Brunel University London
Department of Economics and Finance

Working Paper No. 2419

Economics and Finance Working Paper Series

Guglielmo Maria Caporale, Luis Alberiko Gil-Alana
and Pedro José Piqueras Martínez

Dynamic Factor Models and Fractional Integration – With an Application to US Real Economic Activity

November 2024

http://www.brunel.ac.uk/economics

Dynamic Factor Models and Fractional Integration – With an Application to US Real Economic Activity

Guglielmo Maria Caporale
Brunel University of London, London, UK

Luis Alberiko Gil-Alana
University of Navarra, Pamplona, Spain and Universidad Francisco de Vitoria, Madrid, Spain

Pedro José Piqueras Martínez
University of Navarra, Pamplona, Spain

November 2024

Abstract

This paper makes a twofold contribution, First, it develops the dynamic factor model of Barigozzi et al. (2016) by allowing for fractional integration instead of imposing the classical dichotomy between I(0) stationary and I(1) non-stationary series. This more general setup provides valuable information on the degree of persistence and mean-reverting properties of the series. Second, the proposed framework is used to analyse five annual US Real Economic Activity series (Employees, Energy, Industrial Production, Manufacturing, Personal Income) over the period from 1967 to 2019 in order to shed light on their degree of persistence and cyclical behaviour. The results indicate that economic activity in the US is highly persistent and is also characterised by cycles with a periodicity of 6 years and 8 months.

Keywords: Fractional Integration; Dynamic Factor Models; persistence; Business Cycle, economic activity; Kalman filter; State-Space models

JEL Classification: C22, E32

Corresponding author: Professor Guglielmo Maria Caporale, Department of Economics and Finance, Brunel University of London, Uxbridge, UB8 3PH, UK. Email: Guglielmo-Maria.Caporale@brunel.ac.uk; https://orcid.org/0000-0002-0144-4135

1. Introduction

In recent decades, dynamic factor models have gained increasing popularity owing to their ease of interpretation and their ability to avoid the curse of dimensionality (Barigozzi et al., 2016), and are widely used by practitioners for prediction purposes, to create indices of economic activity and inflation (Stock and Watson, 1988), and to capture regime changes (Hamilton, 1989 and 1994). Such models were initially specified assuming stationarity (Stock and Watson, 2002; Bai and Ng, 2002; Forni et al., 2005; Luciani, 2014). Since most macroeconomic variables in fact do not appear to be stationary, first differences have been used in empirical applications to remove non-stationarity from the series. However, this approach by construction implies that shocks to the variables in the system will have permanent effects, which is a restrictive assumption to make. For that reason, Barigozzi et al. (2016) introduced more general Non-Stationary Dynamic Factor models for Large Datasets that explicitly address the presence of unit roots in the data.

Their work in the time domain has been complemented by the contributions of Bai and Ng (2002 and 2007) and Bai (2004) in a panel context; specifically, the former have proposed methods to test for unit roots in panel dynamic factor models, whilst the approach taken by the latter requires the assumption of stationary idiosyncratic components. In addition, Banerjee et al. (2014) set up a model where cointegration between the common factors and the data, as well as stationarity of the idiosyncratic components, are assumed.

The classical dichotomy between I(1) (unit root or difference-stationary, also called stochastic-trend) models or I(0) trend-stationary became very popular after the influential paper by Nelson and Plosser (1982) based on this approach. Modelling trends correctly is obviously crucial for economic analysis: both removing a (typically linear) deterministic trend from time series that are in fact integrated, and incorrectly differencing can result in spurious behaviour of the series (Chan et al., 1977; Nelson and Kang, 1981, 1984; Durlauf and Phillips, 1988). Early studies had mainly used the Box and Jenkins (1970) approach by estimating autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) models and deterministic trends. However, in their seminal study Nelson and Plosser (1982) applied the newly developed Dickey and Fuller (1979) unit root tests and provided evidence of unit roots in 14 US macroeconomic series over a long-time span. Stock (1991) then highlighted the inadequacy of reporting only test outcomes or point estimates and showed that in the case of the Nelson and Plosser (1982) data set confidence intervals were generally wide and included unit values for the largest autoregressive root ($\rho = 1$) of all series except unemployment and bond yield, but also values significantly different from one.

Subsequently, Campbell and Mankiw (1987) and Cochrane (1988) examined the persistence of macroeconomic series. The former used ARIMA models and non-parametric spectral methods, and concluded that shocks to US GNP are mostly permanent, consistently with the Nelson and Plosser's (1982) findings based on stochastic differencing. Cochrane (1988) obtained different results using a non-parametric variance ratio statistic and other measures based on the spectral density at zero frequency, although such measures might not accurately identify the magnitude of the permanent component unless it follows a random walk (Quah, 1992).

Despite their wide use, standard unit root tests (Dickey & Fuller, 1979; Phillips and Perron, 1988; Elliot et al., 1996; etc.) have been shown to have very low power (see DeJong et al., 1992; Ng and Perron, 2001; Leybourne and McCabe, 1994). For this reason, fractional integration models have been developed in recent decades as an alternative. This approach offers much greater flexibility in modelling low-frequency dynamics that cannot be adequately captured by the Box-Jenkins methodology

(Robinson, 1994; Gil-Alaña and Robinson, 1997); in particular, it allows the differencing parameter to take any real value, including fractional ones (as opposed to only integer ones, as in the classical approach); as a result, a much wider range of stochastic processes can be modelled, and valuable information obtained on persistence and mean reversion. For instance, Gil-Alana and Robinson (1997) used Robinson's (1994) tests on an extended version of the Nelson and Plosser's (1982) data set and obtained mixed results, with the consumer price index and money stock appearing to be the most nonstationary series, and industrial production and unemployment rate being the closest to stationarity; they also showed that the findings are sensitive to the model chosen for the disturbances (e.g., Bloomfield, 1973).

