# Analysis of Non-Stationary Time-Series Business Data

**George Saridakis**
*Kingston University, UK*

**Grammatoula Papaioannou**
*Loughborough University, UK*

## INTRODUCTION

Availability and collection of data are often key restrictions for researchers in formulating research questions and choosing appropriate econometric methodologies and techniques to answer them. In many respects, however, a number of important research questions related to business studies can be answered by a collection of time-series data, which are usually freely available from national statistical services: for example, in the United Kingdom (UK) the National Statistics Hub provides rich information on economic, business and labour market statistics over many decades. Hence, researchers can easily extract single (random) series of observations (time-series) to each time period (e.g year, quarter) considered in their study. A convenient notation for such a variable is $y_t$ for observations $t=1,….,T$; for example, business start-ups ($y_{1,t}$) and closures rates ($y_{2,t}$) for the years 1970 ($t=1$) through to 2012 ($t=T$).

Nevertheless, time-series data are non-stationary and therefore, they can cause serious problems (e.g. spurious regression) when the applied econometrician/statistician fails to explore and address these in the modelling (see econometrics textbooks by Hill et al., 2008; Wooldridge, 2000; Patterson, 2000 among others) process. This chapter discusses important issued related to the analysis of non-stationary business data, and provides some well- established methods of testing and modelling such data in an efficient way to assist time-series analysts and researchers. Specifically, the next section discusses the statistical notion of non-stationarity. We then provide statistical ways of testing for stationarity and an overview of the concept of cointegration, which can be viewed as an effective framework for testing, modelling and estimating long-run relationships among non-stationary time-series. Following this section, we discuss some trends and advances in this area. The last section concludes the paper by providing a brief description of some key econometric packages that can be used to test for non-stationarity and cointegration.

Overall, our chapter puts together a collection of advanced techniques and methodologies employed in time-series econometrics that can be used either for revision purposes by experienced quantitative analysts or as a helpful guide for early time-series researchers and graduate students who wish to enrich their knowledge and skills in time-series analysis and locate relevant and recently published work in this area.

## BACKGROUND

In time-series analysis of business and economic data (e.g. stock index data; corporate dividend payments; corporate profits; business start-ups; business survival rates) the statistical concept that has received considerable attention and gained much popularity among applied researchers is the one related to *non-stationarity*. As discussed in a number of econometrics textbooks (see Verbeek, 2000; Charemza & Deadman, 1997 among others), quantitative analysts are generally concerned

with the concept of w*eak stationarity* (or *covariance stationarity*) i.e. the mean, variances and autocovariances of the series are independent of time; that is $E(y_t)=c$ remains constant for all $t$; $var(y_t)=E(y_t-c)^2=\sigma^2$ remains constant for all $t$; and $cov(y_t, y_{t+g})=E[(y_t-c)(y_{t+g}-c)]=\theta_g$ remains constant for all $t$ and $g\neq0$. If one or more of these conditions are not fulfilled the time-series is called *non-stationary* (this is discussed more analytically in Seddighi et al., 2000).

Standard regression techniques (e.g. OLS) require that the variables included in the regression analysis are stationary and, as it is known in the econometric literature stationary series, are said to be integrated of order zero, namely *I(0)*. Otherwise researchers may obtain significant regression results from unrelated data, and such regressions are said to be spurious (Hill et al., 2008; Granger & Newbold, 1974). A series, however, usually becomes stationary after first-differencing and this series is said to be integrated of order one, namely *I(1)*. Nevertheless, transforming a non-stationary series to a stationary one causes important long-run information loss in the study data under analysis.

Below we explain how researchers can statistically examine whether a series is stationary or non-stationary and determine its order of integration. We then continue by discussing how non-stationary time-series data can be used in regression models to estimate long-run economic associations, which can then be used for policy purposes.

## MAIN FOCUS

### The Augmented Dickey-Fuller Test

Perhaps a simpler way of 'testing' for stationarity is by observing the plot between the sample autocorrelation function (ACF) at lag $p$ against $p$, which is commonly known in the literature as a *sample correlogram* (for examples, see Gujarati, 1995). Over the years, however, more advanced

tests have been developed to assist quantitative researchers to test more methodologically the order of integration of a time-series variable (i.e. how many times a variable should be differenced in order to become a stationary series), and these tests are commonly referred among econometricians and statisticians as *unit root tests of stationarity*.

Undoubtedly one of the most common unit root tests is the (augmented) Dickey-Fuller (ADF) test proposed by Dickey & Fuller (1979). The ADF test considers the following regression (with the optional inclusion of both an intercept and a linear trend: for further discussion see Charemza & Deadman, 1997, pp. 112-114; Hill et al., 2008, pp. 335-338):

$$\Delta y_t = \mu + \gamma t + ay_{t-1}+\delta_1\Delta y_{t-1} + ... + \delta_k\Delta y_{t-k} + \varepsilon_t \tag{1}$$

If the time series $y_t$ has a unit root the coefficient $a$ should be zero (null hypothesis). The test can be easily performed by estimating an Ordinary Least Square (OLS) and comparing the $t$-statistic on the coefficient $a$ - for appropriate critical values, see MacKinnon (1991).

The Equation (1) is augmented by lagged first differences of $y_t$ up to an order $k$ to correct for autocorrelation in the error terms - different model selection criteria such as the Akaike Information Criterion (AIC) and the Shwarz Bayesian Criterion (SBC) could be also employed for choosing the order of $k$. In the case where no lagged differences are included in Equation (1), the standard DF test can be applied.

### Alternative Unit Root Tests

To examine the robustness of the ADF results, researchers also make use of the Phillips-Perron (PP) test (Phillips & Perron, 1988), which is simply a nonparametric method of controlling for serial correlation when testing for a unit root. The PP method estimates the non-augmented DF test equation and then the $t$-statistic of the coefficient

## Related Content

Scheduling of Extract, Transform, and Load (ETL) Procedures with Genetic Algorithm
Vedran Vrbaniand Damir Kalpi (2015). *International Journal of Business Analytics (pp. 33-46).*
www.irma-international.org/article/scheduling-of-extract-transform-and-load-etl-procedures-with-genetic-algorithm/126832

Stock Market Prediction Using Elliot Wave Theory and Classification
Saeed Tabar, Sushil Sharmaand David Volkman (2021). *International Journal of Business Analytics (pp. 1-20).*
www.irma-international.org/article/stock-market-prediction-using-elliot-wave-theory-and-classification/269484

Strategies for Large-Scale Entity Resolution Based on Inverted Index Data Partitioning
Yinle Zhouand John R. Talburt (2014). *Information Quality and Governance for Business Intelligence (pp. 329-351).*
www.irma-international.org/chapter/strategies-for-large-scale-entity-resolution-based-on-inverted-index-data-partitioning/96158

Integrating Business Intelligence With Cloud Computing
Shivani Jaswal (2021). *Impacts and Challenges of Cloud Business Intelligence (pp. 41-56).*
www.irma-international.org/chapter/integrating-business-intelligence-with-cloud-computing/269808

Performance Comparison of Two Recent Heuristics for Green Time Dependent Vehicle Routing Problem
Mehmet Soysal, Mustafa Çimen, Mine Ömürgönülenand Sedat Belba (2019). *International Journal of Business Analytics (pp. 1-11).*
www.irma-international.org/article/performance-comparison-of-two-recent-heuristics-for-green-time-dependent-vehicle-routing-problem/238062