

Article

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

Guglielmo Maria Caporale ^{1,*} , Luis Alberiko Gil-Alana ^{2,3}  and Pedro Jose Piqueras Martinez ³

¹ Department of Economics and Finance, Brunel University of London, Kingston Lane, Uxbridge UB8 3PH, UK

² Facultad de Derecho, Empresa y Gobierno, Universidad Francisco de Vitoria, 28223 Madrid, Spain; alana@unav.es

³ Faculty of Economics and NCID, University of Navarra, 31080 Pamplona, Spain; pedrojosepiquerasmartinez@gmail.com

* Correspondence: guglielmo-maria.caporale@brunel.ac.uk

Abstract: This paper makes a twofold contribution. First, it develops the dynamic factor model of 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



Citation: Caporale, Guglielmo Maria, Luis Alberiko Gil-Alana, and Pedro Jose Piqueras Martinez. 2024.

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

Econometrics 12: 39. <https://doi.org/10.3390/econometrics12040039>

Academic Editor: Julien Chevallier

Received: 12 November 2024

Revised: 5 December 2024

Accepted: 17 December 2024

Published: 19 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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, 1994).

Dynamic factor models were introduced by Geweke (1977) in the context of time series, focusing on the decomposition of time series into common and idiosyncratic components. Early applications of such models to macroeconomics can be found in Sargent and Sims (1977), who emphasised their usefulness in economic forecasting. Watson and Engle (1983) proposed methods for estimating dynamic factor models, highlighting their potential for improving economic predictions, and Quah and Sargent (1993) extended the application of dynamic factor models to analyse economic fluctuations and the impact of policy actions. Carter and Kohn (1994) introduced Gibbs sampling for Bayesian inference in linear state space models, simultaneously generating the entire distribution of the hidden factors and coefficients, leveraging the Kalman filter for efficient state generation. Finally, Doz et al. (2012) developed a quasi-maximum likelihood estimation approach, which is particularly effective for large panels, since it allows for more flexible assumptions regarding the cross-sectional and serial correlations of the idiosyncratic components, thereby enhancing its applicability to real-world data, improving the efficiency and accuracy of parameter estimates, and producing robust estimates even in the presence of various forms of misspecification.

Such models were initially specified assuming stationarity (Stock and Watson 2002; Bai and Ng 2002; Forni et al. 2005; Luciani 2015). 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, 2004, 2007) 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) and $I(0)$ trend-stationary models 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) dataset, confidence intervals were generally wide and included unit values for the largest autoregressive root ($p = 1$) 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, consistent with 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).¹

Despite their wide use, standard unit root tests (Dickey and 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-Alana 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 Nelson and Plosser's (1982) dataset and obtained mixed results, with the consumer price index and money stock appearing to be the most non-stationary 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. (2021) 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 macroeconomic theories and the need for stabilisation policies.²

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 and Watson (1988) dynamic factor model, which is also used in Barigozzi et al. (2021), 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 and Halbert 2017) one obtains the following specification:

$$y_{it} = \gamma f_t + e_{it} \quad (1)$$

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

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

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

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

$$y_t = H \cdot h_t + w_t \quad (4)$$

$$h_t = F \cdot h_{t-1} + v_t \quad (5)$$

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 as follows:

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

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:

$$\begin{aligned} e_{it} &= \psi_{i1}e_{i,t-1} + \psi_{i2}e_{i,t-2} + \varepsilon_{it} \\ e_{it} [1 - \psi_{i1}L - \psi_{i2}L^2] &= \varepsilon_{it}, \end{aligned} \quad (6)$$

which leads to the following state-space representation:

$$\begin{aligned} y_{i,t} &= \gamma_i f_t + e_{i,t} \\ y_{i,t} (1 - \psi_{i1}L - \psi_{i2}L^2) &= (\gamma_i f_t + e_{i,t}) (1 - \psi_{i1}L - \psi_{i2}L^2) \\ y_{i,t}^* &= \gamma_i f_t - \gamma_i \psi_{i1} f_{t-1} - \gamma_i \psi_{i2} f_{t-2} + \varepsilon_{it} \end{aligned} \quad (7)$$

The measurement equation is then given by the following:

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

and the transition equation can be written as follows:

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

2.2. Fractional Integration

An $I(0)$ process, denoted as $u_t, t = 0, \pm 1, \dots$, 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 Autoregressive Moving Average (ARMA)-form.