Given the limitations of setups which only allow for non-stationarity in the form of unit roots, this paper aims to go further than Barigozzi et al. (2016) by developing a dynamic factor model which incorporates fractional integration for the analysis of hidden variables. The proposed framework is then used to analyse the stochastic behaviour of US real economic activity. This is important to assess the empirical relevance of different macroecnomic theories and the need for stabilisation policies.

The layout of the paper is as follows. Section 2 reviews standard dynamic factor models and their estimation methods, and then introduces the concept of fractional integration. Section 3 presents the proposed framework which incorporates fractional integration into a dynamic factor model. Section 4 discusses the empirical application to five US Real Economic Activity series. Section 5 offers some concluding remarks.

2. A Review of the Existing Models

2.1 Dynamic Factor Models

The original Stock-Watson's (1988) dynamic factor model decomposes the dynamics of a set of n time series into a common factor and an idiosyncratic part. With series in first differences and modelled as second-order autoregressive Gaussian processes $AR(2)$ (Kim & Halbert, 2000) one obtains the following specification:

$$y_{it} = \gamma f_t + e_{it} \tag{1}$$

$$e_{it} = \psi_{i1} e_{i,t-1} + \psi_{i2} e_{i,t-2} + \varepsilon_{it} \tag{2}$$

$$f_t = \phi_1 f_{t-1} + \phi_2 f_{t-2} + u_t \tag{3}$$

with $i = \{1, \ldots n\}$; $u_t \sim N(0,1)$ and $\varepsilon_i \sim N(0, \sigma_i)$.

The above model can be expressed in a state-space form as:

$$y_t = H \cdot h_t + w_t \tag{4}$$

$$h_t = F \cdot h_{t-1} + v_t \tag{5}$$

where A, H and F are parameter matrices, H is the measurement matrix and F is the transition matrix containing the parameters that determine the dynamics of the system.

The dimensions of the matrices are:

$$F = (r \times r); A' = (n \times k); H' = (n \times r),$$

whilst the other components are vectors:

$$y_t = (n \times 1) \ \wedge h_t = (r \times 1) \wedge x_t = (k \times 1).$$

The model can be estimated by maximum likelihood through the application of the Kalman filter or by Bayesian methods using *Gibbs Sampling* with the *Carter-Kohn* algorithm (Kim and Halbert, 2017; Blake and Mumtaz, 2017). We follow the Bayesian approach because it has the advantage of providing estimates of the complete distribution of both the parameters and the underlying variables.

For ease of computation, following Kim and Halbert (2017), we modify the state-space representation isolating the disturbance of the dynamics of idiosyncratic terms as follows:

$$e_{it} = \psi_{i1} e_{i,t-1} + \psi_{i2} e_{i,t-2} + \epsilon_{it}$$

$$e_{it}[1 - \psi_{i1} L - \psi_{i2} L^2] = \epsilon_{it}\ , \tag{6}$$

which leads to the following state-space representation:

$$y_{i,t} = \gamma_i f_t + e_{i,t}$$

$$y_{i,t}(1 - \psi_{i1} L - \psi_{i2} L^2) = \left(\gamma_i f_t + e_{i,t}\right)(1 - \psi_{i1} L - \psi_{i2} L^2)$$

$$y_{i,t}^* = \gamma_i f_t - \gamma_i \psi_{i1} f_{t-1} - \gamma_i \psi_{i2} f_{t-2} + \epsilon_{it} \tag{7}$$

The measurement equation is then given by:

$$\begin{bmatrix} y_{1,t}^* \\ y_{2,t}^* \\ y_{3,t}^* \\ y_{4,t}^* \end{bmatrix} = \begin{bmatrix} \gamma_1 & -\gamma_1 \psi_{11} & -\gamma_1 \psi_{12} \\ \gamma_2 & -\gamma_2 \psi_{21} & -\gamma_2 \psi_{22} \\ \gamma_3 & -\gamma_3 \psi_{31} & -\gamma_3 \psi_{32} \\ \gamma_4 & -\gamma_4 \psi_{41} & -\gamma_4 \psi_{42} \end{bmatrix} * \begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \end{bmatrix}$$

and the transition equation can be written as:

$$\begin{bmatrix} c_t \\ c_{t-1} \\ c_{t-2} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} c_{t-1} \\ c_{t-2} \\ c_{t-3} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \\ 0 \end{bmatrix}.$$

2.2 Fractional Integration

An $I(0)$ process denoted as $u_t, t = 0, \pm 1 \ldots$ is a covariance stationary one with positive and finite spectral density at the zero frequency. For instance, it could be a white noise, $u_t \sim N(0, \sigma)$, though one can also allow for weak autocorrelation of the Auto Regressive Moving Average (ARMA)-form.

An $I(d)$ process denoted $x_t, t = 0, \pm 1 \ldots$ is defined as:

$$(1 - L)^d x_t = u_t, t = 1,2, \ldots \tag{1}$$

$$x_t = 0, t \leq 0\ , \tag{2}$$

where L is the lag operator. $x_t \cdot L = x_{t-1}$.

