Silva Lopes, Artur

Preprint

Introduction to the Univariate Analysis of Trends in Economic Time Series

Introduction to the Univariate Analysis of Trends in Economic Time Series

Artur Silva Lopes

This version: January 14, 2025

To the memory of my father,
Carlos Nunes da Silva Lopes

Contents

Preface

Almost a century ago, Wesley Mitchell (1927) justified the scarce comprehension of trends with the rather unusual combination or "fusion" of knowledge on several distinct subjects that the topic required to be grasped: "statistical technique with historical learning and theoretical finess" (p. 232). At the time, Mitchell also considered that the trend topic was little understood and even "obscure".

About 90 years ago, Edwin Frickey (1934) further clarified the complexity that the topic involved stating that it was a problem i) " in mathematical curve fitting", ii) "in statistical description", iii) "in historical description", iv) "in the analysis of causation" and v) "in economic theory" (see more on this in section 2.5).

Although much progress has been made, the subject still remains somewhat "obscure", so that the situation described by Peter Phillips (2005) about 20 years ago — "no one understands trends but everyone sees them in the data" — remains largely valid. Nevertheless, the days when trends were equated with deterministic polynomial time trends and hence these were almost the only adopted models to represent trends belong to a distant past.

The purpose of this book is to contribute to improve this state of affairs, surveying and systematizing most of the literature on the subject. Despite the recommendation by Frickey (1934, F) and by Halbert White and Clive Granger (2011 WG), *not* to study each series "in isolation" (WG), but rather "looking at the whole picture" (F), this contribution is limited to the univariate approach. This constraint was realistically imposed by the vast amount of literature that has been published mainly in

the last 40 years but it also logically stems from a basic principle established by the "unit root and cointegration revolution": sound univariate analysis is a necessary first step on which multivariate analysis must be grounded to be valid.

Despite this limitation, this work is not an exhaustive, all-encompassing univariate approach to trends. Several topics are not even addressed, as Markov chain models, fractional integration or the Bayesian approach to trends. Nonparametric methods are also barely mentioned. The reader aiming to grasp the subject of stochastic trends/unit roots in detail must look elsewhere, as my approach here is merely introductory. Fortunately, there are several excellent alternatives available.

As it requires somewhat advanced knowledge in statistics, econometrics and time series analysis, this book is aimed for graduate students and for practitioners and researchers in several fields where trends are present. It may be useful as a reference, not as a textbook designed for a particular course, though some of its chapters may be used in courses at the master and PhD levels.

I wish to express my gratitude mostly to neurologist doctors Anabela Valadas and Ana Rita Simões. I wish to thank also neurosurgeon Begoña Cattoni and neurologist doctors Ana Raquel Barbosa, Duarte Salgado and Miguel Coelho. My thanks go as well to physiotherapists Gabriela Fonseca, Rita Gama, Iara Almeida, Laura Antunes, Francisco Martins, Beatriz Santos, Maria Inês Preto, Beatriz Alves and Catarina Silva.

I have not used any tool of generative artificial intelligence in writing this book. I would be grateful for any constructive comment about this text and I thank in advance.

Artur Silva Lopes, independent researcher
(arturslopes@proton.me), November 2024

Chapter 1

Introduction

This introductory chapter contains preliminary material about the subject of this book, addressing such basics as the (non-existent) concept of trend, the main characteristics and types of trends, as well as their economic drivers. Some of the issues of the following chapters are introduced and hence this chapter serves as a basic framework for the remainder of the book.

1.1 What is a trend?

Hopefully, the title of this section has a clear message: I consider that there is no single definition of trend that could be simultaneously general and precise to be widely accepted. Choosing "a trend" rather than "the trend" means that I am allowing from the outset that several definitions might coexist. But I also think that some limits must be put forward.

For a long period of time, not long ago, trend was synonymous of a linear deterministic process in the time variable t, given by

$$y_t = \alpha + \beta t + u_t, \quad t = 1, 2, \ldots, T,$$

where y_t denotes the series of interest (that may be logarithmized) and u_t represents a zero mean stationary process. The parameter β is often as-

sumed to be positive to reflect the most usual case of variables which tend to grow with time. Despite its simplicity, even naivety, this "definition" was somewhat successful: it worked reasonably well with short samples, it was very easy to estimate and, above all, it made the measurement of business cycles extremely easy, as these were represented with the OLS residuals, $\widehat{u}_t$. Increasing sample sizes and, mostly, its use as a forecasting device made clear that this definition could not be taken seriously for many purposes: the growth rate of the series should be approximately constant for rather long periods of time and, worse still, it promised that perfect forecasts could be made for time horizons arbitrarily far into the future.

While the first problem can be patched with a few modeled structural breaks, the second cannot. Rather, the identification of trend breaks in the past makes future predictability even more doubtful[1].

Augmenting the order of the deterministic polynomial in time simply increases the peril of the trap: a more sophisticated function appears less naive, fits the data better, and its twists may even dispense structural breaks; but it may also give rise to even more disastrous forecasts.

Although it ceased to correspond to the only notion of trend, the linear trend model still provides one of the two dominant ideas of trend and of trending behaviour, usually associated with deterministic and monotonically increasing functions of time. Even today, "trending behaviour" is commonly understood as meaning that the variable tends to exhibit a systematic upward or downward evolution with time, though not necessarily in a linear way. I call this notion the traditional one.

The alternative notion is that of a long-run evolution characterized by wandering behaviour, with long, erratic and asymmetric cycles of upswings and downswings, separated by shorter periods where the series seems to hesitate and may change direction. This type of trend usually contains several segments exhibiting trending behaviour in the traditional

[1] Altough the broken trend model is often associated with Perron (1989), already in 1954 Burns mentioned it as a commonly adopted model for trends: "*.... the series are split into subperiods, and a trend line fitted separately to the logarithms of the data in each subperiod*" (o. cit., p. 291).

sense, either monotonically increasing or decreasing, which appear to contradict each other, and which are sometimes called "local trends" to mean that a change in direction will occur sooner or later[2].

While the first type of trend is typical of macroeconomic variables such as national income and its components, the second often appears associated with variables that are measured as rates: the inflation rate and interest rates are two prominent examples.

Although it is not widely adopted, this second notion was explicitly mentioned about 20 years ago in Phillips (2005), and it finds its source in the random walk model and more generally in first order integrated processes ($I(1)$) and in the trend component of unobserved component (UC) models (see section 3.8). The perils of simply extrapolating the recent past appear more clearly with this type of trend: since in the past the trend has already changed direction, sometimes several times, why not will that occur again in the future?

A somewhat similar distinction must be made between two types of definitions or notions of cycles (or "business cycles"), as these will be frequently mentioned. Very often the concept of trend appears immersed in the framework of the decomposition of any time series into components with rather different characteristics. And indeed this appears to be its origin (see section 2.2). In its most general form, this traditional decomposition is

$$y_t = \tau_t + c_t + s_t + i_t,$$

where τ_t denotes the trend, c_t the cycle, s_t the seasonal component and i_t the irregular (unexplainable) component, with the order of the components inversely related with the frequency of their variation[3]. To be clear, when I mention a trend in this context I am referring to this component, and the cycles that are associated with it are called "growth cycles" in Zarnowitz (1992) and "output-gap cycles" in Morley and Piger (2012).

[2]Local trends "are trends only for some finite time interval", White and Granger (2011), where the meaning of trend is inequivocally associated with what I call the traditional notion.

[3]And directly related with the periodicity of its variations, though no claim is made that they are strictly periodic.

The famous Beveridge-Nelson (1981) decomposition is associated with these notions (see section 4.2).

There is, however, an alternative notion of cycles, which was the one adopted by Arthur Burns and Wesley Mitchell, which is simply defined in terms of alternation of expansions and recessions in the level of economic activity. These are often called the "classical cycles", corresponding to the "alternating-phases definition", as labeled by Morley and Piger (2012). In this case, the trend is not separated from the cycle, it is embedded in it. It this view that dominates the traditional NBER approach to business cycles, where seasonality is removed but data are not detrended.

The "output-gap" notion of business cycles is also closely related to another notion of trend, that of potential output. I rule out this notion here because the view that I hold for the trend, although acknowledging that it is unobservable, is that of a real variable, of something that really exists, not of some level that is imagined as achievable only under certain conditions. Moreover, as far I am aware, there are at least two rather different notions of potential output: one as that output that will be achieved when the resources of the economy are fully employed, and the other as the "maximum level of durable sustainable production without tensions in the economy, and more precisely without acceleration of inflation" (Ladiray, Mazzi and Sartori, 2003), for short, the non-accelerating inflation level of output.

A notion that Burns (1934, p. 32) mentions only to rule out immediately is that of trends as "equilibria paths which variables tend to approach". This type of notion very seldom appears in the literature and it will be ruled out here as well.

1.2 Characterizing trends

In this section I will discuss the major characteristics of trends, those that may be considered as common to most cases and that sometimes are even considered as indispensable to identify a trend. Therefore this

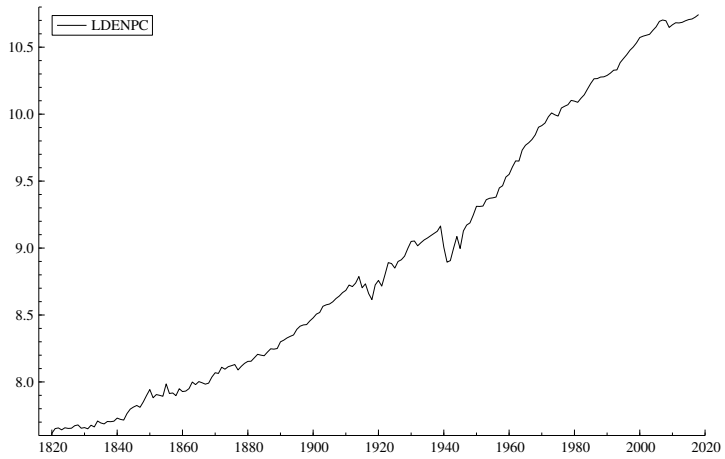

Figure 1.1: An illustration of the traditional type of trend: the logarithm of Denmark's per capita GDP, from 1820 to 2018

section is empirically based. The discussion about the economic forces that underlie trends is postponed to section (1.4).

Visually, trends are often the *dominant* feature of economic time series. This is the case particularly for traditional trends, mostly when they are positive. An illustration of this type of trend is provided in figure 1.1 where the time series for the logarithm of per capita GDP of Denmark, from 1820 to 2018 is represented (taken from Maddison's database updated by Bolt and van Zanden, 2020)

However, this characteristic is almost absent for the random-walk type trend as characterizing the whole sample. In this case, it may be visible in certain stretches only, which however are often contradicted by others with opposing direction. What is common to both cases is the regularity of the increasing or decreasing movement for several contiguous periods, but usually not for the whole sample. In the case of traditional trends, however, this type of movement may characterize all or almost all the sample (depending on the sample size).

Therefore, the requirement that trends should *have a direction* is valid only in the case of traditional trends. It is not a general condition for

all trends. A similar remark can be made about the requirement that a trend needs to be *monotonic throughout*, i. e., over the entire sample. Even traditional growing trends often display large downward level shifts, usually associated with major negative events (e.g., wars, pandemics, large crisis, etc.). Considering both types of trends, it is true that they both exhibit monotonic behaviour, but sometimes the periods when they display such behaviour is not very long.

Also, strict *continuity* is not an attribute of all trends, as their breaks are often abrupt. What remains in many cases is only piecewise continuity. Moreover, a continuous first derivative is often but not always observed. On the other hand, even random walk type trends often have first derivatives with only a few changes in sign over the whole sample.

What is valid in general is a feature that characterizes trends for its absence: trends *do not* display cyclical, oscillatory-type movement which is relatively regular and which is typical of both cyclical and seasonal fluctuations (though with different periodicities). Sometimes, as previously mentioned, random-walk type trends exhibit somewhat cyclical behaviour, but this is usually far from regular or symmetric, as the adjacent cycles typically vary much in length and in amplitude. A real example of this type of trend is presented in figure 1.2 for the series of the long-run interest rate for Portugal, from 1870 to 2020 (taken from the database associated with Jordà, Schularick and Taylor, 2017).

Rather, trends are *slowly varying processes*, very rarely showing any jaggedness or any oscillatory behaviour. Therefore, it should be no surprise that the most often nominated characteristic of trends is their *smoothness*[4], and this more fully characterizes traditional trends than random walk type ones, but it is a dominant characteristic of them as well, at least when compared with stationary series.

This is another way to say that trends correspond to low frequency phenomena, to low frequency variation, those with really long, imperceptible cycles, or without any cycle at all (corresponding to the zero

[4]Quast and Wolters (2022) provide several explanations for the reason why potential output should be smooth; recall, however, that I have ruled out this notion of trend.

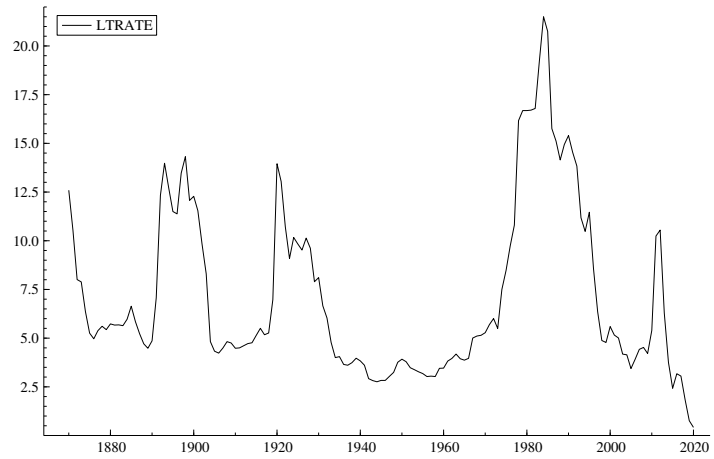

Figure 1.2: An illustration of the random-walk type trend: the long-run interest rate for Portugal, from 1870 to 2020

frequency). For, e.g. Müller and Watson (2008), these low frequencies are those that correspond to periodicities that are greater than those of business cycles, i.e., greater than 8 years, but this is a non-demanding lower limit.

This type of behaviour makes trends also the best representatives of *persistent* forces, not of those variables that tend to change their course frequently. In other words, trends are viewed as reflecting effects that are highly persistent, that tend to be permanently or almost permanently present. For this reason, trends are often called the *permanent*, non-perishing component of economic time series.

Finally, only mathematical convenience justifies two major features usually but incorrectly pointed out to trends: *deterministic functions of time*, making them perfectly predictable. The epitome of this double inadequacy is the linear (deterministic) time trend, which may sometimes serve as a useful approximation but which cannot continue to reign alone. Instead, it is only a rather simple and pedagogical example. Deterministic functions of time are useful, but nonlinear functions may be better approximations.

1.3 The forecasting approach

A common or popular notion of trend is that it says something about the future. Not necessarily about the long term future but about the near future. Something that simply extrapolates the past and the present, possibly only the recent past. This intuitive notion of trend makes the question "what is the trend?" when it really wishes to ask "what is the forecast?", "how will things develop?". Moreover, an accurate, precise forecast is not usually required. Not even a quantitative one. Rather, a general direction, a sign of the possible change, a simple qualitative forecast will do.

This notion has also been adopted in the literature, more formally and rigorously of course. It even has already some history, as Ball and Woods (1996, p. 145) stated that "many modern protagonists of detrending methods argue that the prime function of a trend is predictive rather than smoothing ... The trend is the best predictor of data at points beyond time t". This was particularly the view held by Harvey (1989, p. 284), who extended the forecast horizon:

> What then is a trend? Viewed in terms of prediction, *the estimated trend is that part of the series which when extrapolated gives the clearest indication of the future long-term movements in the series* (italics as in the original).

Actually, although perhaps not dominant, the forecasting approach is very clearly present in two of the most important methods for the decomposition of economic time series, those of Beveridge and Nelson (1981) and Hamilton (2018), see sections 4.2 and 4.6, respectively. In the first case the trend is the level of a forecast function for a far way future horizon of an $I(1)$ process and in the second it is a linear projection of a future value of the series on a constant and its four most recent values.

A more recent example is that of the definition of trend inflation in Nakajima (2023):

> We define trend inflation as the expected infinite-horizon forecast of π_t (the inflation rate), given the information set in t,

i.e., formally as $\mu_t = \lim_{j\to\infty} \mathrm{E}(\pi_{t+j}|\Omega_t)$, Ω_t denoting the information set at time t. Incidentally, as the variable involved is the inflation rate, this is also an example of the non-traditional notion of trend.

1.4 Economic drivers

What are the economic drivers of trends? Which forces cause economic variables to display trends? Are they different and independent from the ones that are responsible for temporary fluctuations? This is the terrain of the last two items in Frickey's (1934) research agenda and important gaps still remain in the literature. Phillips (2005, 2010b) summarizes the situation complaining that economic theory has been offering "little guidance" about the origin and the nature of trend behaviour.

Apparently, there is a large consensus in the literature about these forces, at least when it is the level of output in an economy that is concerned. Three major forces figure prominently in this consensus:

a) population growth;

b) technological progress and

c) capital accumulation.

However, this list can be easily improved after some reflection: a) demographic forces in general extend beyond the number of people pointed in a) to include as well its age composition or structure; b) scientific progress must be added to b), and c) human capital must be added to c) above which typically refers to physical capital only. To these, one might even add some other, intangible and institutional as, e.g., changes in tastes and the rule of law as contributing to the trends that have been observed in western countries since the end of the 18th century. All these are usually seen as forces that cause little or no change in the short and medium terms, and hence the separation from other components of time series would become completely justified. Further still, from a different perspective, while transitory fluctuations could be associated with

demand-side forces, trends could be seen as the outcome of supply-side ones[5].

However, the consensus is only apparent and there are still many issues that are far from settled. Even the pioneers of business cycle analysis were very careful stating clearly that they were not claiming that the causes of these fluctuations were independent of those that lie behind long-run growth. In particular, citing Frickey (1942), Burns (1944) clearly states:

> nor can we assume that secular trends and cyclical variations are due to independent causes. "We should think here, not in terms of distinct sets of causal forces, but in terms of *lines of causal influence*".

That is, the forces that explain trends do not need to be different from those that lie behind business cycles. It is their long-run effects, not necessarily their source, that explains trends.

Rather than independent of those causing business cycles, the forces behind trends may be exactly the same. Fernald, Inklaar and Ruzic (2024) provide a brief survey of endogenous growth models where business cycle fluctuations, and particularly deep recessions, significantly affect long-run GDP growth. The channels through which these effects are produced are mostly labour markets and productivity.

In economics, highly persistent effects are sometimes known as "hysteresis" or, to use an even more updated terminology, "scarring", for negative effects (since recessions leave "scars" that do not tend to disappear with time). Cerra, Fatás and Saxena (2023) provide an extensive survey of recent macroeconomic models containing these types of hysteresis effects, which defy the traditional trend-cycle decomposition by questioning the independence or the orthogonality of their origins. While conventional drivers are not questioned, several other sources of trends, including the once considered inadmissible demand shocks, are considered as potential sources of long-run effects.

[5]For a critical vision of this see, e. g., section 3.8.

1.5 How many observations?

The question is now: how many observations are needed to identify trends in data? What is the minimum sample size for this purpose? And a common and sensible answer is: there is no simple answer. It depends.

For instance, when "analysing 70 years of post-WWII quarterly data, decadal variation is low-frequency"(Müller and Watson 2024, MW24), i. e., since the relation between frequencies and periodicities is inverse, 10 years of quarterly data of a 70-year sample may be considered already as representing the long-run (although not really a very long one). Or as another example, "when studying a decade of daily return data, yearly variation is low-frequency" (MW24 again), i.e., only one year of daily data of a 10-year sample is also considered as representing the long-run. Hence, the answer depends at least on the total *sample size* and on the *frequency* of the data.

Notice that in the first example the above "rule" supersedes another one that considers that periods greater than 8 years are sufficient to represent the long-run because these correspond to frequencies that are inferior to those of business cycles. An even more demanding and intuitive but common one consists of doubling the maximum periodicity allowed for business cycles: this requires that at least 16 years of quarterly observations are needed to identify a trend in economic data, for a sample that contains a minimum of say 40 to 50 years of observations. Nevertheless, the main general rule stated above still holds: it depends on the total sample size and on the frequency of the data.

It also depends on the variability of the data, which sometimes is mentioned as randomness. It is easy to understand the reason: more volatile data requires larger samples to control for the increased uncertainty and hence, to become more confident that a certain (upward, downward or simply smooth) sequence of movements represents a segment of a trend, we require more data. Some authors sometimes also mention the type of model as another factor determining the answer but I will not pursue this topic here.

What is of utmost importance is the nature of the phenomenon un-

der study as well as the purpose of the study. The problem that most practitioners wish to avoid when they make the question is to be fooled by, e.g., an upward drift for several observations that later reveals to be only a part of a cycle with a longer period. This problem of paucity of sample information about the long-run is particularly acute when the traditional notion of trend is the one being considered, but a smooth behaviour for some periods in a row may occur also by chance, leading to an erroneous inference for a random walk type trend. The first case can be simply illustrated with the situation reported in Seater (1993) about the possibility of an increase in average annual world temperatures (possibly originated by greenhouse gases emissions).

Using two samples with "only" about 100 annual observations ending in the late 1980s, Seater (1993) finds evidence for a significant upward trend in temperatures. However, using a much longer sample starting in 500 and ending in 1980 (inferred from tree-ring data), the statistical significance of the trend vanishes[6]. Hence, the rising trend appeared not to be a true feature of the data generation process but only a local upswing, "an artifact of small sample periods (...), too short to capture the true behaviour of the process" (Seater, 1993). In summary, the answer is context-dependent, samples of "only a century" duration can be insufficient to make correct inferences for certain phenomena.

1.6 Detrending

Detrending appears to be a beneficial business for both economists and statisticians (econometricians). The first become acquitted of explanations about the long-run behaviour of economic variables and about the interactions between several types of economic forces, and the second become also acquitted of explaining the non-stationary behaviour of eco-

[6]Regardless of the adequate inference, notice that, so far as I am aware, global warming concerns were at their infancy at the time, initially appearing in the 1990s, with the first assessment report by the IPFC (Intergovernmental Panel of Climate Change) in 1990 and the first COP only in 1995.

nomic time series, as they are trained mostly to deal with stationary data. Indeed, detrending allows:

a) economists to focus on the most tangible and (apparently at least) most easily manageable economic forces and

b) econometricians to employ their usual toolkit mostly designed for stationary time series but often invalid for non-stationary ones.

Detrending is not confined to macroeconomics. Other areas where the technique is investigated are, e.g. climatology[7], infectiology and agricultural economics[8]. But most demand is originated in macroeconomics, where detrending is sometimes used rather mechanically and routinely, resembling the calculation of, e.g., a sample average, but neglecting the way the procedures really work as well as their properties, and hence their potential drawbacks.

The major pitfall of detrending methods is not that their diversity produces a wide variety of stylized business cycle facts, in particular radically different second moments of the cyclical component, as shown in Canova (1998). This dispersion may be even more beneficial than detrimental, as the different methods play the role of alternative windows that may provide different perspectives on macroeconomic subjects (Burnside, 1998, as well as Canova, 1998, albeit to a lesser extent; see also section 4.3).

Rather, the major peril consists of relying on a single method — typically the "Hodrick-Prescott (HP) filter" — to derive the properties of business cycles (and implicitly that of trends). This is one of the lessons to draw from Burnside (1998). However, whereas Burnside prefers to focus on the consequences that any filtering method implies for the testing properties of economic models, my concern is different. Although I leave the details to chapter 4, the main issue I wish to stress here lies

[7]See, e.g., Dakos, Scheffer, van Nes, Brovkin, Petoukhov and Held (2008) and Lenton, Livina, Dakos, van Nes and Scheffer (2012).

[8]See, e.g., respectively, Dessavre, Southall, Tildesley and Dyson (2019), and Lu, Carbone and Gao (2017).

in the distorting features of the HP filter, which may often lead to find misleading traits in economic data, even to the point of spurious cyclical properties.

1.7 A definition of trend?

One reason why there are so many diverse detrending methods is the absence of a unique, widely accepted and precise definition of trend. Not that it is hard to find a definition in the literature, quite the contrary. The following are some examples:

i) "the secular trend of an industry's production may be considered as the persistent, underlying movement of its output over a period which is 'long' in relation to the changes associated with the 'business cycle' " (Burns, 1934, p. 31).

ii) "Trend (...) may be loosely defined as 'long-term change in mean level' " (Chatfield, 2004, p. 12).

iii) "A trend is a persistent long-term movement of a variable over time" (Stock and Watson, 2015, p. 597).

iv) "By trend we mean the low-frequency variability evident in a time series after forming moving averages such as low-pass (cf. Baxter and King, 1999) or Hodrick and Prescott (1997) filters" (Müller and Watson, 2017, p. 53).

A common trait of these definitions is their vagueness. Moreover, while containing a flavour of the forecasting approach, definition ii) seems to imply that a trend necessarily requires some differencing of the data to obtain their "change". On the other hand, definition iv) easily lends itself to a question begging type critique. Actually, it even inverts the problem, subordinating the definition of trend to the operation of a few questionable detrending filters. That is, in the ironic words of White and Granger (2011), trends are defined as "that which trend filters remove".

Given the difficulties that have already been mentioned in this introduction, this absence is hardly surprising. But these difficulties arise also themselves from the absence of a sufficiently precise and commonly accepted definition: "is it possible to measure and discuss with clarity any quantity that is undefined?" (Phillips, 2010b). That is, causality runs in both directions.

A further type of circularity lies behind this indeterminacy. Part of the problem appears to be originated in the ownership of the concept. The concept of trend belongs to both economics and statistics and a clear agreement between the two areas appears ever more distant. Econometricians complain that they were not provided with enough guidelines, that economics has not proposed a sufficiently tight theoretical framework of what a trend is, and economists claim that they need more statistical evidence about the phenomenon to be able to do just that (see, e.g., Canova, 1998).

It is also likely that the common perception of trend has changed through time and continues to change. For instance, it is rather implausible that trends that were observed in living conditions for many countries in certain periods of the 20th century are similar to those trends that were first mentioned in economics, by the end of the 19th century. Even the economic variables to which most trends are referred to are likely to be very different. The absence of a tight definition thus allows that similar, though far from exactly equal phenomena, carries the same qualification, which is a positive feature. Perhaps this absence is more beneficial than negative.

Meanwhile, I believe that the following informal, intuitive and graphical description is the best that one can find about what a trend is (or should be):

> A trend line summarizes these primitive requirements. It summarizes where we have been, shows where we are now in relation to the past, and most of all, reveals a hint of where we are going (Phillips, 2010b).

1.8 Outline of the remaining chapters

Frickey's (1934) list, briefly presented in the preface, is also useful as a brief guide to the remaining chapters of this book.

Adopting a clearly dominant statistical perspective, this book is not concerned with the last two items of that list: causality analysis and economic theory. And as regards "historical description", Frickey is obviously referring to the case of each specific trend and to the work of historians in helping economists and statisticians in the explanation and statistical specification of trends. Related to this perspective, the next chapter provides a historical account covering the pioneering work on trends and the progression that followed until the 1960s. More recent developments are presented with more detail in the following chapters, albeit usually not using a historical perspective.

The first item of Frickey's list, that of "mathematical curve fitting", is addressed in chapter 3, which I consider the most important of the book. In the absence of a clear and tight economic concept of trend, the statistical specifications try to fill the void offering a variety of different ways to think about trends, though some of them might be seriously flawed in some cases. This chapter provides a view to those statistical windows. Notice that some of the models for the trend are presented only in chapter 4, as they are often used as detrending methods.

Chapter 4 addresses "the problem in statistical description" which is the second item of Frickey's list, resuming the issue of detrending previously introduced. The framework is that of decomposition methods and the second equation of the book is the basic departure point for these methods, which aim to isolate or segregate the trend in order to remove it from the data, thereby providing an indirect method for its estimation. The famous Hodrick-Prescott and the recent Hamilton filter are presented and discussed in this chapter, as well as the Beveridge-Nelson decomposition and the unobserved component models.

Chapter 5 contains a different approach to the "problem in statistical description", that of statistical testing. Assessing whether a trend is really present in the data is often the first problem that practitioners face

but it will be the last to be considered here. Since I do not address tests for stochastic trends here, the statistical approach is centered around two major themes, those of testing for linear and nonlinear trends. Linear segmented trend models serve as the most simple version of nonlinear ones. There is some intersection with the literature on structural breaks that is kept limited.

In a separate appendix I provide a small introduction to frequency analysis destined to help readers understanding some of the filters of chapter 4.

Chapter 2

Historical Perspective

It is not my intention to describe in detail, thoroughly, all the historical evolution of the notion of trend and of its practical implementation. Readers interested in such details should consult, e.g., Morgan (1990) and Klein (1997), as well as Shook (2020). Instead, in this chapter I will focus almost exclusively on the initial stages of the historical process. Later, in chapters 3 and 4, most of the models that have been proposed more recently will be presented in some detail, allowing the reader to fill some gaps of the picture of the historical process.