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

$$(1 - L)^d x_t = u_t, t = 1, 2, \dots \tag{8}$$

$$x_t = 0, t \leq 0, \tag{9}$$

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

A covariance stationary process $\{x_t, t = 0, \pm 1, \dots\}$ 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:

$$\sum_{u=-\infty}^{\infty} |\gamma(u)| = \infty, \tag{10}$$

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):

$$(1 - L)^d = 1 + \sum_{j=1}^{\infty} \frac{\Gamma(d + 1)(-L)^j}{\Gamma(d - j + 1)\Gamma(j + 1)} \tag{11}$$

which allows one to specify the model as follows:

$$(1 - L)^d y_t = \mu + \gamma \cdot t + u_t; t = 1, 2, \dots \tag{12}$$

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 \(1983\)](#), which was later refined by [Robinson \(1995\)](#) and [Kim and Phillips \(2006\)](#), among others. Examples of parametric methods are [Sowell's \(1992\)](#) maximum likelihood estimator and the [Robinson \(1994\)](#) test, 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 in [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:

$$y_{it} = \gamma f_t + e_{it} \quad (13)$$

$$(1 - L)^{d_i} \cdot e_{it} = \varepsilon_{it} \quad (14)$$

$$(1 - L)^d \cdot f_t = u_t, \quad (15)$$

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_{it} , 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:

$$y_{it} = \gamma f_t + e_{it} \quad (16)$$

$$e_{it} = \sum_{j=1}^{\infty} \psi_{ij} e_{i,t-j} + \varepsilon_{it} \quad (17)$$

$$f_t = \sum_{j=1}^{\infty} \phi_j f_{t-j} + u_t. \quad (18)$$

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.

An $I(d)$ process allows for fractional differencing, accommodating more complex behaviours than in the case of $I(0)$ or $I(1)$ processes. The parameter (d) can take any real value, leading to various model specifications: trend stationarity, unit roots, anti-persistence, covariance stationarity, mean reversion, and explosive behaviour. Estimation and testing methods for (d) include non-parametric (Hurst exponent, R/S statistic), semi-parametric (log-periodogram estimator), and parametric approaches (maximum likelihood estimator, Robinson test). When using first-differenced series $I(d - 1)$, one aims to determine $d - 1$. Consequently, the corresponding undifferenced series is $I(d - 1 + 1) = I(d)$.

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 mentioned, 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) Lagrange Multiplier test for the null hypothesis ($H_0 : d = d_0$) until we find a value of d which does not reject H_0 . 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:

$$f(t) = a + bt + x(t); \quad (19)$$

$$(1 - L)^d x(t) = u(t), \quad (20)$$

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 Pareja et al. 2017; Gómez Loscos 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 USD, monthly, seasonally adjusted annual rate.
- Real Manufacturing and Trade Industries Sales: millions of chained 2009 USD, monthly, seasonally adjusted.
- Real Energy Consumption: price index 1982 = 100, monthly, seasonally adjusted, deflated from personal consumption expenditures, energy goods and services, billions of USD.

In the first instance, we apply first differences to achieve stationarity. Then Augmented Dickey and Fuller (1979) and ERS (Elliot et al. 1996) tests are carried out. The results are reported in both Tables 1 and 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 and Hornik (2020).

Table 2. ERS test for the Economic Activity Series.

Series	ERS Statistic	Critical Values
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 and 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. Shapiro–Wilk test for the Economic Activity Series.

Series	ERS Statistic	Critical Values
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 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 and comprises a total of 635 observations.

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).

