similar to that used by Stock and Watson (1999) and McCracken and Ng (2016).

A particular model m uses a set of variables $\mathbf{x}_t^m$ that are observed at date t to try to forecast the value of a variable of interest y_{t+h}^h that will not be observed until $t+h$:

$$y_{t+h}^h = \boldsymbol{\pi}^{m\prime}\mathbf{x}_t^m + u_{t+h}^{m,h}. \tag{21}$$

We estimated the value of $\boldsymbol{\pi}^m$ by OLS regression on the subsample $t = T_0, T_0+1, ..., T_1 - h - 1$ and used these coefficients to forecast $y_{T_1}^h$. We then augmented the sample by one observation, estimating the regression over $t = T_0, T_0+1, ..., T_1 - h$ and using those coefficients to forecast $y_{T_1+1}^h$. We repeated this for an evaluation period T_1 to T_2 and calculated the average squared forecast error over this evaluation period. We performed this analysis using three different evaluation periods. The first evaluation period is specified by $T_1 = 1970{:}1$ to $T_2 = 1996{:}12$, which was the evaluation period in the original study by Stock and Watson (1999). The second evaluation period is $T_1 = 1997{:}1$ to $T_2 = 2014{:}12$, which corresponds to the new data used by McCracken and Ng (2016) that were not available to Stock and Watson (1999). The third evaluation period is $T_1 = 2015{:}1$ to $T_2 = 2024{:}9$, which is the new data available since publication of McCracken and Ng (2016). The models we considered were a pure autoregressive model,

$$\mathbf{x}_t^{AR} = (1, y_t^1, y_{t-1}^1, ..., y_{t-5}^1)',$$

and models that add to the autoregressive model six lags of one of the principal components estimates. For example,

$$\mathbf{x}_t^{CF} = (\mathbf{x}_t^{AR\prime}, \hat{f}_t^{CF}, \hat{f}_{t-1}^{CF}, ..., \hat{f}_{t-5}^{CF})'.$$

This differs a little from the forecast evaluations performed by Stock and Watson (1999) and McCracken and Ng (2016) in that these authors used BIC to select different lag lengths for the autoregressive and principal components and for each subsample, whereas we set the lag length to six for every evaluation. Also, since $\hat{f}_t^{CF}$ is only available beginning in 1967:3, we used $T_0 = 1967{:}9$ as the first date for estimation of all models. In every case, for each T we re-estimated the coefficients $\hat{\boldsymbol{\pi}}^m$ for the forecasting regression (21) using an expanding data set ending h periods before the variable being forecast.

A separate question is the data set used to estimate the factors $\hat{f}_t$ themselves. Insofar as the factors are only identified up to sign, the meaning of coefficients multiplying $\hat{f}_t$ could change across expanding samples. We formed data sets to calculate the factors $\hat{f}_t$ for each of the three methods as follows. (i) For the explanatory variable $\hat{f}_t^{MN}$, for the first two evaluation samples we followed McCracken and Ng (2016) in calculating the factor $\hat{f}_t^{MN}$ using the full historical vintage of the FRED-MD database available as of 2015:4.[15] For the third evaluation sample, we estimated $\hat{f}_t^{MN}$ using the full database available as of 2024:12. (ii) For the explanatory variable $\hat{f}_t^{CF}$ we used the value of the CFNAI as it was reported in 2024 for all historical dates.[16] (iii) For the explanatory variable $\hat{f}_t^{HX}$, we need to specify both (a) the data set used to estimate the coefficients $\hat{\alpha}_i$ for the detrending regression (7) and (b) the data set used to calculate principal components of the forecast errors $\hat{c}_{it} = y_{it} - \hat{\alpha}_i' z_{i,t-h}$. For (a), for all three evaluation samples, we estimated $\hat{\alpha}_i$ from the 2015:4 vintage data set. For (b), for the first two evaluation samples, we calculated principal components of $\hat{c}_{it}$ using the 2015:4 vintage data set. For the last evaluation sample, we calculated principal components of $\hat{c}_{it}$ using the 2024:12 data set. Thus our forecasting exercise for the third subsample is not affected by the potential concern that the detrending regression used information that was not available at the time the forecast was made. This is because the coefficients for the detrending regression were all known as of 2015. We also repeated the exercise using instead for (b) trend coefficients estimated using all the data through 2024. This alternative specification produced similar results (not reported here).

For our first set of evaluations we set y_{t+h}^h to be the average inflation rate between month t and $t+h$, quoted at an annual rate,

$$y_{t+h}^h = (1200/h)\log(CPI_{t+h}/CPI_t),$$

where CPI_t denotes the level of the consumer price index in month t.[17] The column labeled AR in

[15]Our series for $\hat{f}_t^{MN1}$ for this subsample is almost (but not quite) identical to the series analyzed by McCracken and Ng (2016). We have been unable to identify the source of the small discrepancies.

[16]We downloaded $\hat{f}_t^{CF}$ on January 13, 2025 from the FRED database at https://fred.stlouisfed.org/series/CFNAI.

[17]Here again our evaluation design differs slightly from that in Stock and Watson (1999) and McCracken and Ng (2016) in that those authors took the object of interest to be to forecast the *change* in the inflation rate as a function of lagged changes:

$$y_{t+h}^h = (1200/h)\log(CPI_{t+h}/CPI_t) - 1200\log(CPI_t/CPI_{t-1}).$$

Table 3: Mean squared forecast errors for different models

sample	horizon	Consumer Price Index					
		AR	CF	MN1	HX1	MN2	HX2
1970-1996	h=1	7.91	1.00	0.99	1.03	0.98	0.90
	h=6	4.26	0.81	0.77	0.80	0.95	0.88
	h=12	5.32	0.70	0.62	0.74	1.02	1.33
1997-2014	h=1	12.26	1.03	1.04	1.02	0.97	1.09
	h=6	6.08	1.23	1.23	1.23	0.96	1.11
	h=12	4.21	1.22	1.22	1.28	0.95	1.17
2015-2024	h=1	8.07	1.51	1.27	1.14	0.94	1.04
	h=6	3.70	1.95	1.48	1.76	1.03	1.01
	h=12	3.64	1.71	1.30	1.49	1.09	0.96

sample	horizon	Industrial Production					
		AR	CF	MN1	HX1	MN2	HX2
1970-1996	h=1	76.73	0.97	0.94	0.96	1.01	1.02
	h=6	38.66	0.92	0.93	0.83	0.71	0.79
	h=12	27.19	1.01	1.06	1.21	0.49	0.87
1997-2014	h=1	58.90	0.85	0.83	0.98	1.04	1.00
	h=6	22.61	0.93	0.94	1.05	1.24	1.12
	h=12	20.11	0.96	1.01	1.06	1.26	1.11
2015-2024	h=1	506.67	1.73	0.93	1.10	1.02	1.00
	h=6	87.37	2.29	1.16	1.03	1.05	0.94
	h=12	38.10	2.31	1.14	0.85	1.03	0.93

Notes to Table 3. AR column reports simulated out-of-sample mean squared forecast error for purely autoregressive model evaluated over three different out-of-sample periods. CF column reports the relative MSE when lags of the Chicago Fed National Activity Index are added to the autoregression, with a value less than one indicating the variable is useful for forecasting. MN1 column reports the relative MSE when lags of the first principal component calculated using the procedures in McCracken and Ng (2016) are used in place of the CFNAI. HX1 reports relative MSE when lags of the first principal component of the estimated cyclical components are used in place of CFNAI. MN2 and HX2 report results when the second principal component is used instead of the first.

Table 3 reports the simulated out-of-sample mean squared error of a purely autoregressive model for each of the three evaluation samples and for forecast horizons of $h = 1, 6$, or 12 months.[18] The first panel reproduces the finding of Stock and Watson (1999) that an index like CFNAI significantly improves forecasts for longer horizons over the 1970-1996 period. The alternative measures $\hat{f}_t^{MN1}$ or $\hat{f}_t^{HX1}$ offer similar improvements. The indexes offer little or no improvement for one-month-ahead forecasts over this period, but again are similar to each other. All three indexes are outperformed by simple autoregressive forecasts over either of the later two evaluation periods. The observation that inflation has become much harder to forecast in data since 1996 has been reported by a number of other researchers, including Atkeson and Ohanian (2001), Fisher et al. (2002), Stock and Watson (2007), and Stock and Watson (2008).[19] Interestingly, the second cyclical factor $\hat{f}_t^{HX2}$ does better than any of the other four indexes at forecasting inflation at the one-month horizon for the 1970-1996 sample and at the 6-12 month horizons for the 2015-2024 sample.