A covariance stationary process $\{x_t, t = 0, \pm 1, \ldots\}$ with mean μ is said to exhibit long memory if the sum of its autocovariances, i.e., $\gamma(u) = E[(x_t - \mu)(x_{t+u} - \mu)]$ is infinite:

$$\sum_{u=-\infty}^{u=\infty} \mid \gamma(u) \mid = \infty, \tag{3}$$

and a typical process satisfying this property is the I(d) one with d > 0. One can define d for all real numbers by using the following expansion (Robinson, 1994):

$$(1-L)^d = 1 + \sum_{j=1}^{\infty} \frac{\Gamma(d+1)(-L)^j}{\Gamma(d-j+1)\Gamma(j+1)} \tag{5}$$

which allows one to specify the model as:

$$(1-L)^d y_t = \mu + \gamma \cdot t + u_t; \qquad t = 1,2, \ldots \tag{6}$$

where y_t is the time series of interest, γ is the coefficient on a deterministic linear trend t, which allows one to test the deterministic against the stochastic approach, and the parameter d provides information about the stochastic behaviour of y_t.

Granger (1980, 1981), Granger and Joyeux (1980), and Hosking (1981) initially proposed these processes after noticing that in the case of many series that appeared to be non-stationary the periodogram of the first differenced data was nearly zero at the zero frequency, which implied over-differentiation. Therefore, they suggested considering fractional values for the differencing parameter d rather than the $I(0)$ and $I(1)$ cases only, which gave rise to fractional integration models.

This framework covers a wide range of specifications, such as (Gil-Alana and Robinson, 1997):

- The classic trend stationary model $I(0)$ if $d = 0$.
- The unit-root case if $d = 1$.
- Anti-persistence if $d < 0$.
- Long memory if d is positive and has a fractional value.

- Covariance stationarity if $0 < d < 0.5$.
- Mean reversion if $d < 1$.
- Explosive and persistent behaviour if $d > 1$.

There are various methods for estimating and testing the differencing parameter d. Some are non-parametric, such as the Hurst exponent and the R/S statistic introduced by Hurst (1951) for assessing long memory. Semi-parametric classical methods include the log-periodogram estimator of d by Geweke and Porter-Hudak (GPH, 1983), which was later refined by Robinson (1995) and Kim and Phillips (2006), among others. Examplesof parametric methods are Sowell's (1992) maximum likelihood estimator and the Robinson test (1994), the latter being the approach used in the present study. It is a testing procedure based on the Lagrange Multiplier (LM) principle for evaluating the null hypothesis $(H_0: d = d_0)$ for any real value (d_0) within an $I(d)$ framework, as specified in equation (1) of this section. It does not require stationarity for its implementation since d_0 is allowed to take values outside the stationary range. Details of the functional form of this test can be found Gil-Alana and Robinson (1997).

3 The Proposed Framework

3.1 Model Specification

Our proposed framework introduces fractional integration into a dynamic factor model. This approach allows us to filter the information contained in the set of n time series to analyse the dynamics of the underlying unobserved factors f_t. Specifically, we consider the following specification:

$$y_{it} = \gamma f_t + e_{it} \tag{1}$$

$$(1 - L)^{d_i} \cdot e_{it} = \varepsilon_{it} \tag{2}$$

$$(1 - L)^{d} \cdot f_t = u_t, \tag{3}$$

with $i = \{1, \dots n\}$ and $u_t \sim N(0,1)$; $\varepsilon_i \sim N(0, \sigma_i)$.

The input series are denoted by y_i and the hidden common factor is by f_t, while γ is the loading parameter for the factor; finally, e_{it} stands for the idiosyncratic part of the series. The main purpose of the model is to estimate the parameters associated with the order d of the lag polynomial in order to analyse the stationarity of the system, in particular of the underlying variable f_t. The lag polynomials can be represented as infinite autoregressive processes:

$$y_{it} = \gamma f_t + e_{it} \tag{4}$$

$$e_{it} = \sum_{j=1}^{\infty} \psi_{ij} e_{i,t-j} + \varepsilon_{it} \tag{5}$$

$$f_t = \sum_{j=1}^{\infty} \phi_j f_{t-j} + u_t. \tag{6}$$

This specification allows us to use the information contained in the time series of interest to examine the dynamics of non-observed underlying factors. We allow for up to 10 lags, which appears to be an appropriate lag length given the fact that d decays at a hyperbolic rate.

3.2 Stationarity Analysis of the Hidden Factor

In the empirical application presented in the next section the hidden factor f_t can be interpreted as an index of the economic activity driving the business cycle (Stock and Watson, 1988), and is filtered from the noise contained in the time series. As already mention, we follow a Bayesian approach to estimate the complete distribution of this variable using the Carter and Kohn's (1994) algorithm.

Once we have fitted f_t we analyse its stationarity. For this purpose we use recursively (considering the interval from 0 to 2, with 0.10 increments) the Robinson (1994) a Lagrange Multiplier test for the null hypothesis (H_0: $d = d_0$) until we find a value of d which does not reject H_o. The chosen value of d is the one for which the test statistic is closest to zero in the range (-1.96, 1.96). Since a linear combination of two $I(1)$ processes will also be $I(1)$ provided that there is no cointegration, f_t will be an $I(1)$ process. By adding 1 to the estimated order of integration d we can obtain the corresponding one for the stationary $I(0)$ counterpart of f_t. Specifically, we consider the following model:

$$f(t) = a + bt + x(t); \tag{7}$$

$$(1 - L)^d x(t) = u(t), \tag{8}$$

where $u(t)$ is a white noise process, and $f(t)$ is the estimated factor.