This historical digression may be also useful to understand current debates and to shed some light on the (un)desirable features of trends in macroeconomic time series. It is expected that it will be useful as well to understand the difficulties with the adoption of a single notion of trend and hence the imprecision of the its most common meaning.

2.1 Secular trend

Even today the word trend is often preceded with the adjective *secular*. Does this means that it is a non-religious or even an anti-religious trend? Obviously not, although this is the common meaning of *secular*. The meaning of secular in this context has nothing to do with this usual meaning, of worldly, of mundane or ordinary as opposed to religious or

spiritual; it is not in any way connected with religion[1].

Rather, the meaning is associated with the latin word *saeculum*[2], which means century, originated in the "*Ludi Saeculares*", the games performed in ancient Rome to celebrate the advent of a new century (whose length could be 110, not 100 years). Starting with astronomy, science appropriated the term to refer to some phenomenon slowly evolving in time, requiring a long period of time to be observed.

Shook (2020) provides an interesting and somewhat detailed description of this research in astronomy, the term secular being used to denote some very slow change or *variation* (sometimes also mentioned as "inequality") in the position of celestial bodies. According to Shook, the first tables with the position of the sun appear to have been published by Erasmus Reinhold (1511-1553), a german astronomer. Although he did not use the term, his tables for the equinoxes from 20 AD to 5000 AD contained a column for the century. Later, in his *Rudolphine Tables* (1627), Johannes Kepler referred to a correction that was necessary to introduce in the previous estimates of the position of the sun and the moon from 4000 BC to 2100 AD as a "*saecular equation*" (*æquationibus secularibus*).

2.2 First references in economics

Before arriving to economics, the idea about the possibility of secular changes was adopted in the natural sciences and, in particular, in geology (Shook, 2020).

[1] According to Shook (2020), this common meaning of the term was borrowed from medieval christians to mention those clergy with duties outside the monastery: "... in the history of the Christian church, some clergy such as the local priest, were called secular because they served parishioners out in the world while others remained cloistered"(Stackhouse, J. G., Religious Diversity, Secularization, and Postmodernity", *in* the Oxford Handbook of Religious Diversity, cited by Shook (2020)).

[2] Although also derived from latin, the word century derives from *centuria*, not from *saeculum*; but siglo (spanish), secolo (italian), secol (romanian), siécle (french) and século (portuguese) are.

In economics the notion of trend is closely connected with the idea of the decomposition of economic time series into separate components according to the type of movement. Hence, to search for the notion of trend is to search for the origin of this idea, as in Nerlove, Grether and Carvalho (1995). And similarly with the *secular* qualification, it emerges that economics borrowed this idea of decomposition from the astronomers of the 17th and 18th centuries, who had applied it to calculate the orbits of planets. In other words, decomposing economic time series into separate components is an idea that is rooted in the work of such scientists as Euler, Lagrange and Laplace.

Although he did not mentioned explicitly the trend component, Augustin Cournot is credited as the first adopter of this idea in economics. Actually, when discussing the price of wheat he says (Cournot, 1838, p. 25):

> Here, as in astronomy, it is necessary to recognize *secular* variations, which are independent of *periodic* variations.

Cournot clearly embraces not only the existence of different types of variations in economic variables observed through time but also the independence between those different variations.

The work of Cournot is supposed to have been read by William Jevons. But according to Shook (2020), Jevons must have been influenced as well by Henry McLeod and by Dionysius Lardner, two political economists who pioneered the application of the astronomy approach to economic time series. In particular, McLeod likens the change in the value of money through time with the slow and persistent movement of distant celestial objects:

> Just as in astronomy the changes of the position of the heavenly bodies, which are at a very great distance from us, are barely perceptible after long continued observation ... and (...) are termed "fixed". So for short periods, the value of money may be considered as fixed ...; and changes in the value of money may be compared to the secular variations

> of the heavenly bodies (McLeod, H., 1858, *The Elements of Political Economy*, cited by Shook, 2020, pp.5-6).

Although mostly known for his theory explaining economic cycles with cyclical sunspots[3], Jevons is credited as "the first to study the problem (of economic fluctuations) seriously in a variety of contexts" (Nerlove et al., 1995, p.9). Sometimes, not often, he refers to secular variations (Jevons, 1878, p. 61):

> the variation is called secular because it proceeds during ages in a similar manner, and suffers no $\pi\epsilon\rho\iota o\delta o\varsigma$ or going round.

His methods, derived from astronomy and meteorology, were carefully designed to estimate the periodic variations in economic and "commercial" fluctuations. But he did not considered these the most important, quite on the contrary:

> Every kind of periodic fluctuation, whether daily, weekly, monthly, quarterly, or yearly, must be detected and exhibited, not only as a subject of study in itself, but because we must ascertain and eliminate such periodic variations before we can correctly exhibit those which are irregular or non-periodic, and probably of more interest and importance (Jevons, W., 1862, On the Study of Periodic Commercial Fluctuations, cited by Shook, 2020, p. 5).

Contrarily to such classical predecessors as Malthus, Ricardo and Marx, *inter alia*, Jevons did not considered long-run tendencies as particularly relevant. He was mainly interested in irregular variations.

2.3 First measurements

Reginald Hooker (1901) did not only coined the term *trend* but he was also the first to propose a measurement of a trend, i.e., an empirical

[3]Morgan (1990) provides a rather interesting description in pages 18-26.

estimate of a trend in a particular case. Inaugurating a much followed route, he used the estimated trend not to analyse it but to detrend the raw data to uncover the real relation between two variables represented by their "oscillations", i.e., their deviations from the trend.

The problem at stake was that of examining a possible relationship between the "marriage rate" and economic prosperity (represented by external trade) which, according to Klein (1997), was conjectured by Thomas Malthus in 1798. Hooker's (1901) suggestion was of

> an elementary method of eliminating the general movement in the particular case — which is of frequent occurrence — of phenomena exhibiting similar regular periodic movements, so as to enable us to correlate the oscillations.

Apparently inspired by a simple (moving average, MA) method previously used by the Bank of England to obfuscate the public divulgation of its cash and bullion holdings after 1832,[4] Hooker (1901, p. 486) proceeded to subtract from each observation its "instantaneous average at the moment":

> The curve or line representing the successive instantaneous averages I propose to call the *trend*. Any point on the trend will be represented by the average of all the observations in the period of which that moment is the central point; *e.g.*, if a curve shows a period of p years, the instantaneous average in any year is the average of the p years of which that particular year is the middle. By working out this instantaneous average for consecutive observations, we obtain the trend of the curve; *i.e.*, the directions in which the variable is really moving when the oscillations are disregarded.

It is better to consider that Hooker's method is mostly a smoothing algorithm designed to remove periodic variations rather than really one

[4]See chapter 4 in Klein (1997) for a detailed description. Calculating moving averages had also been previously introduced in statistics by John H. Poynting (1884) to "free" the data "from accidental irregularities" (p. 35).

for estimating a long-run or *secular* trend. Hooker argued that the average length of British's exports cycle was of nine years and therefore he calculated simple nine year moving averages. But his method could not be applied to series with no periodic variations and in 1905 he proposed an alternative method to that case.

The case at hand was now that of measuring the strength of the relation between the prices of corn at different markets and Hooker's concern was to avoid "*erroneous*" correlations arising from the presence of trends, that is, of what he now called *"secular" changes due to "world influence"* (Hooker, 1905, p. 698). As Jevons, he was mostly interested in the irregular components, that he called the *"smaller rapid changes"*, and according to him (p. 696),

> the "secular", in fact, may entirely mask the other changes...

Although the concept of stationarity did not even existed yet, he proposed to difference the series to remove their trends before calculating the correlation coefficient. At least apparently independently and although his motivations were not so clear, Lucien March (1905) proposed the same method. He is also credited for being the first to refer explicitly to the term "decomposition". Moreover, he clearly justified the lack of interest in the secular trend (March, 1905, p. 270):

> La plupart des statistiques sont de date trop récent pour que l'on ait à s'occuper des changements séculaires ...

A method close to Hooker's (1901) was proposed by John Norton (1902). Instead of moving averages, Norton proposed to use graphical interpolations, that he called "*growth axis*", for three banking series (reserves, deposits and loans). He never mentions neither the word *trend* nor even the more common *secular change*, but his growth axis, "*three steadily ascending smooth curves*", performed the role of trends, serving the detrending purpose.

Norton does not also mention the term decomposition but the idea clearly surfaces in his (*o. cit.*, p. 23) "*three elements at work or elements of the polygons"* (the plots of the time series):

a) "*the growth element*";

b) "*periodic elements"*, which is what he calls to the seasonal fluctuations, and

c) a "*dynamic element*", including "*a cycle*", "*a catastrophe, i. e. a crisis with violent change*", and "*minor dynamic changes*".

2.4 The rise of business cycles

After the appearance of the first ideas, the period that followed was not one of any significant advancement concerning trend analysis. Instead, business cycles became the main and almost exclusive focus of study and research. Some technical advancement in estimating trends was made but, as summarized in, e.g., Mills (2009), although sometimes considered as "*fundamental*", trends were estimated with the only purpose of isolating them, so that they could be removed.

Several reasons might explain this. First, statistical analysis of economic time series were much inspired in the methods used in meteorology; but since at the time trending climate time series were very rare, the help from meteorology was unavailable (Klein, 1997, pp. 228-9). Second, because during this period the U.S. faced a deep economic crisis, much deeper than in Europe or Canada, *"the problem (of business cycles) was especially intriguing to Americans"* (Wesley Mitchell, 1927, p. 201). Moreover, Mitchell (1927) also points out that in the U.S. available data for business cycles analysis was accumulating and became abundant. Finally, still according to Mitchell (1927), theoretical advancements needed to explain and interpret trends was missing.

At the time, this *detrend and forget* approach was based on a(n implicit or explicit) view that there was no relation between the forces of trend growth and those lying behind the shorter run fluctuations. For instance, Kitchin (1923) claims that trends "*are doubtless dependent upon the changing amount of the world's total money*", while "*business and price cycles are due to cyclical recurrences in mass psychology reacting*

through capitalistic production". As Mills (2009) emphasizes, "*instability in the trend component and/or the use of an incorrect procedure for detrending will complicate the separation of trend from cycle*". This would become clear only much later and, even today, it has not became common or popular knowledge.

I consider that this period roughly corresponds to the first 30 years of the twentieth century. The famous clear enunciation on the decomposition of (macro)economic time series dates from this period, and I find the version by Warren Persons (1919, p. 8) the most revealing[5]:

> ... each series is a composite consisting of four types of fluctuations. The four types are:
> 1. A long-time tendency or secular trend; in many series, such as bank clearing or production of commodities, this may be termed the growth element;
> 2. A wave-like or cyclical movement superimposed upon the secular trend; these waves appear to reach their crests during periods of industrial prosperity and their troughs during periods of industrial depression, their rise and fall constituting the business cycle;
> 3. A seasonal movement within the year with a characteristic shape for each series;
> 4. Residual variations due to developments which affect individual series, or to momentous occurrences, such as wars or national catastrophes, which affect a number of series simultaneously.

Mitchell (1927, p. 213) lists five methods usually employed in this period to detrend data from which I select three:

1. "*fitting a mathematical curve" (for example a straight line or a third-degree parabola) to the data, or to the logarithms of the data by the method of least squares or of moments"*;

[5] Alternative but similar formulations may be found in, e.g., Moore (1914, p. 106), Yule (1921) and Kitchin (1923).

2. computing moving averages with the number of data points seeming "*to give satisfactory results*", and

3. "*drawing a free-hand curve trough the data representing the investigator's impression*".

Therefore, the still very popular linear deterministic trend model fitted to the y_t series,

$$y_t = \alpha + \beta t + u_t,\ t = 1, 2, \ldots, T,$$

u_t usually representing a zero mean stationary process, dates from this period and it is the simplest example of the methods in 1.

Although this model is certainly naive, some of its limitations were acknowledged at the time. In particular, Persons (1919) made extensive comments about the possibility that a "*trend satisfactorily determined for a past period does not always warrant us in extending the same trend into the present and the future*" (p. 9). In other words, he emphasized the period-dependency of fitted trends and warned about the possibility of their non-homogeneity, particularly for extended periods, with a sentence that shows that the analysis of structural breaks in time series is likely to be much older than we could think:

> If it is so long as to include heterogeneous subperiods of widely different fundamental conditions, the problem of determining the secular trend is greatly complicated and its solution of questionable value.

The theory of correlation has been also introduced not many years before this period, where a much debated issue was that of the "time-correlation problem", as it was designated by Udney Yule (1921). Simply put, the problem was of large correlation coefficients between series whose only common characteristic was the presence of a trend. Yule (1921) summarizes the problem but, contrarily to what is nowadays the popular perception about his work:

a) denied that the problem was one of a spurious or misleading correlation, and

b) criticized the differencing practice recommended by "Student" and Anderson and instead used the issue to further strengthen the case to apply detrending methods.

Again it must be stressed that the concept of stationarity was not even yet formulated. Instead, Yule (1921, 1926) preferred to attack the problem with the harmonic analysis approach. On the other hand, it is easy to recognize in b) the origin of the recent debate between difference stationarity and trend stationarity.

The original version of a famous article by Eugene Slutsky (1937) dates also from this period[6]. It represents an extreme but representative example of the issues considered relevant at the time: while there is no reference to economic trends, much attention is devoted to the demonstration that the accumulation of random shocks may produce cyclical processes.

The procedure of *graduation*, that will reappear later in the context of the popular Hodrick-Prescott filter, dates also from this period. It was introduced in 1904 by John Spencer[7] as a procedure to obtain smoothed estimates of mortality rates and it consists of an iterated application of several moving averages. Denoting with u_t the original series, Spencer's 21-term graduate, or smoothed of filtered series is

$$\begin{aligned} v_t \;=\; & 0.148u_t + 0.138(u_{t-1} + u_{t+1}) + 0.122(u_{t-2} + u_{t+2}) + 0.097(u_{t-3} + u_{t+3}) \\ & +0.068(u_{t-4} + u_{t+4}) + 0.039(u_{t-5} + u_{t+5}) + 0.013(u_{t-6} + u_{t+6}) \\ & -0.005(u_{t-7} + u_{t+7}) - 0.015(u_{t-8} + u_{t+8}) - 0.016(u_{t-9} + u_{t+9}) \\ & -0.011(u_{t-10} + u_{t+10}) - 0.004(u_{t-11} + u_{t+11}), \end{aligned}$$

[6]The original version, in russian, dates from 1927. Although page 127, on section 6, contains a brief description of a covariance stationary process, according to Morgan (1990) all the sections that followed section 5 were rewritten for the english version.

[7]On the graduation of the rates of sickness and mortality presented by the experience of the Manchester Unity of Oddfellows during the period 1893-–97, *Journal of the Institute of Actuaries*, 38, pp. 334–343.

which can be represented more elegantly resorting to a specific notation, where, for instance, $[2k+1]u_t$ represents the sum of $2k+1$ observations of u_t centered on u_t. Another well known filter is the Henderson's 23-term MA, designed to minimize the variance of the third difference of the filtered series; see, e.g., Mills and Patterson (2015).

2.5 Maturity at 30

A significant change concerning the analysis of trends appears to have occurred around 1930. Both the concept of trend and its measurement became a subject of interest in themselves, a more thoughtful business cycle analysis emerged, the detrend and forget approach of business cycle statisticians began facing some criticism and gaps in economic theory relating to trends and to business cycles were identified and became clearly visible.

Obviously, this partition is not completely precise. On the one hand, the book by Simon Kuznets (1930), already in this period, is mostly an extensive sample of the practices of the previous period, containing a much disaggregated analysis of many economic time series (for many sectors and for several countries)[8]. On the other hand, although still dated from 1927, the second of Mitchell's books contains some important reflections and criticisms about those practices.

Indeed the change is already perceptible in this book, differing from the first one, from 1913, in many respects. Even the change in the wording and in the meaning of expressions concerning the trend is clear: while in the first book *trend* is used with no qualification, with its most usual or common meaning and it is often preceded by "general" — e.g., "*the general trend of the market*" (p. 171) —, in the second the meaning is often statistical and in this case *trend* is preceded by *secular*, an expression that was never used in his 1913 book. In the second book, recognizing the

[8]Perhaps my observation is overly demanding because, according to Mitchell (1927, p. 213, ft. 1), "*so far as I know, the only one working upon secular trends as a problem in its own right is Dr. Simon Kuznets...*".

scope and complexity of the subject, Mitchell (1927) summons economic theorists, economic historians and statisticians to address the problem of trends. His criticism to the prevailing approach is clear (Mitchell, 1927, p. 212):

> secular trends of time series have been computed mainly by men who where concerned to get rid of them.

This change towards trends appears to attain its zenith in Arthur Burns' (1934) book, where a method to *decycle* data, i.e., to estimate, isolate and remove cyclical fluctuations is described with the purpose of analysing the trend.

Nevertheless, the change was neither revolutionary nor completely widespread. For instance, despite several attempts or approximations the concept of trend remained rather imprecise. Burns (1934, p. 30) admits that "*there is probably no concept in the whole field of contemporary 'quantitative' economics that is vaguer than that of secular trend*". One of the best definitions appears to be the one by Mitchell (1927, p. 230) precisely because it admits some ignorance:

> ... lines of secular trend show the effects of causes which, though subject to change at any moment, have influenced an economic process in some regular, or regularly changing, way through periods of time long in comparison with business cycles. What these causes have been, and whether they are still in operation, are matters for further inquiry.

Despite its negative character, Burns (1934) would identify with its essence defining trends as "*non-cyclical movements*". He made, however, several comments attempting to improve its precision, e.g., rejecting the idea that trends might represent "*equilibrium paths*", and stating that they "*measure growth*"[9].

[9]The difficulty and complexity of the problem explains this comment by Burns (1934, p. 32), important but contradictory with Mitchell's words and with some of its own: "*this concept is free from the question-begging notion that secular trends are measures of the effects of causes independent of those generating the cyclical, seasonal, and random variations in production series*".

One of the most important contributions of the time is the one by Edwin Frickey (1934), which I believe is worthy of detailed attention. Frickey begins by listing the "*various points of view*" concerning "*the meaning of secular trend*", classifying them as "*problems*": 1) "*in mathematical curve fitting*"; 2) "*in statistical description*"; 3) "*in historical description*"; 4) "*in the analysis of causation*" and 5) *"in economic theory"*.

The first problem partially derives from the absence of a precise definition of trend; its notion is reduced to the idea that a trend is something that should evolve slowly and smoothly with time[10]. Therefore, one should not aim that the trend line fits closely the observed data. Instead, "*we must "choose a curve which reproduces the underlying movement of the data without bending or twisting itself so as to conform to the extreme sinuosities, and which at the same time gives a good fit as judged by some arbitrary criterion ..."* (Frickey, 1934, p. 199, citing Henry Schultz, 1928, *Statistical Laws of Demand and Supply*, Chicago).

As Frickey notes, however, these requirements are contradictory and some "*arbitrary judgment is required in effecting a compromise between them*" (p. 199). That is, to what line should the fitted trend adjust? How much bending is allowed? How can the closeness of the estimated trend to the "true" one be assessed? Which criteria should be used?

The problem in statistical description is also closely related to the definition of trend, which is essentially statistical, performing the role of "*describing a general tendency*". But this requires that non-homogeneous periods are not combined and that the cyclical movements at the beginning and end of the sample period should not be allowed to distort the calculations. Moreover,

> such a statistical description can have little or no value unless the calculations are preceded by an extensive casual analysis of the forces behind the movements of the series — such analysis to be obtained through a theoretical examination of the causal forces in question, or through historical investigation,

[10]Even these requirements are somewhat subjective and arbitrary.

> or both; and still further, that in the absence of an exhaustive analysis of this sort, the determination and representation of secular movements should not be undertaken" (p. 199).

However, Frickey himself offers an alternative to such a radical decision: in case that causal knowledge is missing, *"there is still room which starts with the statistical summarization and perhaps works back from that summarization".* And Frickey finds admissible an even less ambitious purpose: *"there is nevertheless justification for the employment of a procedure which will yield the answer to some significant question".* One of these purposes may be only the calculation of the average rate of growth. Or it can be to fit a series with a "*declining percentage rate"*, or a bit further in this case, to measure the "*average rate of retardation per time unit over the period covered".* For each of these cases Frickey recommends a particular functional form of the trend function.

The third problem is that of historical description, which represents another perspective of the trend problem. Here it consists in writing, in the form of "*lines or curves*", the "*best description possible of the fundamental tendencies*" represented by "*historical materials*", these consisting of economic time series "*suplemented by certain non-statistical information*" (p. 201).

A fundamental question then arises at this point:

> "What part of the fluctuation of the series is secular and what portion is cyclical?" This question cannot be evaded, for our computed representations of secular and cyclical movements are palpably interdependent (p. 201).

An obvious statement clearly presented but often neglected, which originates another difficult question:

> ... every set of lines and curves labeled "secular trend" and "cyclical fluctuations" represents the use of someone's discretionary judgment. This poses the question: "On what basis should this discretionary judgment be made?" (p.201)

To demonstrate empirically these points Frickey provides two illustrations, both using the series for pig iron production in the U.S. . In the first he collects 23 published studies with different time periods and with the trends fitted with several methods — the linear being the most frequent but including also, e.g., the logistic and various parabolas — to derive the corresponding cycles. As expected, the average length of the estimated cycles presents a very wide diversity, ranging from 3.3 to 40-45 years. Since, as also expected, the results also vary a lot with the sample period, Frickey fixes this period to 1877 to 1914 and repeats the exercise. The dependence on the method becomes more clear: again the average length of the cycles exhibits a wide diversity, ranging from 3.6 to 40-45 years.

In the second exercise Frickey presents his own calculations using trends estimated with several centered moving averages, ranging from 3 to 20 years periods. Again the estimated cycles exhibit great variation, with their averages ranging from 3.5 to 10.1 years when the sample is fixed in the 1854–1933 period. Obviously, the dependence is not confined to the average lengths. As Frickey (p. 205) notes, "*the whole form of the supposed cyclical picture*" may vary a lot as well, *"depending upon the kind of secular trend which was previously been fitted."*

The fourth problem, that of "*the analysis of causation*" consists of the rationalization of the statistical decomposition which is made through the correspondence established between the empirical components and particular economic causes or forces. More precisely,

> in computing lines of secular trends and curves of cyclical variation we are trying to separate, by statistical devices, the effects of two sets of causal influences — one set operating in general gradually and over comparatively longs periods of time; the other set operating in a more oscillatory manner and producing fluctuations of shorter length (p. 201).

While this transfer to a causation problem could be helpful to solve the general problem, in practice it faces several objections and criticisms that Frickey himself provides. One of the simplest is that "*the rate of change*

exhibited by the secular trend is in part a result of cyclical variations"(p. 201). A mild one is that this "*attempted separation of causal influences into two types (...) very much oversimplifies a problem which is in fact highly complex*"(p. 202). Nevertheless, Frickey maintains that this point of view should not be discarded.

The fifth and last problem, in economic theory, resembles the previous one but Frickey rejects that they are identical on the basis of a "*lack of unanimous agreement among economists*" on the causal analysis. From my understanding it is far more general, and for clarification Frickey resorts to Schumpeter[11], whom he cites (p. 203):

> "if trend-analysis is to have any meaning, it can derive it only from previous theoretical considerations, which must not only guide us in interpreting results, but also in choosing the method. Failing this, a trend is no more than a descriptive device summing up past history with which nothing can be done. It lacks economic connotation. It is in fact merely formal".

Indeed, in the absence of a sound economic interpretation, an estimated trend will resemble a body with its skeleton and skin but no flesh and blood to fill in.

Although dated from the same year as Frickey's important paper, an article by Working (1934) is also related with the trend topic, albeit much more tenuously. Working was a forerunner of processes integrated of order one ($I(1)$) or more simply of martingale processes, that he labels as "*random-difference series*", and which are "*cumulations of purely random changes*". He notes the resemblance of these series with price series and, although he never arrives to mention that they contain *stochastic trends*, he emphasizes that "*in these series conspicuous trends will be found*". His intention appears to be to note that such processes give rise to several easily observed local trends, not to global trends.

[11]Schumpeter, J. (1930), Mitchell's business cycles, *Quarterly Journal of Economics*, 55, pp. 166-67.

2.6 Some Further Developments

This is a miscellaneous section, without any pretence of unity or completeness. I will briefly review two important works on business cycle analysis, those of Frickey (1942) and Burns and Mitchell (1946), the notion of the "great ratios of economics" of Kaldor (1957, 1961) and Klein and Kosobud (1961), and the important but not well known contribution of Leser (1961) to filtering (graduation) methods.

Two books on business cycles in the 1940s

The books by Frickey (1942) and Burns and Mitchell (1946) provided a new but somewhat skeptical impulse to business cycle research. Although unsurprisingly the book by Frickey is more inclined towards studying the trend, the ultimate purpose is the same and there are several common aspects in both books. As expected, particularly in Frickey's book, in what concerns detrending both diverge from the "traditional method" [12]. This was made much more carefully, much less mechanically, then was current practice particularly in the 1920s.

In the case of Burns and Mitchell, this was not due to any significant increase in interest in the trend. Instead, it became ever more obvious that detrending could affect dramatically business cycle analysis. Indeed, Burns and Mitchell (1942, p. 270) acknowledge that

> cyclical fluctuations are so clearly interwoven with these secular changes in economic life that important clues to the understanding of the former may be lost by mechanically eliminating the latter.

Frickey (1942, p. 256, and later also Burns, 1944) is (are) particularly vigorous attacking the underpinning argument of (different and) independent causes: "*we should think here, not in terms of distinct sets of causal forces, but in terms of lines of causal influence*". This is not to say

[12] "Curve fitting", mostly with the linear or the exponential trend model, and using moving averages were considered the two traditional methods.

that these lines cannot be treated separately. Rather, as Benjamin Higgins (1955, p. 595) would later emphasize: *"the essential causal factors of fluctuations and growth can, and should, be analysed separately"*.

Burns and Mitchell are particularly cautious with detrending[13], using previous knowledge about the cycles and only then choosing a conforming trend, i.e., "*making the trend depend upon the cycles"*. Moreover, they make no attempt to remove that portion of the trend lying "*within the limits of a single cycle*"., i.e., they detrend each series eliminating the inter-cycle portions of the trend, but not the intra-cycle ones, so that cyclical properties are the least distorted by detrending.

Frickey (1942) goes even further and devises a rather elaborate, complex and iterative process, with successive approximations to the trend and cyclical components, marked from the outset with an inversion in the traditional orders of detrending and "decycling": cyclical variation is removed first and the trend is obtained as a residual. Moreover, as he had recommended in his 1934 paper, he attacks "*the problem as a unified whole*". This means that after selecting 13 important series he searches for "*consistencies and uniformities of behavior*" (p. 53), allowing him to arrive at a *"standard pattern"* representing the cycle. Only then the trend is obtained in a first stage of an iterative process where, for each series, both the standard pattern and the trend are recalculated at each step. The process stops when the revision in the standard pattern is insignificant.

After a few more transformations, trends are fitted to the series but for only two a "*simple mathematical curve"* provides reasonable results. For the remaining, there are "*few, if any, for which the secular variations are capable of satisfactory representation by a single simple mathematical curve"* (p. 292). That is, it is necessary to split the sample period (1866–1914) into subperiods and fit a separate trend to the (logarithms) of the data for each subperiod. Some series require only two different trend functions, which may differ even in the functional form, but others require

[13]And caustic to many others as well: *"a least squares trend line (...) may move majestically on a chart, but the analytic significance of the trend line is obscure"*(p. 38).

three or even four subperiods. In summary, rather than the exception, broken trends are the most frequent outcome of trend modeling.

As is well known, with a celebrated paper Tjalling Koopmans (1947) initiated a debate that came to be known by its title: "*measurement without theory*". In a few words, Koopmans criticized the empiricism of business cycle researchers, and particularly the work by Burns and Mitchell. The following is only one of the many examples (p. 161):

> The various choices as to what to 'look for', what economic phenomena to observe, and what measures to define and compute, are made with a minimum of assistance from theoretical conceptions or hypotheses regarding the nature of the economic processes by which the variables studied are generated.

Although there is some truth in this, Koopmans considered that Burns and Mitchell should base their work on "*structural equations*" modeling the economic behaviour of individual units. That is, for Koopmans the only valid approach was the one of the Cowles Comission.

Without entering in much detail, what appears to pass unnoticed was that Koopmans' eagerness to criticize led him to adopt the empiricist views he so violently attacked. This is the case when he accused Burns and Mitchell of analysing the change in the cyclical behaviour of pig iron production and freight car orders "*without prior elimination of secular trend*" (p. 168). That is, he invoked an empirical, poorly defined concept, and an even less theoretically based practice to support his criticism.