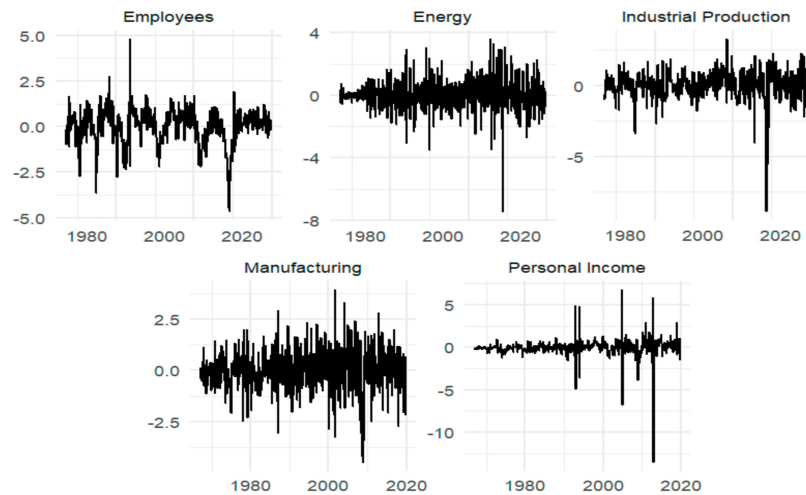


Figure 1. Real activity variables. Source: FRED (2024). The series depicted in the graphs are seasonally adjusted, first differenced, centred around the mean and scaled by the standard deviation.

Table 4. Statistical summary of the real activity series.

Series	Mean	Median	SD	Min	Max	Q1	Q2	IQR
Employees	136	171	204	−820	1118	50	256	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	174	−0.13	0.38	0.51
Manufacturing	1630	1804	7363	−31,453	30,611	−2662	6356	9017
Personal Income	18	17	57	−761	4.08	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.

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.

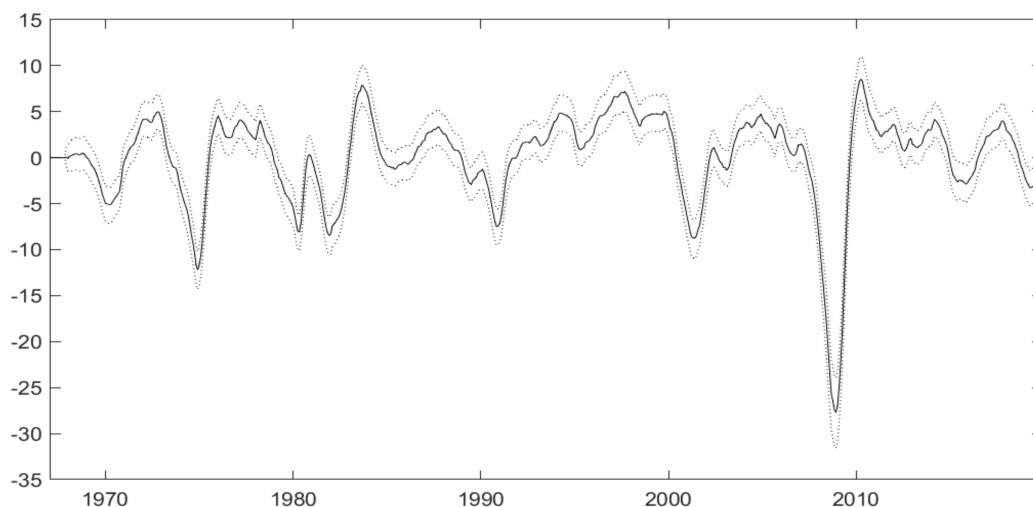


Figure 2. Monthly index of economic activity. The monthly index of economic activity shows the median together with the first and third quartile of the factor distribution (dashed). This figure follows Figure 1 in Stock and Watson (1988) and is based on FRED (2024) data.

Table 5. Statistical summary of the parameter distributions.