3.3. Possible Extensions of the Model

This procedure can be generalised to accommodate multiple latent factors f_t, which requires an adjustment to the state-space model, namely an increase in the dimensions of both the measurement and transition equations to capture the interactions between the multiple factors.

Further, more complex processes than the Gaussian white noise or basic autocorrelation ones can also be considered for the error u_t – for instance, non-linear, heteroscedastic, or regime-switching ones. This introduces greater flexibility and is crucial when analysing differenced processes, as it enables the model to capture more complex, real-world dynamics that simple autocorrelation structures may fail to represent adequately (Robinson, 1994).

Possible structural breaks represent an additional significant challenge, as ignoring them may lead to biased parameter estimates and ultimately inaccurate predictions. Appropriate break tests and time-varying parameter models might therefore be required to capture the behaviour of the series.

Finally, incorporating a deterministic trend as an exogenous variable in the state-space framework introduces a non-stationary I(1) hidden factor that captures long-term trends in the data. This factor could be modelled as a combination of f_t and a deterministic trend, allowing for more accurate modelling of the non-stationary components. To model the deterministic trend, a flexible and computationally efficient approach is the one involving the use of Chebyshev polynomials, as proposed by Gil-Alana and Cuestas (2012, 2016). These polynomials provide a parsimonious representation of non-linear trends.

4 Data and Empirical Results

4.1 Data Description and Sources

We select the series for the empirical application following the paper by Stock and Watson (1988) as well as more recent ones estimating the *EURO-STING* model (Camacho and Pérez-Quiros, 2008; Pacce and Pérez-Quirós, 2019), and the *SPAIN-STING* model (Camacho and Pérez-Quiros, 2009; Arencibia *et al.,* 2017*;* Gómez *et al.*, 2024), and also include the variables used to create the Coincident Economic Activity Index for the United States (USPHCI). This set of variables has been shown to produce accurate GDP forecasts by capturing the latent factor representing economic activity.

More specifically, for the analysis we use the following series retrieved from FRED (2024) over the period from 1967 to 2019 (which excludes the Covid-19 pandemic with the resulting structural changes), for a total of 635 observations:

- Industrial Production Index (Index 2012=100): Monthly, Seasonally Adjusted.
- All Employees: Total Nonfarm Payrolls, Thousands of Persons, Monthly, Seasonally Adjusted.
- Real personal income excluding current transfer receipts: Billions of Chained 2009 Dollars, Monthly, Seasonally Adjusted Annual Rate.
- Real Manufacturing and Trade Industries Sales: Millions of Chained 2009 Dollars, Monthly, Seasonally Adjusted.
- Real Energy Consumption, Price Index 1982=100, Monthly, Seasonally Adjusted. Deflated from Personal consumption expenditures: Energy goods and services, Billions of Dollars.

In the first instance, we apply first differences to achieve stationarity. Then Augmented Dickey-Fuller (1979) and ERS (Elliot *et al,* 1996) tests are carried out. The results are reported in both Table 1 and Table 2. It can be seen that all the series have ADF statistics that are highly negative and p-values that are significantly below 0.05. This suggests that all of them are stationary in first differences. Additionally, the series have ERS test statistics that are below the critical value of 1.99. This indicates that the null hypothesis of a unit root (non-stationarity) can be rejected for all the differenced series.

Table 1. ADF test for the Economic Activity Series

Series	ADF statistic	ADF p-value
Employees	-4.59	0.01
Energy	-9.60	0.01
Industrial Production	-6.02	0.01
Manufacturing	-6.12	0.01
Personal Income	-5.82	0.01

Computed using the tseries package on R from Trapletti & Hornik (2020)

Table 2. ERS test for the Economic Activity Series

Series	ERS Statistic	Critical Value
Employees	0.673	1.990
Energy	0.092	1.990
Industrial Production	0.657	1.990
Manufacturing	0.258	1.990
Personal Income	0.102	1.990

Computed using the urca package on R from Pfaff (2008)

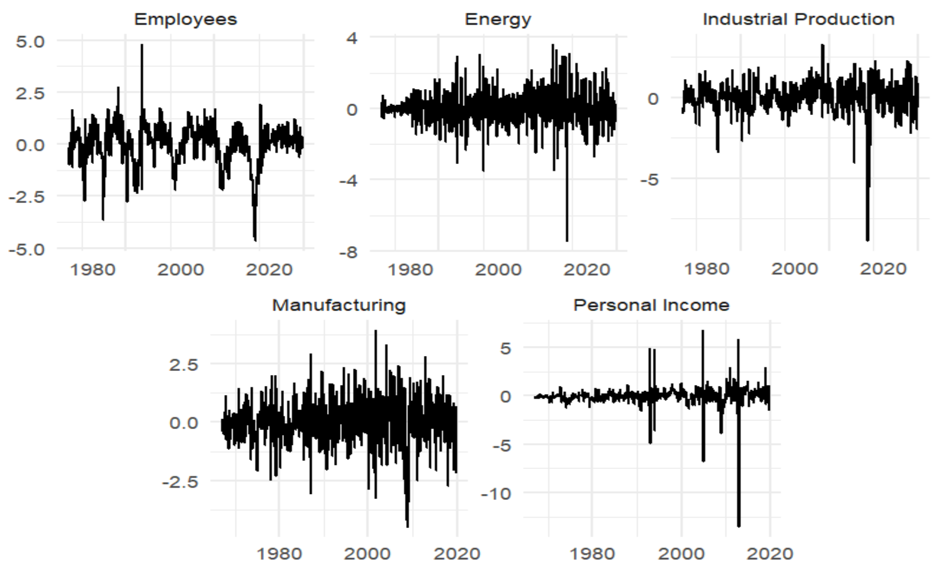
The Shapiro-Wilk (1965) test was then employed to assess the normality of the various economic indicators. The results are displayed in Table 3. The p-values for all economic indicators are 0, which is significantly below the 0.05 threshold. This implies that the null hypothesis of normality is rejected for all series.