Later, Vining (1949) would present several counter-arguments "*in defense of empiricism as a fundamental part of scientific procedures*" (p. 79), as well of, e.g., the aggregate or macroeconomic perspective that "*Commissar Koopmans*" (p. 85) so violently rejected. According to Kydland and Prescott (1990), however, in the 1950s and 1960s the change of focus to the modeling of structural systems of equations advocated by Koopmans led to the abandonment of research on business cycles.

"The great ratios of economics"

This is only a brief and momentary incursion in multivariate analysis and its motivation has nothing to do with the linear trend model which, incidentally, will appear below. Instead, it arises from the growth modeling, long-run perspective that is involved here, therefore closely related with the purpose of this investigation. Moreover, this research is also considered as a forerunner of the important topic of decomposition, i.e., of long-run equilibrium relationships between (macroeconomic) variables. In this case, it is confined to pairs of variables only.

The "great ratios of economics" were mentioned firstly by Nicholas Kaldor (1957) as "*remarkable historical constancies*" and later in Kaldor (1961) again, as "*stylized facts*". The title adopted here is due to Lawrence Klein and Richard Kosobud (1961, KK61), who also label them as the "*celebrated ratios of economics*", and use them to build a growth model. Following KK61 these ratios are:

a) the savings-income ratio;

b) the capital-output ratio;

c) labour's share of income;

d) income velocity of circulation, and

e) the capital-labour ratio.

Considering, for instance, the first one, a decomposition interpretation may be readily available provided that the two variables are integrated of order one $(I(1))$[14].

KK analyse the stability of the ratios for the 1900–1953 period running regressions of the logarithmized ratios on a deterministic linear trend

[14] A recent survey on the empirical evidence about this perspective is provided by Chudik, Pesaran and Smith (2022).

and focusing on the (in)significance of the trend terms. For instance, considering the first two, they get

$$\widehat{\log\left(\frac{C_t}{Y_t}\right)} = -0.039 + 0.00054\,t, \text{ and}$$

$$\widehat{\log\left(\frac{K_t}{Y_t}\right)} = 0.547 - 0.0015\,t,$$

where C_t denotes consumption, Y_t national income and K_t the capital stock.

Although Kaldor (1961) denies that these ratios can "*be plausibly explained by the theoretical construction of neo-classical theory*", according to Smith (2019) there is some confusion about them in the literature. Often it is not clear whether they should be regarded as empirical facts that growth models should match or as theoretical conditions for steady state growth. The empirical evidence gathered by KK does not lend much support to the first possibility because most ratios appear to be trending. In their "numerical model", estimated equation-by-equation and containing trends, the rate of growth of income varies (declines) over time.

Leser and the "Hodrick-Prescott filter"

Mills (2009) considers that 1961 was an "*annus mirabilis for the modelling of trends and cycles*". Besides the paper by Klein and Kosobud, three more papers published in 1961 address these issues: one on exponential smoothing by David Cox, one on Kalman filtering by Rudolph Kalman and Richard Bucy, and the paper by Leser (1961), on trend extraction or smoothing[15].

The decomposition model of Leser is very simple because the series

[15]To locate this work on the topic of graduation methods see, e.g., Phillips and Jin (2021).

y_t is decomposed only in the sum of a trend f_t and a disturbance c_t,

$$y_t = f_t + c_t,$$

the first component capturing the low frequency or long-run growth of the series and the second the business cycle component. Leser proposes that the trend is estimated solving

$$\hat{f}_t = \arg\min_{f_t} \left\{ \sum_{t=1}^{T} (y_t - f_t)^2 + \lambda \sum_{t=2}^{T-1} \Delta^2 f_{t+1} \right\},$$

that is, using a penalised least squares principle. The first term of the minimand aims to maximize the goodness of fit while the second penalizes departures from smoothness, λ denoting the smoothing parameter. The estimate of the cycle is obtained residually: $\hat{c}_t = y_t - \hat{f}_t$. This is a filter that is very close to the one by Whittaker, proposed in 1923 as a refinement of those of Spencer and Henderson mentioned previously, but in Whittaker's case the penalising term is made of third differences of the trend.

Leser called the resulting estimated trends as a family of "quasi-linear" trends" because the linear trend case emerges when $\Delta^2 f_t = 0$. He also presents rather laborious solutions for some particular cases with very small sample sizes and proposes an approximate solution for $T \geq 8$ but this should not concern us here. What must be emphasized is that Leser's method is precisely the method that 20 years later would be called the Hodrick-Prescott (filtering) method (see section 4.3).

Chapter 3

Modeling the Trend

This is perhaps the most important chapter of the book: in the presence of a trending time series, the primary purpose must be to identify the type of trend, i.e., the appropriate model (using also tools from hypothesis testing), to estimate it and to describe and explain it. But a model for the trend may also be estimated in order to subtract it from the data when the purpose is to analyse business cycles (i.e., growth cycles) and/or seasonality. Moreover, lacking a precise and widely accepted definition of trend, browsing the menu of the available functional forms may provide useful insights about its behaviour throughout time and foster a reflection about the properties that the model must exhibit, as well as motivate a further study along the lines of economic history.

A warning for less experienced practitioners concerns the goodness of fit. It is by now widely acknowledged that, despite fitting many time series rather well, the linear trend model may be a spurious representation of the data (see section 3.8) and may produce dramatically wrong forecasts (see the next section). But, though not so well documented, higher order polynomials in the time regressor may be even more misleading due to their sometimes amazing fit. The no less amazing extremely poor quality of ensuing forecasts usually follows and hence these models should be very rarely considered in empirical work.

A further warning concerns linear segmented trend models. Although

it is often relatively easy to find breaks in linear trending models, as well as reasons to impose them, practitioners must use of parsimony in specifying the number of breaks, at the risk of ending up with a meaningless model. As White and Granger (2011) advise, "chopping practically any time series into a succession of trend breaks" is "unlikely to be informative". Again, goodness of fit is rarely the best criteria to select a model for the trend.

3.1 The linear and the exponential trend models

The (deterministic) linear trend model corresponds to the most traditional and popular notion of trend, both graphically and analytically. Facing a time series plot with distinct values for the terminal points, most economists immediately draw with their eyes a (virtual) straight line joining those two points, increasing or decreasing with time, even when the series exhibits large, violent and possibly very erratic fluctuations in between. Actually, for many of us, a linear approximation to the behaviour of a series that most often grows with time is the best intuitive synonym of what a trend is.

It is also the oldest and simplest model for the trend[1], its best representative (already previously introduced) is given by

$$y_t = \alpha + \beta\, t + u_t,\ t = 1, 2, \ldots, T, \tag{3.1}$$

with u_t a zero mean stationary process. To emphasize that the parameters are assumed fixed for all the sample (and "*forever*") some authors label this model as the *global* trend model. It easily complies with the two most basic properties that a trend should possess, according to White and Granger (2011, p. 3): i) it certainly has a direction and ii) it is

[1]Rivas and Gonzalo (2020) interpret this model as "the best linear least squares approximation to an unknown trend function" and base their test for the trend on it; see section 5.2 on this subject.

"somewhat smooth". In the particular case that $u_t \equiv \epsilon_t \sim iid(0, \sigma_\epsilon^2)$ I consider that the y_t process is a *purely* (deterministic) linear trend process.

Apparently, about 100 years ago the time variable was sometimes given the status of a causal factor, i. e., some economic variables were viewed as functions of time. But although the model of equation (3.1) was the most popular for detrending at that time (Frickey, 1934), most economists and statisticians did not considered that time could be considered as a causal factor (e.g., Yule, 1926, p. 4). Instead, it was used as a proxy for some unobserved variable, as was (then) the case of population; later it has been used to represent the evolution of technical progress. Smith (1925, p. 543) is particularly clear about this:

> It is merely a statistical convenience to group all these factors together and predicate that their combined effect may be defined as a mathematical function of time when time is numerically described.

Either way, in case the model is considered the detrending device, its errors, u_t, are rather naively interpreted as the cyclical component of the series, the object of study of business cycles analysts. Later I will come back to point out the pitfalls of this approach.

The model itself is also rather naive and it is rarely adequate for economic time series. Indeed, either silencing the error term and taking first differences, $\Delta y_t = \beta$, or simply taking the expectation and differencing, $\mathrm{E}(\Delta y_t) = \beta$, it is easy to conclude that it is rarely the case that the model is appropriate in economics because it requires that the series should increase ($\beta > 0$) or decrease ($\beta < 0$) approximately the same amount in every period.

Instead, most often in economics it is the rate of growth rather than growth itself that is considered approximately constant through time. Then, the appropriate model is the exponential growth model

$$y_t = \exp(\alpha + \beta t + u_t),$$

which becomes the linear trend model for the (natural) logarithm of the series,

$$\log(y_t) = \alpha + \beta t + u_t, \tag{3.2}$$

and therefore, $\beta \approx (y_t - y_{t-1})/y_{t-1}, \forall t$. The smaller is β the better the approximation; that is, for small values of $\widehat{\beta}$, $100\widehat{\beta}\,\%$ is a good approximation to the average rate of growth of y_t. For instance, if y_t represents investment observed in a sample of quarterly data for the 1990:1 – 2022:4 period and the estimated equation is $\widehat{\log(y_t)} = \widehat{\alpha} - 0.008\,t$, then -0.8% is a very good approximation for the average quarterly growth rate of investment in that period.

If either in (3.1) or (3.2), besides white noise the errors are Gaussian, i.e., $u_t \equiv \epsilon_t \sim iid\mathcal{N}(0, \sigma^2)$, then OLS is the best unbiased estimator (BUE) and usual inference methods associated with it are valid exactly. Indeed this is a case which goes even further than the classical regression model with exogenous regressors[2] because the only regressor is deterministic and hence, in the style of old textbooks, "fixed in repeated samples". The practical relevance of this result is however very limited due to the white noise assumption; it even conflicts with the interpretation of u_t as the cyclical component because it requires that cycles cannot be neither serially correlated (i.e., really non-cyclical) nor heteroskedastic.

In case the error term u_t is serially correlated (and possibly non-Gaussian as well), OLS is not even BLUE but, provided u_t is only a very general stationary process, OLS is asymptotically efficient (Fuller, 1996, pp. 476-80 or Canjels and Watson, 1998)[3]. However, the usual estimator for its variance is generally biased and inconsistent and therefore inferences based on that estimator cease to be valid, even asymptotically. Hence, the OLS estimators for the coefficients of both (3.1) and (3.2) need to be complemented with some HAC estimator for its covariance matrix (see section 5.3)..

Returning to the case of the purely linear (or exponential) determin-

[2]See, e.g.,chapter 8 in Hamilton (1994) or chapter 1 in Hayashi (2000).

[3]This result is originally attributed to Grenander, Ulf, and Murray Rosenblatt (1957), *Statistical Analysis of Stationary Time Series*, New York, Wiley.

istic trend, the OLS estimator can be shown to be superconsistent — and at a rate that is even faster than in the case of decomposition — despite non-conformity with classical asymptotic assumptions. The interest of this case results precisely from this ("benign") violation since in this case the typical matrix to invert to obtain the OLS estimator is

$$\begin{aligned}\mathbf{X'X} = & \sum_{t=1}^{T} \mathbf{x_t x_t}' = \begin{bmatrix} \sum_{t=1}^{T} 1 & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{bmatrix} \\ = & \begin{bmatrix} T & T(T+1)/2 \\ T(T+1)/2 & T(T+1)(2T+1)/6 \end{bmatrix},\end{aligned}$$

where $\mathbf{x_t} = (1,\ t)'$, and hence, even when divided by T three of the elements of this matrix diverge when $T \to \infty$. This means that the usual assumption that

$$\operatorname{plim}\left(\frac{1}{T}\mathbf{X'X}\right) = \mathbf{\Sigma_{xx}},$$

a finite and positive definite matrix does not hold. However, provided the DGP is really (3.1) or (3.2) with white noise errors ($u_t \equiv \epsilon_t$) and provided their fourth order moment is finite ($\mathrm{E}(\epsilon_t^4) < \infty, \forall t$), it can be shown that

$$\begin{bmatrix} \sqrt{T}(\hat{\alpha} - \alpha) \\ T^{3/2}(\hat{\beta} - \beta) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \sigma^2 \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}^{-1} \right),$$

see, e. g. Hamilton 1994, pp. 455-60 or Martin, Hurn and Harris, 2013, pp. 594-6. That is, while the OLS estimator of α is consistent at the usual rate, the OLS estimator of the slope, $\widehat{\beta}$, is super-consistent[4]: it is not enough to scale its sampling error with $T^{1/2}$, as usual; it is necessary to scale it with $T^{3/2}$ to get a non-degenerate limit distribution; while $\widehat{\alpha}$ is $O_p(T^{-1/2})$ as usual, $\widehat{\beta}$ is $O_p(T^{-3/2})$.

Hassler (2000) proves a result similar to this in the case of a bivariate

[4]The calculations and the intuition for the simpler case where there is no intercept may be seen in, e.g., Stewart and Gill (1998), pp. 227-8.

trend stationary process. That is, consider that x_t is generated as in equation (3.1),

$$x_t = \gamma + \delta t + u_t,$$

with u_t a zero mean stationary process. And that y_t is linearly related with x_t through the equation

$$y_t = \alpha + \beta x_t + v_t,$$

with v_t a zero mean stationary process as well. Then, as $T \to \infty$,

$$\sqrt{T^3}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, 12\frac{\omega_v^2}{\delta^2}),$$

with $\widehat{\beta}$ the OLS estimator of β and where ω_v^2 is the long-run variance of v_t, $\omega_v^2 = \mathrm{Var}(v_t) + 2\sum_{j=1}^{\infty} \mathrm{cov}(v_t, v_{t-j})$. Since x_t is clearly dominated by the linear trend, this result is hardly surprising.

Although the main criticism to this type of models for the trend is postponed to section 3.8, one of its most important pitfalls will be mentioned here. It concerns the apparently high predictability that it conveys, which is implied by its (deterministic) nature and which can be very misleading. Actually, the rather strong criticism made by Stock and Watson (1988) to these models begins with an empirical illustration where, perhaps even more important than their "dramatic forecasting errors", is the severe underestimation of uncertainty regarding the future (as reflected in the narrow bands associated with the point forecasts). This is obviously an inescapable implication of these models, where the only source of uncertainty lies in the deviations from the fixed, non-changing trend. Such predictability must be always seen with much suspicion. In figure 3.1 I present an example which is similar to the one of Stock and Watson (1988) for the case of Australian *per capita* GDP.

In the left panel the observations of the series and their corresponding fitted values until 1993 are presented. The fit is very good ($R^2 = 0.988$ but a Durbin Watson statistic of only 0.456), likely originating an ex-

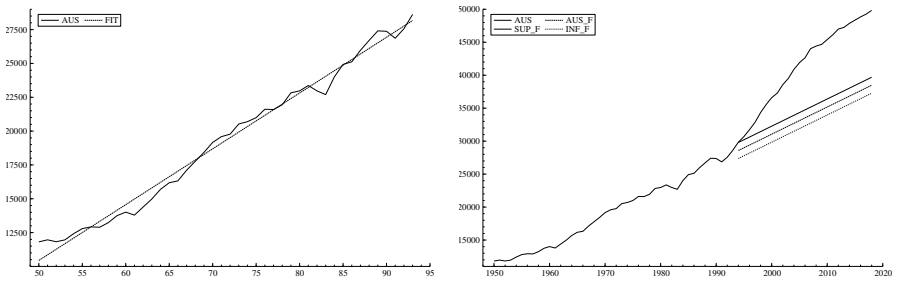

Figure 3.1: Australian *per capita* GDP: linear trend fitted until 1993 and corresponding forecasts for 1994–2018 with their (aprox.) 95% band

pectation that forecasts produced by the model will be equally good. However, the right panel shows that this expectation fails completely as the forecasts turn out to be dramatically wrong. Already the first observation after the base year for the forecasts (i.e., for 1994) jumps above the upper limit of the (approximate) 95% forecast band and the path that actual observations follow is ever more divergent from that band, rendering ever growing positive forecast errors.

These days the linear trend model is rarely considered adequate to model the trend of economic time series. Linear detrending is also seldom used explicitly, though it is often employed implicitly in multivariate analysis, both in regression and in VAR models, through the inclusion of a trend term. A somewhat recent exception is Hafner (2003), who argues that the linear trend model is the most adequate to model the trend of cereal yields for many of the 188 nations whose data were gathered by FAO (Food and Agriculture Organization) for the 1961–2001 period.

3.2 Polynomial trend models

It is often assumed that although with unknown functional form the trend must be a continuous and smooth function of time. But since according to the Weierstrass approximation theorem any continuous function defined on a compact interval can be approximated by a polynomial, a function

such as

$$\tau(t) = \alpha + \beta_1 t + \beta_2 t^2 + \ldots + \beta_p t^p$$

appears to be a good candidate to approximate the trend. Moreover, since the smoothness of the function requires that its first derivative changes sign only a few times, the order of the polynomial must be low, say $p = 2$ or 3.[5] Further still, a p too high will very likely imply some overfitting, with cyclical movements becoming incorporated in the trend (Mills, 2003, p. 15).

But even quadratic and cubic polynomials may be viewed with much suspicion:

a) because there is usually no plausible economic justification for them and their coefficients are not clearly interpretable;

b) since they tend to plus or minus infinity as time increases, sometimes very rapidly, they may produce disastrous forecasts for the medium and long terms;

c) because often "the effect of altering the degree of the polynomial by one will be to alter the direction of one or other of the branches of the fitted function", which is a "highly unsatisfactory circumstance" (Pollock, 1998, p. 5); figure 3.2 illustrates this with dramatic consequences for forecasting.

Indeed, in the left panel of the figure it is clear that the fit of the cubic model is better than the one of the quadratic, particularly in the initial and final observations of the sample. Adding the cubic term increases the R^2 from 0.90 to 0.97 and its coefficient is highly significant by standard criteria, with a t-ratio statistic of -4.63. However, as can be observed in the right panel, its forecasts are catastrophically, diverging to $-\infty$ while the actual series grows. This explains why even low order polynomials are rarely used. And when they are, they should be viewed as approximating *local* rather than global trends.

[5] Fuller (1996, p. 481) provides an alternative that will be addressed later, that of "grafted polynomials".

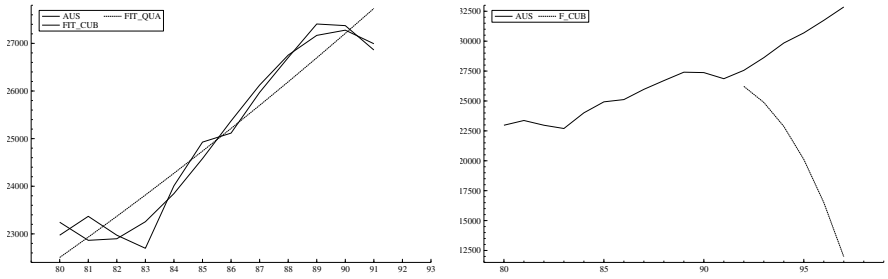

Figure 3.2: Australian *per capita* GDP again, 1980–1991: in the left panel the cubic trend ($\bar{R}^2 = 0.97$) fits much better than the quadratic ($\bar{R}^2 = 0.90$), but (right panel) its forecasts are catastrophically wrong, diverging to $-\infty$

A recent extensive empirical application of the quadratic trend is provided by Arata, Fabrizi and Sckokai (2020) who consider trend polynomials with degree between zero and two for 8088 country crop yields series from the FAO database (covering e.g., cereals, fruits, nuts, vegetables, etc.). The quadratic trend is considered the most adequate in 53.8% of the cases. An interesting feature of this study is the employment of an estimation method which is more robust to outliers than OLS, the MM method. In Cuestas and Garratt (2011) a cubic trend is considered adequate to approximate a possibly nonlinear trend in the time series of GDP per capita for a panel of 19 developed countries between 1870 and 2003 (and it is used to provide the stationary alternative hypothesis for a unit root test).

3.3 Further nonlinear classical models

In some cases one may consider that the trend is limited by an upper bound. The most common example is that of the penetration rate of a consumer durable — say, TV sets — in households. But the idea can be extended to the proportion of the general population owing a certain item, or of companies possessing some capital good. In that case, the so-called S-shaped (or sigmoid) growth curves must be considered as good candidates. The logistic and the Gompertz models have been used

in such cases and they may be considered as "classical" because they are both known for almost a century (see, e.g., Mitchell, 1927). Meade (1984) and Meade and Islam (1995) contain many examples and further references. Here I am mostly based in Pollock (1998).

The most basic version of the logistic function is

$$g(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}.$$

It varies between zero, which is its limit when $x \to -\infty$, and unity, the limit when $x \to +\infty$, and its midpoint , when $x = 0$ is $g(0) = 1/2$, which is an inflexion point because the growth rate begins to decline (the second derivative is zero and then becomes negative). The logistic trend function may be presented as

$$\tau(t) = \frac{\gamma}{1+e^{h(t)}} = \frac{\gamma e^{h(t)}}{1+e^{h(t)}}, \quad \text{with} \quad h(t) = \alpha + \beta t,$$

where the parameter γ is the asymptote or upper limit of the function, and therefore represents the "saturation level". While β determines the growth rate, α determines the mid point of the function.

Since

$$\log\left[\frac{\tau(t)}{\gamma - \tau(t)}\right] = h(t),$$

the regression equation can be written as

$$\log\left[\frac{y_t}{\gamma - y_t}\right] = \alpha + \beta t + u_t,$$

which, for each value of γ can be used to calculate the value of the dependent variable. That is, for each value of γ, α and β can be estimated by OLS. This means that the minimization of the sum of squares of the residuals can be implemented through a grid search procedure to find the optimal value for γ (and for α and β). An alternative presentation for the logistic regression model is (see, e.g., Franses, 1998, and Meade,

1984)

$$y_t = \frac{\gamma}{1 + \alpha e^{-\beta t}} + u_t,$$

where all parameters are positive. Estimation must resort to a non-linear least squares (NLLS) algorithm.

On the other hand, the Gompertz growth trend is

$$\tau(t) = \gamma \exp[-\alpha \exp(-\beta t)],$$

where again γ represents the saturation level. Franses (1998, p. 77) provides a simple selection method to choose between the logistic and the Gompertz curves.

3.4 Segmented linear trend models

The piecewise or segmented linear trend model lies somewhere between the linear trend model of section 3.1 and the stochastic trend or unit root model that will be addressed below, in section 3.8. It is also one of the simplest versions of the standard switching regression model in the terminology of Teräsvirta, Tjøstheim and Granger (2010). Although recently it has been associated mostly with the work originating in Perron (1989), in practice it has been used for a rather long time, dating back at least to Frickey (1942) (see subsection 2.6). It is a rather simple and inexpensive alternative to more sophisticated non-linear models that may be viewed as an entrance door and an approximation to those models, and its coefficients are very easily interpretable, particularly when the variable is logarithmized. It also makes the linearity assumption more tenable, in so far as it is assumed to hold for (much) shorter periods. Clearly, reducing the time interval where fixed linearity holds makes it more flexible and local (instead of rigid and global).

A simple example of a segmented linear trend is

$$\tau(t) = \alpha_1 + \alpha_2 \mathbf{1}(t \geq T_\alpha) + \beta_1 t + \beta_2 \mathbf{1}(t \geq T_\beta)\, t,$$

where $\mathbf{1}(.)$ is the indicator function (e.g., $\mathbf{1}(t \geq T_\beta)$ is 1 for $t \geq t_\beta$ and 0 otherwise), and where T_α may be different from T_β, that is, the time of change or the break date in the intercept can be different from the one of the slope. As is well known, this trend function can be also represented using dummy or binary variables:

$$\tau(t) = \alpha_1 + \alpha_2 \, D_{\alpha,t} + \beta_1 \, t + \beta_2 \, D_{\beta,t} \, t$$

where

$$D_{\alpha,t} = \begin{cases} 1, & \text{for } t \geq T_\alpha \\ 0, & \text{otherwise,} \end{cases}$$

and

$$D_{\beta,t} = \begin{cases} 1, & \text{for } t \geq T_\beta \\ 0, & \text{otherwise.} \end{cases}$$

This allows that the intercept is α_1 until time $T_\alpha - 1$ and $\alpha_1 + \alpha_2$ thereafter; the slope of the trend is β_1 until time $T_\beta - 1$ and then changes to $\beta_1 + \beta_2$. It may occur that there is only one break, occurring simultaneously in the intercept and in the slope, that is, that $T_\alpha = T_\beta$. It is also admissible that either α_2 or β_2 is zero, in which case only one of the coefficients changes. Note additionally that if $\alpha_2 \neq 0$ the trend is discontinuous at the point $t = T_\alpha$, i.e., it has a jump at that point.

A further generalization of this model consists in the increase in the number of break dates, for both or for only one of the coefficients. However, this increase must be sensible and only changes that are qualitatively important must be object of modeling. The choice of these dates for modeling purposes, however, is not so much constrained as when the purpose is testing the significance of those breaks. Provided the purpose is not this last one, they may be specified endogenously to the data, i.e., on the basis of some previous narrative which may be itself based in the data or simply by observing a plot of the series (as, e.g., in Crafts and Mills, 2017). Preferably, however, it must be based only in qualitative information, to preserve the size properties of the tests assessing the significance of the breaks.

Admitting the presence of at least one break during the sample pe-

riod does not allow ruling out the possibility that a similar problem might affect forecasts, quite on the contrary. And as Stock (1994, p. 2821) emphasizes, "treating the break as a one-time nonrandom event presumably leads to understating the uncertainty of multistep forecasts".

Gao and Hawthorne (2006) present a model of this type, which they call bilinear, as a crude approximation to a non-linear and semiparametric model estimated over a series of global temperatures for the 1867–2001 period with a turning point in 1890.

3.5 Linear spline trend models

A close alternative to the models of the previous section is the linear spline trend model which, on one hand, ensures continuity of the trend function but, on the other hand, does not allow changes in level, i.e., only the slope of the trend may change, not its intercept. Continuity is preserved at the cost of some flexibility that is lost.

In its simplest version, the trend function with only two linear segments that are joined at the point called knot is

$$\tau(t) = \alpha + \beta_1 t + \beta_2(t - T_\beta)\mathbf{1}(t \geq T_\beta),$$

whose intercept is always α but which has two different slopes (provided $\beta_2 \neq 0$), joined at the point $t = T_\beta$, the breakpoint: in the first subperiod the slope is β_1 but in the second it changes, somewhat smoothly, without any jump, to $\beta_1 + \beta_2$, i.e., β_2 again measures the change in the slope. Therefore, now the function is constrained to be continuous also at $t = T_\beta$. Moreover, notice also that the model is more parsimonious than those of the previous section because it contains one less parameter. However, as in those models, the break is still somewhat abrupt, sudden, occurring in one period only[6].

[6]To make the change more gradual, Franses (1998, p. 143) proposes a model that, besides the change in slope, introduces two breaks in level, one at the time of the break and the other in the following period.

In case there are several changes in slope, the model may be generalized to

$$\tau(t) = \alpha + \beta\, t + \sum_{i=1}^{m} \delta_i (t - T_{\beta_i}) \mathbf{1}(t \geq T_{\beta_i}),$$

with m denoting the number of breaks. Drake and Mills (2010) estimate a trend function of this type for the log of the GDP of the Euro area, for the period 1983Q1 to 2006Q1, imposing five breakdates or knots at 1983Q1, 1992Q1, 1993Q1, 2000Q4 and 2001Q1, this one corresponding to the entrance of Greece; for instance, the estimated slope for the segment of the period 1993Q1 to 2000Q3 is $\widehat{\beta} + \widehat{\delta}_1 + \widehat{\delta}_2 + \widehat{\delta}_3$. More generally, the slope of the trend for the i-th segment is given by $\beta + \delta_1 + \ldots + \delta_i$ $(i = 1, \ldots, m.)$. More plausibly, Crafts and Mills (2017) also estimate a model that is mostly a linear spline trend with five breaks for the case of per capita GDP in England and Britain for the rather long sample initiated in 1270 and ending in 1913. However, they consider also a change in level corresponding to the period of the Black Death (1348 to 1352), which they find that has left the rate of growth unchanged [7].

Although these models may often appear attractive, particularly to a layman in economic history, they must also be subject to a further critical, non-econometric perspective, as, e.g., in Evans and Quigley (1995). They argue that a segmented trend representation of time series can be used to support many alternative, even contradictory hypotheses about the economic history of a country. More concretely, they argue that since it is only one of the possible representations of the data, conclusions relying exclusively on it may be spurious. On the other hand, they recognize that segmented trends can perform a rather positive role provoking the debate about the nature of long-run economic growth and the mechanisms that might explain major periods of expansion and contraction.[8].

[7] The remaining breakdates considered by Crafts and Mills (2017) are 1663, 1702 and 1823. Crafts and Mills consider that this model is preferable to the Hodrick and Prescott filter to produce the estimated trend.