Table 3 also reports forecasts of industrial production, setting

$$y_{t+h}^h = (1200/h)\log(IP_{t+h}/IP_t),$$

for IP_t the level of the industrial production index in month t. The indexes $\hat{f}_t^{CF}$, $\hat{f}_t^{MN1}$, and $\hat{f}_t^{HX1}$ all help forecast industrial production over near horizons in the first two evaluation periods. The CFNAI does particularly poorly at forecasting either inflation or industrial production at any horizon for 2015-2024. Both $\hat{f}_t^{HX1}$ and $\hat{f}_t^{HX2}$ do significantly better than CFNAI in every case over this period.

We conclude that our approach offers similar benefits to conventional PCA when evaluated in terms of simulated out-of-sample forecasts, and does much better than measures like the CFNAI for recent data. We share the conclusion of the earlier literature that the usefulness of any principal-component-based measure for purposes of forecasting depends on the variable, evaluation period, and horizon of the forecast.

[18] Stock and Watson (1999) reported results for $h = 12$ months whereas McCracken and Ng (2016) reported results for $h = 1, 6$, and 12.

[19] The result is also revealed in the numbers reported in Table 5 of McCracken and Ng (2016) , though the authors did not comment on it.

7 Conclusion

Calculating principal components of medium-horizon forecast errors is a viable approach to identifying the common cyclical factors that drive a large collection of potentially nonstationary economic indicators. This avoids the need to decide how to detrend each individual series and is much more promising than approaches such as the Chicago Fed National Activity Index for handling data that include the large outliers of 2020.

References

Antolin-Diaz, J., Drechsel, T., and Petrella, I. (2017). Tracking the slowdown in long-run GDP growth. *Review of Economics and Statistics*, 99(2):343–356.

Atkeson, A. and Ohanian, L. E. (2001). Are Phillips Curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1):2–11.

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

Bai, J. and Ng, S. (2004). A PANIC attack on unit roots and cointegration. *Econometrica*, 72(4):1127–1177.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Ng, S. (2008). *Large Dimensional Factor Analysis*. Now Publishers Inc.

Bai, J. and Ng, S. (2010). Instrumental variable estimation in a data rich environment. *Econometric Theory*, 26(6):1577–1606.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18–29.

Bańbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160.

Barigozzi, M., Lippi, M., and Luciani, M. (2021). Large-dimensional dynamic factor models: Estimation of impulse–response functions with i(1) cointegrated factors. *Journal of Econometrics*, 221(2):455–482.

Beghin, L., Nikitin, Y., and Orsingher, E. (2005). Exact small ball constants for some Gaussian processes under the L^2-norm. *Journal of Mathematical Sciences*, 128(1):2493–2502.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120(1):387–422.

Beveridge, S. and Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics*, 7(2):151–174.

Bry, G. and Boschan, C. (1971). Cyclical analysis of time series: Selected procedures and computer programs. *NBER Technical Paper*, 20.

Chen, X. and Li, W. V. (2003). Quadratic functionals and small ball probabilities for the m-fold integrated brownian motion. *The Annals of Probability*, 31(2):1052–1077.

Çinlar, E. (2011). *Probability and Stochastics*. Springer.

Crump, R. K. and Gospodinov, N. (2022). On the factor structure of bond returns. *Econometrica*, 90(1):295–314.

D'Agostino, A., Giannone, D., Lenza, M., and Modugno, M. (2016). Nowcasting business cycles: A Bayesian approach to dynamic heterogeneous factor models. In *Dynamic Factor Models*, pages 569–594. Emerald Group Publishing Limited.

Doz, C., Giannone, D., and Reichlin, L. (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics*, 94(4):1014–1024.

Federal Reserve Bank of Chicago (2021). Background on the Chicago Fed national activity index.

Fisher, J. D., Liu, C. T., and Zhou, R. (2002). When can we forecast inflation? *Economic Perspectives Federal Reserve Bank of Chicago*, 26(1):32–44.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, 25(5):1319–1347.

Gao, F., Hannig, J., and Torcaso, F. (2003). Integrated brownian motions and exact l_2-small balls. *The Annals of Probability*, 31(3):1320–1337.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics*, 100(5):831–843.

Harris, D. (1997). Principal components analysis of cointegrated time series. *Econometric Theory*, 13(4):529–557.

Joslin, S., Singleton, K. J., and Zhu, H. (2011). A new perspective on Gaussian dynamic term structure models. *Review of Financial Studies*, 24(3):926–970.

Komlós, J., Major, P., and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 34:33–58.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Onatski, A. and Wang, C. (2018). Alternative asymptotics for cointegration tests in large VARs. *Econometrica*, 86(4):1465–1478.

Onatski, A. and Wang, C. (2021). Spurious factor analysis. *Econometrica*, 89(2):592–614.

Piazzesi, M. (2010). Affine term structure models. In *Handbook of Financial Econometrics: Tools and Techniques*, pages 691–766. Elsevier.

Shumway, R. H. and Stoffer, D. S. (2017). *Time Series Analysis and Its Applications*. Springer.

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

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39:3–33.

Stock, J. H. and Watson, M. W. (2008). Phillips Curve inflation forecasts. *NBER Working Paper*, 14322.

Stock, J. H. and Watson, M. W. (2014). Estimating turning points using large data sets. *Journal of Econometrics*, 178:368–381.

Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of Macroeconomics*, volume 2, pages 415–525. Elsevier.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

A Proofs

A.1 Proof of Theorem 2

Recall that the operator norm of a symmetric matrix S is defined by

$$\|S\|_{op} = \sup_{\gamma\in\Gamma}\frac{\gamma' S\gamma}{\gamma'\gamma}.$$

When S is positive semidefinite, this is the largest eigenvalue: $\|S\|_{op} = \varrho_{\max}(S)$. Since the trace of S (the sum of the diagonal elements) is equal to the sum of the eigenvalues, it follows immediately that for any symmetric positive semidefinite matrix S,

$$\sup_{\gamma\in\Gamma}\frac{\gamma' S\gamma}{\gamma'\gamma} \leq \textstyle\sum_{i=1}^{N} s_{ii}. \tag{A-1}$$

Proof of Theorem 2(i)-(ii)

Note first that

$$\begin{aligned}
& \left\|(NT)^{-1}\textstyle\sum_{t=1}^{T}\hat{C}_t\hat{C}_t' - (NT)^{-1}\sum_{t=1}^{T}C_tC_t'\right\|_{\text{op}} \\
= & \sup_{\gamma\in\Gamma}\frac{\left|\gamma'\left((NT)^{-1}\sum_{t=1}^{T}\hat{C}_t\hat{C}_t' - (NT)^{-1}\sum_{t=1}^{T}C_tC_t'\right)\gamma\right|}{\gamma'\gamma} \\
= & \sup_{\gamma\in\Gamma}\left|(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^{T}\hat{C}_t\hat{C}_t'\gamma - (N^2T)^{-1}\gamma'\sum_{t=1}^{T}C_tC_t'\gamma\right| \\
\leq & \sup_{\gamma\in\Gamma}(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^{T}\hat{V}_t\hat{V}_t'\gamma + 2\sup_{\gamma\in\Gamma}(N^2T)^{-1}\left|\gamma'\sum_{t=1}^{T}C_t\hat{V}_t'\gamma\right|,
\end{aligned} \tag{A-2}$$

where the last line uses the relation $\hat{C}_t = C_t + \hat{V}_t$. We first show that the two terms in the last line of (A-2) converge in probability to 0. To show this for the first term, notice from (A-1) that

$$\begin{aligned}
\sup_{\gamma\in\Gamma}(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^{T}\hat{V}_t\hat{V}_t'\gamma &= \sup_{\gamma\in\Gamma}(NT)^{-1}\frac{\gamma'\sum_{t=1}^{T}\hat{V}_t\hat{V}_t'\gamma}{\gamma'\gamma} \\
&\leq (NT)^{-1}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^2 \leq \max_{1\leq i\leq N} T^{-1}\sum_{t=1}^{T}\hat{v}_{it}^2,
\end{aligned} \tag{A-3}$$

which tends to zero in probability by Assumption 3(ii).