Table 3. Saphiro-Wilks test for the Economic Activity Series

Series	Shapiro Wilk	p-value
Employees	0.92	0.00
Energy	0.94	0.00
Industrial Production	0.90	0.00
Manufacturing	0.98	0.00
Personal Income	0.62	0.00

Computed using Royston's (1982) algorithm.

Figure 1. Real Activity Variables



Source: FRED (2024). The series depicted in the graph are seasonally adjusted, first differenced, centred around the mean and scaled by the standard deviation.

Figure 1 shows the time series for all five economic indicators, retrieved from FRED (2024) over the period from 1967 to 2019. This period excludes the Covid-19 pandemic and its resulting structural changes, comprising a total of 635 observations.

Table 4. Statistical Summary of the Real Activity Series

Series	Mean	Median	SD	Min	Max	Q1	Q3	IQR
Employees	136	171	204	-820	1118	50	265	216
Energy	0.09	0.22	5.61	-42	20 -	2.90	2.86	5.76
Industrial Production	0.11	0.15	0.49	-4	1.74	-0.13	0.38	0.51
Manufacturing	1630	1804	7363	-31453	30611	-2662	6356	9017
Personal Income	18	17	57	-761	408	0.60	36	35

Source: FRED (2024). The series are seasonally adjusted and first differenced. Q1 is the first quartile, Q3 is the third quartile and IQR is the interquartile distance. Elaboration by author.

Table 4 reports descriptive statistics for the time series analysed, such as the mean, median, standard deviation, minimum and maximum values, as well as the quantiles and interquartile range (IQR).

4.2 Empirical Results.

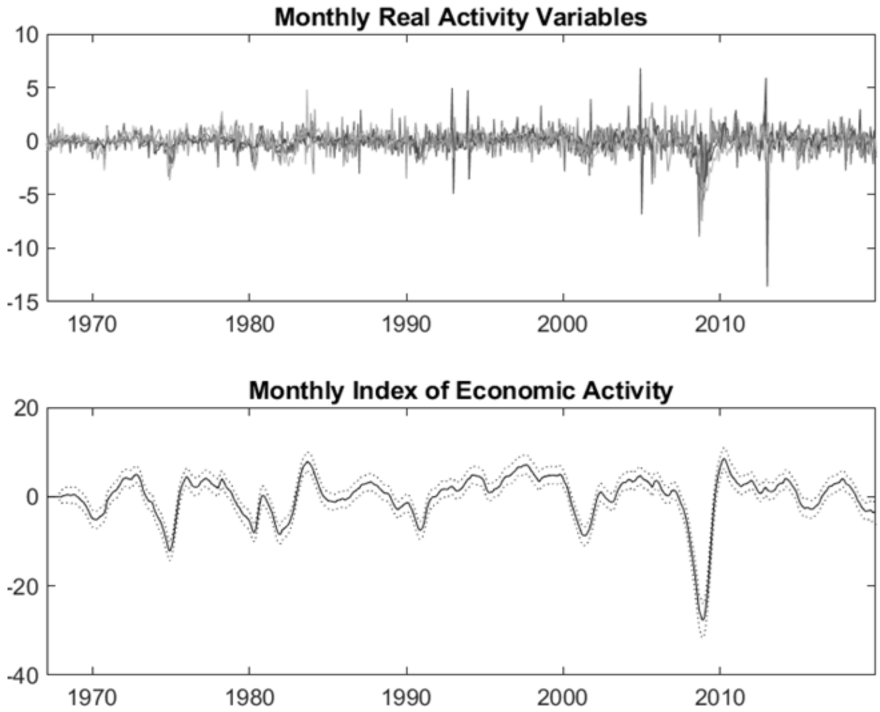
The estimated distribution of the parameters which describe the hidden factor and its relationship with the input variables are reported in Table 5. Figure 2 shows the hidden factor series together with the input series and helps to understand the hidden factor f_t as an Index of Economic Activity. This variable closely follows the trends of the input series, which suggests that it is a good representation of the underlying economic activity.

Table 5. Statistical summary of the parameter's distributions.

Parameter	Q1	Median	Q3	Average	SD
φ1	1.103	1.141	1.178	1.141	0.055
φ2	-0.090	-0.032	0.021	-0.034	0.082
φ3	-0.095	-0.039	0.015	-0.040	0.082
φ4	-0.093	-0.041	0.018	-0.039	0.082
φ5	-0.087	-0.031	0.025	-0.031	0.083
φ6	-0.077	-0.023	0.034	-0.023	0.082
φ7	-0.066	-0.013	0.043	-0.012	0.082
φ8	-0.068	-0.014	0.043	-0.013	0.082
φ9	-0.061	-0.004	0.049	-0.005	0.083
φ10	-0.040	-0.004	0.031	-0.004	0.053
λ1	0.067	0.073	0.081	0.074	0.011
λ2	0.100	0.111	0.123	0.113	0.018
λ3	0.112	0.124	0.136	0.125	0.017
λ4	0.043	0.048	0.053	0.048	0.008
λ5	0.018	0.023	0.028	0.023	0.008
σ1	0.689	0.719	0.751	0.721	0.046
σ2	0.621	0.651	0.682	0.652	0.046
σ3	0.370	0.386	0.402	0.387	0.024
σ4	0.866	0.901	0.937	0.902	0.052
σ5	0.953	0.990	1.029	0.992	0.057

The variables are in the same order as described in the data. The φ parameters are the autoregressive coefficients of the factor, the λ ones are the loadings and the σ ones are the variances of the idiosyncratic disturbance terms. Q stands for quantile and SD for standard deviation.