Parameter	Q1	Median	Q3	Average	SD
ϕ_1	1.103	1.141	1.178	1.141	0.055
ϕ_2	−0.090	−0.032	0.021	−0.034	0.082
ϕ_3	−0.095	−0.039	0.015	−0.040	0.082
ϕ_4	−0.093	−0.041	0.018	−0.039	0.082
ϕ_5	−0.087	−0.031	0.025	−0.031	0.083
ϕ_6	−0.077	−0.023	0.034	−0.023	0.082
ϕ_7	−0.066	−0.013	0.043	−0.012	0.082
ϕ_8	−0.068	−0.014	0.043	−0.013	0.082
ϕ_9	−0.061	−0.004	0.049	−0.005	0.083
Φ_{10}	−0.040	−0.004	0.031	−0.004	0.053
λ_1	0.067	0.073	0.081	0.074	0.011
λ_2	0.100	0.111	0.123	0.113	0.018
λ_3	0.112	0.124	0.136	0.125	0.017
λ_4	0.043	0.048	0.053	0.048	0.008
λ_5	0.018	0.023	0.028	0.023	0.008
σ_1	0.689	0.719	0.751	0.721	0.046
σ_2	0.621	0.651	0.682	0.652	0.046
σ_3	0.370	0.386	0.402	0.387	0.024
σ_4	0.866	0.901	0.937	0.902	0.052
σ_5	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 parameters are the loadings and the σ parameters are the variances of the idiosyncratic disturbance terms. Q stands for quantile and SD for standard deviation.

For the computation of d , we first allow for a linear time trend as is common in the unit roots literature (Bhargava 1986; Schmidt and Phillips 1992), such that the model becomes a combination of (19) and (20), as follows,

$$f(t) = \alpha + \beta t + x(t), (1 - L)^d x(t) = u(t), \quad (21)$$

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 is 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 a but not for a linear trend, namely $\beta = 0$. We then obtain the result $d = 2.10$ (2.02, 2.19) with $\alpha = 0.930$ which is 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. The results

were not affected by the inclusion of dummy variables corresponding to various outliers such as the one produced by the 2007–2008 financial crisis.

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 (see Figure 3).

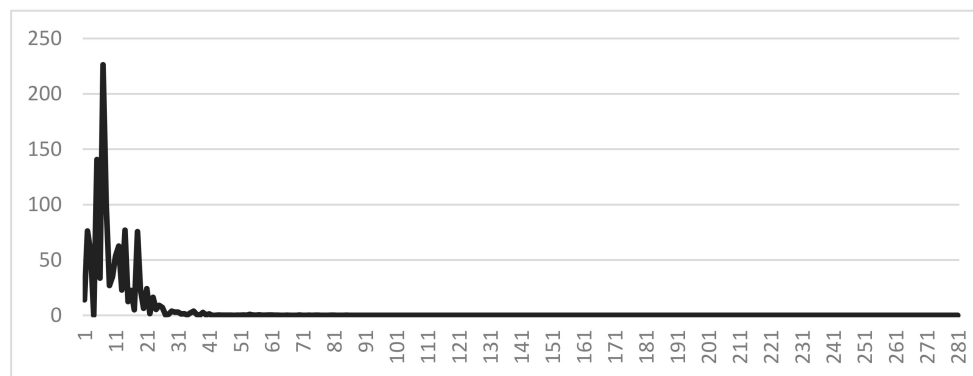


Figure 3. Periodogram of the Index of Economic Activity.

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 obtain valuable information about the dynamics of the series, such as their degree of persistence and mean reversion. Future work will also analyse its asymptotics and its finite sample behaviour by means of Monte Carlo simulations, both of which are beyond the scope of the present study. 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 generated by demand shocks (Clarida et al. 1999; Woodford 2003; Blanchard and Raggi 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.