For the last term in (A-2),

$$\sup_{\gamma\in\Gamma}\left|(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^{T} C_t\hat{V}_t'\gamma\right| \leq \left[\sup_{\gamma\in\Gamma}(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^{T} C_tC_t'\gamma\right]^{1/2}\left[\sup_{\gamma\in\Gamma}(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^{T} \hat{V}_t\hat{V}_t'\gamma\right]^{1/2}.$$

The first term converges in probability to a number no larger than $\omega_{11}^{1/2}$ from Theorem 1, and the second converges in probability to 0 from (A-3). Thus $\sup_{\gamma\in\Gamma}\left|(N^2T)^{-1}\gamma'\sum_{t=1}^{T} C_t\hat{V}_t'\gamma\right| \xrightarrow{p} 0$. We thus conclude from (A-2) that

$$\sup_{\gamma\in\Gamma}\left|(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^{T} \hat{C}_t\hat{C}_t'\gamma-(N^2T)^{-1}\gamma'\sum_{t=1}^{T} C_tC_t'\gamma\right| \xrightarrow{p} 0, \tag{A-4}$$

or equivalently

$$\left\|(NT)^{-1}\textstyle\sum_{t=1}^{T}\hat{C}_t\hat{C}_t'-(NT)^{-1}\sum_{t=1}^{T} C_tC_t'\right\|_{\text{op}} \xrightarrow{p} 0.$$

Recall that $T^{-1}\sum_{t=1}^{T}\hat{f}_{jt}^2$ and $T^{-1}\sum_{t=1}^{T}\tilde{f}_{jt}^2$ denote the j-th largest eigenvalues of $(NT)^{-1}\sum_{t=1}^{T}\hat{C}_t\hat{C}_t'$ and $(NT)^{-1}\sum_{t=1}^{T}C_tC_t'$, respectively, and that $T^{-1}\sum_{t=1}^{T}\tilde{f}_{jt}^2$ converges in probability. Therefore, by Weyl's inequality on perturbation, it is true for all j that

$$|T^{-1}\textstyle\sum_{t=1}^{T}\hat{f}_{jt}^2-T^{-1}\sum_{t=1}^{T}\tilde{f}_{jt}^2| \leq \left\|(NT)^{-1}\sum_{t=1}^{T}\hat{C}_t\hat{C}_t'-(NT)^{-1}\sum_{t=1}^{T} C_tC_t'\right\|_{\text{op}},$$

which implies that $T^{-1}\sum_{t=1}^{T}\hat{f}_{jt}^2$ converges in probability to ω_{jj} for $j = 1,2,\ldots,r$, and to 0 for $j = r+1,\ldots,k$.

Proof of Theorem 2(iii)

Notice that $(N^2T)^{-1}\hat{\Lambda}'\sum_{t=1}^{T}\hat{C}_t\hat{C}_t'\hat{\Lambda}$ is a diagonal matrix for all N and T by the definition of $\hat{\Lambda}$ with diagonal elements converging in probability to ω_{jj} by result (i):

$$(N^2T)^{-1}\hat{\Lambda}'\textstyle\sum_{t=1}^{T}\hat{C}_t\hat{C}_t'\hat{\Lambda} \xrightarrow{p} \Omega_{FF}. \tag{A-5}$$

Equation (A-4) then establishes that $(N^2T)^{-1}\hat{\Lambda}'\sum_{t=1}^{T}C_tC_t'\hat{\Lambda} \xrightarrow{p} \Omega_{FF}$. We also know from results (R2)-(R6) in Stock and Watson (2002) that

$$(N^2T)^{-1}\gamma'\sum_{t=1}^{T}C_tC_t'\gamma - (N^2T)^{-1}\gamma'\sum_{t=1}^{T}\Lambda'F_tF_t'\Lambda\gamma \xrightarrow{p} 0$$

where the convergence is uniform for $\gamma \in \Gamma$, meaning

$$\text{plim}\left\{(N^2T)^{-1}\hat{\Lambda}'\sum_{t=1}^{T}C_tC_t'\right\} = \text{plim}\left\{(N^2T)^{-1}\hat{\Lambda}'\sum_{t=1}^{T}\Lambda F_tF_t'\Lambda'\hat{\Lambda}\right\} = H\Omega_{FF}H' \qquad \text{(A-6)}$$

for $H = \text{plim }(N^{-1}\hat{\Lambda}'\Lambda)$. Combining results (A-4)-(A-6),

$$\Omega_{FF} = H\Omega_{FF}H'. \qquad \text{(A-7)}$$

Let $\hat{h}_j'$ denote the jth row of $\hat{\Lambda}'\Lambda/N$,

$$\underset{(1\times r)}{\hat{h}_j'} = \underset{(1\times N)}{\hat{\lambda}_j'}\underset{(N\times r)}{\Lambda}/N,$$

for $\hat{\lambda}_j'$ the jth row of $\hat{\Lambda}'$. Then

$$\hat{h}_j'\hat{h}_j = \frac{\hat{\lambda}_j'}{\sqrt{N}}\frac{\Lambda\Lambda'}{N}\frac{\hat{\lambda}_j}{\sqrt{N}}.$$

This is less than or equal to the largest eigenvalue of $\Lambda\Lambda'/N$, which converges to 1. Letting $h_j' = (h_{j1}, h_{j2}, ..., h_{jr})'$ denote the jth row of H, we thus have

$$\hat{h}_j'\hat{h}_j \xrightarrow{p} h_{j1}^2 + h_{j2}^2 + \cdots h_{jr}^2 \leq 1.$$

The (1,1) element of (A-7) states

$$h_1'\Omega_{FF}h_1 = h_{11}^2\omega_{11} + h_{12}^2\omega_{22} + \cdots + h_{1r}^2\omega_{rr} = \omega_{11}.$$

Since $\omega_{11} > \omega_{22} > \cdots > \omega_{rr} > 0$, this requires $h_{11}^2 = 1$ and $h_{12} = \cdots = h_{1r} = 0$. Thus the (1,1) element of $\hat{\Lambda}'\Lambda/N$ converges in probability to ± 1 and other elements of the first row converge to zero.

The (2,2) element of (A-7) states

$$h_{21}^2\omega_{11} + h_{22}^2\omega_{22} + \cdots + h_{2r}^2\omega_{rr} = \omega_{22} \tag{A-8}$$

where $h_{21} = \text{plim } \hat{\lambda}_2\lambda_1/N$. Regress λ_1 on $\hat{\lambda}_1$ with residual q_1:

$$\lambda_1 = \hat{k}_1\hat{\lambda}_1 + q_1 \tag{A-9}$$

$$\hat{k}_1 = (\hat{\lambda}_1'\hat{\lambda}_1/N)^{-1}(\hat{\lambda}_1'\lambda_1/N)$$

$$q_1'\hat{\lambda}_1 = 0$$

$$\lambda_1'\lambda_1/N = \hat{k}_1^2(\hat{\lambda}_1'\hat{\lambda}_1/N) + q_1'q_1/N.$$

We saw above that $\hat{k}_1^2 \xrightarrow{p} 1$, which along with $\lambda_1'\lambda_1/N \to 1$ and $\hat{\lambda}_1'\hat{\lambda}_1/N = 1$ establishes $q_1'q_1/N \xrightarrow{p} 0$. Premultiply (A-9) by $\hat{\lambda}_2'/N$:

$$\hat{\lambda}_2'\lambda_1/N = \hat{k}_1\hat{\lambda}_2'\hat{\lambda}_1/N + \hat{\lambda}_2'q_1/N = \hat{\lambda}_2'q_1/N.$$

But from Cauchy-Schwarz

$$(\hat{\lambda}_2'q_1/N)^2 \leq (\hat{\lambda}_2'\hat{\lambda}_2/N)(q_1'q_1/N) \xrightarrow{p} 0.$$

Thus $\hat{\lambda}_2'\lambda_1/N \xrightarrow{p} h_{21} = 0$ and (A-8) becomes

$$h_{22}^2\omega_{22} + h_{23}^2\omega_{33} + \cdots + h_{2r}^2\omega_{rr} = \omega_{22}.$$

Since $\omega_{22} > \omega_{33} > \cdots > \omega_{rr}$ and $h_{22}^2 + h_{23}^2 + \cdots + h_{2r}^2 \leq 1$, this requires $h_{22}^2 = 1$ and all other elements of the second row of H to be zero, establishing the second row of the claim in Theorem 2(*iii*). Proceeding iteratively through rows 3,4,..,r establishes the rest of the result in (*iii*).