Figure 2. The Monthly Index of Economic Activity



The monthly index of economic activity shows the median together with the first and third quartile of the factor distribution (dashed). This figure follows Figure 1 in Stock and Watson (1988) and is based on FRED (2024) data.

For the computation of d we first allow for a linear time as is common in the unit roots literature (Bhargava, 1986, Schmidt and Phillips, 1992), such that the model becomes a combination of (7) and (8), i.e.,

$$f(t) = \alpha + \beta t + x(t), \qquad (1-L)^d x(t) = u(t), \qquad (6)$$

where α and β are jointly estimated with d, and $u(t)$ follows a white noise process with zero mean and constant variance.

The Lagrange multiplier test for the differencing parameter of the hidden factor d are carried out using three different model specifications and under the assumption of white noise residuals; the results can be summarised as follows:

(i) In the first case, we include a constant and a linear trend, and thus α and β are estimated together with d. The test provides the following value and confidence interval for the differencing parameter: $d = 2.09\ (2.01, 2.18)$. However, β is non-significant, therefore we remove the linear trend.

(ii) In the second case, we allow for a constant α but not for a linear trend, namely $\beta = 0$. We then obtain the result $d = 2.10\ (2.02, 2.19)$ with $\alpha = 0.930$ statistically significant with a t-value of 4.25.

(iii) In the third case, neither a constant nor a trend are included, i.e. $\alpha = \beta = 0$ a priori. We obtain the same result as in case (ii), namely $d = 2.10\ (2.02, 2.19)$.

Next, we allow for autocorrelation in $u(t)$ and estimate the model using the non-parametric approximation of Bloomfield (1973) for AR structures. The results are now the following:

(i) With a constant and a linear time trend, $d = 1.93$ $(1.72, 2.16)$. However, the linear trend is statistically insignificant.

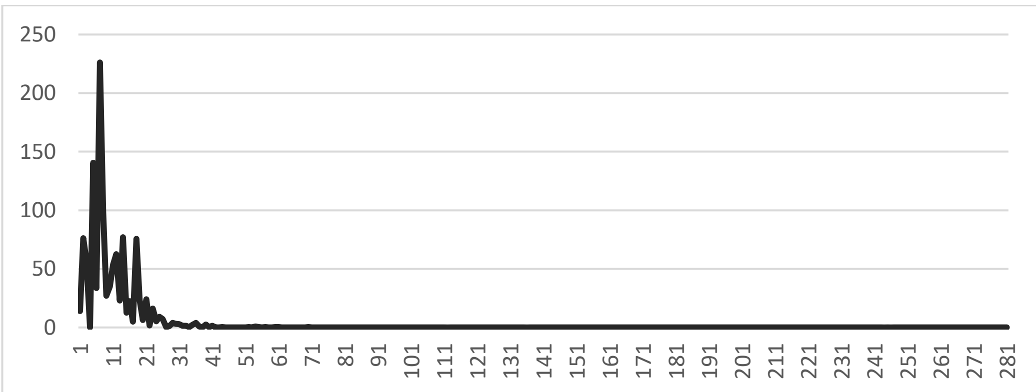
(ii) With a constant but without a trend, $d = 1.94$ $(1.73, 2.16)$. Note that the constant is significant $a = 0.937$ (with a t-value of 4.06).

(iii) Without either a constant or a trend, $d = 1.94$ $(1.73, 2.16)$.

Regardless of the assumption made about the disturbances, the estimated values of d suggest high persistence in the dynamic behaviour of the economic activity.

Finally, the periodogram of $f(t)$ was estimated using the squared coefficients of the Discrete Fourier Transformation applied to the mean of the estimated common factor and scaled by the length of the signal. It can be seen that the biggest value does not correspond to the zero frequency ($j = 1$) but to $j = 7$ instead, which suggests the existence of cycles of periodicity $T/j = 563/7 = 80$ months or 6 years and 8 months.

Figure 2. Periodogram of the Index of Economic Activity.



5. Conclusions

This paper makes a twofold contribution, First, it develops the dynamic factor model of Barigozzi et al. (2016) by allowing for fractional integration instead of imposing the classical dichotomy between I(0) stationary and I(1) non-stationary series. This more general setup is applicable in a variety of contexts and enables one to consider a much wider range of stochastic processes and to obtain valuable information about the dynamics of the series, such as their degree of persistence and mean reversion. Second, the proposed framework is used to analyse the behaviour of five annual US Real Economic Activity series (Employees, Energy, Industrial Production, Manufacturing, Personal Income) over the period from 1967 to 2019 in order to shed light on their persistence and cyclical behaviour. The results indicate that economic activity in the US is highly persistent and is also characterised by cycles with a periodicity of 6 years and 8 months.

Our findings have important policy implications. Specifically, the evidence that shocks have long-lived effects suggests that they originate from the supply side. It is well known that traditional stabilisation policies have an important role to play in smoothing the amplitude of fluctuations associated with the cyclical behaviour of economic activity and generated by demand shocks (Clarida et al., 1999; Woodford, 2003; Blanchard and Riggi, 2013). By contrast, effective policy responses to supply shocks require structural reforms and investment in productivity-enhancing technologies to achieve sustained growth (Kydland and Prescott, 1982). Given the evidence presented above it appears that it is the latter set of policies that are most appropriate in the case of the US.