Author Contributions: Conceptualization, P.J.P.M. and L.A.G.-A.; methodology, P.J.P.M. and L.A.G.-A.; software, P.J.P.M.; validation, L.A.G.-A. and G.M.C.; formal analysis, P.J.P.M. and G.M.C.; investigation, P.J.P.M. and L.A.G.-A.; resources, L.A.G.-A.; data curation, P.J.P.M. and L.A.G.-A.; writing—original draft preparation, G.M.C.; writing—review and editing, G.M.C.; visualization, L.A.G.-A. and G.M.C.; supervision, P.J.P.M.; project administration, P.J.P.M.; funding acquisition, L.A.G.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the project from ‘Ministerium de Economía, Industria y Competitividad’ (MINEIC), ‘Agencia Estatal de Investigación’ (AEI) Spain and ‘Fondo Europeo de Desarrollo Regional’ (FEDER), Grant number D2023-149516NB-I00.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available from the authors on request.

Acknowledgments: Comments from the Editor and two anonymous reviewers are gratefully acknowledged. And the support from an internal Project of the Universidad Francisco de Vitoria.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ An interesting application of a dynamic factor model to central banks and climate change can be found in Braga et al. (2024).
- ² Another $I(d)$ factor approach to modelling time series data can be found in Morana (2014).

References

- Arencibia Pareja, Ana, Ana Gómez Loscos, Mercedes de Luis López, and Gabriel Pérez Quirós. 2017. Un Modelo de Previsión del PIB y de sus Componentes de Demanda. Available online: <https://repositorio.bde.es/handle/123456789/8292> (accessed on 6 May 2024).
- Bai, Jushan, and Serena Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70: 191–221. [CrossRef]
- Bai, Jushan, and Serena Ng. 2004. A PANIC Attack on Unit Roots and Cointegration. *Econometrica* 72: 1127–77. [CrossRef]
- Bai, Jushan, and Serena Ng. 2007. Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics* 25: 52–60. [CrossRef]
- Banerjee, Anindya, Massimiliano Marcellino, and Igor Masten. 2014. *Structural FECM: Cointegration in Large-Scale Structural FAVAR Models*. Working Paper 9858. London: CEPR.
- 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.
- Barigozzi, Matteo, Marco Lippi, and Matteo Luciani. 2021. Non-Stationary Dynamic Factor Models for Large Datasets. *Journal of Econometrics* 221: 455–82. [CrossRef]
- Bhargava, Alok. 1986. On the theory of testing for unit roots in observed time series. *Review of Economic Studies* 53: 369–84. [CrossRef]
- Blake, Andrew, and Haroon Mumtaz. 2017. *Applied Bayesian Econometrics for Central Bankers*. London: Centre for Central Banking Studies, Bank of England.
- Blanchard, Olivier J., and Marianna Riggi. 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: 1032–52. [CrossRef]
- Bloomfield, Peter. 1973. An Exponential Model for the Spectrum of a Scalar Time Series. *Biometrika* 60: 217. [CrossRef]
- Box, George, and Gwilym Jenkins. 1970. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Braga, Joao, Pu Chen, and Willi Semmler. 2024. Central Banks, Climate Risks, and Energy Transition—A Dynamic Macro Model and Econometric Evidence. Available online: <https://ssrn.com/abstract=4794049> (accessed on 6 May 2024).
- Camacho, Maximo, and Gabriel Pérez-Quirós. 2008. *Introducing the Euro-STING: Short Term Indicator of Euro Area Growth*. Documento de Trabajo No. 0807. Madrid: Banco de España, Eurosistema.
- Camacho, Maximo, and Gabriel Pérez-Quirós. 2009. *Spain-STING: Spain Short Term Indicator of Growth*. Documento de Trabajo No. 0912. Madrid: Banco de España, Eurosistema.
- Campbell, John Y., and N. Gregory Mankiw. 1987. Are output fluctuations transitory? *Quarterly Journal of Economics* 102: 857–80. [CrossRef]
- Carter, Chris K., and Robert Kohn. 1994. On Gibbs Sampling for State Space Models. *Biometrika* 81: 541–53. [CrossRef]
- Chan, Hung K., Jack C. Hayya, and Keith J. Ord. 1977. A note on trend removal methods: The case of polynomial trend versus variate differencing. *Econometrica* 45: 737–44.
- Clarida, Richard, Jordi Gali, and Mark Gertler. 1999. The Science of Monetary Policy: A New Keynesian Perspective. *Journal of Economic Literature* 37: 1661–707. [CrossRef]
- Cochrane, John H. 1988. How big is the random walk in GNP? *Journal of Political Economy* 96: 893–920. [CrossRef]
- DeJong, David N., John C. Nankervis, N. E. Savin, and Charles H. Whiteman. 1992. The power problems of unit root tests in time series with autoregressive errors. *Journal of Econometrics* 53: 323–43. [CrossRef]
- Dickey, David A., and Wayne A. Fuller. 1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* 74: 427–31.
- Doz, Catherine, Domenico Giannone, and Lucrezia Reichlin. 2012. A quasi-maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics* 94: 1014–24. [CrossRef]
- Durlauf, Steven, and Peter C. B. Phillips. 1988. Trends versus random walks in time series analysis. *Econometrica* 56: 1333–54. [CrossRef]
- Elliot, Graham, Thomas J. Rothenberg, and James H. Stock. 1996. Efficient Tests for an Autoregressive Unit Root. *Econometrica* 64: 813–36. [CrossRef]
- Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. 2005. The Generalized Dynamic Factor Model: One sided estimation and forecasting. *Journal of the American Statistical Association* 100: 830–40. [CrossRef]