Proof of Theorem 2(iv)

Write

$$\begin{aligned}
\hat{\Xi}\hat{F}_t - F_t &= N^{-1}\hat{\Xi}\hat{\Lambda}'\hat{C}_t - F_t \\
&= N^{-1}\hat{\Xi}\hat{\Lambda}'(\Lambda F_t + e_t + \hat{V}_t) - F_t \\
&= (N^{-1}\hat{\Xi}\hat{\Lambda}'\Lambda - I_r)F_t + N^{-1}\hat{\Xi}\hat{\Lambda}'e_t + N^{-1}\hat{\Xi}\hat{\Lambda}'\hat{V}_t. \qquad \text{(A-10)}
\end{aligned}$$

The task is to show that all three terms in (A-10) have plim 0. That $(N^{-1}\hat{\Xi}\hat{\Lambda}'\Lambda - I_r)F_t \xrightarrow{p} 0$ follows immediately from result (*iii*). For the second term,

$$N^{-1}\hat{\Xi}\hat{\Lambda}'e_t = N^{-1}(\hat{\Xi}\hat{\Lambda}' - \Lambda')e_t + N^{-1}\hat{\Xi}\Lambda'e_t. \qquad \text{(A-11)}$$

Consider the square of the jth element of the first term in (A-11):

$$\left[\frac{(\hat{\Xi}_j\hat{\lambda}_j' - \lambda_j')e_t}{N}\right]^2 \leq \left[\frac{(\hat{\Xi}_j\hat{\lambda}_j' - \lambda_j')(\hat{\Xi}_j\hat{\lambda}_j - \lambda_j)}{N}\right]\left[\frac{e_t'e_t}{N}\right]. \qquad \text{(A-12)}$$

The first term in (A-12) is

$$\frac{(\hat{\Xi}_j\hat{\lambda}_j' - \lambda_j')(\hat{\Xi}_j\hat{\lambda}_j - \lambda_j)}{N} = \frac{\hat{\Xi}_j^2\hat{\lambda}_j'\hat{\lambda}_j}{N} - \frac{\lambda_j'\hat{\Xi}_j\hat{\lambda}_j}{N} - \frac{\hat{\Xi}_j\hat{\lambda}_j'\lambda_j}{N} + \frac{\lambda_j'\lambda_j}{N},$$

which converges in probability to zero by Theorem 2(*iii*). The second term in (A-12) is $O_p(1)$, by result (R1) in Stock and Watson (2002), meaning the plim of (A-12) is zero. The second term in (A-11) also converges in probability to zero as in Stock and Watson (2002) Result (R15). Hence $N^{-1}\hat{\Xi}\hat{\Lambda}'e_t \xrightarrow{p} 0$.

For the third term in (A-10), $N^{-1}\hat{\Xi}\hat{\Lambda}'\hat{V}_t$, note that the jth element is $N^{-1}\hat{\Xi}_j\hat{\lambda}_j'\hat{V}_t$ whose square is

$$N^{-2}\hat{\lambda}_j'\hat{V}_t\hat{V}_t'\hat{\lambda}_j \leq N^{-1}\textstyle\sum_{i=1}^N \hat{v}_{it}^2 \xrightarrow{p} 0 \qquad \text{(A-13)}$$

due to our Assumption 3.

A.2 Proof of Theorem 3

We first verify Assumption 3(ii) that $\max_{1\le i\le N} T^{-1}\sum_{t=1}^{T}\hat{v}_{it}^2 \xrightarrow{p} 0$. Write (16) as $\sum_{t=1}^{T}\hat{v}_{it}^2 = q_{iT}' Q_{iT}^{-1} q_{iT}$ where

$$q_{iT} = \sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} c_{it} \text{ and } Q_{iT} = \sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1}.$$

Then

$$\max_{1\le i\le N} \frac{1}{T}\sum_{t=1}^{T}\hat{v}_{it}^2 \le \left(\frac{1}{T}\max_{1\le i\le N}\|q_{iT}\|^2\right)\cdot\left(\min_{1\le i\le N}\varrho_{\min}(Q_{iT})\right)^{-1}. \tag{A-14}$$

By Condition 4(i) and a standard maximal inequality argument (see, for example, Lemma 2.2.2 of van der Vaart and Wellner 1996),

$$\frac{1}{T}\max_{1\le i\le N}\|q_{iT}\|^2 = O_p(N/T). \tag{A-15}$$

To control the minimum eigenvalue in (A-14), we employ Weyl's inequality on perturbation, which bounds the difference in minimum eigenvalues by the norm of the matrix difference:

$$\left|\varrho_{\min}(Q_{iT}) - \varrho_{\min}(Q_i)\right| \le \|Q_{iT} - Q_i\|.$$

An implication of this is that if $\min_{1\le i\le N}\varrho_{\min}(Q_{iT}) < \delta_N$ for some δ_N and $\max_{1\le i\le N}\|Q_{iT}-Q_i\| \le \mathfrak{c}_6(T,N)$, then $\min_{1\le i\le N}\varrho_{\min}(Q_i) < \delta_N + \mathfrak{c}_6(T,N)$. We can then use Condition 4(iii) to conclude

$$\begin{aligned}
Prob\Big(\min_{1\le i\le N}\varrho_{\min}(Q_{iT}) < \delta_N\Big) = \; & Prob\Big(\min_{1\le i\le N}\varrho_{\min}(Q_{iT}) < \delta_N \text{ and } \max_{1\le i\le N}\|Q_{iT}-Q_i\| > \mathfrak{c}_6(T,N)\Big) \\
& + Prob\Big(\min_{1\le i\le N}\varrho_{\min}(Q_{iT}) < \delta_N \text{ and } \max_{1\le i\le N}\|Q_{iT}-Q_i\| \le \mathfrak{c}_6(T,N)\Big) \\
\le \; & Prob\Big(\max_{1\le i\le N}\|Q_{iT}-Q_i\| > \mathfrak{c}_6(T,N)\Big) \\
& + Prob\Big(\min_{1\le i\le N}\varrho_{\min}(Q_i) < \delta_N + \mathfrak{c}_6(T,N)\Big) \\
= \; & o(1) + Prob\Big(\min_{1\le i\le N}\varrho_{\min}(Q_i) < \delta_N + \mathfrak{c}_6(T,N)\Big) \\
\le \; & o(1) + N\cdot\max_{1\le i\le N} Prob\Big(\varrho_{\min}(Q_i) < \delta_N + \mathfrak{c}_6(T,N)\Big).
\end{aligned}$$

Now set $\delta_N = C\log^{-\mathfrak{c}_4}(N)$ for some small constant C satisfying $\mathfrak{c}_3/(2C)^{1/\mathfrak{c}_4} > 1$, where $\mathfrak{c}_4$ is defined in Condition 4(ii). Since $\mathfrak{c}_6(T, N) = o(\delta_N)$, the above is further bounded by

$$
\begin{aligned}
Prob\Big(\min_{1\le i\le N} \varrho_{\min}(Q_{iT}) < C\log^{-\mathfrak{c}_4}(N)\Big) \le\ & o(1) + N\cdot \max_{1\le i\le N} Prob\Big(\varrho_{\min}(Q_i) < 2\delta_N\Big) \\
&\le o(1) + N\cdot \mathfrak{c}_2 \exp\Big\{-\frac{\mathfrak{c}_3}{(2C\log^{-\mathfrak{c}_4}(N))^{1/\mathfrak{c}_4}}\Big\} \\
&= o(1) + \mathfrak{c}_2 \exp\Big\{\Big(-\frac{\mathfrak{c}_3}{(2C)^{1/\mathfrak{c}_4}} + 1\Big)\log(N)\Big\} \to 0.
\end{aligned}
$$

As a result, we have that

$$
\Big(\min_{1\le i\le N} \varrho_{\min}(Q_{iT})\Big)^{-1} = O_p\Big(\log^{\mathfrak{c}_4}(N)\Big). \tag{A-16}
$$

Results (A-14), (A-15), and (A-16) establish that Assumption 3(ii) follows from Assumption 4 provided that $N\log^{\mathfrak{c}_4}(N)/T = o(1)$.