[8] See also Inwood and Stengos (1995) for a detailed reply to the criticisms of Evans and Quigley (1995), particularly about the number and the nature of the structural breaks.

3.6 Segmented and spline polynomial trend models

The models of the previous section can be generalized to higher order polynomials, i.e., with order higher than one (but usually not exceeding three). Some smoothness is achieved with the spline form: a quadratic spline has continuous first derivative and a cubic spline has continuous first and second derivatives. For instance, a quadratic spline trend with only one knot (or break point) is

$$\tau(t) = \alpha + \beta_1 t + \beta_2 t^2 + \beta_3 (t - T_\beta)\mathbf{1}(t \geq T_\beta),$$

where it can be noticed that for $t \leq T_\beta$ the function is the quadratic $\alpha + \beta_1 t + \beta_2 t^2$ and its second derivative is $2\beta_2$, but for $t \geq T_\beta$ this derivative is $2(\beta_2 + \beta_3)$, i.e., with a change of $2\beta_3$. For Mills and Crafts (1996) the preferred model for British industrial output over the period 1700 to 1913 is a quadratic spline trend where they consider three possible intervals for the knots: $1764 < T_1 < 1786$, $1814 < T_2 < 1836$ and $1854 < T_3 < 1876$; the first corresponds to the start date of the Industrial Revolution and the last two to alternative locations for its ending. Minimizing the sum of squares of the residuals of the model $y_t = \tau(t) + \epsilon_t$, ϵ_t assumed white noise, they estimate these dates as $\widehat{T}_1 = 1776$, $\widehat{T}_2 = 1834$ and $\widehat{T}_3 = 1874$.

More generally, a pth-order spline trend function with m knots $T_{\beta_1} < T_{\beta_2} < \ldots < T_{\beta_m}$ is

$$\tau(t) = \sum_{j=0}^{p} \beta_j t^j + \sum_{i=1}^{m} \beta_{p+i}(t - T_{\beta_i})\mathbf{1}(t \geq T_{\beta_i}).$$

Crafts and Mills (2017) estimate cubic spline models for both industrial production and GDP, for England pre-1700 and Britain post-1700, over the period 1270 to 1913, with a single knot for the slope of the trend located in 1871 (and also with four changes in level, two of them for the beginning and end of the Black Death).

The name that Fuller (1996, pp. 480-4) gives to these spline polyno-

mials for the trend is "grafted polynomials". Moreover, he also provides a few examples where they are specified obeying to certain restrictions on the length of the subsamples or even on the form adopted for some of those subsamples.

According to Pollock (1998), often the end of these splines is left free, with second derivatives equal to zero, and forecasting is made with a simple linear trend. This is perhaps the best practice to follow with these models, despite his recommendation in favour of cubic splines; indeed, these can really provide excellent fits to the data, as in Crafts and Mills (2017), but Pollock (1998, p. 10) himself warns that

> the device which is most appropriate to the extrapolative forecasting of a trend is rarely the best means of representing it within the sample. An extrapolation is usually based upon a simple analytic function; and any attempt to make the function reflect the local variations of the sample will endow it with global characteristics which may affect the forecasts adversely.

Cubic splines may fit the data very well but most researchers will be hardly convinced to use them outside the sample, to produce forecasts. At least they will hesitate a lot, and with good reasons.

3.7 Modern nonlinear models: the smooth transition trend

Despite a significant outburst in research on nonlinear models for the mean for a period of about two decades, starting around 1990, the interest in modeling the trend with a nonlinear approach lagged much behind. This was because most efforts centered on stationary time series. Hence, very often the series to model was previously (first or seasonally) differenced to become stationary.

Moreover, some models for non-stationary, trending time series were proposed with the sole purpose of serving as alternative hypotheses to

the unit root/stochastic trend null model; in some of these cases the role of such models was almost virtual, not even arriving at the estimation stage[9]. In some other cases the presentation of the estimated nonlinear trend was omitted because attention was directed towards analysing the long-run properties of the residuals. It is therefore rather difficult to find a modern empirical nonlinear model for the trend as a result of this research.

Kapetanios (2003) provides an interesting example of these modeling efforts, where the basic approach does not differ radically from the one that was used in empirical research dated a century ago. A threshold autoregressive (TAR) model for U.S. GDP based in ideas from Hicks is proposed with the attractive feature that fluctuations in output are endogenous, i.e., there is no need to assume that they result from exogenous shocks. But modeling the trend is relinquished because the model is built for the detrended series, the trend previously estimated with the Hodrick-Prescott filter.

The most important and representative model of this research period is the smooth transition (ST) trend model, derived from the regression model originally proposed by Bacon and Watts (1971) for general regressors[10], and which nests the segmented trend model of section 3.4. Later, Granger and Teräsvirta (1993), Lin and Teräsvirta (1994) and Teräsvirta (1994) further developed the general smooth transition regression model. The main feature of this type of model is the smoothness, the continuity of states or regimes that it allows, which is coupled with the complete endogeneity in selecting the timing and the speed of the transition between the "extreme" regimes. In other words, the information about these issues that is usually required by the previous models is now totally redundant as the estimation of the parameters that reflect them

[9]This is because they may become unidentified under a certain null hypothesis and therefore testing needs to rely on a regression obtained with a Taylor series expansion.

[10]Albeit with a transition function that would be abandoned. Bacon and Watts (1971) emphasize that the particular form of the transition function is "*of secondary importance*" (p. 527). According to Granger and Teräsvirta (1993), in a textbook published in 1977, G. S. Maddala was also an early proponent.

is completely data-dependent. This is most often an advantage but it can be seen as unappealing as well when the researcher wishes to impose a specific date for the break (corresponding to, e.g., a major political or institutional change). But even in such a case the ST trend model is admissible because there is usually no simultaneity in the reaction of economic agents to the shocking event; i.e., though individually some of them may react swiftly, the aggregate response that follows may be rather protracted. Therefore, allowing the data speak for themselves is usually a good idea.

Some versions of the ST trend model were firstly presented in Leybourne, Newbold and Vougas (1998, LNV)[11] as a non-linear alternative to the unit root hypothesis. The simplest version, the logistic ST trend for the case where there is only a change in level (and the slope of the trend is constrained to remain fixed), labeled as model B in LNV, is

$$\tau(t) = \alpha_1 + \beta\, t + \alpha_2 S_t(\gamma, c),$$

where

$$S_t(\gamma, c) = \{1 + \exp[-\gamma(t - cT)]\}^{-1},\ \gamma > 0,$$

is the (logistic) transition function, which monotonically traverses the interval (0, 1) as $t \to \infty$, and which controls the transition between the two "extreme" regimes. Therefore, the model permits a smooth transition between the two submodels. It can also be viewed as a simple linear trend model but with deterministically changing parameters. Most often, it is viewed as a regime-switching model with two regimes, corresponding to the extremes of the interval. Alternatively, it can be interpreted as a model that allows for a continuum of regimes, each associated with each value taken by the transition function between 0 and 1.[12] That is, the combined intercept fluctuates between α_1 and $\alpha_1 + \alpha_2$.

Since, for $\gamma > 0$,

$$S_{-\infty}(\gamma, c) = 0, \quad S_{+\infty}(\gamma, c) = 1, \quad \text{and} \quad S_{cT}(\gamma, c) = 0.5,$$

[11] See also Mills (2003) for a very clear presentation and for a few examples.

[12] See, e.g., van Dijk, Teräsvirta and Franses (2002, vDTF) and Teräsvirta (1998).

the parameter c determines the timing of the transition mid-point. The smoothness or the speed of the transition is determined by the parameter γ:

a) if γ is small the function needs a long period of time to traverse the interval (0, 1) and, in the case that $\gamma \to 0$, then $S_t(0, c)$ approaches $0.5, \forall t$, that is,

$$\tau(t) = (\alpha_1 + 0.5\,\alpha_2) + \beta\,t,$$

i.e., there is only one regime, the model collapsing to the simple and single linear trend.

b) if γ is large, the function traverses the interval (0, 1) very rapidly and, as γ goes to $+\infty$, the limiting behaviour of the S_t function is that of a change from 0 to 1 instantaneously, at time $t = c\,T$, that is, the S_t function reduces to the indicator function. Instead of smooth, the change in regime becomes abrupt, as with the models of the previous sections.

If, on the other hand, $\gamma < 0$, then the order of the initial and final regimes or states is reversed but the interpretation of the parameters remains unchanged.

A more general model for the trend corresponds to model C in LNV and allows the change in regime to affect both the level and the slope of the trend:

$$\tau(t) = \alpha_1 + \beta_1\,t + (\alpha_2 + \beta_2\,t)\,S_t(\gamma, c),$$

where the interpretation of the parameters remains the same except in the case of the collapse to a single regime, which is now

$$\tau(t) = (\alpha_1 + 0.5\,\alpha_2) + (\beta_1 + 0.5\,\beta_2)t,$$

but still only one change is permitted, the initial regime now given by $\alpha_1 + \beta_1 t$ and the final one by $(\alpha_1 + \alpha_2) + (\beta_1 + \beta_2)t$. Moreover, the restrictions that the changes in intercept and in slope are simultaneous and occur with the same speed are also imposed. Using the form of

presentation adopted in vDTF, it can be also written as

$$\tau(t) = (\alpha_1 + \beta_1 t)[1 - S_t(\gamma, c)] + (\alpha_2 + \beta_2 t)\, S_t(\gamma, c).$$

Estimation of these models needs to resort to a non-linear least squares (NLLS) optimization algorithm. A popular option is the BFGS (Broyden, Fletcher, Goldfarb and Shanno) algorithm. However, noting that for fixed values of the parameters γ and c the transition function becomes linear in the remaining parameters, as suggested in LNV estimation may also be made in two steps. In the first step, conditional upon γ and c, the sum of squares of residuals function can be concentrated with respect to the αs and the βs, which are estimated by OLS. In the second step, estimation of γ and c is made with a two-dimensional grid search. Teräsvirta (1994) and vDTF offer some practical tips that may be very helpful for estimation, e.g., to choose the starting values or to overcome difficulties when γ is large.

Using the series of quarterly U.S. logged GDP for the period 1947:1 to 2004:4, Vougas (2007) estimates the model

$$\widehat{y}_t = \underset{(0.018)}{7.370} + \underset{(0.0004)}{0.0087}\, t + \underset{(0.025)}{(0.160} - \underset{(0.0004)}{0.0011}\, t)\, S_t\, (\underset{(0.453)}{0.979},\ \underset{(0.004)}{0.325}),$$

complemented with an AR(3) model for the errors. The negative (and significant) estimate for β_2 provides evidence for a productivity slowdown in the second regime. And since $\widehat{c} = 0.325$ corresponds approximately to the 75th observation, the estimated mid-point of the transition is at about 1965:2 (which Vougas, 2007, finds that accords well with estimates from other models but which is much different from the dates associated with the first oil shock, in 1973).

Still in the context of testing for unit roots, Harvey and Mills (2002) extend the previous models to the case of two smooth changes in regime, with

$$\tau(t) = \alpha_1 + \beta_1\, t + (\alpha_2 + \beta_2\, t)\, S_{1t}(\gamma_1, c_1) + (\alpha_3 + \beta_3\, t)\, S_{2t}(\gamma_2, c_2)$$

with

$$S_{it}(\gamma_i, c_i) = \{1 + \exp[-\gamma_i(t - c_i T)]\}^{-1},\ i = 1, 2,$$

where the mid-points are now at $c_1 T$ and $c_2 T$ and the two transition speeds may be different. This model was used by Salamalikis and Venutis (2014) to study the long-run properties of the series for labour supply in the U.S., for 1948:1 to 2011:1.

In Teräsvirta, Tjøstheim and Granger (2010, p. 38) the general ST model is further generalized to the case of several changes in regime, and adapting it to a model for the trend is easy. However, the speed of the transitions is constrained to be same for all of them. For instance, the model that (adapting from, e.g., Teräsvirta, 1998) may be labeled as the LT2 for the trend has the transition function

$$S_t(\gamma, c) = \{1 + \exp[-\gamma(t - c_1 T)(t - c_2 T)]\}^{-1},\ \gamma > 0,\ c_1 \leq c_2,$$

which is also symmetric, now about $(c_1 T + c_2 T)/2$, and which produces three regimes. However, contrary to the simplest LT (or LT1) model, now the S_t function is not monotonic. A transition function with similar properties is obtained with the exponential function,

$$S_t(\gamma, c) = 1 - \exp[-\gamma(t - cT)^2],\ \gamma > 0,$$

originating the ET model for the trend. To the best of my knowledge, this model was proposed by Sollis (2005) and used by him to produce unit root test statistics for 20 series of quarterly real exchange rates against the U.S. dollar. Sollis graphically presents some estimated trends.

3.8 Unit root/stochastic trend model

Although the basic idea of the unit root/stochastic trend model dates back at least to Working's (1934)[13], it started to be seriously considered

[13] You may need to review section 2.5, but notice also that according to Klein (1997, p. 272) the random walk model was initially proposed in 1900 by Louis Bachelier in

in macroeconomics only in the beginning of the 1980s, and only after the divulgation of the work of time series analysts through Box-Jenkins (1976) influential book and the proposal of the first unit root tests by Dickey and Fuller (1979, 1981). Moreover, the introduction of this subject in econometrics marks the beginning of a revolution of its methods for macroeconomic analysis — the unit root and decomposition revolution —, and indeed the beginning of macroeconometrics as a really separate branch of study as well. In this section I assume that the typical reader is aware of the basic material about unit root tests; in case I am wrong I recommend the following survey papers, monographs and textbooks: Campbell and Perron (1993), Enders (2010, chapter 4), Hamilton (1994, chapters 15-17), Hatanaka (1996, chapters 1-7) and Maddala and Kim (1998, chapters 3-4)[14]. My purpose is only to highlight certain points more closely related with the subject of this book.

The simplest member of the family is the driftless random walk, $y_t = y_{t-1}+\epsilon_t$, $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ which, solving backwards, can be also represented as

$$y_t = y_0 + \sum_{i=1}^{t} \epsilon_i, \tag{3.3}$$

where y_0 represents the initial value and the second term represents the stochastic trend, the accumulation of small (and possibly not so small) random shocks, that is, of fluctuations without any systematic nature. Although time appears only as the upper index of the summation, it is indeed appropriate to call this term a trend because it is the main (and almost the only) ingredient of forecasts, whatever their horizon, even in the very long-run. Actually, note that the best (in MSE) forecast of the process for an horizon that may lie arbitrarily far in the future, say $t+s$ with $s > 0$, made with the information about the process that is available

his doctoral thesis.

[14] A concise presentation is provided by Phillips (2010a). More advanced treatments are provided in, e.g., Martin, Hurn and Harris (2013, chapters 16-17) and in Stock (1994). Choi (2015) provides an extensive and updated survey. My own little book (Lopes, 2015) is an alternative in case portuguese is preferred.

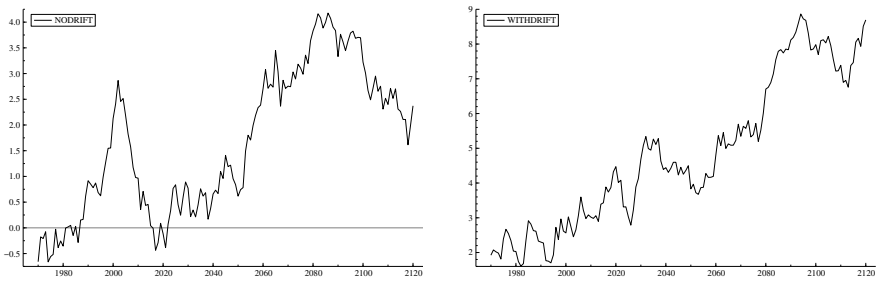

Figure 3.3: Two simulated random walks, without (left) and with (right) drift, both with zero as initial value and with $\epsilon_t \sim iid\mathcal{N}(0, 0.3^2)$

at time t, is

$$\begin{aligned} y_{t+s|t} &= \mathrm{E}_t(y_{t+s}) \\ &= \mathrm{E}_t(y_t + \epsilon_{t+s} + \epsilon_{t+s-1} + \ldots + \epsilon_{t+1}) \\ &= y_t. \end{aligned}$$

But if the forecast horizon changes, say h periods into the future, for any $h > 0$ (that can be very large), and the forecast is now made at time $t + s$, the best forecast is now given by

$$\begin{aligned} y_{t+s+h|t+s} &= \mathrm{E}_{t+s}(y_{t+s+h}) \\ &= \mathrm{E}_{t+s}(y_t + \epsilon_{t+s+h} + \ldots + \epsilon_{t+s+1} + \epsilon_{t+s} + \epsilon_{t+s-1} + \ldots + \epsilon_{t+1}) \\ &= y_t + \epsilon_{t+s} + \epsilon_{t+s-1} + \ldots + \epsilon_{t+1} = y_{t+s}. \end{aligned}$$

Hence, the random shocks that have occurred between periods t and $t+s$ are completely included in the new forecast, even for an horizon that may lie in the very far future. The change in the forecasts is completely made up of these shocks. It is therefore really a trend, because it may concern really long-run forecasts, which is made of stochastic fluctuations, which changes stochastically.

On the other and, if the random walk contains a drift parameter, say γ, $y_t = \gamma + y_{t-1} + \epsilon_t$, $\sim iid(0, \sigma_\epsilon^2)$, then it may be also represented (again

solving backwards) as

$$y_t = y_0 + \gamma\, t + \sum_{i=1}^{t} \epsilon_i, \tag{3.4}$$

where the presence of a linear deterministic trend emerges due to the drift. That is, besides the stochastic trend, the random walk contains also a deterministic trend. As is well known, the origin of the relevant non-stationarity lies, however, in the stochastic trend. In figure 3.3 plots of 150 observations of two simulated random walks are presented, one without drift and the other with a positive drift (in both cases, discarding the first 50 simulated observations, to reduce the dependence on the initial value, which was set equal to zero in both cases). Notice that although it has no drift and hence no deterministic trend, if the observation of the driftless random walk is made within certain stretches of the sample only, the presence of a deterministic trend is clearly suggested. This is the case particularly for the long period between 2000 and 2080, with the series appearing to be dominated by a strongly growing deterministic trend. Unfortunately, such a misleading graphical behaviour occurs very often. This is indeed unfortunate because one is expecting that such a process always wanders without showing any systematic tendency to increase or to decrease.

The previous random walk may be presented in a way that makes more clear the presence of the unit root in the autoregressive polynomial:

$$(1 - L)y_t = \gamma + \epsilon_t,$$

where L denotes the usual lag or backward operator ($L^k y_t = y_{t-k}$), that is, $(1 - L)y_t = \Delta y_t$ and, for macroeconomic time series, γ is usually positive. But this is only the simplest unit root or $I(1)$ process, i.e., first order difference stationary process (DSP)[15]. More generally, a unit root

[15]To be precise, a (weakly) stationary process with positive long-run variance is said to be integrated of order 0, $y_t \sim I(0)$. And it is said to be integrated of order d, positive integer, $y_t \sim I(d)$, if $\Delta^d y_t \sim I(0)$.

process may be represented as

$$(1 - L)y_t = \gamma + \psi(L)\epsilon_t, \tag{3.5}$$

where $\psi(L) = 1+\psi_1 L+\psi_2 L^2+\ldots$, with $\psi(1) \neq 0$ (to rule out the possibility that the original level process is already stationary) and $\sum_{j=0}^{\infty} |\psi_j| < \infty$, the usual condition for the coefficients of the MA representation of the error process to be absolutely summable. This ensures that this error process, say $u_t = \psi(L)\epsilon_t$ is stationary, $I(0)$. Thus, instead of equation (3.4) we now have

$$y_t = y_0 + \gamma\, t + \sum_{i=1}^{t} u_i,$$

that is, although stationary $I(0)$, the accumulating shocks may be serially correlated. For the simplest and particular case of the random walk $\psi(L) = 1$.

But the same process (3.5) can equivalently be represented as

$$y_t = \alpha + \gamma t + v_t, \tag{3.6}$$

with v_t a non-stationary, unit root $I(1)$ process, $(1-L)v_t = \psi(L)\epsilon_t$, with ARIMA representation

$$\phi^*(L)(1 - L)v_t = \theta(L)\epsilon_t,$$

with all the roots of $\phi^*(L)$ outside the unit circle, $\theta(L)$ a finite and invertible MA polynomial and $\psi(L) = \theta(L)\psi^*(L)^{-1}$. That is, instead of appearing directly in the autoregressive polynomial of the original series y_t, the unit root is lurking in the autoregressive polynomial of the error process of (3.6). Therefore, the presence of a unit root in the level series y_t is equivalent to the presence of a unit root in the autoregressive polynomial of its deviations from a deterministic trend. One must therefore not be tempted to say that deciding for the presence of an autoregressive unit root in a series is the same as deciding whether the series contains a deterministic *or* a stochastic trend (as can be found

sometimes). Rather, very often macro time series contain a strongly growing or positive deterministic trend and the issue is whether, besides this, its deviations around that trend are stationary or not (in which case there is a stochastic trend as well).

As became clear only much later, the importance of unit root tests for economic theory was greatly exaggerated. It started with a paper by Nelson and Plosser (1982) whose main ideas have been followed by many others as, for instance, Shapiro and Watson (1998)[16]. Their work suggested that the presence or absence of a stochastic trend in the series of real output should decide whether real business cycles (RBC) theory or Keynesian theory should be accepted as the best explanation for output fluctuations. This was made with the association of supply shocks (especially technological shocks) with stochastic trends, and of demand shocks (especially monetary shocks) with purely transitory, stationary fluctuations.

But as noted by Hansen (2022, p. 581), "*fundamentally, the unit root/stationarity distinction says little about the RBC/Keynesian debate*". Actually, standard Keynesian models of the 1970s resorted to the deterministic linear trend model with stationary disturbances around that trend — the trend stationary process (TSP) — because it was traditional and the most plausible among available econometric models of the time to describe the behaviour of output through time. Then in the 1980s, rapidly associating with the unit root revolution, RBC models linked supply side shocks with permanent but stochastic effects and prohibited, by assumption, that demand shocks could have such kind of effects; for instance, Shapiro and Watson (1988, p. 112) stated that: "*it [our model] only excludes the possibility that aggregate demand shocks permanently affect the level of output*". Hence, the strong evidence supporting the unit root model found by Nelson and Plosser (1982) could be only interpreted as favouring RBC theory and excluding the possibility that aggregate demand policies could permanently affect the level of

[16]Some issues of this debate are summarized in, e.g., section 1.1 of Blanchard and Fisher (1989).

Table 3.1: Comparison between TSPs with DSPs with drift

	TSP	DSP or UR
transformation to stationarize	subtracting the deterministic trend	differencing
dynamic multipliers	the effect of shocks is transitory (goes to zero as time passes)	shocks have a permanent effect on the level of the process
forecasts	forecasts, $y_{t+s\|t}$, converge to a linear function of their horizon, s	
	forecasts converge to a line whose intercept is independent of y_t	the intercept of the line of the limiting forecast varies with y_t
MSE of forecasts	the limiting MSE when the horizon increases is finite	MSE grows linearly with the forecast horizon

output. The fallacy was firmly rooted in a predetermined allocation of assumptions.

Incidentally, to defend the Keynesian position from the attacks of RBC theorists, rejecting that it implied the association of "*technical progress and capital accumulation*" (p. 449) with the linear deterministic trend, De Long and Summers (1998) also adopted the particular view of the trend in output as its potential level. Therefore, cycles corresponded to output gaps.

Perhaps the main motivation for the study of unit roots was indeed the choice between the difference stationary (DSP) or unit root process of this section and the trend stationary process (TSP), as represented by (3.1) or (3.2), as the most adequate model to represent macro time series. Summarizing the most important issues, in table (3.1) I provide a comparison between the two models that is entirely based in chapter 15 of Hamilton (1994), whose consultation I strongly recommend for a detailed presentation[17].

With the obvious exception of the required transformation to station-

[17]This table is a translation from a similar table in Lopes (2015).

arize, the two models differ in their limiting or asymptotic properties. That is, their implications are rather extreme but one has to consider infinite time horizons. Here lies perhaps the main reason why the choice is so difficult in practice, with limited and often really small samples.

The radical difference concerning dynamic multipliers is particularly meaningful. For a TSP as the one of (3.1), $y_t = \alpha + \beta t + u_t$, u_t zero mean stationary with $u_t = \psi(L)\epsilon_t$ its MA representation, with absolutely summable coefficients ($\sum_{j=0}^{\infty} |\psi_j| < \infty$), the dynamic multiplier for the effect s periods ahead of a transitory unit shock occurring in period t is

$$\frac{\partial y_{t+s}}{\partial \epsilon_t} = \psi_s.$$

But precisely because u_t is stationary, the ψ_j coefficients decline to zero with j and hence

$$\lim_{s\to\infty} \frac{\partial y_{t+s}}{\partial \epsilon_t} = 0,$$

that is, the effect of *all* shocks vanishes with time, it is purely transitory. There is not a single shock with a permanent effect.

Radically opposing to this, the UR model postulates that all the shocks have permanent effects; there is no single shock with only a transitory effect. To see this, since now it is Δy_t that admits a MA representation with absolutely summable coefficients, say ψ_j again to simplify, it is easy to show that the effect s periods ahead of a transitory unit shock in period t is

$$\frac{\partial y_{t+s}}{\partial \epsilon_t} = 1 + \psi_1 + \psi_2 + \ldots + \psi_s.$$

And hence,

$$\lim_{s\to\infty} \frac{\partial y_{t+s}}{\partial \epsilon_t} = 1 + \psi_1 + \psi_2 + \ldots = \psi(1) \neq 0.$$

Thus, both views are overly extreme; neither is plausible in this respect.

Concerning forecasts, the stochastic or "changing" or "variable" trend model provides the most reasonable framework, allowing uncertainty to

grow as the forecast horizon grows. The main motivation for this model in Stock and Watson (1988) relies precisely in ridiculing and dismissing long-run forecasts for U.S. macroeconomic time series based in the linear deterministic trend model, which are catastrophically wrong.

Another frequently used motivation is the failure of classical asymptotic theory, designed for stationary processes, to provide valid inference methods in models with unit root processes. Besides the well known problem of spurious regressions with independent $I(1)$ series in multivariate analysis, firstly highlighted by Granger and Newbold (1974), a more pertinent issue here is the one of spurious regressions of $I(1)$ series on a deterministic time trend regressor. This is the case where data is generated by a unit root or stochastic trend model but a deterministic linear (or exponential) trend model is assumed as adequate, and a regression as (3.1) or (3.2) is estimated with OLS.

This problem was first addressed by Chan, Hayya and Ord (1977), and although they have found some serious consequences, the case they addressed is not the most worrisome because the data generation process (DGP) they have considered was of a random walk with drift, and hence, although the stochastic trend was left without a proper consideration, the data really contained a linear deterministic trend originated by the drift. The most serious case was addressed by Nelson and Kang (1984), who have shown that the regression of a driftless random walk on a deterministic time trend very often produces spurious results, "finding" erroneous or misleading evidence supporting the presence of such a deterministic trend. Conventional standard errors and t-statistics associated with the coefficient of the trend term will very often support the presence of that trend in the data although it is not really present. Provided one remembers the graphical behaviour of the driftless random walk of figure (3.3), such a frequent outcome is not completely surprising. Nevertheless, it is very disturbing that deterministic trends appear to fit so well to stochastic ones, falsifying completely their nature.

On this regard, my own Monte Carlo study for samples with size $T = 200$, based in 10000 replications, produced the following results:

a) an estimate of the real size of the test for the insignificance of the trend coefficient of 91.6% for 5% nominal size tests;

b) a sample average of 0.434 for the R^2 of those 10000 regressions.

Moreover, Nelson and Kang (1984) also demonstrated that the residuals of those regressions will tend to exhibit pseudo-periodicity, that is, they will often mislead researchers into "finding" cyclical behaviour that is nonexistent in the data.

The simplest example of a process which is $I(2)$ is the second order random walk, where the process that accumulates is itself a first order random walk process. Denoting with y_t this one, $y_t = y_{t-1} + \epsilon_t$, it is given by

$$\begin{aligned} z_t &= z_{t-1} + y_t \\ &= z_{t-1} + y_{t-1} + \epsilon_t \\ &= 2z_{t-1} - z_{t-2} + \epsilon_t, \end{aligned}$$

which requires differencing twice to become stationary:

$$\Delta^2 z_t = (1 - L)^2 z_t = \epsilon_t.$$

It can be interpreted as a trend where the slope — its first difference — is a random walk. If the variance of the shocks is small, this random walk changes slowly, meaning that the slope of the trend changes slowly. Hence, as noted by Pollock (1998), for long periods it may appear to follow a linear time trend.

More generally, variables that are $I(2)$ usually exhibit even more smoothness than $I(1)$ variables. Although the $I(2)$ model appears to be only rarely adequate to describe macroeconomic time series[18], Clements and Hendry (1999) recommend its use for forecasting because its forecasts are more robust to structural breaks, tending to adapt more quickly to them, even when there is no statistical evidence (e.g., favourable unit root test statistics) supporting that practice.