References

Arencibia Pareja, A., Gómez Loscos, A., de Luis López, M., & Pérez Quirós, G. (2017). *Un modelo de previsión del PIB y de sus componentes de demanda*. Boletín Económico, 4/2017. Banco de España.

Barigozzi, Matteo, Marco Lippi, and Matteo Luciani (2016). *Non-Stationary Dynamic Factor Models for Large Datasets*, Finance and Economics Discussion Series 2016-024. Washington: Board of Governors of the Federal Reserve System

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

Bai, J. and S. Ng (2007). *Determining the number of primitive shocks in factor models*. Journal of Business and Economic Statistics 25, 52–60.

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

Banerjee, A., M. Marcellino, and I. Masten (2014). *Structural FECM: Cointegration in large-scale structural FAVAR models*. Working Paper 9858, CEPR.

Bhargava, A. (1986), On the theory of testing for unit roots in observed time series, Review of Economic Studies 53, 369-384.

Blake, A., & Mumtaz, H. (2000). *Applied Bayesian econometrics for central bankers*, Centre for Central Banking Studies, Bank of England.

Blanchard, O.J. and Riggi, M. (2013). *Why are the 2000s so different from the 1970s? a structural interpretation of changes in the macroeconomic effects of oil prices*. Journal of the European Economic Association, 11: 1032-1052.

Bloomfield, P. (1973). An Exponential Model for the Spectrum of a Scalar Time Series. Biometrika 60(2), 217.

Box, G. and Jenkins, G. (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Camacho, M., & Pérez-Quiros, G. (2008). *Introducing the Euro-STING: Short term indicator of Euro area growth* (Documento de Trabajo No. 0807). Banco de España. Eurosistema.

Camacho, M. & Perez-Quiros, G. (2009). *Spain-STING: Spain short term indicator of growth*. (Documento de Trabajo No. 0912). Banco de España. Eurosistema.

Campbell and N.G. Mankiw, (1987). *Are output fluctuations transitory?*, Quarterly Journal of Economics 102,857-880.

Carter, C. K., & Kohn, R. (1994). *On Gibbs Sampling for State Space Models*. Biometrika, 81(3), 541–553.

Chan, H.K., J.C. Hayya, and K.J. Ord, (1977), *A note on trend removal methods: the case of polynomial trend versus variate differencing*, Econometrica 45, 737-744.

Clarida, R., Gali, J., & Gertler, M. (1999). *The Science of Monetary Policy: A New Keynesian Perspective*. Journal of Economic Literature, 37(4), 1661-1707.

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

Dickey, D. A., & Fuller, W. A. (1979). *Distribution of the Estimators for Autoregressive Time Series With a Unit Root*. Journal of the American Statistical Association, 74(366), 427–431.

DeJong, D.N. and C.H. Whiteman, (1989). *Trends and cycles as unobserved components in US real GNP: A Bayesian perspective*, Proceedings of the American Statistical Association, 63-70.

DeJong, D. N., Nankervis, J. C., Savin, N. E., & Whiteman, C. H. (1992). *The power problems of unit root tests in time series with autoregressive errors*. Journal of Econometrics, 53(1-3), 323-343.

Durlauf, S. and P.C.B. Phillips, (1988). *Trends versus random walks in time series analysis*, Econometrica 56, 1333-1354.

Elliot, G., T.J. Rothenberg and J.H. Stock (1996), Efficient Tests for an Autoregressive Unit Root, Econometrica 64, 4, 813-836.

(FRED) Board of Governors of the Federal Reserve System (US). (2024*). Industrial Production: Total Index [INDPRO]*, retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/INDPRO.

(FRED) Federal Reserve Bank of St. Louis. (2024). *Real Manufacturing and Trade Industries Sales [CMRMTSPL]*, retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CMRMTSPL.

(FRED) U.S. Bureau of Labor Statistics. (2024). *All Employees, Total Nonfarm [PAYEMS]*, retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/PAYEMS.

(FRED) U.S. Bureau of Labor Statistics (2024). *Consumer Price Index for All Urban Consumers: Services Less Energy Services in U.S. City Average [CUSR0000SASLE]*, retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CUSR0000SASLE

(FRED) U.S. Bureau of Economic Analysis (2024). *Real personal income excluding current transfer receipts [W875RX1]*, retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/W875RX1.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). *The Generalized Dynamic Factor Model: One sided estimation and forecasting*. Journal of the American Statistical Association 100, 830–840.

Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009). *Opening the black box: Structural factor models versus structural VARs*. Econometric Theory 25, 1319–1347.

Geweke, J., Porter-Hudak, S. (1983). *The estimation and application of long memory time series models*. Journal of Time Series Analysis, 4 (4), 221-238.

Gil-Alana, L.A. and Robinson, P.M. (1997), *Testing of unit root and other nonstationary hypotheses in macroeconomic time series*, Journal of Econometrics, 80, issue 2, p. 241-268.

Gil-Alana, L. A., & Cuestas, J. C. (2012). *A non-linear approach with long range dependence based on Chebyshev polynomials*. University of Navarra. Retrieved from unav.edu.

Gil-Alana, L. A., & Cuestas, J. C. (2016). *Testing for long memory in the presence of non-linear deterministic trends with Chebyshev polynomials*. *Studies in Nonlinear Dynamics & Econometrics, 20*(1), 57-74. De Gruyter.