We next verify Assumption 3(i) that $N^{-1}\sum_{i=1}^N \hat{v}_{it}^2 \xrightarrow{p} 0$ for each t. To start, we have

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^N \hat{v}_{it}^2 &\le \frac{1}{N}\sum_{i=1}^N \Big\{\Big\|\sqrt{T}\Upsilon_{iT}^{-1} z_{i,t-h}\Big\|^2 \cdot \Big\|\Big(\sum_{t=1}^T \Upsilon_{iT}^{-1} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1}\Big)^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^T \Upsilon_{iT}^{-1} z_{i,t-h} c_{it}\Big)\Big\|^2\Big\} \\
&\le \Big(\max_{1\le i\le N}\Big\|Q_{iT}^{-1}\Big(\frac{1}{\sqrt{T}} q_{iT}\Big)\Big\|^2\Big)\Big(\frac{1}{N}\sum_{i=1}^N \|\sqrt{T}\Upsilon_{iT}^{-1} z_{i,t-h}\|^2\Big) \\
&\le \Big(\min_{1\le i\le N} \varrho_{\min}(Q_{iT})\Big)^{-2}\Big(\frac{1}{T}\max_{1\le i\le N}\|q_{iT}\|^2\Big)\Big(\frac{1}{N}\sum_{i=1}^N \|\sqrt{T}\Upsilon_{iT}^{-1} z_{i,t-h}\|^2\Big).
\end{aligned}
$$

From (A-15) and (A-16), we have

$$
\Big(\min_{1\le i\le N} \varrho_{\min}(Q_{iT})\Big)^{-2}\Big(\frac{1}{T}\max_{1\le i\le N}\|q_{iT}\|^2\Big) = O_p\Big(\frac{N\log^{2\mathfrak{c}_4}(N)}{T}\Big).
$$

Finally, Markov's inequality and Condition 4(i) together imply that

$$
\frac{1}{N}\sum_{i=1}^N \|\sqrt{T}\Upsilon_{iT}^{-1} z_{i,t-h}\|^2 = O_p\Big(\frac{1}{N}\sum_{i=1}^N E(\|\sqrt{T}\Upsilon_{iT}^{-1} z_{i,t-h}\|^2)\Big) = O_p(1),
$$

which concludes the proof.

A.3 Proof of Theorem 4

In this proof, we will verify the conditions of Assumption 4 in three settings, (1) stationary y_{it} ($d_i = 0$); (2) unit-root y_{it} without any time trend ($d_i = 1$ and $\mu_i = 0$); (3) unit-root y_{it} with time trend ($d_i = 1$ and $\mu_i \neq 0$). For ease of exposition, we will partition $\{1, 2, \ldots, N\}$ into $\mathcal{N}_1$, $\mathcal{N}_2$ and $\mathcal{N}_3$, corresponding to the three settings. Recall that $u_{it} = (1-L)^{d_i} y_{it} - \mu_i = \sum_{\ell=0}^{\infty} \psi_{i\ell} \eta_{i,t-\ell}$.

Case 1, stationary y_{it}

In the stationary case, there is no need to introduce any additional scaling or rotation of the regressors, so $u_{it} = y_{it} - \mu_i$ and $\Upsilon_{iT}^{-1} = T^{-1/2} I_{p+1}$ where I_{p+1} is the identity matrix. For this case Assumption 5 guarantees that

$$Q_{iT} = \textstyle\sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1} = T^{-1} \sum_{t=1}^{T} z_{i,t-h} z_{i,t-h}' \xrightarrow{p} E(z_{i,t-h} z_{i,t-h}') := Q_i.$$

Below we will verify Conditions 4(i–iii).

For the first part of 4(i), our Assumption 5 directly implies that $E(y_{it}^2)$ is uniformly bounded. To show the second part of 4(i), it suffices to bound the following terms individually:

$$E\Big(\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^{T} c_{it}\Big|^2\Big) = var\Big(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} c_{it}\Big),$$
$$\text{and } E\Big(\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{i,t-l} c_{it}\Big|^2\Big) = var\Big(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{i,t-l} c_{it}\Big), \text{ for } l = h, \ldots, h+p-1.$$

As c_{it} is a linear combination of $u_{it}, u_{i,t-h}, \ldots, u_{i,t-h-p+1}$, the above can further be bounded by the following variances:

$$var\Big(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it}\Big), \text{ and } var\Big(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{i,t-l} u_{i,t-l'}\Big),$$

for $l = h, \ldots, h+p-1$ and $l' = 0, h, \ldots, h+p-1$. Given our linear process specification, the above is bounded. See, for example, Appendix A of Shumway and Stoffer (2017) for an exact calculation. The same argument can be applied to show that $E(\|Q_{iT} - Q_i\|^2) = O(1/T)$.

Condition 4(ii) holds automatically given our assumption that the minimum eigenvalue of $\tilde{Q}_i$ is bounded away from zero (i.e., it is satisfied for any $\mathfrak{c}_4 > 0$).

Finally we verify 4(iii). We adopt Markov's inequality with a union bound, which implies that

$$Prob\left(\max_{i\in\mathcal{N}_0}\left\|Q_{iT}-Q_i\right\| > \sqrt{\frac{N\log^{1/2}(N)}{T}}\right) \le \sum_{i\in\mathcal{N}_0} Prob\left(\left\|Q_{iT}-Q_i\right\| > \sqrt{\frac{N\log^{1/2}(N)}{T}}\right)$$
$$\le C\frac{|\mathcal{N}_0|}{N}\frac{1}{\log^{1/2}(N)} \le C\frac{1}{\log^{1/2}(N)},$$

where C is some constant and $|\mathcal{N}_0|$ is the cardinality of $\mathcal{N}_0$. Thus, we may take $\mathfrak{c}_6(T,N) = \sqrt{N\log^{1/2}(N)/T}$.

Case 2, unit-root y_{it} without time trend

For these observations, we use $\Upsilon_{iT}^{-1} = \tilde{\Upsilon}_{iT}^{-1}R_i^{-1}$, where $\tilde{\Upsilon}_{iT}$ is a diagonal scaling matrix

$$\tilde{\Upsilon}_{iT} = \begin{bmatrix} \sqrt{T}I_p & 0 \\ 0 & T \end{bmatrix}$$

and R_i is a nonsingular rotation matrix such that $\Upsilon_{iT}^{-1}z_{i,t-h} = \tilde{\Upsilon}_{iT}^{-1}\tilde{z}_{i,t-h}$ with

$$\tilde{z}_{i,t-h} = (u_{i,t-h},\ u_{i,t-h-1},\ \ldots,\ u_{i,t-h-p+2},\ 1,\ y_{i,t-h}).$$

Notice that if $p = 1$, we simply set $\tilde{z}_{i,t-h} = z_{i,t-h} = (1,\ y_{i,t-h})$ so no rotation of the regressors is needed. Assumption 5 guarantees that as in Hamilton (1994, eq. (17.7.18)),

$$Q_{iT} = \textstyle\sum_{t=1}^{T}\Upsilon_{iT}^{-1}z_{i,t-h}z_{i,t-h}'\Upsilon_{iT}^{-1} \xrightarrow{d} \begin{bmatrix} \tilde{Q}_i & 0 & 0 \\ 0 & 1 & \psi_i(1)\int_0^1 W_i(r)dr \\ 0 & \psi_i(1)\int_0^1 W_i(r)dr & \psi_i(1)^2\int_0^1 W_i(r)^2dr \end{bmatrix} := Q_i \qquad \text{(A-17)}$$

where $\psi_i(1) = \sum_{l=0}^{\infty}\psi_{il}$, and $W_i(r)$ is a standard Brownian motion for each i.