[18]This appears to be the case only for some price indexes as well as for some nominal variables.

3.9 Revisiting the segmented trend model

Starting with the highly influential work by Perron (1989) as well as with the one by Rappoport and Reichlin (1989), a large branch of literature developed defying or at least examining in more detail the evidence about unit roots in macroeconomic time series. In a simple and intuitive but imprecise way, the problem was of a lack of power of unit root tests when there was a break or a shift in an otherwise deterministic linear trend. Such segmented or broken trend processes confounded tests, producing misleading evidence supporting stochastic trends.

Although Campbell and Perron (1991) were very clear expositing the importance of correctly specifying the deterministic components in testing regressions for consistent and efficient inference, in some cases the problem was not exactly this one. Rather, as early acknowledged in both seminal papers, it was instead one of choosing the frequency with which large shocks, those shocks with permanent, long-run effects occur. Indeed, the rather extreme and implausible opposition between the TS and the DS models of the previous section is ill-posed. The choice must not be between a trend that never changes and a trend that is always changing with each new shock. Large shocks, with permanent effects, may occur more often than the TS model admits but less often than the DS model imposes. Such large shocks include, for instance, the 1929 crash, World War II, the oil shocks of 1973 and the beginning of the 1980s, the recent Global Financial Recession and the covid19 pandemic, etc. The problem is that modeling these large shocks is difficult because they possibly arise from a distribution that is different from the one that generates regular or normal shocks. Hence, as acknowledged in Perron (1989), breaks in the deterministic component were used as a device to remove those shocks from the noise function and into the deterministic trend, without having to model them explicitly. The segmented deterministic linear trend model gained new momentum but it could *not* be considered an "unconditional representation of the time series properties of the various variables" (Perron, 1989, p. 1387).

Perron (1989) built a rather systematic and exhaustive framework of

analysis, considering several distinct cases and the single break situation, as well as different transition mechanisms inspired in the time series literature. More recent surveys summarizing Perron's approach as well as some others are, *inter alia*, Haldrup, Kruse, Teräsvirta and Varneskov (2013) and Perron (2006). His work was further extended in several directions but, besides a renewed interest in the segmented trend model and in non-linear models in general, no grand, all-encompassing or very useful model emerged for the trend. Moreover, in some cases the model of the alternative hypothesis was not even really considered as an alternative because it had to be merged with the null model to form a nesting regression, statistically useful to provide the test but meaningless.

A model that appears to be particularly useful for the case of level shifts is the general model proposed by Lanne, Lütkepohl and Saikkonen (2002) for the case where the time of the break is known:

$$y_t = \alpha + \beta t + f_t(\theta)'\gamma + u_t,$$

where θ and γ represent $(m \times 1)$ and $(k \times 1)$ vectors respectively, and $f_t(\theta)$ is a $(k \times 1)$ vector of deterministic sequences. In the most interesting case, $f_t(\theta)$ defines a slowly evolving trend after the time of the break (T_b) and is given by

$$f_t(\theta) = \begin{cases} 0, & < T_b \\ 1 - \exp\{-\theta(t - T_b + 1)\}, & t \geq T_b, \end{cases}$$

which allows for a smooth deterministic shift resembling the (exponential) transition function proposed by Sollis (2005) and presented in section 3.7. Figure 3.4 presents a series simulated with a segmented trend model with a break of this type in 2020. The smoothness of the break makes its dating by visual inspection very hard or even impossible.

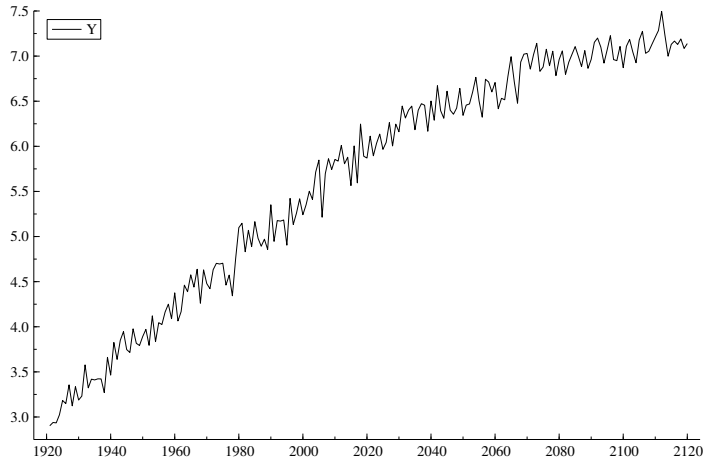

Figure 3.4: A series simulated with the segmented trend model with a break of the type suggested by Lanne et al. (2002) in 2020

3.10 Pure and mixed autoregressive models

Hansen (2022, p. 474) attributes to Nobert Wiener and to Pesi Masani, in a paper published in 1958,[19] the authorship of a theorem ensuring the existence of an infinite order autoregressive representation for any stationary process whose coefficients of its Wold representation satisfy a summability condition. But regardless of this result, the practical importance of autoregressive models to model the behaviour of economic time series has been established a long time ago, independently of the so called Box-Jenkins approach, not as a particular case of their general model. Autoregressive models are the most important econometric models for time series. Their interpretation is simple, their estimation is easy and the methods available to specify them have well established properties. Their basic role in the definition of Granger causality is also well known. They are usually considered as the best instrument to get

[19] "The prediction theory of multivariate stochastic processes, II", *Acta Mathematica*, 99, pp. 93-137.

rid of serially correlated regression errors, i.e., as the best available tool to capture dynamic relationships in the economy.

According to Hansen (2022, p. 477), "*the AR(1) model is probably the most important model in econometric time series analysis*". Further, in one of the most important textbooks in econometrics for undergraduates, Stock and Watson (2015) begin the presentation of the part on "regression analysis of economic time series data" with autoregressive models, which they also consider the preferable tool to use in forecasting [20]. Therefore, although it is true that, apparently eclipsed by unit roots, autoregressive models have not been often used recently to model trends, it is somewhat surprising that Mills (2003) implicitly rules out this possibility, reserving them only to model cyclical behaviour. At least, they might be helpful to improve the precision in the estimation of the coefficients of deterministic trend terms.

The autoregressive model of order p, AR(p), is

$$y_t = \alpha + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \epsilon_t, \quad \text{or}$$

$$\phi(L) y_t = \alpha + \epsilon_t,$$

where $\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ is the autoregressive polynomial. Stationarity demands that all the roots of $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$ lie outside the unit circle. In this case the OLS estimator of the autoregressive coefficients is consistent and if, besides white noise, ϵ_t is Gaussian as well, then it is the same as the conditional maximum likelihood estimator, and hence it is asymptotically efficient.

A polynomial in the deterministic linear trend term may be also added, as in the "augmented" Dickey-Fuller test regressions, but in most applications a simple linear trend will be sufficient, as in the following example, which is one of the final empirical models estimated by Seater (1993) with data series of annual average global temperatures over the

[20] And significantly, in the subject index of this textbook, autoregressive models are implicitly designated as *the* time series regression models (p. 825).

period from 1854 to 1989:

$$\widehat{y}_t = \underset{(1.220)}{-5.287} + \underset{(0.0006)}{0.0027t} + \underset{(0.086)}{0.494y_{t-1}}, \quad R^2 = 0.669.$$

Seater started his work on model specification with a cubic polynomial in t and an autoregressive polynomial of order $(p =)11$ and subsequently simplified his models using a general-to-specific (GTS) modeling strategy resorting to (asymptotically valid) F-tests until he arrived at models as the one presented as example. Instead of F tests he could have used equivalently individual t tests, that is, the popular $t-sig$ testing strategy. Both this strategy and the one that minimizes the AIC statistic are not consistent since they tend to overfit, leading to autoregressions which are too lengthy or liberal; on the other hand, and provided the starting order (usually called p_{max}) is at least as large as the true one, the strategy that minimizes the BIC (or SIC) statistic is consistent, that is

$$\lim_{T\to\infty} \widehat{p}_{BIC} = p,$$

where p now denotes the true lag length (see, e.g., Hayashi, 2000, pp. 394-7). However, my own empirical experience indicates that although it works well when the purpose is forecasting, in small samples the BIC criterion tends to produce autoregressions that are too short, too much parsimonious, and that sometimes fail in capturing all the relevant dynamics; both the GTS $t-sig$ method (usually with a liberal significance level, say $\alpha = 0.10$) and the AIC criterion appear to be preferable on these grounds.

Along the way to the final models, Seater (1993) reports also some collinearity problems which are typical of these models. To reduce them, we may resort to the useful polynomial decomposition

$$\phi(L) = \phi(1)L + \gamma(L)(1 - L),$$

where $\gamma(L)$ is a new polynomial in L of order $p - 1$, with coefficients $\gamma_0 = \phi_0 = 1$ and $\gamma_i = -\sum_{j=i+1}^{p} \phi_j$, $i = 1, \ldots, p - 1$. Considering the

"mixed version" with only one deterministic trend term, one may then write

$$\Delta y_t = \delta + \beta t + \phi y_{t-1} + \sum_{i=1}^{p-1} \Delta y_{t-1} + \epsilon_t,$$

where $\phi = -\phi(1)$, which is a typical augmented Dickey-Fuller (ADF) test regression as well. Since most regressors are now differenced, collinearity problems must be now almost completely absent.

On the other hand, modeling trends seems to dispense with the resource to threshold (TAR) and to smooth transition (STAR) autoregressive models as primary mechanisms. These models are particularly adequate to model asymmetric features which are common in cycles but that do not appear to be present in trends. Nevertheless, a logistic smooth transition autoregressive model (LSTAR) proposed by He and Sandberg (2006) to serve as alternative to the unit root hypothesis is worth mentioning:

$$y_t = \mathbf{x_t'}\pi_\mathbf{1} + \mathbf{x_t'}\pi_\mathbf{2} F(t) + u_t,$$

where $\mathbf{x_t} = (1, y_{t-1}, \ldots, y_{t-p}, t)'$, $\pi_\mathbf{1}$ and $\pi_\mathbf{2}$ are $(p+2) \times 1$ vectors of coefficients, e.g., $\pi_\mathbf{1} = (\pi_{10}, \pi_{11}, \ldots, \pi_{1,p+1})$, and $F(t)$ is a logistic transition function defined by

$$F(t) = \{1 + \exp[-\gamma(t-c)]\}^{-1} - \frac{1}{2},$$

where γ is the parameter that determines the speed of the transition from one regime to another, and c indicates the point about which the transition is symmetric, i.e., the midpoint of the transition[21]. The model then allows one transition over time for each parameter. But it does not necessarily impose that all parameters change; if, for instance, only the coefficients of the deterministic regressors are allowed to change, the restrictions $\pi_{21} = \ldots = \pi_{2,p} = 0$ must be imposed. A model similar to this one was already presented in section (3.7), but the current model

[21] He and Sandberg (2006) use a notation that is slightly different from the one I have used in section (3.7)

allows that dynamics may also change, which may be useful to improve the precision of estimation of the changing trend. To the best of my knowledge, this model was not yet estimated with any real data. It is not difficult, however, to find applications over differenced, non-trending data (an example is mentioned at the beginning of section 3.7).

3.11 Other models

Besides the previous models, there are other approaches that may be used to estimate trends. Some of them do not really rely on models, justifying that I use the term "approach" instead. For instance, Mills and Patterson (2015) describe in some detail three of these approaches proposed in the 1920s: rolling window estimation, the "arctan trend" and the concept of quadrature. The first approach is well known and it may indeed sometimes represent an effective way to track a growing but slowly changing trend. The resource to the inverse of the tangent function was a way to approximate a non-linear trend that has been rarely used and the concept of quadrature may be seen as a precursor of the random walk/white noise pair, i. e., of a process that is defined as the accumulation of another, much simpler.

Markov switching (MS) models are addressed in Mills (2003) but, like the TAR and STAR models of the previous section, they are not really used to model trends, at least directly. Instead, they may be very useful to capture nonlinearities that are usually reported on business cycles. The difference to the TAR and STAR models lies in the nature of the state variable that rules the regime switching of the series, which is unobservable (and discrete) in the case of MS models.

As acknowledged in Mills (2003), nonparametric trend modeling remains rather uncommon. Nonparametric estimation may be quite effective smoothing the observed series but the meaning of the estimated trend is indefinite. As the presentation in Mills (2003) suggests, nonparametric methods seem to follow closely the first approaches to the trend, aiming only to remove it from the data rather than to describe or

explain it. This appears to be the case as well with the method proposed recently in Fritz, Gries and Feng (2019), which contains some obscure aspects. For instance, the reason why one should estimate the "ν-th" derivative of a smooth function of time is simply not explained.

Finally, Granger, Inoue and Morin (1997) proposed a class of models for non-stationary processes that is wider than the linear unit root model, which is defined in general by

$$y_t = y_{t-1} + g(y_{t-1}) + \epsilon_t,$$

with $g(y) > 0$ and

$$\mathrm{E}(\epsilon_t^2 | y_{t-1}, y_{t-2}, \ldots) = \sigma^2(y_{t-1}) > 0,$$

which they call the nonlinear stochastic trend model. Several of the models of this class are presented and attention is focused on the "$K(\alpha, \beta)$ model", characterized by

$$g(y) = cy^\alpha, \ \alpha < 1, \quad \text{and} \quad \sigma^2(y) = \nu y^\beta$$

that is

$$y_t = y_{t-1} + cy_{t-1}^\alpha + \sqrt{\nu y_{t-1}^\beta}.\epsilon_t,$$

with $\epsilon_t \sim iid\mathcal{N}(0, \sigma_\epsilon^2)$. Its estimation is addressed, involving a three-step procedure, mostly based in maximum likelihood but requiring NLS to obtain initial estimates of α and β.

Several empirical applications are provided with monthly data over relatively long periods that provide large samples, with sizes in the order of 400-600 observations, rendering useful the conditional heteroskedasticity assumption of the model. However, no significant gains over simpler models appear to emerge clearly in the study. Perhaps this explains the modest number of empirical applications that followed. Neusser (1999) is an exception, providing a presentation focused in the $K(\alpha, \beta)$ model as well as several empirical examples.

Chapter 4

Decomposition Methods

Under the title of this chapter – decomposition methods– I subsume the procedures that are designed to remove or nullify components of the series which are not considered as relevant, as worthy of further attention, and that permit segregating the trend. That is, they are aimed at either removing the trend or, on the contrary, removing the cyclical (and possibly the seasonal) component(s) and isolate the trend. As is well known, most often their aim is to isolate the cycle, i.e., they usually belong to the toolbox of business cycles methods, but as mentioned in chapter 2, their aim may be instead to isolate the trend (as in Burns'1934 book). This is however seldom the case. Generally, these methods do not propose any model for the underlying trend, and they rarely suggest a way to predict their values, but there are important exceptions.

The most well known of these procedures are the so called filters, and their fundamental characteristic is precisely that one: they do not generate any model for the trend. Much of the material in this chapter refers to these filters, and readers less familiarized with their approach are invited to spend some time previously reading the appendix that is provided to introduce them.

Although the approach is mostly statistical, the economic perspective may be helpful for the case of GDP, particularly if the (restrictive) notion of the trend as representing potential output is adopted. Poten-

tial output is a concept from economic theory and since it represents an unobservable magnitude, it is often made to coincide with the trend or permanent component of GDP. Hence, methods that are devised to estimate it — or its difference to effective output, i.e., the output gap — are indeed methods allowing to estimate the trend, often with the purpose of subtracting it from output.

Even when considered jointly with the previous chapter, the present chapter does not exhaust the presentation of the wide spectra of trend estimation methods. For instance, Álvarez and Gómez-Loscos (2018), Chalmovianský and Nêmec (2022) and Ladiray, Mazzi and Sartori (2003) address several other methods used in economics, particularly multivariate ones, that lie beyond the scope of this work. Curiously, the common ultimate purpose of the methods surveyed in these papers is to measure the output gap. Nachane and Chaubal (2022) provide a pedagogical introduction to the digital signal processing filters, rarely used in economics, as well as a comparison between their properties and those of the two band-pass filters that will be addressed below. In other research areas, a recently proposed methodology which is mostly empirical, that of "empirical mode decomposition", is gaining popularity as a tool to remove nonstationary and nonlinear trends. For instance, it is employed in Li, Yao and Chau (2017) to uncover effects that appear to be masked by an increasing trend in temperatures on reference-crop evotranspiration (whose estimation is required as an input to schedule field irrigation, and to drought assessment and climate change research).

4.1 Moving averages

This section briefly reviews the simplest and oldest method of detrending, that of using moving averages, not to be confused with moving average models.

Simple moving averages

Detrending with a symmetric, simple, two-sided moving average is the oldest method for detrending, proposed by Hooker (1901), that can be considered as a smoothing method, i.e., one which attenuates or even completely removes the high frequency components of the series, while preserving those that are low frequency. Thus, it closely agrees with the idea of the trend as the smooth component of the series. And it is for this reason that Brockwell and Davies (1987) consider that it is a low-pass filter. It is also one of the most basic methods of detrending, dominating practice in the 1980s and 1990s within government and quasi-government organizations according to Osborn (1995).

A simple (symmetric, two-sided) moving average of the y_t series is given by

$$z_t = \frac{1}{2m+1} \sum_{j=-m}^{m} y_{t+j}, \quad t = m+1, m+2, \ldots, T-m,$$

with m a non-negative odd integer[1], which is sometimes denoted simply as MA($2m+1$), and called a MA($2m+1$) filter or smoother, the MA really corresponding to its *stricto sensu* (i.e., not as in "moving average" models). Hence, it is a particular case of a linear filter:

$$\sum_{j=-m}^{m} a_j y_{t+j} = (a_{-m}L^{-m} + a_{-m+1}L^{-m+1} + \ldots + a_0 + \ldots + a_{m-1}L^{m-1} + a_m L^m) y_t,$$

with weights $a_j = 1/(2m+1)$, $j = -m, \ldots, m$, $\sum_{j=-m}^{m} a_j = 1$. Therefore, it is indeed an average that moves along the original series, until z_t is computed at each observation for which all elements of the average that are available; as it advances, the oldest observation is dropped and the next observation is included in the computation.

[1]The case where m is even and/or the moving average is one-sided is addressed in e.g., Hyndman (2009) and Mills (2003). Hyndman (2009) addresses also the equivalence between weighted moving averages and kernel regression.

A well known problem affecting this method is the "end-effects" problem, the estimated trend becoming available only for observations from $t = m + 1$ to $T - m$ for obvious reasons. To remedy this problem the method is sometimes complemented with backcasting and forecasting procedures but I will not address these here.

An alternative patch consists in employing an asymmetric moving average with a gradually expanding window for the first and last m observations of the series (as in, e.g. Chalmovianský and Nêmec (2022)):

a) for $t \leq m$, $z_1 = \frac{\sum_{i=0}^{m} y_{t+i}}{m+1}$, $z_2 = \frac{\sum_{i=-1}^{m} y_{t+i}}{m+2}, \ldots, z_m = \frac{\sum_{i=-m+1}^{m} y_{t+i}}{2m}$;

b) for $t > T - m$, $z_T = \frac{\sum_{i=-m}^{0} y_{t+i}}{m+1}$, $z_{T-1} = \frac{\sum_{i=-m}^{1} y_{t+i}}{m+2}$, $\ldots$, $z_{T-m+1} = \frac{\sum_{i=-m}^{m-1} y_{t+i}}{2m}$.

Denoting the (estimated) detrended series with $\widehat{D}_t$, one may write it as

$$\begin{aligned} \widehat{D}_t &= y_t - a(L)y_t \\ &= [1 - a(L)]y_t \\ &= d(L)y_t, \end{aligned}$$

where $a(L)$ is the filtering polynomial of the previous equation and $d(L) = 1 - a(L)$ represents the detrending polynomial. Since the weights sum to unity, $a(1) = 1$, implying that $d(1) = 0$. Therefore, $d(L)$ can be factorized as

$$d(L) = (1 - L)b(L),$$

that is, as the detrending filter incorporates a differencing operator it will remove a single stochastic unit root; therefore, it will render any $I(1)$ process stationary.

Actually, it is easy to go even further and show that simple moving averages remove two unit roots that the series might have and hence they also reduce the order of any deterministic trend polynomial by 2 (Osborn, 1995). This also means that they introduce two moving average unit roots if the original series is $I(0)$, and one moving average unit root if it is $I(1)$. This is the most likely case given the prevalence of the $I(1)$ property among macroeconomic time series.

This is not, however, the most negative aspect of this method. Using tools from frequency domain analysis, Osborn (1995) also shows that the method is rather inadequate for business cycle analysis because it may induce spurious cyclicality. In particular, for the case of $I(1)$ processes, Osborn shows that the strength of this effect increases with the length of the moving average filter that is applied. This explains why the method appears having been abandoned in empirical applications.

Weighted moving averages

Weighted moving averages are obtained with weights that sum to one, that are symmetric, that is, $a_j = a_{-j}$, but that are not equal in absolute value, producing an estimated trend that can be much smoother because, instead of observations suddenly entering and leaving the calculations of the average, they can be slowly downweighted.

The Spencer 15-point moving average is a weighted moving average that can be obtained applying successively an equal weighted MA(4) twice to an equal weighted MA(5) and finally applying a weighted 5-term MA with weights. It can be shown that its weights [2] are:

$$a_0 = \frac{74}{320}, \quad a_{\pm1} = \frac{67}{320}, \quad a_{\pm2} = \frac{46}{320}, \quad a_{\pm3} = \frac{21}{320},$$

$$a_{\pm4} = \frac{3}{320}, \quad a_{\pm5} = -\frac{5}{320}, \quad a_{\pm6} = -\frac{6}{320}, \quad a_{\pm7} = -\frac{3}{320}.$$

Henderson's weighted moving averages, on the other hand, have been particularly popular in seasonal adjustment methods. They rely on an idea of local polynomial regression, with a (deterministic trend) cubic polynomial fitted to rolling samples of size $2m + 1$. Hence, polynomial trends of degree less than or equal to 3 pass undistorted through the filter. For further details see, e.g., Pollock (2016).

[2] The weights of the 21-point Spencer's MA are presented in section 2.4.

4.2 The Beveridge-Nelson (BN) decomposition

Beveridge and Nelson (1981, BN) proposed that the measure of the trend of a series should be given by its long-run forecast, i.e., "*the best estimate of where the variable will be in the distant future*" (Nelson, 2008). For this purpose, they rejected that the (then prevailing) linear trend model could be a valid possibility, as it implied a perfectly predictable evolution for the series and, as shown by Nelson and Kang (1984), could strongly distort the cyclical properties of the series. The BN decomposition is, however, also a model based approach to decompose a time series into a permanent or trend component and a transitory or cyclical component.

To obtain the original forecasting version decomposition, assume that the y_t series is integrated of order one, $I(1)$, with Wold representation for its first difference given by

$$\Delta y_t = \delta + \psi(L)\epsilon_t = \delta + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \tag{4.1}$$

with $\psi(0) = 1$ (i.e., $\psi_0 = 1$), $\psi(1) \neq 0$ (to rule out the possibility that $y_t \sim I(0)$), and $\sum_{j=0}^{\infty} j^{1/2}|\psi_j| < \infty$, where this $1/2-$summability condition, together with the Wold decomposition guarantee the existence and uniqueness of the decomposition[3].

The optimal prediction of y_{t+k} conditional on the information dataset available at time t (I_t) is

$$\mathrm{E}(y_{t+k}|I_t) = \widehat{y}_{t+k|t} = y_t + \widehat{\Delta y}_{t+1|t} + \widehat{\Delta y}_{t+2|t} + \ldots + \widehat{\Delta y}_{t+k|t}, \tag{4.2}$$

[3]See Oh, Zivot and Creal (2008) and note that this condition is less demanding than the usual of one-summability (e.g., Hamilton, 1994, p. 504, or Hayashi, p. 564).

with terms given by

$$\begin{aligned}
\widehat{\Delta y}_{t+1|t} &= \delta + (\psi_1\epsilon_t + \psi_2\epsilon_{t-1} + \psi_3\epsilon_{t-2} + \ldots), \\
\widehat{\Delta y}_{t+2|t} &= \delta + (\psi_2\epsilon_t + \psi_3\epsilon_{t-1} + \psi_4\epsilon_{t-2} + \ldots), \\
\widehat{\Delta y}_{t+3|t} &= \delta + (\psi_3\epsilon_t + \psi_4\epsilon_{t-1} + \psi_5\epsilon_{t-2} + \ldots), \\
&\vdots \\
\widehat{\Delta y}_{t+k|t} &= \delta + (\psi_k\epsilon_t + \psi_{k+1}\epsilon_{t-1} + \psi_{k+2}\epsilon_{t-2} + \ldots)\,.
\end{aligned}$$

Substituting into (4.2) gives

$$\widehat{y}_{t+k|t} = y_t + k\,\delta + \left(\sum_{j=1}^{k}\psi_j\right)\epsilon_t + \left(\sum_{j=2}^{k+1}\psi_j\right)\epsilon_{t-1} + \left(\sum_{j=3}^{k}\psi_j\right)\epsilon_{t-2} + \ldots.$$

Considering a very long horizon, an approximation is given by

$$\widehat{y}_{t+k|t} \approx y_t + k\,\delta + \left(\sum_{j=1}^{\infty}\psi_j\right)\epsilon_t + \left(\sum_{j=2}^{\infty}\psi_j\right)\epsilon_{t-1} + \left(\sum_{j=3}^{\infty}\psi_j\right)\epsilon_{t-2} + \ldots.$$

Therefore, what may be called the forecast function is a linear function of the forecast horizon, k, with slope δ, which is the drift parameter. It contains also a level or intercept, denoted with $\overline{y}_t$, that BN consider as the permanent or trend component of the series, which is given by

$$\overline{y}_t = y_t + \left(\sum_{j=1}^{\infty}\psi_j\right)\epsilon_t + \left(\sum_{j=2}^{\infty}\psi_j\right)\epsilon_{t-1} + \left(\sum_{j=3}^{\infty}\psi_j\right)\epsilon_{t-2} + \ldots, \quad (4.3)$$

and which is a pure random walk with drift:

$$\begin{aligned}
\overline{y}_t - \overline{y}_{t-1} &= \delta + \left(\textstyle\sum_{j=0}^{\infty}\psi_j\right)\epsilon_t \\
&= \delta + \psi(1)\epsilon_t,
\end{aligned}$$

where $\psi(1) = \sum_{j=0}^{\infty}\psi_j$ is the very long-run effect of a transitory shock ϵ_t, i.e., the limit of the impulse-response function as the horizon goes to infinity (and recall that it is not zero to rule out the possibility of

stationarity of y_t).

Returning to equation (4.3), one can write it as

$$y_t = \overline{y}_t + (\gamma_0\epsilon_t + \gamma_1\epsilon_{t-1} + \gamma_2\epsilon_{t-2}\ldots),$$

where $\gamma_j = -\sum_{i=j+1}^{\infty}\psi_j$, that is, y_t may be decomposed as the sum of a random walk — the trend —, accumulating the elements of a white noise process, $\epsilon_{t-j}, j = 0, 1, 2, \ldots$, and a stationary stochastic process, which is driven by exactly the same white noise process, the cycle.

The alternative derivation of the decomposition can be derived from the polynomial factorization[4]

$$\psi(L) = \psi(1) + (1-L)\gamma(L),$$

where $\gamma(L) = \sum_{j=0}^{\infty}\gamma_j L^j$, with $\gamma_j = -\sum_{j=0}^{\infty}\psi_j$. Using this factorization in equation (4.1), it may be written as

$$\Delta y_t = \delta + \psi(1)\epsilon_t + (1-L)\gamma(L)\epsilon_t.$$

Integrating and denoting with η_t the stationary process $\eta_t = \gamma(L)\epsilon_t$, one obtains

$$y_t = y_0 + \sum_{s=1}^{t}\Delta y_s = y_0 + \delta\, t + \psi(1)\sum_{s=1}^{t}\epsilon_s + \eta_t - \eta_0,$$

which is the version of the decomposition appearing in, e.g., Hayashi (2000): any linear $I(1)$ process can always be written as the sum of a linear time trend, a pure random walk or stochastic trend, a stationary process representing the cycle, η_t, and an initial condition $(y_0 - \eta_0)$.

It should be also noted that:

a) even if the process is driftless (i.e., $\delta = 0$), it contains a long-run

[4]This factorization is similar but it is not identical to another factorization of the previous chapter. To prove it, one must note that $\psi(L) - \psi(1) = -(1-L)[\psi_1 + \psi_2(1+L) + \psi_3(1+L+L^2) + \ldots]$. Pollock (2011) names this result as the "the ordinary remainder theorem of polynomial algebra".

or permanent or trend component, which is the stochastic trend. Only the deterministic trend disappears.

b) Since they are driven by the same shocks, the BN trend and cycle are perfectly correlated.

c) It is usually assumed that Δy_t follows an ARMA(p, q) process, so that

$$\psi(L) = \theta(L)/\phi(L),$$

with p and q denoting the orders of $\phi(L)$ and $\theta(L)$, the autoregressive and moving average polynomials, respectively, whose roots are assumed to lie outside the unit circle.