Granger, C.W.J. (1980). *Long memory relationships and the aggregation of dynamic models*, Journal of Econometrics 14, 2, 227-238.

Granger, C.W.J. (1981). *Some properties of time series data and their use in econometric model specification*, Journal of Econometrics 16, 121-130.

Granger, C. W. J., Joyeux, R. (1980). *An introduction to long-memory time series models and fractional differencing.* J. Time Series Analysis. 1, 15–29.

Gómez Loscos, A., González Simón, M. Á., & Pacce, M. J. (2024). *Short-term real-time forecasting model for Spanish GDP (Spain-STING): New specification and reassessment of its predictive power* (Documentos Ocasionales No. 2406). Banco de España.

Hamilton, J. D. (1989). *A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle*. Econometrica, 57(2), 357–384.

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

Hosking, J.R. (1981). *Fractional differencing*. Biometrika, 68, 165-176

Hurst, H.E. (1951). *Long-term storage capacity of reservoirs*. *Transactions of the American Society of Civil Engineers*. 116

Kim, C.-J., & Halbert, D. C. R. (2017). *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications*. The MIT Press.

Kim, C.S., Phillips, P.C.B. (2006). *Log Periodogram Regression: The Nonstationary Case*. Cowles Foundation Discussion Paper No. 1587

Kydland, F. E., & Prescott, E. C. (1982). Time to Build and Aggregate Fluctuations. Econometrica, 50(6), 1345–1370.

Leybourne, S. J., & McCabe, B. P. M. (1994). *A consistent test for a unit root.* Journal of Business & Economic Statistics, 12(2), 157-166.

Luciani, M. (2015). *Monetary policy and the housing market: A structural factor analysis*. Journal of Applied Econometrics 30, 199–218.

Nelson, C.R. and H. Kang, (1981). *Spurious periodicity in inappropriately detrended time series*, Econometrica 49,741-751.

Nelson, CR. and H. Kang, (1984). *Pitfalls in the use of time as an explanatory variable in regression*, Journal of Business and Economics Statistics 2,73-82.

Nelson, C.R. and C.I. Plosser, (1982). *Trends and random walks in macroeconomic time series*, Journal of Monetary Economics 10,139-162.

Ng, S., & Perron, P. (2001). *Lag length selection and the construction of unit root tests with good size and power.* Econometrica, 69(6), 1519-1554.

Pacce, M., & Pérez-Quirós, G. (2019). *Predicción en tiempo real del PIB en el área del euro: Recientes mejoras en el modelo Euro-STING*. Boletín Económico, Notas Económicas, 1(2019), 28 de febrero. Banco de España.

Perron, P. (1989). *The great crash, the oil price shock, and the unit root hypothesis*, Econometrica 57, 1361-1401.

Perron, P. (1993). *Trend, unit root and structural change in macroeconomic time series*, Unpublished manuscript (University of Montreal, Montreal).

Pfaff, B. (2008). *Analysis of integrated and cointegrated time series with R* (2nd ed.). Springer.

Phillips, P.C.B., (1991). *To criticize the critics: an objective Bayesian analysis of stochastic trends*, Journal of Applied Econometrics 6,333-364.

Phillips, P.C.B. and P. Perron (1988), Testing for a Unit Root in Time Series Regression, Biometrika 75, 2, 335-346.

Quah, D. (1992). *The relative importance of permanent and transitory components: identification and some theoretical bounds*, Econometrica 60, 107-18.

Robinson, P.M. (1994), *Efficient tests of nonstationary hypotheses*, Journal of the American Statistical Association, 89, 1420-1437.

Robinson, P.M. (1995). *Log periodogram regression of time series with long range dependence*, Annals of Statistics 23, 3, 1048-1072.

Royston, P. (1982). *An extension of Shapiro and Wilk's test for normality to large samples*. Applied Statistics, 31, 115--124. 10.2307/2347973.

Schmidt, P. and P.C.B. Phillips (1992), *LM tests for a unit root in the presence of deterministic trends*, Oxford Bulletin of Economics and Statistics 54(3), 257-287.

Shapiro, S. S.; Wilk, M. B. (1965). *An analysis of variance test for normality (complete samples)*. Biometrika. 52 (3–4): 591–611.

Sowell, F. (1992). *Maximum likelihood estimation of stationary univariate fractionally integrated time series models*. Journal of Econometrics, 53(1-3), 165-188.

Stock, J.H. (1991). *Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series*, Journal of Monetary Economics 28,435-460.

Stock, J-H., (1994). *Deciding between I(1) and I(O)*, Journal of Econometrics 63, 105-131.

Stock, J. H. and M. W. Watson (1991). *A Probability Model of the Coincident Economic Indicators*. In: Moore G, Lahiri K The Leading Economic Indicators: New Approaches and Forecasting Records. Cambridge University Press; pp. 63-90.

Stock, J. H. and M. W. Watson, (1988). *A Probability Model of The Coincident Economic Indicators*, NBER Working Papers 2772, National Bureau of Economic Research, Inc.

Stock, J. H. and M. W. Watson (2002). *Macroeconomic forecasting using diffusion indexes*. Journal of Business and Economic Statistics 20, 147–162.

Stock, J. H. and M. W. Watson (2005). *Implications of dynamic factor models for VAR analysis*. Working Paper 11467, NBER.

Trapletti, A., & Hornik, K. (2020). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-48. https://CRAN.R-project.org/package=tseries

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.

Zivot, E. and D.W.K. Andrews (1992). *Further evidence on the great crash, the oil price shock and the unit root hypothesis*, Journal of Business and Economic Statistics 10,251-270.