For Condition 4(i), we first note that for any stationary component of $\tilde{z}_{i,t-h}$, the condition holds by standard variance calculation (see the proof of the stationary case). For the nonstationary component of $\tilde{z}_{i,t-h}$, namely $y_{i,t-h}$, we rewrite it as

$$y_{i,t-h} = \psi_i(1)\sum_{s=1-h}^{t-h}\eta_{is} - \tilde{\epsilon}_{i,t-h} + \tilde{\epsilon}_{i,-h}, \text{ where } \tilde{\epsilon}_{i,t-h} = \sum_{l=0}^{\infty}\tilde{\psi}_{il}\eta_{i,t-h-l} \text{ and } \tilde{\psi}_{il} = \sum_{l'=l+1}^{\infty}\psi_{il'}; \qquad \text{(A-18)}$$

see for example Hamilton (1994, eq. (17.5.3)). Thus to bound moments of $T^{-1}\sum_{t=1}^{T} y_{i,t-h}u_{it}$ it is sufficient to note that

$$E\left(\left|\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1-h}^{t-h}\eta_{is}\right)\left(\sum_{l=0}^{\infty}\psi_{il}\eta_{i,t-l}\right)\right|^2\right) = O(1),$$

where the bound is uniform for $i \in \mathcal{N}_1$.

We next verify Condition 4(ii). Assumption 5(v) bounds the smallest eigenvalue of the upper-left block of (A-17), while Assumption 5(ii) allows us to ignore the additional scaling $\psi_i(1)$ for the Brownian motion. Thus it is sufficient to establish a bound on the smallest eigenvalue of the (2×2) matrix

$$Q_i = \begin{bmatrix} 1 & \int_0^1 W_i(r)dr \\ \int_0^1 W_i(r)dr & \int_0^1 W_i(r)^2 dr \end{bmatrix}$$

as a slight abuse of notation.

To provide a probabilistic bound on its minimum eigenvalue, notice that the determinant of Q_i is given by

$$|Q_i| = \int_0^1 \left(W_i(r) - \int_0^1 W_i(s)ds\right)^2 dr = \varrho_{\min}(Q_i)\varrho_{\max}(Q_i)$$

while the norm $\|Q_i\| \geq \varrho_{\max}(Q_i)$. This means

$$\varrho_{\min}(Q_i) \geq \|Q_i\|^{-1} \int_0^1 \left(W_i(r) - \int_0^1 W_i(s)ds\right)^2 dr.$$

Beghin et al. (2005) showed in their Section 3 that

$$Prob\left(\int_0^1 \left(W_i(r) - \int_0^1 W_i(s)ds\right)^2 dr < \epsilon\right) = O(\exp\{-C\epsilon^{-1}\}).$$

To bound the matrix norm $\|Q_i\|$, we notice that the supremum of a Brownian motion is sub-Gaussian by the reflection principle (see, for example, Çinlar 2011, Chapter VIII, 1.21), and therefore the entries of Q_i are at most sub-exponential, which leads to the following:

$$Prob\left(\|Q_i\| > \frac{1}{\epsilon}\right) = O(\exp\{-C\epsilon^{-1}\}).$$

As a result, Condition 4(ii) holds with $\mathfrak{c}_4 = 2$.

Finally, we verify Condition 4(iii) by establishing a uniform distributional approximation result. We can write

$$Q_{iT} = \sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} x_{i,t-h}x_{i,t-h}' & \frac{1}{T}\sum_{t=1}^{T} x_{i,t-h} & \frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{i,t-h}x_{i,t-h} \\ \frac{1}{T}\sum_{t=1}^{T} x_{i,t-h}' & 1 & \frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{it} \\ \frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{i,t-h}x_{i,t-h}' & \frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{i,t-h} & \frac{1}{T^{2}}\sum_{t=1}^{T} y_{i,t-h}^2 \end{bmatrix},$$

where $x_{i,t-h} = (u_{i,t-h}, u_{i,t-h-1}, \ldots, u_{i,t-h-p+2})$. We will only illustrate convergence of the bottom-right 2×2 block of Q_{iT}, as the other entries converge in probability, which can be demonstrated by standard variance calculation.

To be specific, we will show that under the condition $N\log^{12}(N)/T \to 0$, the following hold:

$$Prob\left(\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-h}}{\sqrt{T}} - \int_0^1 \tilde{W}_i(t)dt\right| > \frac{1}{\log^3(N)}\right) = o(1) \tag{A-19}$$

$$Prob\left(\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{i,t-h}}{\sqrt{T}}\right)^2 - \int_0^1 \tilde{W}_i(t)^2dt\right| > \frac{1}{\log^{9/4}(N)}\right) = o(1), \tag{A-20}$$

where $\tilde{W}_i(t)$ denotes a standard Brownian motion scaled by $\psi_i(1)$. Since $\log^3(N) > \log^{9/4}(N)$, we conclude that if we set

$$\mathfrak{c}_6(T,N) \propto \frac{1}{\log^{9/4}(N)} = o\left(\frac{1}{\log^{\mathfrak{c}_4}(N)}\right) \text{ with } \mathfrak{c}_4 = 2,$$

then Condition 4(iii) would hold.

We first prove (A-19). Write

$$\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-h}}{\sqrt{T}} - \int_0^1 \tilde{W}_i(t)dt\right| \le \underbrace{\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-h}}{\sqrt{T}} - \frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)\right|}_{\text{(I)}} + \underbrace{\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T) - \int_0^1 \tilde{W}_i(t)dt\right|}_{\text{(II)}}.$$

The second term on the right-hand side is simply the approximation error of the Riemann sum of the Brownian path, which satisfies

$$\text{(II)} \le \max_{i\in\mathcal{N}_1} \frac{1}{T}\sum_{t=1}^{T} \sup_{\frac{t}{T}\le s<\frac{t+1}{T}} |\tilde{W}_i(t/T)-\tilde{W}_i(s)|.$$

To provide a probabilistic bound, we notice that the supremum $\sqrt{T}\sup_{\frac{t}{T}\le s<\frac{t+1}{T}} |\tilde{W}_i(t/T)-\tilde{W}_i(s)|$ is sub-Gaussian by the reflection principal. Therefore, by a maximal inequality argument, one has

$$Prob\left(\text{(II)} > \frac{\log(N)}{\sqrt{T}}\right) = o(1).$$

Now consider (I). We will employ the decomposition in (A-18) to provide a nonasymptotic distributional approximation to y_{it}. To start, our assumption implies that $\tilde{\psi}_{il}$ are absolutely summable, from which we conclude that $\tilde{\epsilon}_{it}$ have bounded fourth moments. As a result,

$$\max_{i\in\mathcal{N}_1}\max_{1\le t\le T}\left|\tilde{\epsilon}_{i,t-h}\right| = O_p((NT)^{1/4}).$$

Next, our nonasymptotic distributional approximation builds on Theorem 4 of Komlós et al. (1976) with a union bound: setting $x=\sqrt{T}/\log^3(N)$ and $H(x)=x^4$ in their theorem, we have

$$Prob\left(\max_{i\in\mathcal{N}_1}\max_{1\le t\le T}\left|y_{i,t-h}-\tilde{W}_i(t)\right| > \frac{\sqrt{T}}{\log^3(N)}\right) = O\left(\frac{N\log^{12}(N)}{T}\right)+o(1)=o(1), \quad \text{(A-21)}$$

which implies

$$Prob\left(\text{(I)} > \frac{1}{\log^3(N)}\right) = o(1).$$

The conclusion (A-19) then follows from the rates we established for (I) and (II) as $\log(N)/\sqrt{T}$ is negligible compared to $\log^{-3}(N)$.