Apparently starting with Campbell and Mankiw (1987), an estimate of $\psi(1)$ based on ARIMA modeling has been used to measure the persistence of economic shocks. That is,

$$\widehat{\psi}(1) = \sum_{j=1}^{\infty} \widehat{\psi}_j = \widehat{\theta}(1)/\widehat{\phi}(1)$$

is the estimate of the long-run effect of shocks, with unity corresponding to the case of the random walk model as the reference value. For a (trend) stationary process $\psi(1) = 0$, as can be easily shown[5], and, if the order of integration is larger than 1, this measure is infinity (see Miller and Newbold, 1995). Estimates between 1.2 and 1.6 are common for series of GDP.

To the traditional, linear trend based decomposition, with no shocks with permanent effects, the BN decomposition opposed radically, resting on the unit root model and hence considering that all shocks have permanent effects. It is not surprising, then, that when applied to output series, BN estimated trend output almost coincides with actual output, leaving only a small role to the cycle. Ascribing much movements to the trend not only makes it much less smooth than customary but also

[5]It may also be helpful to revisit section 3.8, in the previous chapter.

implies that business cycles are very small, unimportant and noisy. This is the main critique that is usually made to the BN decomposition and indeed it may be noted that if $\phi(1) > 1$ the variance of the innovations of the trend ($\phi(1)^2\sigma^2$) is larger than the variance of the original innovations (σ^2). Another common critique is that since the decomposition depends on ARIMA modeling and different acceptable models may possess very different long-run properties, alternative specifications may imply very different decompositions[6].

Empirical implementation of the decomposition again may resort to ARIMA modeling. As reviewed in Mills (2003), a few simple calculations allow obtaining the BN trend as

$$\begin{aligned} \tau_t &= \tfrac{\phi(L)}{\theta(L)}\psi(1)y_t, \\ &= \omega(L)y_t, \end{aligned}$$

with $\omega(1) = [\phi(1)/\theta(1)]\psi(1) = 1$, i.e., the trend is a weighted average of current and past observations. Notice also the crucial role performed by the persistence measure of the series in the calculation. This equation permits the recursive estimation of τ_t. For instance, if the adequate model for Δy_t is an ARMA(1,1),

$$\Delta y_t = \delta + \phi\Delta y_{t-1} + \epsilon_t + \theta\epsilon_{t-1},$$

then $\phi(L) = 1 - \phi L$ and $\theta(L) = 1 + \theta L$, and hence

$$\tau_t = \left(\frac{1-\phi L}{1+\theta L}\right)\psi(1)y_t,$$

that is,

$$\begin{aligned} (1-\theta L)\tau_t &= (1-\phi L)\psi(1)y_t, \\ \tau_t &= \psi(1)y_t - \phi\,\psi(1)y_{t-1} - \theta\tau_{t-1}, \quad t = 2,3,\ldots,T. \end{aligned}$$

[6]See Álvarez and Gómez-Loscos (2016). Oh et al. (2008) contest the idea that BN cycles are always too noisy.

Therefore, it must be assumed that $\tau_1 = y_1$ to initiate the recursive calculations. The cycle is obtained residually ($c_t = y_t - \tau_t$) and the previous assumption now implies that $c_1 = 0$.

4.3 The Hodrick-Prescott filter

As mentioned in chapter 2, the so-called Hodrick-Prescott (1981, 1997)[7] procedure uses penalized least squares and coincides with the one proposed by Leser (1961). Hodrick and Prescott (HP) are very clear from the outset: their purpose is not to study the trend; the fluctuations that are aimed "*are those that are too rapid to be accounted for by the slowly changing demographic and technological factors and changes in the stocks of capital that produce smooth secular growth in output per capita*" (e.g., 1981, p.1).

Since the 1990s the procedure has become widely adopted and perhaps it is still the most popular method used for detrending in statistical and government agencies and in academic research. The main reason for this dominance may be found in the intuitive appeal of the procedure: the trend must be the best fitting curve to the data which, simultaneously, is as smoothest as possible. This idea corresponds very closely to the common, graphical or intuitive notion of the trend among economists: a curve that fits the data very well but which must sacrifice some goodness of fit for the sake of smoothness, as this is an indisputable feature that a trend must possess[8]. Moreover, since it is based in an optimizing procedure, it is in close harmony with the behaviour of rational, optimizing, economic agents.

In close similarity with Leser's (1961) decomposition (see subsection

[7]The history of this paper is succintly but delightfully described in footnote 2 of Hodrick's (2020) working paper.

[8]E.g., Canova (1998, p. 485):"... the implied trend resembles what an analyst would draw by hand through the plot of the data"; see also Kydland and Prescott (1990).

??, Hodrick and Prescott start from the decomposition

$$y_t = g_t + c_t, \tag{4.4}$$

where the trend, g_t, is called the "growth component" and c_t denotes the cycle or cyclical component (but notice that in Leser there is no cycle, only disturbances besides the trend). The data are assumed to be seasonally adjusted but, relatively to some traditional decompositions (e.g., Persons, 1919), there is no irregular or noise component; one must therefore presume that it is included in c_t.

As in Leser, the growth component is the solution to the problem

$$\min_{\{g_t\}_{t=-1}^{T}} \left\{ \sum_{t=1}^{T} c_t^2 + \lambda \sum_{t=1}^{T} [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \right\}, \tag{4.5}$$

where $c_t = y_t - g_t$ and λ is a positive parameter which penalizes variability in the growth component, i.e., it is a smoothing parameter: the larger it is the smoother the solution, which will be a linear time trend in the limit, when $\lambda \to \infty$. On the other hand, when $\lambda \to 0$ the solution will be the series itself. The terms of the second sum are easily recognizable as the second differences of the trend ($\Delta^2 g_t = \Delta\Delta g_t = \Delta(g_t - g_{t-1}) = (g_t - g_{t-1}) - (g_{t-1} - g_{t-2})$).

While maximizing the goodness of fit of the trend to the data appears to be a sensible purpose, one which is common sense or simply intuitive for most economists, minimizing the variability of the cyclical component, is questionable as a desirable property that one must search, unless that component is essentially noise. Viewing g_t as potential output — and hence c_t as the output gap — does not improve this state of affairs: potential output is a supply variable and seeking its optimal adherence to observed output is hardly an undisputable objective.

As regards the smoothing or penalizing parameter λ, under restrictive assumptions that both c_t and $\Delta^2 g_t$ follow white noise processes, with variances σ_c^2 and σ_g^2 respectively, HP find that its value should be given by the ratio σ_c^2/σ_g^2 and proposed that a reasonable and empirically based

value should be $\lambda = 1600$. This is the value typically used with quarterly data, but for annual data the usual choice is $\lambda = 100$.

Solving the matrix equation of the first order conditions produces the solution of (4.5) as (see e.g., Hamilton, 2018):

$$\mathbf{g}^* = (\mathbf{H}'\mathbf{H} + \lambda\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{H}'\mathbf{y},$$

where $\underset{(\tilde{T}\times 1)}{\mathbf{g}} = (g_T, g_{t-1}, \ldots, g_1)'$, with $\tilde{T} = T+2$, $\underset{(T\times 1)}{\mathbf{y}} = (y_T, y_{T-1}, \ldots, y_1)'$, $\underset{(T\times\tilde{T})}{\mathbf{H}} = [\underset{(T\times T)}{\mathbf{I}_T} \ \underset{(T\times 2)}{\mathbf{0}}]$, and

$$\underset{(T\times\tilde{T})}{\mathbf{Q}} = \begin{bmatrix} 1 & -2 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 1 & -2 & 1 \end{bmatrix}.$$

Hence, the rows of the matrix $(\mathbf{H}'\mathbf{H} + \lambda\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{H}'$ contain the filter weights that allow obtaining the trend estimates. And the rows of the matrix $\mathbf{I} - (\mathbf{H}'\mathbf{H} + \lambda\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{H}'$ represent the weights of the cyclical filter.

It is not easy to reconcile the wide popularity of the filter with the theoretical analyses published about it throughout time, beginning with the work of King and Rebelo (1993). These have been mostly critical, pointing to drawbacks and frailties that most practitioners have been reluctant to acknowledge. The reason for this lies very likely in the intuitive attractiveness of the filter, particularly in its smoothness, "*that shows where trending economic activity has been, is now, and where it may be heading*" (Phillips and Jin, 2021, p. 501).

The first criticism has been already introduced, implicitly at least: the framework of the procedure reduces the importance of the subject that it claims to be its object of study, the cycle, to the point of denying its existence. Although Canova (1998) considers that the HP filter is an economic procedure, there is no basis in economic theory that supports the notion of cyclical fluctuations as a component with minimum

variability. Although this is desirable and a goal of economic policy, what is required here is a measure of observed, actual, not of desired cyclical fluctuations. Searching for the decomposition that aims at this minimization logically stems from the replacement of Leser's disturbance term with the cyclical component but there is no further justification in this besides convenience reasons.

Actually, Leser's framework goes back at least to the work of Wiener and Kolmogorov in the 1940s, aiming to extract a signal from a series of observations when that signal is contaminated or corrupted by noise. In equation (4.5) the trend performs the role of the signal and, contradicting the stated purpose, the cycle corresponds to noise. Gómez (1998, 2001) and Pollock (2013, 2016) provide brief presentations of the Wiener and Kolmogorov's approach and explicitly consider the HP filter as one of its particular cases; however, both are silent about the adequacy of Hodrick and Prescott's approach to the decomposition of macroeconomic time series.

In a sense, however, the previous criticism is not completely new. Formally, it can be related with a result by Harvey and Jaeger (1993) according to which the HP filter is optimal for a DGP given by a structural or unobserved components model where the trend has a random walk growth rate, i.e., it is $I(2)$, and the cycle is reduced to a simple white noise process. This point will be resumed in section (4.5), where these models will be addressed.

Resorting to a simpler framework, in his proposition 1 Hamilton (2018) proves a similar optimality result for the filter, implying a serially uncorrelated cyclical component, i.e., with "*no discernible pattern*". As Hamilton (2018, p. 832) observes with some irony, this obviously and strongly contrasts with what "*users of the HP filter hope to see*": "*suggestive patterns in plots of the series that are supposed to be interpreted as the cyclical component of* y_t".

The work of Gómez (1999, 2001) is also important in this regard. Gómez showed that the HP filter is a Butterworth filter[9] for a signal

[9]Butterworth filters are filters used mainly in electrical engineering. The gain

extraction problem with the signal following an $IMA(2,0)$ process and the noise a white noise process uncorrelated with the innovations of the signal (as in Harvey and Jager, 1993). Besides the stringent and partially inappropriate conditions for optimality, this also opens the possibility of finding a better Butterworth filter. Gómez (2001) further shows that the HP filter for the cycle — that is, 1 minus the trend filter — is a rather good approximation for the ideal filter that passes components with periods smaller than eight years. Notwithstanding the general positiveness of this result, this is a serious shortcoming as well, because these frequencies include those with periodicities between six and two quarters, "*which cannot be considered to belong to a proper cyclical component*"[10]; that is, "*the cycle estimated with the HP filter will be more noisy than the one estimated with a band pass filter*" (Gómez, 2001, p. 367).

King and Rebelo (1993, KR) had also previously derived the inappropriate optimality conditions of the filter but directed their attention mainly to its distorting consequences, both in univariate analysis and for the operation of multivariate dynamic models. Some of their examples showed that cyclical fluctuations could be "*dramatically altered*" (p. 210) by the filter and that, as a consequence, gathered stylized facts could become corrupted. From the first order conditions for optimization and assuming an infinite sample, they derived the (approximate or asymptotic version of the) trend filter as

$$\tau(L) = [\lambda(1-\lambda)^2(1-L^{-1})^2 + 1]^{-1},$$

which is a two-sided moving average, and from which the one of the

---

function of the filter based on the sine function is $G(\omega) = \left[1 + \left(\frac{\sin(\omega/2)}{\sin(\omega_c/2)}\right)^{2d}\right]^{-1}$, where ω_c is a frequency such that $G(\omega_c) = 1/2$ and d is a positive integer. See also Pollock (2013, 2016).

[10]Recall that for quarterly data the classical approach by Burns and Mitchell (1946) defines business cycles as those components with no less than six quarters (1.5 years) and no more than 32 quarters (8 years) in duration.

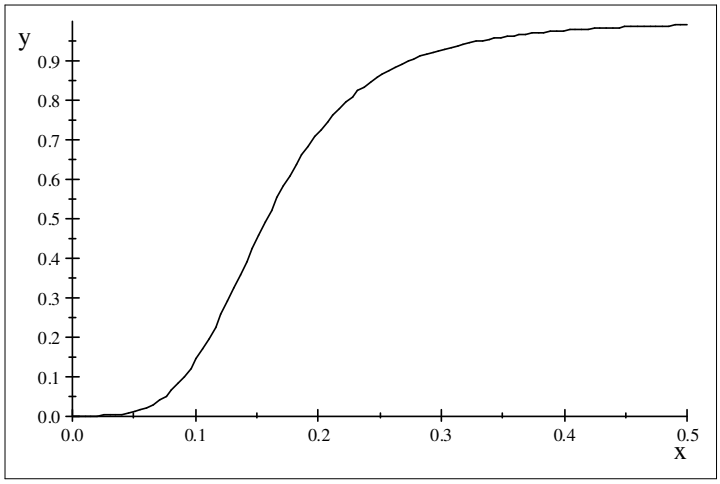

Figure 4.1: The gain function of the Hodrick-Prescott filter for $\lambda = 1600$

cyclical filter can be obtained:

$$c(L) = \frac{\lambda(1-L)^2(1-L^{-1})^2}{1+\lambda(1-L)^2(1-L^{-1})^2}. \tag{4.6}$$

Since the gain of this filter is

$$G(\omega) = \frac{4\lambda(1-\cos\omega)^2}{1+4\lambda(1-\cos\omega)^2},$$

the cyclical filter assigns zero power to the zero frequency (because $\cos 0 = 1$) and, since $\cos(\pi) = -1$, power close to unity to the high frequencies for large values of λ ($G(\omega) = 16\lambda/(1+16\lambda)$). Hence, as previously mentioned, the cyclical filter approximates an ideal high pass filter. Figure 4.1 presents the plot of the gain of the usual HP filter.

Cogley and Nason (1995) extend the critique of King and Rebelo (1993) to the point of showing that the filter may induce spurious effects — both univariate and multivariate, through RBC models — on persistent data, exaggerating the importance of fluctuations at the business cycle frequencies. The severity of the problem is particularly acute

when the original data is generated by random walks, in which cases business cycle dynamics may emerge as a complete artifact of the procedure (as clearly there were none originally). In this case, there is thus a close analogy with the spurious deterministic linear detrending of random walks described by Nelson and Kang (1984). Harvey and Jaeger (1993) prove the same result with a different method and, although agreeing with the spurious label, they also refer to the case as an example of the Yule-Slutsky effect (see section A.11).

Hamilton (2018) uses equation (4.6) to shed more light on the operation of the filter on random walks. Considering observations near to the middle of the sample, so that the asymptotic approximation performs well, and naming the polynomial of the denominator as $F(L)$, i.e., $F(L) = 1 + \lambda(1 - L^{-1})^2(1 - L)^2$, the estimated cyclical component can be approximated with

$$\begin{aligned} c_t &= \frac{\lambda(1-L^{-1})^2(1-L)^2}{F(L)} y_t \\ &= \frac{\lambda(1-L)^4}{F(L)} y_{t+2} \\ &= \frac{\lambda(1-L)^3}{F(L)} \epsilon_{t+2}, \end{aligned}$$

where ϵ_t denotes the (white noise) innovations of the random walk ($(1 - L)y_t = \epsilon_t$). Then, still following Hamilton (2018) closely, for the usual choice $\lambda = 1600$, it becomes

$$\begin{aligned} c_t = \; & 89.72 \Big\{ -q_{0,t+2} + \textstyle\sum_{j=0}^{\infty} (0.8941)^j [\cos(0.1117j) \\ & + 8.916 \sin(0.1117j)] \, (q_{1,t+2-j} + q_{2,t+2+j}) \Big\}, \end{aligned}$$

where

$$\begin{aligned} q_{0t} &= \epsilon_t - 3\epsilon_{t-1} + 3\epsilon_{t-2} - \epsilon_{t-3}, \\ q_{1t} &= \epsilon_t - 3.79\epsilon_{t-1} + 5.37\epsilon_{t-2} - 3.37\epsilon_{t-3} + 0.79\epsilon_{t-4}, \text{ and} \\ q_{2t} &= -0.79\epsilon_{t+1} + 3.37\epsilon_t - 5.37\epsilon_{t-1} + 3.79\epsilon_{t-2} - \epsilon_{t-3}. \end{aligned}$$

Therefore, starkly contrasting with the original innovations, the estimated cycle is highly predictable because it depends on past innovations

and, since it also depends on future innovations, it is capable of predicting the future as well. Hence, as the coefficients of $F(L)^{-1}$ depend only on the value assigned to λ, "*these patterns in the cyclical component are entirely a feature of having applied the HP filter to the data rather then reflecting any true dynamics of the data-generating process itself*" (Hamilton, 2018, p. 833). Hamilton further illustrates these spurious effects with the cases of the series of U.S. stock prices and real consumption.

Hamilton (2018) also points out that it is the two-sided nature of the filter that lies at the center of its attractiveness because it is its use of unknown future values that allows it to display the expected or imagined smoothness that many economists believe a trend must possess. For instance, in early 2009 it would be reasonable to expect that much of the 2008's crash of the stock market would be permanent or at least that it would take a long time to recover. With its "foresight", the filter was able to smooth the crash, as well as the rapid and strong appreciation that followed. The smoothness of the "long-run" component that most researchers enjoy observing is not a feature of the DGP; instead, it "*is just something that their imagination has imposed on the data*" (*ibidem*, p. 835).

Another drawback of the filter that many practitioners know well from applied work is the failure to stationarize integrated time series, in particular $I(1)$ series. This is a feature that runs against expectations because the filter is a detrending one and $I(1)$ series contain a stochastic trend. But series detrended with the HP filter often continue to fail rejection of the unit root hypothesis by, say, ADF tests. Moreover, the intuitive expectation is also theoretically grounded in KR's work, in particular in equation (4.6): since the numerator contains a fourth difference, the filter should render stationary any time series integrated up to the fourth order.

The problem with this inference, according to DeJong and Sakarya (2016, DS), is that it is based on the approximate version of the filter, which neglects the effect of the observations in the beginning and end of the sample. Indeed the first order conditions from which equation (4.6)

is derived cannot hold for those observations. DS prove that the HP filter is capable of removing a unit root from a DGP but only provided the resulting estimated cyclical component is limited to observations away from the beginning and end of the sample. The influence of these observations in practice may explain that the observed original nonstationary behaviour is passed to the estimated cycle.

Phillips and Jin (2021) also criticize the approach by KR and all that followed, labeling it as "*superficial*", pointing out that in equation (4.6) the numerator cannot be considered in isolation from the denominator; more precisely, they stress that the cyclical filter is a nonlinear rational function filter, where the "*denominator operator produces a smoothing operation on the data that accompanies and partially reverses the fourth difference operator in the numerator of the filter*" (p. 473). Although their approach is purely asymptotic, they claim that the popular value adopted for the smoothing parameter in quarterly data samples of common size ($\lambda = 1600$) is too large to remove a unit root, and that the price to pay for the smoothness of the trend is a estimated cycle "*that inevitably imports the random wandering character of a stochastic trend*" (ibidem, p. 486). Their asymptotic analysis, where they make $\lambda \to \infty$ as $T \to \infty$, leads them to state that it is not so much the value of λ that is a concern but rather that value in relation with the sample size; and $\lambda = 1600$ is "*extremely large*" for the usual sample sizes of quarterly data. Unfortunately, their recommendation to effectively remove a unit root has limited practical value because it requires only that, as $T \to \infty$, the smoothing parameter must be of order $o(T)$.

With such an important role, one would expect that the choice of λ would obey some rationality, that its value would be selected on the basis of some principle. As Harvey and Jaeger (1993) observed, however, if λ were estimated in the framework of the structural model where the filter is optimal, the cycle would be reduced to a white noise process. Imposing a value selected *a priori* counteracts this effect but has no support in empirical evidence. Hamilton (2018) obtains estimates for λ for some U. S. macroeconomic time series and finds that they are hugely at odds with the popular choice, much smaller than 1600, close to unity,

producing estimated trends that would be much choppier, barely differing from the original series.

When facing the choice between the smooth look and statistical optimality, economists understandably often tend to favour the first. What is questionable, even dangerous however, is to base all inferences in such unique, possibly flawed estimates (both for the trend and the cycle). Canova (1998) tried to compile stylized facts for U. S. business cycles applying a variety of detrending methods to 7 macroeconomic time series. One of his main conclusions was that the outcomes vary widely across detrending methods, making impossible to derive one single set of uncontroversial stylized facts. Another was that the HP filter may produce results that are not much dissimilar from a few other methods. He also found that the popular version of the HP filter tended to emphasize cycles with average duration between 4 and 6 years, which may be "*inappropriate*" in some cases. But, as Burnside (1998) pointed out in the first sentence of his paper, "*there is nothing misleading in the fact that different filtering techniques lead to different facts about macroeconomic time series*". On the contrary, given the absence of commonly accepted notions of trend and cycle, it is rather beneficial that different methods illuminate different angles of the subjects. The HP filter may continue being used, provided one bears in mind its drawbacks and limitations, and hence avoids basing the analysis uniquely on its outcome.

Despite the wide popularity of the method, in most applications the objective is to estimate the cyclical component; moreover, often the main purpose is multivariate. Crafts and Mills (2017) provide an interesting example for British GDP and industrial production trends for the period 1270-1913. However, they prefer to estimate the trend with a segmented trend model.

Recently, Phillips and Shi (2021, PS) proposed an iterated procedure with the HP filter, addressing its inadequacy to remove effectively stochastic trends, which they have named the boosted HP filter, bHP. The procedure really consists of the repeated application of the HP filter, with all the iterations following the first applied to the residual, i. e., to the cyclical component. In all these iterations it is now the trend that is

obtained residually, with

$$\widehat{\tau}_t^{(i)} = y_t - \widehat{c}_t^{(i)},$$

the superscript denoting the iteration. The purpose is to gradually purge the estimated cycle from any stochastic trend effect. PS claim that, asymptotically, the bHP filter is capable of removing trend mechanisms that might contain integrated processes, deterministic trends and even multiple structural breaks. In this analysis, not only $\lambda \to \infty$ with $T \to \infty$ but also the second tuning parameter, m, the number of iterations, grows to infinity as $T \to \infty$. Two data-based stopping rules are considered, one based on the outcome of an ADF unit root test, bHP-ADF, and the other on an information criterion, $IC(m)$, similar to BIC, the bHP-BIC.

In their simulation study, PS show that the new method performs much better than the standard HP filter in a MSE sense. This is the case particularly for the BIC version of the boosted filter which, however, usually requires much more iterations than the bHP-ADF.

Even more recently, Lu and Pagan (2023) made two criticisms to PS. First, using the asymptotic solution for the HP filter previously presented, they show that if the (true) trend and cycle are $I(1)$ and $I(0)$, respectively, the HP cycle is also $I(0)$, so that there is no integration leakage *per se*, and so boosting is not necessary to remove the unit root from the estimated cycle; rather, they ascribe the empirical evidence of a unit root in HP cycles to beginning and end of sample effects in small samples. Second, they contest the PS recommendation based on their simulation results, arguing that there are plausible DGPs whose boosting may not work well. This is particularly the case when permanent shocks are relatively unimportant in the first differenced series (Δy_t).

4.4 The exponential smoothing filter

The exponential smoothing (ES) filter was presented by King and Rebelo (1993, KR) as another example of a filter to compare with the HP filter but has not gained much popularity since then. The work of Lucas

(1980) provides an empirical illustration. To derive the filter, one may begin again with an optimization problem, which is now

$$\min_{g_t} \sum_{t=1}^{T} \left[(y_t - g_t)^2 + \lambda (g_t - g_{t-1})^2 \right], \tag{4.7}$$

where g_t continues to denote the growth component or trend and, therefore, since the terms of the second sum are its simple first differences (Δg_t), it is now the changes in the trend that are penalized.

From the first order condition, the solution is

$$g_t = \frac{1}{1 + 2\lambda - \lambda L - \lambda L^{-1}} y_t,$$

and hence the asymptotic version of the trend filter is now

$$\begin{aligned} \tau(L) = &\ (1 + 2\lambda - \lambda L - \lambda L^{-1})^{-1} \\ = &\ \left[\lambda (1 - L)(1 - L^{-1}) + 1 \right]^{-1}. \end{aligned} \tag{4.8}$$

Similarly to the HP filter, a matrix form solution can be provided for this infinite sample version:

$$\mathbf{g}^* = (\mathbf{I}_T + \lambda \mathbf{D}_\mathbf{1}' \mathbf{D}_\mathbf{1})^{-1} \mathbf{y},$$

where $\mathbf{g}^*, \mathbf{I}_T$ and $\mathbf{y}$ maintain their previous definitions and $\mathbf{D}_1$ is the $(T-1) \times T$ first difference matrix such that $\mathbf{D_1 y} = [\Delta y_2, \ldots, \Delta y_T]'$ [11].

The cyclical filter is easily derived from (4.8):

$$c(L) = \frac{\lambda (1 - \lambda L)(1 - L^{-1})}{1 + \lambda (1 - L)(1 - L^{-1})},$$

and, using an identity that results from Euler's identity (see the ap-

[11] See Yamada (2020), where this result, together with many other algebraic results, is presented. Yamada (2018) provides an alternative expression that shows that the estimated trend can be decomposed into the sum of a deterministic linear trend and a low-frequency component of the linearly detrended series (the residuals from the regression of the series on the linear deterministic trend).

pendix), the gain is

$$G(\omega) = \frac{2\lambda(1-\cos\omega)}{1+2\lambda(1-\cos\omega)}.$$

Since $\cos 0 = 1$, this filter has zero power at the zero frequency and conversely the trend filter has unity power. Moreover, since $\cos\pi = -1$, the cyclical filter ascribes power close to unity to the high frequencies.

Compared to the HP filter, the ES cyclical filter is slightly further away from the ideal filter because its gain is not so close to zero for frequencies below $\pi/16$ (i.e., for periodicities longer than 8 years) and it is not so close to 1 for frequencies larger than $\pi/16$ (i.e., for periodicities shorter than 8 years).

Unfortunately I am not aware of any recent empirical application that has resorted to the ES filter[12]. Tödter (2002) proposes an "extended" exponential smoothing filter obtained with a slight modification of the ES filter, which is now the solution to

$$\min_{g_t} \sum_{t=1}^{T} \left[(y_t - g_t)^2 + \lambda(\Delta g_t - \beta)^2\right],$$

where β denotes the sample average of the change in the trend component. Yamada (2020) proposes also a (slightly) "modified" HP (mHP) filter that he considers to be closer to the ES filter.

4.5 Unobserved components models

For an obvious reason, structural unobserved components (UC) models could have been presented in the previous chapter, dedicated to trend modeling. Moreover, they are the models that are the closest heirs of the historical idea that lies at the origin of the theme of this book —

[12]Moreover, the filter is usually omitted in both methodological surveys (as, e.g., Álvarez and Gómez-Loscos, 2018) and empirical comparisons (as, e.g., Chalmovianský and Němec, 2022).

that of the decomposition of economic time series into rather distinct components —, which could justify their presentation earlier. However, they are most often used with the cycle as their primary interest, with the purpose of isolating and analysing the cyclical component only, i.e., simply to detrend the data. Therefore, often they are mostly competitors of the Beveridge-Nelson decomposition and of the HP filter. And despite allowing a direct and immediate association with those economic components, as well as with the seasonal component, in what concerns the trend that association is seldom made with thorough descriptive and analytical purposes.

The UC approach appears to have been introduced chiefly in two papers, those of Harvey (1985) and Clark (1987) and, slightly later, in Harvey's (1989) book and in another paper, by Harvey and Jaeger (1993). To defy the then prevailing "Box-Jenkins" or ARIMA approach two main arguments were put forward:

a) the ARIMA approach does not allow a direct correspondence with the usual economic components;

b) both the ARIMA and the unit root testing approach are too much concerned with parsimony, and rely too much on (typically) small sample sizes, leaving some explainable but unexplained variation in the data; hence, it is argued that sometimes they may even yield misleading conclusions.

As an example of this last critique, Watson (1986) contrasts the (much larger) estimated long-run effect of an innovation using the ARIMA approach with the (much smaller) UC one. In the same vein, Clark (1987) argues against the large proportion of the variation in economic time series that is attributed to the trend by the unit root/Beveridge-Nelson approach. On the other hand, a drawback of UC modeling that is rarely acknowledged lies in its identification issues (see, in particular, Oh and Zivot, 2006).