- FRED. 2024. Board of Governors of the Federal Reserve System (US): Industrial Production: Total Index [INDPRO]. Available online: <https://fred.stlouisfed.org/series/INDPRO> (accessed on 6 May 2024).
- Geweke, John. 1977. The Dynamic Factor Analysis of Economic Time Series. Available online: <https://cir.nii.ac.jp/crid/1571980075928307712> (accessed on 6 May 2024).
- Geweke, John, and Susan Porter-Hudak. 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4: 221–38. [CrossRef]
- Gil-Alana, Luis A., and Juan C. Cuestas. 2012. *A Non-Linear Approach with Long Range Dependence Based on Chebyshev Polynomials*. Navarra: University of Navarra.
- Gil-Alana, Luis Alberiko, and Juan Carlos Cuestas. 2016. Testing for long memory in the presence of non-linear deterministic trends with Chebyshev polynomials. *Studies in Nonlinear Dynamics & Econometrics* 20: 57–74.
- Gil-Alana, Luis A., and Peter M. Robinson. 1997. Testing of unit root and other nonstationary hypotheses in macroeconomic time series. *Journal of Econometrics* 80: 241–68. [CrossRef]
- Gómez Loscos, Ana, Miguel Ángel González Simón, and Matías José Pacce. 2024. *Short-Term Real-Time Forecasting Model for Spanish GDP (Spain-STING): New Specification and Reassessment of Its Predictive Power*. Documentos Ocasionales No. 2406. Madrid: Banco de España.
- Granger, Clive W. J. 1980. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14: 227–38. [CrossRef]
- Granger, Clive W. J. 1981. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16: 121–30. [CrossRef]
- Granger, Clive W. J., and Roselyne Joyeux. 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1: 15–29. [CrossRef]
- Hamilton, James D. 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* 57: 357–84. [CrossRef]
- Hamilton, James D. 1994. *Time Series Analysis*. Princeton: Princeton University Press.
- Hosking, J. R. 1981. Fractional differencing. *Biometrika* 68: 165–76. [CrossRef]
- Hurst, Harold Edwin. 1951. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* 116: 770–99. [CrossRef]
- Kim, Chang-Jin, and Daniel C. R. Halbert. 2017. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge: The MIT Press.
- Kim, Chang S., and Peter C. B. Phillips. 2006. *Log Periodogram Regression: The Nonstationary Case*. Cowles Foundation Discussion Paper No. 1587. New Haven: Cowles Foundation.
- Kydland, Finn E., and Edward C. Prescott. 1982. Time to Build and Aggregate Fluctuations. *Econometrica* 50: 1345–70. [CrossRef]
- Leybourne, Stephen J., and Brendan P. M. McCabe. 1994. A consistent test for a unit root. *Journal of Business & Economic Statistics* 12: 157–66.
- Luciani, Matteo. 2015. Monetary policy and the housing market: A structural factor analysis. *Journal of Applied Econometrics* 30: 199–218. [CrossRef]
- Morana, Claudio. 2014. Factor Vector Autoregressive Estimation of Heteroskedastic Persistent and Non-Persistent Processes Subject to Structural Breaks. *Open Journal of Statistics* 4: 292–312. [CrossRef]
- Nelson, Charles R., and Charles I. Plosser. 1982. Trends and random walks in macroeconomic time series. *Journal of Monetary Economics* 10: 139–62. [CrossRef]
- Nelson, Charles R., and Heejoon Kang. 1981. Spurious periodicity in inappropriately detrended time series. *Econometrica* 49: 741–51. [CrossRef]
- Nelson, Charles R., and Heejoon Kang. 1984. Pitfalls in the use of time as an explanatory variable in regression. *Journal of Business and Economic Statistics* 2: 73–82. [CrossRef]
- Ng, Serena, and Pierre Perron. 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69: 1519–54. [CrossRef]
- Pacce, Matías, and Gabriel Pérez-Quirós. 2019. Predicción en tiempo real del PIB en el área del euro: Recientes mejoras en el modelo Euro-STING. Available online: <https://repositorio.bde.es/handle/123456789/8443> (accessed on 6 May 2024).
- Pfaff, Bernhard. 2008. *Analysis of Integrated and Cointegrated Time Series with R*, 2nd ed. Berlin and Heidelberg: Springer.
- Phillips, Peter C. B., and Pierre Perron. 1988. Testing for a Unit Root in Time Series Regression. *Biometrika* 75: 335–46. [CrossRef]
- Quah, Danny. 1992. The Relative Importance of Permanent and Transitory Components: Identification and Some Theoretical Bounds. *Econometrica* 60: 107–18. [CrossRef]
- Quah, Danny, and Thomas J. Sargent. 1993. A dynamic index model for large cross sections. In *Business Cycles, Indicators, and Forecasting*. Chicago: University of Chicago Press, pp. 285–310.
- Robinson, Peter M. 1994. Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* 89: 1420–37. [CrossRef]
- Robinson, Peter M. 1995. Log periodogram regression of time series with long range dependence. *Annals of Statistics* 23: 1048–72. [CrossRef]
- Royston, Patrick. 1982. An extension of Shapiro and Wilk's test for normality to large samples. *Applied Statistics* 31: 115–24. [CrossRef]