We now show (A-20). We first write

$$\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{i,t-h}}{\sqrt{T}}\right)^2-\int_0^1 \tilde{W}_i(t)^2dt\right| \le \left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{i,t-h}}{\sqrt{T}}\right)^2-\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2\right| + \left|\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2-\int_0^1 \tilde{W}_i(t)^2dt\right|.$$

Next, we further bound the first term on the right-hand side by

$$\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{i,t-h}}{\sqrt{T}}\right)^2-\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2\right|\le\frac{2}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}}-\tilde{W}_i(t/T)\right|\tilde{W}_i(t/T)+\frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}}-\tilde{W}_i(t/T)\right|^2$$

$$\le 2\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}}-\tilde{W}_i(t/T)\right|^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2}+\frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}}-\tilde{W}_i(t/T)\right|^2.$$

Therefore,

$$\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{i,t-h}}{\sqrt{T}}\right)^2-\int_0^1\tilde{W}_i(t)^2dt\right|\le 2\underbrace{\sqrt{\max_{i\in\mathcal{N}_1}\frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}}-\tilde{W}_i(t/T)\right|^2}}_{\text{(III)}}\underbrace{\sqrt{\max_{i\in\mathcal{N}_1}\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2}}_{\text{(IV)}}$$

$$+\underbrace{\max_{i\in\mathcal{N}_1}\frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}}-\tilde{W}_i(t/T)\right|^2}_{\text{(V)}}+\underbrace{\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2-\int_0^1\tilde{W}_i(t)^2dt\right|}_{\text{(VI)}}.$$

Clearly from (A-21),

$$Prob\left(\text{(III)}>\frac{1}{\log^3(N)}\right)=o(1),\quad Prob\left(\text{(V)}>\frac{1}{\log^6(N)}\right)=o(1).$$

By a maximal inequality for sub-Gaussian random variables, one has

$$Prob\left(\text{(IV)}>\log^{3/4}(N)\right)=o(1).$$

Term (VI) is again the approximation error of a Riemann sum, which is bounded by

$$\text{(VI)}\le\max_{i\in\mathcal{N}_1}\frac{1}{T}\sum_{t=1}^{T}\sup_{\frac{t}{T}\le s<\frac{t+1}{T}}|\tilde{W}_i(t/T)^2-\tilde{W}_i(s)^2|\le\max_{i\in\mathcal{N}_1}\left(\frac{1}{T}\sum_{t=1}^{T}\sup_{\frac{t}{T}\le s<\frac{t+1}{T}}|\tilde{W}_i(t/T)-\tilde{W}_i(s)|\right)\left(2\sup_{1\le s\le T}|\tilde{W}_i(s)|\right),$$

and therefore it satisfies

$$Prob\left(\text{(VI)}>\frac{\log^2(N)}{\sqrt{T}}\right)=o(1).$$

Finally, we note that the rate for the product (III)·(IV) is $\log^{-9/4}(N)$, which dominates the rates for (V) and (VI).

Case 3, unit-root y_{it} with linear time trend

With a time trend, $\Upsilon_{iT}^{-1} = \tilde{\Upsilon}_{iT}^{-1} R_i^{-1}$ with

$$\tilde{\Upsilon}_{iT} = \begin{bmatrix} \sqrt{T} I_p & 0 \\ 0 & T^{3/2} \end{bmatrix}$$

and $\Upsilon_{iT}^{-1} z_{i,t-h} = \tilde{\Upsilon}_{iT}^{-1} \tilde{z}_{i,t-h}$ with $\tilde{z}_{i,t-h} = (u_{i,t-h},\ u_{i,t-h-1},\ \ldots,\ u_{i,t-h-p+2},\ 1,\ y_{i,t-h})$. Here

$$\textstyle\sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1} \xrightarrow{p} \begin{bmatrix} \tilde{Q}_i & 0 & 0 \\ 0 & 1 & \mu_i/2 \\ 0 & \mu_i/2 & \mu_i^2/3 \end{bmatrix} := Q_i$$

To verify Condition 4(i), it suffices to compute $E((\sqrt{T} T^{-3/2})^2 y_{i,t-h}^2)$, which is uniformly bounded in i and t, and

$$\begin{aligned} E\left(\left| \frac{1}{T^{3/2}} \sum_{t=1}^{T} t \left(\sum_{l=1}^{\infty} \psi_{il} \eta_{i,t-l} \right) \right|^2 \right) &= \frac{1}{T^3} \sum_{t=1}^{T} \sum_{t'=1}^{T} t t' cov(u_{it}, u_{it'}) \\ &\leq \frac{1}{T^2} \sum_{t=1}^{T} t \sum_{t'=1}^{T} cov(u_{it}, u_{it'}) \leq C \frac{1}{T^2} \sum_{t=1}^{T} t \end{aligned}$$

for some constant C, making the above bounded.

Condition 4(ii) is satisfied by Assumption 5, and we take $\mathfrak{c}_4 = 1/2$.

Finally we check Condition 4(iii). The upper diagonal block of Q_{iT} converges to $\tilde{Q}_i$, which has been discussed in Case 1 in this proof. Take any stationary component $u_{i,t-l}$ of $x_{i,t-h}$, then we can consider

$$E\left(\left| \frac{1}{T^2} \sum_{t=1}^{T} y_{i,t-h} u_{i,t-l} \right|^2 \right) \leq 2E\left(\left| \frac{1}{T^2} \sum_{t=1}^{T} (y_{i,t-h} - \mu_i t) u_{i,t-l} \right|^2 \right) + 2\mu_i^2 E\left(\left| \frac{1}{T^2} \sum_{t=1}^{T} t u_{i,t-l} \right|^2 \right).$$

The first term on the right-hand side is of order T^{-2} by our discussion in Case 2, and the second term is of order T^{-1} by Condition 4(i), which we verified earlier. Therefore, $T^{-2} \sum_{t=1}^{T} y_{i,t-h} u_{i,t-l} = O_p(T^{-1/2})$, where the bound is uniform for $i \in \mathcal{N}_2$. This establishes the convergence of too-right and bottom-left entries of Q_{iT}.

Now consider $T^{-2}\sum_{t=1}^{T} y_{i,t-h}$. We write it as the sum $T^{-2}\sum_{t=1}^{T}(y_{i,t-h} - \mu_i t)$ and $T^{-2}\sum_{t=1}^{T}\mu_i t$. Clearly the former is of order $T^{-1/2}$ by standard variance calculation.

Finally, we consider

$$\frac{1}{T^3}\sum_{t=1}^{T} y_{i,t-h}^2 = \frac{1}{T^3}\sum_{t=1}^{T}\Big((y_{i,t-h} - \mu_i t) + \mu_i t\Big)^2 = \frac{1}{T^3}\sum_{t=1}^{T}(y_{i,t-h} - \mu_i t)^2 + \frac{2\mu_i}{T^3}\sum_{t=1}^{T}(y_{i,t-h} - \mu_i t)t + \frac{\mu_i^2}{T^3}\sum_{t=1}^{T} t^2.$$

The first term on the right-hand side is of order T^{-1} by mean calculation. The second term has order $T^{-1/2}$ by variance calculation.

B Data appendix

Table B1: R^2 for each variable explained by first and second principal components, 1962:3 to 2023:6

Group 1. Output and income

Index	FRED	Description	PC1	PC1&2
1	RPI	Real Personal Income	0.23	0.40
2	W875RX1	Real personal income ex transfer receipts	0.61	0.75
6	INDPRO	IP Index	0.77	0.85
7	IPFPNSS	IP: Final Products and Nonindustrial Supplies	0.81	0.88
8	IPFINAL	IP: Final Products (Market Group)	0.78	0.82
9	IPCONGD	IP: Consumer Goods	0.52	0.83
10	IPDCONGD	IP: Durable Consumer Goods	0.49	0.82
11	IPNCONGD	IP: Nondurable Consumer Goods	0.41	0.55
12	IPBUSEQ	IP: Business Equipment	0.71	0.71
13	IPMAT	IP: Materials	0.65	0.73
14	IPDMAT	IP: Durable Materials	0.64	0.75
15	IPNMAT	IP: Nondurable Materials	0.62	0.69
16	IPMANSICS	IP: Manufacturing (SIC)	0.77	0.87
17	IPB51222S	IP: Residential Utilities	0.02	0.03
18	IPFUELS	IP: Fuels	0.07	0.07
19	CUMFNS	Capacity Utilization: Manufacturing	0.68	0.73
		Median	0.63	0.74

Table B1 (continued)