Although the basic local linear trend model is the most appropriate to introduce UC modeling due to its simplicity, the absence of an explicit

cyclical component makes it somewhat unappealing. Instead, I present the basic equations of the *trend plus cycle model*, proposed by Harvey (1985):

$$y_t = \tau_t + c_t, \tag{4.9}$$

where, as usual, y_t denotes the log of the series (usually output), τ_t represents the trend, and c_t the cycle. Sometimes an irregular component, assumed to be Gaussian white noise is also added but, for instance in Clark (1987), it is deleted at a later stage due to an insignificant estimate of its variance. The model for the trend is the *local linear trend model*, a random walk,

$$\tau_t = \tau_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim iid\mathcal{N}(0, \sigma_\eta^2), \tag{4.10}$$

with stochastic drift or slope β_t, which is itself also a random walk

$$\beta_t = \beta_{t-1} + \xi_t, \quad \xi_t \sim iid\mathcal{N}(0, \sigma_\xi^2). \tag{4.11}$$

Therefore, it is a double random walk model[13], which means that y_t is assumed to be an integrated process of order 2, $y_t \sim I(2)$. This is a rather flexible trend model, one which can allow for a smooth structural break with some ease. Further, it is usually assumed that η_t and ξ_t are independent of each other. Combining these components so far and hence their disturbances gives rise to a single equation model with a single disturbance, which is called the reduced form model. The reduced form of the trend here is an ARIMA(0,2,1) process.

If $\sigma_\xi^2 = 0$ the trend is reduced to a random walk with constant drift and if, additionally, $\sigma_\eta^2 = 0$ it becomes a deterministic trend, $\tau_t = \tau_0 + \beta\, t$. On the contrary, the larger these variances the greater the stochastic movement in the trend. Even if $\sigma_\eta^2 = 0$ but $\sigma_\xi^2 > 0$, the process is still $I(2)$.

As regards the (stochastic) cycle, in e. g. Harvey (1985) and Harvey

[13]It is more often known as a double-drift trend model but, though not strictly correct, the label that I adopt appears to be more appropriate.

and Jaegger (1993) it is assumed to be generated (trigonometrically) with

$$c_t = \rho \cos \lambda \, c_{t-1} + \rho \sin \lambda \, c^*_{t-1} + \omega_t \tag{4.12}$$

$$c^*_t = -\rho \sin \lambda \, c_{t-1} + \rho \cos \lambda \, c^*_{t-1} + \omega^*_t \tag{4.13}$$

where c^*_t appears by construction and is unimportant, ρ is a damping factor on the amplitude of the cycle such that $0 \leq \rho < 1$ (so that it is stationary), λ is the frequency of the cycle in radians (so that the period is $2\pi/\lambda$), and ω_t and ω^*_t are both $iid\mathcal{N}(0, \sigma^2_\omega)$. The disturbances of all the components are independent of each other and of the irregular component.

Harvey (1985) provides the reduced form for c_t that shows that it is an ARMA(2,1) process, which becomes an AR(2) if $\sigma^2_\omega = 0$. But it is a special case of an AR(2) with the restriction $0 < \lambda < \pi$, because the parameters are restricted to imply that its roots are complex, which is a desirable feature because it implies "pseudo-cyclical" behaviour[14]. Equations (4.9) to (4.13) imply that the reduced form of the model is an ARIMA(2,2,4) representation which, since it is heavily parametrized, would very unlikely emerge as the final result of a typical exercise in ARIMA modeling for a common macroeconomic time series.

Harvey (1985) proposed also a different model for the trend, incorporating the cycle within it, which he called the *cyclical trend model*. Equations (4.9) and (4.10) are replaced by

$$y_t = \tau_t + \epsilon_t \quad \text{and}$$

$$\tau_t = \tau_{t-1} + \beta_{t-1} + c_t + \eta_t,$$

where $\epsilon_t \sim iid\mathcal{N}(0, \sigma^2_\epsilon)$, and equation (4.11) remains unchanged. This model appears also in Harvey (1989) but, apparently at least, it was subsequently abandoned.

On the other hand, as previously mentioned, the HP filter corresponds

[14] See, e.g., Harvey (1993), p. 18 or Mills (2003), pp. 29-31. See also Harvey (1993), pp. 182-86 for a more detailed treatment of this model for the cycle.

to model (4.9)-(4.11) with $c_t = \epsilon_t \sim iid\mathcal{N}(0, \sigma_\epsilon^2)$, i.e., with the cyclical component reduced to an irregular, white noise process and the very smooth trend resulting from the restriction $\sigma_\xi^2 = 0.025^2\sigma_\epsilon^2$ or $\sigma_\xi^2/\sigma_\epsilon^2 = 1/1600$ for quarterly data.

In practice, as proposed originally by Clark (1987), the most popular version of the UC model uses an autoregressive representation for the cyclical component instead of the sinusoidal one. Thus, equations (4.12) and (4.13) are replaced by

$$\phi(L)c_t = \omega_t, \;\; \omega_t \sim iid\mathcal{N}(0, \sigma_\omega^2), \tag{4.14}$$

where $\phi(L)$ is a stationary autoregressive polynomial, often of second order, i.e., $\phi(L) = (1 - \phi_1 L - \phi_2 L^2)$, with $\phi + \phi_2 < 1$. If $\phi_1^2 + 4\phi_2 < 0$ its roots are complex and hence it displays pseudo-cyclical behaviour. The reduced form model corresponding to equations (4.9)–(4.11) and (4.14) is an ARIMA(2,2,3)[15].

Estimation is usually carried out in the time domain framework and resorting to the Kalman filter. The first step consists in casting the model in state space form. For the model of equations (4.9)– (4.11) and (4.14) the state vector is $\alpha_t = (\tau_t, \beta_t, c_t, c_{t-1})'$, the measurement equation is

$$y_t = [1\ 0\ 1\ 0]\,\alpha_{\mathbf{t}},$$

and the transition equation is

$$\alpha_{\mathbf{t}} = \mathbf{T}\alpha_{\mathbf{t-1}} + \nu_{\mathbf{t}}, \quad \nu_{\mathbf{t}} \sim iid\mathcal{N}(\mathbf{0}, \mathbf{V}), \quad \text{where}$$

[15]Indeed, differentiating twice equation (4.9) produces $\Delta^2 y_t = \Delta^2\tau_t + \Delta^2 c_t$. Substituting from equations (4.10) and (4.11) yields: $\Delta^2 y_t = \xi_{t-1} + \Delta\eta_t + \Delta^2\phi(L)^{-1}\omega_t$. Multiplying by $\phi(L)$ gives $\phi(L)\Delta^2 y_t = \phi(L)\xi_t + \phi(L)\Delta\eta_t + \Delta^2\omega_t$. Since the right hand side of this equation is a MA(3) process (see, e.g., Oh and Zivot, 2006) and since $\phi(L)$ is a second order autoregressive polynomial, the process follows an ARIMA(2,2,3).

$$\mathbf{T} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \phi_1 & \phi_2 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \nu_{\mathbf{t}} = (\eta_t, \xi, \omega_t, 0)', \quad \text{and} \quad \mathbf{V} = \begin{bmatrix} \sigma_\eta^2 & 1 & 0 & 0 \\ 0 & \sigma_\xi^2 & 0 & 0 \\ 0 & 0 & \sigma_\omega^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Application of the Kalman filter to these equations provides first estimates of the trend and cycle that are based only on current and past observations and that are known as filtered estimates. The Kalman filter provides also one-step-ahead predictions and prediction errors or innovations that, together with the Gaussian assumption, allow constructing the likelihood function. The likelihood is maximized numerically to estimate the parameters σ_η^2, σ_ξ^2, σ_ω^2, ϕ_1 and ϕ_2. Finally, a smoothing algorithm is run backwards, starting from the estimates produced by the Kalman filter at time T, to produce smoothed estimates of the trend and cyclical component. These are the estimates that are usually retained and are indeed smoother than the filtered estimates because they use all the observations in the sample.

The model of (4.9)– (4.11) and (4.14) is very close to the Clark (1987) model and will be referred as such in what follows. The observation of the Q matrix reveals that the model contains many restrictions. In particular, several orthogonality restrictions are imposed affecting all the shocks or innovations of the model: $\sigma_{\eta\xi} = \sigma_{\eta\omega} = \sigma_{\xi\omega} = 0$. Indeed, without imposing *a priori* some assumption(s) concerning the correlations between these shocks it would not be identified.

In the context of UC models identification is often not addressed but its nature is the usual: identification of the UC, structural model requires that its parameters are solved uniquely from the knowledge of the reduced form parameters. But only somewhat recently it became clear that so many restrictions as previously imposed were not really necessary to identify the model, i.e., the model is over-identified. In particular, although usually considered non-realistic[16], the assumption that the correlation between the shocks to the trend and to the cycle ($\sigma_{\eta\omega}$) is

[16] Clark (1987, p. 800) himself acknowledged that "*strictly speaking, this assumption is clearly false*".

zero, despite not strictly necessary to achieve identification, is frequently adopted in UC models. It is known to produce a smooth trend, not very far from a linear one, and large and relatively smooth business cycles, i.e., decompositions that differ radically from the Beveridge-Nelson ones.

Oh and Zivot (2006) considered the following two cases of exact identification:

$$\text{I) } \sigma_{\eta\xi} = \sigma_{\xi\omega} = 0 \ (\sigma_{\eta\omega} \neq 0) \quad \text{and} \quad \text{II) } \sigma_{\eta\xi} = \sigma_{\eta\omega} = 0 \ (\sigma_{\xi\omega} \neq 0).$$

Han, Liu and Ma (2020, HLM) label case I as the *trend-cycle case*: the shocks to the trend and to the cycle can be correlated but the shock to the drift must be orthogonal to the other two; and they label case II as the *drift-cycle case*: the shocks to the drift and the cycle are allowed to be correlated but the shocks to the trend need to be independent from the remaining two.

HLM estimate these two models as well as the original ("*independent*") one and, although the estimated models share some features, others differ significantly, demonstrating that identification issues are indeed very far from innocuous in these models. The models can produce Beveridge-Nelson-like decompositions, with volatile trends and small and noisy cycles, as well as radically different ones, depending on the restrictions that are imposed. These restrictions can be tested with likelihood ratio tests but, in some cases, as those of cases I and II above, tests for non-nested hypotheses are needed and, as far as I am aware, these have not yet been used in this context. Moreover, in some cases, as appears to be the case in HLM, data may not contain enough information to discriminate between rival assumptions[17].

Oh and Zivot's (2006) paper is an extension to an influential paper by Morley, Nelson and Zivot (2003, MNZ), who considered a restricted version of Clark's model, with constant (deterministic) drift ($\sigma_\xi^2 = 0$) and showed that, instead of being restricted, the correlation between the shocks of the trend and the cycle could be identified from the data

[17] In the current context, given the imprecision associated with the concepts of trend and cycle, this is not necessarily a "bad thing".

alone. Moreover, in that case they also reconciled the BN decomposition with the UC one, in sharp contrast with popular wisdom based on this one, which assigned most of the variation in output to transitory/cyclical shocks[18]. That is, the UC decomposition is also able to yield much more (less) volatile trends (cycles) than was previously thought, provided their correlation is left to be determined freely.

Although some of them are somewhat "dry", all the papers mentioned so far in this section contain also interesting empirical illustrations. In Mills and Crafts (1996) the version of the model with the sinusoidal cycle is applied to the series of the British industrial sector output, the sample ranging from 1700 to 1913. In an interesting paper, Ball and Wood (1996) criticize some of the work of Mills and Crafts. They adopt two alternative specifications differing only in the restrictions imposed on the trend with little to choose empirically according to statistical criteria. However, the models imply radically different estimated trends and, consequently, radically different estimated cyclical properties as well, demonstrating that statistical analysis alone may not be able to act as a neutral arbiter of disputes concerning historical questions.

Han, Liu and Ma (2020) use the autoregressive version of the cycle to model annual GDP data for China, for the period between 1952 and 2017. Their analysis is complemented with qualitative, historical information, to "*make sense*" of the estimated trend and cycles. They interpret the cyclical component as the output gap and propose a simple method for its empirical validation, analysing its relation with the inflation rate (which they later include in a bivariate UC model). Another interesting application is that of de Bonis and Silvestrini (2014), who estimate several versions of Harvey's (1985) and Clark's (1987) models with annual Italian data from 1861 to 2011 of the credit-to-GDP ratio. The statistical analysis both benefits and feeds a detailed historical analysis of the Italian banking sector.

A rather distinct use of Clark's model is made in Hodrick (2020),

[18] Although this reconciliation was incomplete because the UC model of MNZ still implies that $y_t \sim I(2)$ (while the BN decomposition assumes that $y_t \sim I(1)$), Oh and Zivot (2006) showed that it still holds for BN decompositions of $I(2)$ processes.

where it serves as one of the DGPs in a simulation study designed to assess its own properties and to compare it with the HP filter, as well as with the Baxter and King and Hamilton filters. Both these procedures will be addressed below.

The previous "classical" models have been extended and/or generalized more recently. For instance, Harvey and Trimbur (2003) generalize equations (4.10)-(4.11) and (4.12)-(4.13) so that the model contains a m^{th}-order stochastic trend and a n^{th}-order stochastic cycle, respectively, with m and n positive integers[19]. Perron and Wada (2009), propose to constrain the drift of the trend function to be deterministic but introduce a change in its slope at an exogenously determined date and, in another alternative model, maintain equations (4.9)-(4.11) and (4.14) but model the shocks to the drift and to the cycle as mixtures of Gaussian white noise processes with weights given by Bernoulli random variables. And Grant and Chan (2017) propose replacing equations (4.10)-(4.11) for the trend with a second order Markov process ($\Delta\tau_t = \Delta\tau_{t-1} + u_t^\tau$, $u_t^\tau \sim iid\mathcal{N}(0, \sigma_{u^\tau}^2)$) and model the cycle with a general stationary AR(p) process.

As regards software, Pedregal (2022) offers a panoramic view of the available options, ranging from the general statistical software (such as `Eviews` or `RATS`) and from popular environments as `R` or `MATLAB`, to `STAMP`, a commercial package dedicated exclusively to UC modeling. In the same article, Pedregal provides also a brief but compelling presentation of `UComp`, a powerful open source library of procedures for building and analysing UC models.

4.6 The Hamilton filter

Pure and mixed autoregressive models have already been addressed in section 10 of the previous chapter. The "Hamilton filter", as has been dubbed in the recent literature, relies on a particular, fixed order, pure

[19] E.g., for the trend: $\tau_{1,t} = \tau_{1,t-1} + \eta_t$ and $\tau_{i,t} = \tau_{i,t-1} + \tau_{i-1,t}, i = 1, 2, \ldots, m$, with $\eta_t \sim iid\mathcal{N}(0, \sigma_\eta^2)$.

autoregressive model, containing only a constant besides the autoregressive terms. Hamilton (2018) proposed it consistently with his critic of the HP filter and recently it has been subject to close scrutiny in Schüler (2021) and in Quast and Wolters (2022). Although it is not immune to criticism, it closely matches the simple but compelling requirements summarized by Phillips and Jin (2021), showing "*where trending economic activity has been, is now, and where it may be heading*". However, as its previous rivals in this chapter, it is chiefly designed to isolate the cyclical component. Strictly speaking, although it is linear, it is not a linear filter in the usual sense because it does not operate always the same way on every series; it is not the same linear transformation that is applied across all series. Instead, as will be seen below, it is data-dependent, that is, the way it operates changes with the dynamic properties of each series and hence its weights are not invariant. For this reason, I prefer to refer to it as Hamilton's method (HM). Nevertheless, as with the HP filter, often I will follow the literature.

The Hamilton filter is clearly inspired in the Beveridge-Nelson decomposition and hence it adopts also a forecasting approach[20]. The cycle is the answer to the question: "*how different is the value at date $t+h$ from the value that we would have expected to see based on its behavior through date t?*" And the trend, the expected values, are the "*very simple forecasts within a restricted class: the population linear projection of y_{t+h} on a constant and the four most recent values of y as of date t*" (Hamilton, 2018, p. 836, h denoting as usual the forecast horizon). Provided that the order of the autoregression, p, is equal or higher than the order of integration of the series (d) or the order of the deterministic polynomial in time that produces a stationary (I(0)) deviation, Hamilton shows that the OLS residuals of the regression of y_{t+h}, the value to forecast, on $(y_t, y_{t-1}, y_{t-p+1}, 1)$, that is,

$$\widehat{v}_t = y_{t+h} - (\widehat{\beta}_0 + \widehat{\beta}_1 y_t + \widehat{\beta}_2 y_{t-1} + \ldots + \widehat{\beta}_{p-1} y_{t-p+1}),$$

[20] As Hodrick (2020) notes, while Beveridge and Nelson's purpose it to forecast for a very distant future, that we may formalize with $\lim_{h\to\infty} \mathrm{E}_t(y_{t+h})$, in Hamilton the horizon is only eight quarters ahead and so the forecast is $\mathrm{E}_t(y_{t+8})$.

converge to a stationary process that can be used to represent the cycle. Assuming that the maximum admissible order of integration is 4, $d_{max} = 4$, Hamilton proposes that the regression is

$$y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \beta_4 y_{t-3} + v_{t+h},$$

so that the fitted values of y_{t+h} provide the estimated trend.

As regards h, the forecast horizon, assuming that the data are quarterly, Hamilton proposes that $h = 8$ because he considers that 8 quarters is the horizon at which cyclical factors, such as a recession or a "*recovery from any downturn*", play the primary role at justifying forecast errors for macro and financial variables. A distinct reason, the frequency of data, is used as a further argument for the convenience that h (and p as well) is a multiple of the number of observations in a year.

Hamilton (2018) does not claim the superiority of his proposal over any other method to estimate the trend. The merits of autoregressive models and their appeal for macroeconomic time series are well known and have been mentioned in the previous chapter. Instead, his goal is to avoid the drawbacks of the HP filter and hence his major concern is to show that, under the very liberal conditions previously mentioned, his filter always provides a stationary estimated cycle. In particular, for the case where the HP filter exhibits a strong vulnerability, that of the random walk process, Hamilton shows that asymptotically his filter converges to a difference filter, i.e., $\widehat{\beta}_1 \xrightarrow{p} \beta_1 = 1$ and all other coefficient estimators converge to zero, and hence the filtered series converges to the overall change between periods t and $t + h$, that is, with $h = 8$, to $\epsilon_{t+8} + \epsilon_{t+7} + \ldots + \epsilon_{t+1}$. This limiting filter, $1 - L^8$, thus "*wipes out any cycles with frequency exactly one year", that is, besides detrending* it also removes any seasonality, aiming to render a clean cyclical component.

Hamilton further argues for the following gains of this method over the HP filter:

a) any forecasting power of the estimated cycle is not an artifact of the method because it is a one-sided filter and hence it does not borrow any information from the future;

b) the estimated cycle is model-free and, as an outcome of a simple autoregression, "*essentially [an] assumption-free summary of the data*" (*o. cit.*, p. 837).

Moreover, one does not need to know the exact form of the non-stationarity of the series nor the correct or true model for forecasting. Further, although equating a trend with a more or less medium to long-term forecast is a common and intuitive practice, the merit of the proposal lies in the simplicity of the forecasting model.

Since Hamilton's chief concern is to ensure that the cycle is stationary, not that the trend contains only the variation that is due to permanent, non-stationary fluctuations, it may contain transient, short-lived fluctuations that may show up as erratic movements. More precisely, Hamilton shows that when the order of integration of the series, d, is smaller than $p(= 4)$, while the first d regressors $(y_t, \ldots, y_{t-d})$ serve to absorb non-stationarity, the remaining $p + 1 - d$ capture part of the variation of $\Delta^d y_t$, which is stationary. Thus, these transitory fluctuations will pass to the estimated trend, which will then tend to show less smoothness than expected. This is perhaps the major criticism to the procedure, which has been empirically confirmed for the case of US GDP by Quast and Wolters (2022). Their aim is to get a measure of potential output and they provide some reasons why it must be smooth, not the "*noisy*" estimate provided by Hamilton's procedure.

Another unsurprising criticism concerns the (fixed) order of the autoregression. Hodrick (2020) notes that the regression of y_{t+8} onto y_t and its first three lags may not be sufficient to capture all the forecasting information that is contained in the past history of the series. Augmenting the order of the autoregression appears to be the obvious and straightforward answer but it risks to further worsen the noisiness of the estimated trend.

Most of the criticisms are formulated by Schüler (2021), but several of them simply originate in a rather different perspective from that of Hamilton, one that is based in frequency domain analysis. While acknowledging that Hamilton's procedure avoids the drawbacks of the HP

filter[21], Schüler bases his critical approach on a bandpass filter view, which radically diverges from the forecasting approach of Hamilton. In particular, Schüler frequently criticizes Hamilton's method (HM) for not being amenable to a strict frequency domain approach. And indeed, although it is linear it is not strictly a linear filter, as previously mentioned. The issue lies in the series- or data-specific nature of HM, with coefficients that are not fixed across series, although they are time invariant for each series. This implies that "*the cyclical properties* of the Hamilton filter vary across data generating processes (DGPs)" (Schüler, 2021, p.2), contrasting with those of band pass filters, that are constant across series. This is obviously due to the fact that the weights or coefficients of HM depend on the dynamic, autoregressive properties of each series, preventing a general characterization of its properties.

Hence, Schüler (2021) correctly points out that "*the Hamilton filter does not separate an underlying time series into a cyclical component and a trend component according to some cutoff frequency*" (*o. cit.*, p. 14). And it is only provided that one assigns cyclical fluctuations as corresponding to frequencies with periods between 1.5 and 8 years that several criticisms arise concerning estimated cycles. For instance, both Schüler and Quast and Wolters (2022) agree that HM tends to provide an uneven coverage of those frequencies. They further agree that it tends to mute short cycles while simultaneously amplifying medium and long term cycles, with periods between 10 and 32 quarters. That is, the HM may distort some features of business cycles as commonly perceived.

Simulation studies appear to be more useful than frequency domain based approaches in this case. Provided they employ a sufficiently wide variety of DGPs, they may shed some light on the general performance of HM. Although their focus is restrained to the properties of the estimated cycles, both Hodrick (2020) and Schüler (2021) offer this type of analysis.

Hodrick compares the HP filter, the bandpass filter of Baxter and King (1999, BK) and the Hamilton filter, while Schüler compares the

[21] "*The Hamilton filter does indeed avoid spurious cycles, ad hoc filter settings, and end-of-sample bias*", Schüler (2021, p.1).

Hamilton and HP filters. Hodrick obtains a quite interesting conclusion: while HM dominates the HP and BK filters in traditional, classical or simple environments where there is no clear-cut separation between trends and cycles (such as those that are generated by random walks and ARIMA models), the reverse is true when data is generated with more complex models, containing growth (trend) components that vary slowly over time.

Schüler's results do not allow him to assert that, in general, the HP filter dominates HM, but he finds that the first tends to outperform the second in matching simulated cycles. His first recommendation for this purpose is, therefore, the HP filter, with HM relegated to a secondary role. He considers, however, that HM may provide a rather useful, complementary perspective to the HP filter.

Quast and Wolters (2022) proposed a simple modification of HM to correct the shortcomings they perceived. Instead of a fixed 8 quarter forecast horizon, they propose taking the average of forecasts with horizons from 4 to 12 quarters as a smoother estimate of the trend (potential output in their case). That is, start by running the regressions

$$y_t = \beta_{0,i}+\beta_{1,i}\, y_{t-i}+\beta_{2,i}\, y_{t-i-1}+\beta_{3,i}\, y_{t-i-2}+\beta_{4,i}\, y_{t-i-3}+v_{t,i}, \quad i = 4, 5, \ldots, 12,$$

and obtain the corresponding forecast errors

$$\widehat{v}_{t,i} = y_t - (\widehat{\beta}_{0,i} + \widehat{\beta}_{1,i}\, y_{t-i} + \widehat{\beta}_{2,i}\, y_{t-i-1} + \widehat{\beta}_{3,i}\, y_{t-i-2} + \widehat{\beta}_{4,i}\, y_{t-i-3}),$$

allowing to estimate the cycle with

$$\widehat{v}_t = \frac{1}{9}\sum_{i=4}^{12} \widehat{v}_{t,i}.$$

According to Quast and Wolters, this produces both a smoother trend and a more comprehensive coverage of the frequencies considered to represent business cycles.

Similarly to Hodrick (2020), Quast and Wolters (2022) perform a sim-

ulation study comparing HM, both the original and their modified version (MHM), with the HP filter and a bandpass filter due to Christiano and Fitzgerald (1999). But instead of a Monte Carlo study, their simulations aim to assess the "*real-time reliability*" of estimated output gaps, computing them on the basis of real-time data vintages that are compared to those that result from revised data. They conclude that both HM and MHM are the most reliable because their revisions are (much) smaller and are mainly caused by data revisions. That is, both the HM and MHM present the most favourable real-time properties[22]. Jönsson (2020) arrives at the same conclusion when comparing the Hamilton and the HP filter with recursive estimates for US trend and cycles of US GDP, as well as with data with several vintages. According to Jönsson, the source for the instability of the HP filter estimates lies in a large volatility of its weights for the last sample observations as more data become available.

4.7 Band-pass filters

For the sake of completeness, I now briefly address band-pass filters (BP), and the Baxter and King (1999, BK) and Christiano and Fitzgerald (2003) filters in particular. While previous filters proposed trend estimators as means to derive business cycle measures, band-pass-pass filters are rather different: trend estimation is pointless as cyclical fluctuations can be isolated directly. As their name indicates, BP filters are filters that pass through components of time series with periodic fluctuations inside a predetermined range or band of periodicities or frequencies, while removing components at higher and lower frequencies. Since Burns and Mitchell identified U.S. business cycles as those cyclical fluctuations with lengths between 6 and 32 quarters, BP filters are now commonly understood as referring to those periodicities. Here lies precisely their main appeal for macroeconomists: they make the notion of business cycle

[22]Chalmovianský and Němec (2022) contain a somewhat similar study considering a larger set of alternative estimation methods; however, readers must be aware of some typographical errors.

operational.

Filters aiming to isolate the trend component are low-pass filters: they retain only the long-term, slow moving components of the data, i.e., they pass through only those components corresponding to low frequencies, and remove those at medium and high frequencies. A well known result is that the weights or coefficients of the ideal symmetric low pass filter are

$$b_0 = \omega_1/\pi, \quad \text{and} \quad b_j = \sin(j\,\omega_1)/j\pi,\ j = \pm 1, \pm 2, ...$$

where ω_1 denotes the upper frequency of the band. But since this filter is an infinite order moving average, it is unfeasible. A feasible filter is an approximation to this ideal filter and it is a finite moving average, $a_J(L) = \sum_{j=-J}^{J} a_j L^j$.

This approximation filter can be determined so that a certain distance (the integrated square error) between its frequency response function and that of the ideal filter is minimized. As noted by BK (p. 577), given a maximum lag length, J, an indeed "*remarkable result*" for this minimization problem is provided by simply truncating the weights of the ideal filter at lag J, i.e., the optimal approximating filter solution has coefficients

$$a_j = b_j, \text{ for } j = 0, 1, \ldots, J, \text{ and } a_j = 0, \text{ for } j \geq J + 1.$$

On the other hand, to stationarize an integrated time series, the sum of the weights of the filter must sum to zero, so that the frequency response function at frequency zero is zero. This implies that the filter may be factored as

$$\begin{aligned} a_J(L) &= -(1-L)(1-L^{-1})\Psi_{J-1}(L) \\ &= L^{-1}(1-L)^2\Psi_{J-1}(L), \end{aligned}$$

where $\Psi_{J-1}(L)$ is a lag polynomial of order $J-1$. Therefore, as the filter contains two differencing operators, it removes both linear and quadratic deterministic trends and up to two unit roots.

Denoting with $\underline{\omega}$ and $\overline{\omega}$ the lower and upper frequencies of the BP fil-

ter, BK note that the ideal BP filter is constructed from the two low-pass filters with those frequencies as their cutoff frequencies. Their respective weights are represented with $\underline{b}_j$ and $\overline{b}_j$, and the weights of the BP filter are simply $\underline{b}_j - \overline{b}_j$.

However, these weights need to be adjusted to comply with the stationarity condition stated above. BK show that these adjusted optimal approximate coefficients must have the form

$$a_j = b_j + \theta,$$

where (b_j denote the weights of the ideal low-pass filter and) θ is a constant that depends on the selected maximum lag length, J, which is given by

$$\theta = \frac{1 - \sum_{j=-J}^{J} b_j}{2J + 1}.$$

The weights of the approximate BP filter must be similarly adjusted, and hence they are given by

$$(\overline{b}_j - \underline{b}_j) + (\overline{\theta} - \underline{\theta}),$$

where $\overline{\theta}$ and $\underline{\theta}$ are the adjustment coefficients of the upper- and lower-cutoff filters, respectively. BP further suggest the notation $BP_J(p, q)$ to denote the approximation to the ideal filter that passes cycles with lengths between p and q periods (usually 6 and 32, for quarterly data).