- Sargent, Thomas J., and Christopher A. Sims. 1977. *Business Cycle Modeling Without Pretending to Have Too Much a Priori Economic Theory*. Working Papers 55. Minneapolis: Federal Reserve Bank of Minneapolis.
- Schmidt, Peter, and Peter C. B. Phillips. 1992. LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics* 54: 257–87. [CrossRef]
- Shapiro, Samuel Sanford, and Martin Bradbury Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52: 591–611. [CrossRef]
- Sowell, Fallaw. 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* 53: 165–88. [CrossRef]
- Stock, James H. 1991. Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics* 28: 435–60. [CrossRef]
- Stock, James H., and Mark W. Watson. 1988. *A Probability Model of the Coincident Economic Indicators*. NBER Working Papers 2772. Cambridge: National Bureau of Economic Research, Inc.
- Stock, James H., and Mark W. Watson. 2002. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20: 147–62. [CrossRef]
- Trapletti, Adrian, and Kurt Hornik. 2020. Tseries: Time Series Analysis and Computational Finance: R Package Version 0.10-48. Available online: <https://CRAN.R-project.org/package=tseries> (accessed on 6 May 2024).
- Watson, Mark W., and Robert F. Engle. 1983. Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics* 23: 385–400. [CrossRef]
- Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton: Princeton University Press.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.