Group 2. Labor market

Index	FRED	Description	PC1	PC1&2
20	HWI	Help-Wanted Index for United States	0.55	0.56
21	HWIURATIO	Ratio of Help Wanted/No. Unemployed	0.55	0.55
22	CLF16OV	Civilian Labor Force	0.25	0.34
23	CE16OV	Civilian Employment	0.75	0.75
24	UNRATE	Civilian Unemployment Rate	0.69	0.71
25	UEMPMEAN	Average Duration of Unemployment (Weeks)	0.24	0.24
26	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	0.38	0.46
27	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	0.65	0.68
28	UEMP15OV	Civilians Unemployed - 15 Weeks and Over	0.66	0.66
29	UEMP15T26	Civilians Unemployed for 15-26 Weeks	0.65	0.66
30	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	0.59	0.59
31	CLAIMSx	Initial Claims	0.42	0.45
32	PAYEMS	All Employees: Total nonfarm	0.81	0.81
33	USGOOD	All Employees: Goods-Producing Industries	0.85	0.85
34	CES1021000001	All Employees: Mining and Logging: Mining	0.03	0.41
35	USCONS	All Employees: Construction	0.67	0.74
36	MANEMP	All Employees: Manufacturing	0.74	0.74
37	DMANEMP	All Employees: Durable goods	0.77	0.77
38	NDMANEMP	All Employees: Nondurable goods	0.52	0.53
39	SRVPRD	All Employees: Service-Providing Industries	0.67	0.68
40	USTPU	All Employees: Trade, Transportation and Utilities	0.80	0.80
41	USWTRADE	All Employees: Wholesale Trade	0.74	0.80
42	USTRADE	All Employees: Retail Trade	0.67	0.68
43	USFIRE	All Employees: Financial Activities	0.42	0.43
44	USGOVT	All Employees: Government	0.08	0.09
45	CES0600000007	Avg Weekly Hours : Goods-Producing	0.33	0.48
46	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	0.38	0.62
47	AWHMAN	Avg Weekly Hours : Manufacturing	0.31	0.53
115	CES0600000008	Avg Hourly Earnings : Goods-Producing	0.04	0.42
116	CES2000000008	Avg Hourly Earnings : Construction	0.00	0.34
117	CES3000000008	Avg Hourly Earnings : Manufacturing	0.02	0.33
		Median	0.55	0.59

Group 3. Housing

Index	FRED	Description	PC1	PC1&2
48	HOUST	Housing Starts: Total New Privately Owned	0.14	0.37
49	HOUSTNE	Housing Starts, Northeast	0.17	0.39
50	HOUSTMW	Housing Starts, Midwest	0.11	0.42
51	HOUSTS	Housing Starts, South	0.12	0.28
52	HOUSTW	Housing Starts, West	0.11	0.26
53	PERMIT	New Private Housing Permits (SAAR)	0.11	0.35
54	PERMITNE	New Private Housing Permits, Northeast (SAAR)	0.13	0.43
55	PERMITMW	New Private Housing Permits, Midwest (SAAR)	0.10	0.47
56	PERMITS	New Private Housing Permits, South (SAAR)	0.08	0.30
57	PERMITW	New Private Housing Permits, West (SAAR)	0.09	0.24
		Median	0.11	0.36

Table B1 (continued)

Group 4. Consumption, orders, and inventories

Index	FRED	Description	PC1	PC1&2
3	DPCERA3M086SBEA	Real personal consumption expenditures	0.54	0.77
4	CMRMTSPLx	Real Manu. and Trade Industries Sales	0.73	0.89
5	RETAILx	Retail and Food Services Sales	0.46	0.50
58	AMDMNOx	New Orders for Durable Goods	0.69	0.69
59	AMDMUOx	Unfilled Orders for Durable Goods	0.26	0.40
60	BUSINVx	Total Business Inventories	0.27	0.75
61	ISRATIOx	Total Business: Inventories to Sales Ratio	0.22	0.27
		Median	0.46	0.69

Group 5. Money and credit

Index	FRED	Description	PC1	PC1&2
62	M1SL	M1 Money Stock	0.02	0.02
63	M2SL	M2 Money Stock	0.03	0.05
64	M2REAL	Real M2 Money Stock	0.03	0.45
65	BOGMBASE	Monetary Base	0.20	0.20
66	TOTRESNS	Total Reserves of Depository Institutions	0.28	0.28
67	NONBORRES	Reserves Of Depository Institutions	0.00	0.00
68	BUSLOANS	Commercial and Industrial Loans	0.10	0.17
69	REALLN	Real Estate Loans at All Commercial Banks	0.23	0.23
70	NONREVSL	Total Nonrevolving Credit	0.29	0.30
71	CONSPI	Nonrevolving consumer credit to Personal Income	0.09	0.17
118	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	0.03	0.03
119	DTCTHFNM	Total Consumer Loans and Leases Outstanding	0.21	0.22
120	INVEST	Securities in Bank Credit at All Commercial Banks	0.02	0.04
		Median	0.09	0.17

Group 6. Interest and exchange rates

Index	FRED	Description	PC1	PC1&2
76	FEDFUNDS	Effective Federal Funds Rate	0.34	0.68
77	TB3MS	3-Month Treasury Bill:	0.37	0.69
78	TB6MS	6-Month Treasury Bill:	0.38	0.71
79	GS1	1-Year Treasury Rate	0.36	0.71
80	GS5	5-Year Treasury Rate	0.16	0.63
81	GS10	10-Year Treasury Rate	0.08	0.59
82	AAA	Moody’s Seasoned Aaa Corporate Bond Yield	0.02	0.60
83	BAA	Moody’s Seasoned Baa Corporate Bond Yield	0.00	0.61
84	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	0.06	0.31
85	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS	0.04	0.27
86	T1YFFM	1-Year Treasury C Minus FEDFUNDS	0.00	0.16
87	T5YFFM	5-Year Treasury C Minus FEDFUNDS	0.12	0.34
88	T10YFFM	10-Year Treasury C Minus FEDFUNDS	0.21	0.41
89	AAAFFM	Moody’s Aaa Corporate Bond Minus FEDFUNDS	0.33	0.49
90	BAAFFM	Moody’s Baa Corporate Bond Minus FEDFUNDS	0.40	0.49
91	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	0.00	0.01
92	EXJPUSx	Japan / U.S. Foreign Exchange Rate	0.00	0.03
93	EXUSUKx	U.S. / U.K. Foreign Exchange Rate	0.07	0.08
94	EXCAUSx	Canada / U.S. Foreign Exchange Rate	0.00	0.04
		Median	0.08	0.49

Table B1 (concluded)

Group 7. Prices

Index	FRED	Description	PC1	PC1&2
95	WPSFD49207	PPI: Finished Goods	0.07	0.74
96	WPSFD49502	PPI: Finished Consumer Goods	0.07	0.72
97	WPSID61	PPI: Intermediate Materials	0.06	0.65
98	WPSID62	PPI: Crude Materials	0.11	0.44
99	OILPRICEx	Crude Oil, spliced WTI and Cushing	0.02	0.50
100	PPICMM	PPI: Metals and metal products:	0.17	0.36
101	CPIAUCSL	CPI : All Items	0.09	0.82
102	CPIAPPSL	CPI : Apparel	0.04	0.40
103	CPITRNSL	CPI : Transportation	0.04	0.56
104	CPIMEDSL	CPI : Medical Care	0.11	0.41
105	CUSR0000SAC	CPI : Commodities	0.08	0.73
106	CUSR0000SAD	CPI : Durables	0.00	0.33
107	CUSR0000SAS	CPI : Services	0.01	0.67
108	CPIULFSL	CPI : All Items Less Food	0.04	0.77
109	CUSR0000SA0L2	CPI : All items less shelter	0.07	0.78
110	CUSR0000SA0L5	CPI : All items less medical care	0.10	0.82
111	PCEPI	Personal Cons. Expend.: Chain Index	0.08	0.76
112	DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	0.00	0.37
113	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	0.06	0.78
114	DSERRG3M086SBEA	Personal Cons. Exp: Services	0.04	0.66
		Median	0.06	0.66

Group 8. Stock market

Index	FRED	Description	PC1	PC1&2
72	S&P 500	S&P's Common Stock Price Index: Composite	0.22	0.33
73	S&P: indust	S&P's Common Stock Price Index: Industrials	0.18	0.28
74	S&P div yield	S&P's Composite Common Stock: Dividend Yield	0.05	0.40
75	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio	0.11	0.44
		Median	0.15	0.36
		Overall median	0.19	0.50

Notes to Table B1. Index refers to the index number of the variable in our database. FRED refers to variable name in the FRED database. PC1 is the fraction of the variance of the cyclical component of that variable that is explained by the first principal component. PC1&2 is the fraction of the variance of the cyclical component of that variable that is explained by the first and second principal components combined.