Finally, an issue still remains: the selection of J. A tradeoff must be balanced: J must be large enough to provide a good approximation but the larger it is the shorter will be the filtered series (since J observations are lost at the beginning and end of the sample). After substantial experimentation, BK recommend using $J = 12$, as lower values incur in serious problems of leakage and compression[23]. Hence, BK recommendation is

[23]Leakage refers to the problem that the filter passes through frequencies that it was built to suppress, adding them to those that it must retain. Compression means that the filter has frequency response less than unity in the range immediately above the cutoff frequency.

to use a $BP_{12}(6, 32)$ filter for business cycle analysis. In table 4 of their article (p. 591) they present the weights for this filter, as well as those for $BP_{12}(2, 32)$ and $BP_3(2, 8)$ filters, this one for annual data.

Similarly to the Schüler *vs.* Hamilton case, Murray (2003) criticizes the BK filter (and BP filters in general) adopting a definition of business cycles grounded on UC models, as stationary deviations from a stochastic trend, not as the stationary component that remains unaltered when the series is subject to an ideal BP filter. Murray (2003) departs from a simple UC model, with trend given by

$$\eta_t = \mu + \tau_{t-1} + \eta_t, \;\; \eta_t \text{ stationary},$$

and shows that the BK filter stationarizes the series removing the unit root but allowing its first difference, η_t, to pass through. In many cases, this implies that the BK filter will tend to exaggerate the importance of transitory dynamics in estimated business cycles.

Canova (2007, p. 96) reports results of a simulation study where the DGP is a simple random walk ($y_t = y_{t-1} + \epsilon_t, \epsilon_t \sim iid(0, \sigma^2)$) where the BK filter induces spurious periodicity in the filtered series because it tends to be serially correlated. Hence, in this case the BK filter performs worse than the simple differencing filter, which renders a serially uncorrelated filtered series (ϵ_t).

Christiano and Fitzgerald (2003) propose also a filter that approximates the ideal BP filter but employ a rather different approach. Instead of truncating the coefficients of the ideal filter, they minimize a distance that is weighted by the estimated spectral density of the series. Their approximate filter is asymmetric, nonstationary (as its weights vary with the observations) and depends on the time series properties of each series. As they state, their method "*is not for everybody*", meaning that when the focus is the usual statistical analysis of business cycles for quarterly data they consider that the "*HP filter appears to do just fine*" (p. 463).

However, they also propose a simpler filter, derived for the case where the data are assumed to be generated by a pure random walk, which they argue performs reasonably well for many macroeconomic time series. In

this case the filtered series is given by

$$\begin{aligned}\widehat{c}_t = \; & B_0 y_t + B_1 y_{t+1} + \ldots + B_{T-1-t} y_{T-1} + \tilde{B}_{T-t} y_T + \\ & B_1 y_{t-1} + \ldots + B_{t-2} y_2 + \tilde{B}_{t-1} y_1\end{aligned}$$

for $t = 2, 3, \ldots, T-1$, where

$$Bj = \frac{\sin(j\, b) - \sin(j\, a)}{\pi j},\ j \geq 1,$$

$$B_0 = \frac{b-a}{\pi}, \quad a = \frac{2\pi}{p_u}, \quad b = \frac{2\pi}{p_l},$$

$$\tilde{B}_{T-t} = \sum_{j=T-t}^{T} B_j, \quad \text{and} \quad \tilde{B}_{t-1} = \sum_{j=t-1}^{T} B_j,$$

where p_l and p_u denote the lower and upper periodicities of the band, corresponding to the upper and lower frequencies, respectively (e.g., for quarterly data, $p_l = 6$ and $p_u = 32$). As for the first and last observation of the filtered series, they are given by

$$\widehat{c}_1 = \frac{1}{2} B_0 y_1 + B_1 y_2 + \ldots + B_{T-2} y_{T-1} + B_{T-1} y_T, \ \text{and}$$

$$\widehat{c}_T = \frac{1}{2} B_0 y_T + B_1 y_{T-1} + \ldots + B_{T-2} y_2 + \tilde{B}_{T-1} y_1,$$

respectively. Drake and Mills (2010) apply this filter to the GDP of the Eurozone for the 1980:1 to 2006:1 period, drawing interesting conclusions about the effect of the adoption of the Euro on trend output.

Chapter 5

Testing for the presence of a trend

As in the case of the Loch Ness monster or in the one of UFOs, establishing whether a trend is present in the data cannot rely simply on sightings. So, why has testing for the presence of a trend not been addressed yet? In many instances, it logically appears to be the first step in a coherent strategy for empirical research. That is, this chapter seems to arrive too late. Actually, the absence of a single, rigorous and generally accepted definition of trend justifies the late positioning of this chapter. Obviously, the aim is not to measure the significance of "'*that which trend filters remove*"' (White and Granger, 2011, p. 14). But postponing the presentation of the tests to a stage preceded by that of the models and the decomposition methods is very useful: only after knowing how diverse trends are understood we are ready and minimally equipped to search for them. At the very worst, this ordering may be useful to highlight a shortcoming of this chapter: it can hardly be considered as thorough as it could be.

A wide variety of tests for the most popular form that trends are perceived, the linear trend, are presented: it starts with the standard textbook approach and ends with recent tests that are robust to the order of integration of the data (provided it is either $I(0)$ or $I(1)$). Besides a

brief presentation of two nonparametric tests employed outside economics at the end of the chapter, the rest of the chapter is devoted to testing for nonlinear trends: first a small sample of approaches to their most popular version — the linear segmented trend — followed by a general test for nonlinearity. Although stochastic trends are obviously one of the most important cases of trends, as previously stated, tests to detect them, i.e., unit root and stationarity tests, will not be addressed here. A very useful reference for this topic is Choi (2015).

5.1 Motivations for testing

In macroeconomics there are several motives underlying the interest in testing for the presence of a trend. First, if the trend corresponds to a special, smooth, permanent or slowly evolving long-run component, which is present in some series but not in others, then it is obviously important to know whether the series at hand really contains such a component. From this investigation another may arise that sheds some light on the root causes of such behaviour.

Second, several testing procedures for the presence of a trend require some previous knowledge about the long-run statistical properties of the series, which often turn out to amount to the knowledge about the properties of the errors in the regression of the series on a linear deterministic trend. That is, testing for a trend often requires submitting the series to a previous unit root or stationarity test. But the properties of these tests are also strongly dependent on the deterministic regressors included in the test regression and, in particular, on the presence or absence of a linear trend term. There is thus a serious circularity problem: to know whether a linear trend is present in a series, we often need to know whether it is $I(1)$ or $I(0)$ (in the most common cases), but the validity of the inference about this choice is itself highly dependent on the previous knowledge we have about that presence.

Moreover, as clearly exposited in Perron, Shintani and Yabu (2017), a similar problem often afflicts tests for nonlinearity in the trend:

a) the limiting distribution of test statistics designed to detect non-linearities in the trend usually depends on the order of integration of the series;

b) but testing to determine this order usually also depends on the exact nature of the deterministic trend.

Thus, the relevance of tests for the trend inflates itself.

To break the first type of circularity problems, Ayat and Burridge (2000) propose an algorithm, a sequential procedure for unit root and trend testing, aiming also at the identification of the degree of the trend polynomial. The algorithm is not particularly complex but it obviously cannot eliminate uncertainty completely.

Third, assessing whether a time series is characterized by (linear) trending behaviour is both a general and a specific issue in macroeconomics. In general terms, it matters whether the GDP of a country, or a stock price index, or even real interest rates tend to evolve approximately along a linear trend in the long-run. For instance, although it may appear somewhat surprising, Rogoff, Rossi and Schmelzing (2024) present some evidence supporting a decreasing trend of global long-maturity real interest rates over the very deep long-run, as they analyse the period 1311–2021.

A more specific issue is, for instance, assessing whether there is a catching-up process in the difference between (logged) income per capita of two countries, which usually requires that the gap between those two (logged) incomes linearly decreases with time (see, e.g., Bernard and Durlauf, 1996). Another specific example concerns the Prebisch-Singer hypothesis, according to which the prices of primary commodities, sold by developing economies, present a downward long-run trend in relation to those of manufactured goods and services, produced by more advanced economies. Therefore, developing countries could face a deterioration in their terms of trade that prevents their advancement. Still another specific example is provided by Papell and Prodan (2014), who try to establish a close connection between several linear segmented models for

the trend and some specific hypotheses about long-run economic growth (more on this in subsection 5.6).

5.2 The standard textbook test

It should be immediately clear that the test that is focused here is the test for the significance of the parameter β in the equation

$$y_t = \alpha + \beta t + \epsilon_t, \quad \epsilon_t \sim iid(0, \sigma_\epsilon^2), \tag{5.1}$$

which is equation (3.1) with white noise errors and where y_t may denote a logged time series. This is the most common and simple test for the trend, but it concerns also the most limited, naive and often inadequate of its notions. This equation corresponds to the linear trend model of section (3.1), or to the exponential trend model of the same section in case y_t represents a logged time series, but particularizes it to the rather restricted case of white noise errors, a case that was labeled as the purely deterministic linear trend model. To comply with this model, y_t must be a very particular $I(0)$ process, a rather inhibited version of a trend stationary process, with no serial correlation in its innovations and hence, for instance, with no cyclical component.

Testing for the trend consists in testing $H_0 : \beta = 0$ and the alternative in macroeconomics is most often $H_1 : \beta > 0$, but in some cases it is $H_1 : \beta < 0$. It is rarely the case that $H_1 : \beta \neq 0$. Recall also that if, instead of $y_t \sim I(0)$, y_t is integrated of order one, $y_t \sim I(1)$, the error term of the equation is also necessarily $I(1)$, and equation (3.1) is a spurious regression equation, of the Nelson and Kang (1984) variety, which was addressed in section (3.8)[1]. As recently as 2022 this test has been used by Lima, Cribari-Neto and Lima-Junior (2022) in the framework of dynamic quantile regression models to analyse trends in time series of river streamflows.

[1]The problem of spurious inference induced by (strongly) serially correlated errors is not confined to macroeconomics; Rybski and Bunde (2009) report a similar problem in the analysis of temperatures.

The test statistic is the simple t-ratio or t-statistic $\widehat{\beta}/se(\widehat{\beta})$, $\widehat{\beta}$ denoting the OLS estimator of β and $se(\widehat{\beta})$ its usual associated standard error. Recall also that if equation (3.1) is indeed the DGP, $\widehat{\beta}$ is superconsistent, it is $O_p(T^{-3/2})$, and that $T^{3/2}(\widehat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0, \sigma_\epsilon^2/6)$, so that under the null hypothesis the statistic is asymptotically standard normal. Still assuming that the DGP coincides with the model and adding the gaussianity assumption of the error term produces an exact, small sample, $t-$Student distribution under the null.

Removing the assumption that the error term of equation (5.1) is white noise and even allowing it to be $I(1)$, Rivas and Gonzalo (2020, RG) propose a questionable "*practical definition*" of trend (p. 158):

> a characteristic C_t of a functional stochastic process X_t contains a trend if in the [O]LS regression
>
> $$C_t = \alpha + \beta t + u_t, \;\; t = 1, \ldots, T,$$
>
> $\beta = 0$ is rejected.

In this definition, equation (3.1) ceases to be considered the DGP and serves only as the "*linear LS approximation of an unknown trend function* $h(t)$". C_t (and hence u_t) is now even allowed to be $I(1)$.[2] RG are resorting to the asymptotic power properties of the standard test to solve the old and difficult problem of trend definition:

a) when $y_t \sim I(0)$, a trend stationary process, the test asymptotically classifies the series as trending because it is consistent;

b) when y_t is a driftless $I(1)$ process, asymptotically the test tends to produce erroneous inferences, rejecting the null, thereby finding evidence of a trend, which actually exists but it is a stochastic one in this case, not the deterministic linear trend of the alternative hypothesis.

[2] Rivas and Gonzalo (2020) extend their analysis to fractionally integrated, near unit root and local models as well.

Actually, RG are taking advantage of two surprising results concerning b) presented in Durlauf and Phillips (1988):

i) when y_t is a driftless $I(1)$ process the OLS estimator for the slope $(\widehat{\beta})$ is surprisingly consistent (at rate $T^{1/2}$) because it converges to the true value of the parameter, which is zero;[3]

ii) despite this, the test statistic for the significance of β diverges, as the previous Monte Carlo study by Nelson and Kang (1983) had indicated.

Furthermore, when $y_t \sim I(1)$ with drift, as the asymptotic behaviour of the OLS slope coefficient estimator and its t-statistic coincide with those of the driftless case, the trend, which is present in both deterministic and stochastic forms in this case, will be detected in the limit as $T \to \infty$.

It appears that RG confound consistency of the test in the detection of a trend — be it either deterministic or stochastic or both — with the definition of trend. However, it does not appear to be admissible that a definition regarding the presence or the absence of a characteristic might depend entirely on the outcome of a statistical test (however powerful it might be). A finding of evidence can surely depend, but that is not the same as an undisputable, uncertainty-free statement about the true nature of a series. To say that there is evidence on the presence of a trend is not the same as asserting unambiguously that one is present.

Moreover, one can easily conceive a simple counter-example: a time series that contains some trend but whose sample size is insufficient for the test to reject the (false) null hypothesis is incorrectly *defined*, not simply tentatively considered, as non-trending. While the result of the test is admissible, as well as the inference that no evidence of trending behaviour was found, using such evidence to unambiguously assert the true nature of the series is not. Paraphrasing White and Granger (2011) this would amount to declare that "the trend is that thing that emerges when the standard test rejects". Unless the standard test never fails, i.e.,

[3]See also theorem 2 of Canjels and Watson (1997).

its probability of making type I or type II errors equals zero, one cannot rely entirely on a decision of a statistical test to make such unambiguous statements about the presence of a trend. Asserting the true nature of a series is impossible on the basis of the outcome of a statistical test only: it would require the observation of all other possible realizations of the stochastic process that underlies that series. There is thus a logical contradiction between the definition and its object, the time series.

Summing up, there are two problems with RG's approach:

a) make a definition (completely) depend on the outcome of a statistical test;

b) assume that the asymptotic behaviour of the test statistic is sufficient to ensure that one always arrives at the correct decision concerning the DGP.

According to a), the standard t-test would possess the power of deciding what the series really is, not simply the capacity to make a decision, possibly wrong, regarding the hypothesis about the nature of a series. RG further recommend using a robust HAC version of the test statistic, not its standard version. But as will become clear in the next section, this recommendation is far from precluding the approach of being flawed.

5.3 Robust versions of the standard test

The assumption that the error term of equation (5.1) is a white noise process is rather unrealistic for most applications in macroeconomics. As previously mentioned, it amounts to impose the absence of a cyclical component or, equivalently, that this component displays no pattern at all as any serial correlation is prohibited. In this section the error term is allowed to be serially correlated, but no so strongly as to be $I(1)$, and it is now denoted with u_t, that is, the test equation is (3.1), which is reproduced for convenience,

$$y_t = \alpha + \beta t + u_t, \tag{5.2}$$

instead of (3.1), with $u_t \sim I(0)$. At the end of this section, a brief reference about methods that attempt to be robust to strongly serially correlated errors, with $u_t \sim I(1)$, will be made. The topic will be further pursued in the next section.

With u_t usually positively autocorrelated, conventional standard errors tend to overestimate the precision of OLS estimators, inflating the respective t-statistics. Hence, the most common effect is a case of over-rejection of the true null hypothesis and, particularly when testing for the significance of the trend slope, a situation of spurious finding of evidence supporting the presence of the linear trend.

Initially with heteroskedasticity only, and later with both heteroskedasticity and serial correlation allowed, inference methods that attempt to preserve the usual asymptotic properties of inferences associated with OLS in the presence of these problems have become common practice and are popularly known as HAC (autocorrelation and Heteroskedasticity Consistent) methods. This is because they simply aim at the consistent estimation of the (asymptotic) covariance matrix of the OLS coefficient estimator under those circumstances, exempting from the knowledge of the particular form of serial correlation.

As in Canjels and Watson (1997) and Bunzel and Vogelsang (2005), the following rather liberal assumptions about the error term are made:

$$u_t = \rho u_{t-1} + \epsilon_t,\ t = 2, 3, \ldots, T, \quad u_1 = \epsilon_1, \quad |\rho| < 1,$$

$$\epsilon_t = d(L)e_t,\ d(L) = \sum_{i=0}^{\infty} d_i Li, \quad \sum_{i=0}^{\infty} i\,|d_i| < \infty, \quad d(1)^2 > 0,$$

with $\{e_t\}$ a martingale difference sequence (m.d.s.) satisfying $\mathrm{E}(e_t^2|e_{t-1}, e_{t-2}, \ldots) = 1$ and $\sup_t \mathrm{E}(e_t^4) < \infty$.

The only nuisance parameter is then $\sigma^2 = d(1)/(1-\rho)^2$, the long-run error variance, and it is estimated non-parametrically from the OLS residuals, $\widehat{u}_t$, with

$$\widehat{\sigma}_u^2 = \widehat{\gamma}_0 + 2\sum_{j=1}^{T-1} k\left(\frac{j}{M}\right)\widehat{\gamma}_j, \tag{5.3}$$

where $\widehat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T}\widehat{u}_t\widehat{u}_{t-j}$ are the estimates of the autocovariances, and $k(x)$ is a kernel function, M denoting its bandwidth or truncation lag. For consistency, this estimator requires that sample autocovariances for longer lags receive less weight or are simply neglected. In particular, as $T \to \infty$, the conditions $M \to \infty$ but $M/T \to 0$ must hold. Most popular kernel functions simply discard autocovariances with order higher than M, so that the kernel function complies with $k(x) = 0$ for $|x| > 1$.

It is this estimator that enters the estimation formula for the variance of $\widehat{\beta}$, to produce the HAC robust t-statistic:

$$t^{HAC} = \frac{\widehat{\beta}}{\sqrt{\widehat{\sigma}_u^2\left[\sum_{t=1}^{T}(t-\bar{t})^2\right]^{-1}}}, \tag{5.4}$$

where the denominator is simply known as the robust HAC standard error of $\widehat{\beta}$ ($se^{HAC}(\widehat{\beta})$) and which is asymptotically standard normal under the null hypothesis.

Despite important research efforts in this area, HAC methods have managed to achieve some attenuation in the tendency for over-rejection but they have failed to attain their complete elimination in empirical work, particularly when ρ, the largest autoregressive root, is positive and large (but less than one, in the case that I am considering). Bunzel and Vogelsang (2005) provide two main reasons for this failure:

a) although the asymptotic distribution of the tests does not depend on specific choices of bandwidth and kernel, finite sample distributions do, and this dependence is high;

b) the possibility of strong positive serial correlation must be seriously entertained more often than previously thought (and even the unit root case must be taken into consideration).

Kiefer and Vogelsang (2005, KV) designed a framework to overcome the deficiencies of the already traditional HAC approach and, in particular, to improve its performance in controlling the size of the tests. Phillips, Wang and Zhang (2019, PWZ) label this approach as HAR

(Heteroskedasticity and Autocorrelation Robust) and stress that it sacrifices consistent estimation of the covariance matrix of the OLS estimator in the interest of achieving an improved performance in testing.

The specification of the bandwidth parameter M as a fixed proportion of the sample size T, i.e., $M = bT,\ b \in (0,\ 1]$, lies at the center of KV's approach. They label it as "*fixed-b*" asymptotic framework, to contrast it with the traditional "*small-b*" asymptotics, because it holds b fixed as a constant, while the standard approach had to consider that $b \to 0$ as T increases (to comply with condition that $M/T \to 0$ as $T \to \infty$). Although the statistics are computed the same way as in the standard HAC approach, the asymptotic approximation to the sampling distribution is changed to reflect the choices of kernel and bandwidth. In particular, the standard normal is considered an inaccurate approximation and it is no longer used to avoid size distortions.

Bunzel and Vogelsang (2005) adapt KV's approach to the tests for the linear trend and recommend using the Daniell kernell, which is given by

$$k(x) = \frac{\sin(\pi\, x)}{\pi\, x},$$

to maximize power[4]. Furthermore, they also recommend using the following data dependent procedure to select the bandwidth. Start by estimating ρ with the OLS regression of $\widehat{u}_t$ on $\widehat{u}_{t-1}$, so that

$$\widehat{\rho} = \frac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^{T} \widehat{u}_{t-1}}.$$

Then compute $\widehat{\overline{\rho}} = T(1 - \widehat{\rho})$, which I denote simply with ρ^* to lighten the notation. With this, estimate the optimal kernel with

$$\begin{aligned} \widehat{b}_{opt} = \quad & .02 + .02 \cdot 1(\rho^* \leq 21) + .02 \cdot 1(\rho^* \leq 20) + .04 \cdot 1(\rho^* \leq 19) + \\ & .02 \cdot 1(\rho^* \leq 18) + .12 \cdot 1(\rho^* \leq 17) + .1 \cdot 1(\rho^* \leq 14) + .1 \cdot 1(\rho^* \leq 12) + \\ & .06 \cdot 1(\rho^* \leq 11) + .12 \cdot 1(\rho^* \leq 10) + .02 \cdot 1(\rho^* \leq 7) + .2 \cdot 1(\rho^* \leq 4), \end{aligned}$$

[4]Recall that the Bartlett and the quadratic spectral kernels are the most popular in the traditional HAC approach of Newey and West.

Bandwidths and 5% critical values for 3 cases

case	$\widehat{\rho}$	$\widehat{b}_{opt}$	$\widehat{M}$	$cv(\widehat{b}_{opt})_{0.95}$
I	0.30	0.02	2	1.71
II	0.80	0.06	6	1.87
III	0.85	0.24	24	3.04

which maximizes a measure of average power over the space of the alternative hypothesis. This allows obtaining the bandwidth that should be used in (5.3) with:

$$\widehat{M} = \max(\widehat{b}_{opt}T,\, 2).$$

Finally, the asymptotic critical value ($cv(b)$) is calculated with a polynomial function

$$cv(\widehat{b}_{opt}) = \theta_0 + \theta_1\widehat{b}_{opt} + \theta_2\widehat{b}_{opt}^2 + \theta_3\widehat{b}_{opt}^3 + \theta_4\widehat{b}_{opt}^4 + \theta_5\widehat{b}_{opt}^5,$$

that Bunzel and Vogelsang estimate with OLS. For instance, for the 95% percentile, the following estimates are taken from their table 2: $\widehat{\theta}_0 = 1.6383, \widehat{\theta}_1 = 3.5083, \widehat{\theta}_2 = 3.1079, \widehat{\theta}_3 = 31.3777, \widehat{\theta}_4 = -16.0674, \widehat{\theta}_5 = 3.6881$.

A simple example illustrates the importance of $\widehat{\rho}$. Suppose that the sample size is $T = 100$ and that the purpose is to do a 5% one-sided test. Further consider three cases: in case I $\widehat{\rho} = 0.30$, in case II $\widehat{\rho} = 0.80$ and in case III $\widehat{\rho} = 0.85$. The most relevant information is summarized in the simple table that follows, which shows that both bandwidth parameters and 5% critical values are highly sensitive to small changes in $\widehat{\rho}$ when its magnitude is already large.

Actually, Bunzel and Vogelsang's approach is more general than this shortened presentation here and encompasses the unit root case too. However, in this case the statistic must be scaled with a factor that depends on a unit root test statistic. When only the stationary case is considered this factor is simply unity, and the statistic reduces to (5.4). Approaches that are designed to be robust to both $I(0)$ and $I(1)$ error terms will be addressed after the next section. On the other hand, PWZ's

approach is meant to address the pure spurious regression case only, i.e., the case of the regressions of stochastic trends on a (deterministic) linear trend.

As regards Rivas and Gonzalo's recommendation to use the standard HAC approach to assist in their definition, it is clear that while it possibly improves it in the $I(0)$ case, it further complicates the case against them in the $I(1)$ case. Indeed, on the one hand, trendless $I(0)$ series will be less often (incorrectly) defined as trending, as size distortions will be alleviated, but on the other hand, in the driftless stochastic trend case, the same effect will allow the (changing) trend to pass undetected more often than with standard OLS inference, thereby implying that more series will be incorrectly defined as non-trending. In this case, as the HAR approach is more effective in controlling size, the problem is further worsened.

5.4 The $I(1)$ case

Assume now that it is known that the series is $I(1)$, or at least that there is some evidence that it can be considered as $I(1)$. Recalling again equation (3.4) from section 3.8, the problem then amounts to know whether there is a (non-zero) drift in the series. In this case, although consistent, the OLS estimator of equation (5.2) ceases to be asymptotically efficient, as it was in the $I(0)$ case. The simplest solution is to difference the series, so that the sample mean of Δy_t, which is the OLS estimate of the coefficient of the regression of Δy_t on a constant only, is asymptotically efficient[5]. The presence of a linear trend can then be assessed with the t-statistic of that coefficient (possibly with a serial correlation robust version). The testing regression is then simply,

$$\Delta y_t = \beta + v_t,$$

[5]This is due to the equivalence of the OLS estimator with the GLS estimator, which in this case, since $\rho = 1$, implies differencing y_t.

and the robust HAC t-statistic is

$$t_1^{HAC} = \frac{\tilde{\beta}}{\sqrt{\tilde{\sigma}_v^2/(T-1)}}, \tag{5.5}$$

where $\tilde{\beta}$ denotes the OLS estimator of β in the previous equation and $\tilde{\sigma}_v^2$ represents the estimate of the long-run variance of its errors, v_t, built as in (5.3) but with the corresponding residuals, $\tilde{v}_t = \Delta y_t - \tilde{\beta}$.

There is however an alternative test that continues to rely on equation (5.2), which is due to Ventosa-Santaulária and Gómez-Zaldívar (2010, V-SG-Z), and which uses the R^2 of that regression, since they show that:

a) while when the null is true ($\beta = 0$) it converges to a non-degenerate and non-standard distribution that is always strictly less than one,

b) under the alternative ($\beta \neq 0$), it converges in probability to one ($R^2 \xrightarrow{p} 1$).

Therefore, a critical or rejection region can be defined for large values of R^2. V-SG-Z obtain the asymptotic critical values by means of Monte Carlo simulation: for instance, for 10% and 5% sizes they are, respectively, 0.84 and 0.89.

V-SG-Z also address the case of a unit root process affected by one structural break in the drift parameter, implying a change in the slope of the linear trend at the same date. Continuing to base inference in equation (5.2) does not allow (asymptotically) valid inferences anymore, as the R^2 no longer converges to unity under the alternative hypothesis, and the test loses all its power. They propose a relatively simple procedure that is robust to these shifts and that allows estimating the date of the break and making asymptotically valid inferences about the presence of the trend. The distribution of the test statistic (R^2) under the null becomes dependent on the location of the break in the sample, and V-SG-Z tabulate the most common critical values for multiples of 0.05 for the fraction representing that location.

5.5 Robustness to the order of integration

As hopefully is more clear now, testing for a linear trend faces a serious circularity problem that can be summarized briefly as follows:

a) while such test requires prior knowledge about the long-run properties of the regression error term in equation (5.2) because different orders of integration imply different asymptotic distributions for the test statistic under the null hypothesis,

b) to preserve the usual properties of any unit root (or stationarity) test one needs to know in advance whether a linear trend term must be considered.

However, starting with the pioneering work of Vogelsang (1998) and Bunzel and Vogelsang (2005, BV05), several test statistics have been proposed that do not require *a priori* such knowledge, breaking the circularity just mentioned. BV05 recommend the test they name as $Dan-J$, which is a scaled version of test statistic (5.4). I will not further address it here because the scale factor depends on a somewhat uncommon unit root test statistic, J, proposed by J. Y. Park and by J. Y. Park and I. Choi[6]. Moreover, its small sample performance appears to be dominated by the tests summarized below: the z_λ test of Harvey, Leybourne and Taylor (2007, HLT07), the t_β^{RQF} test of Perron and Yabu (PY07), and the TR tests test of Elliot (2020). Astill, Harvey, Leybourne and Taylor (2014) propose two test procedures that modify the tests of Vogelsang (1998) and Bunzel and Vogelsang (2005) in a way that is similar to the HLT07 test but, rather than switching between two test statistics as will be seen below, they switch between the two critical values that are appropriate for the $I(1)$ and $I(0)$ cases, according to the value of the J test statistic.

[6]Park, J. Y. (1990), Testing for unit roots and decomposition by variable addition, *in Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots*, eds. T. Fomby and F. Rhodes, London: Jai Press, pp. 107–134 and Park, J. Y. and Choi, I. (1988), A new approach to testing for a unit toot, *Working Paper 88-23*, Center for Analytic Economics, Cornell University.

The z_λ test

Besides asymptotically valid regardless of the order of integration of u_t, the test statistic proposed in HLT07 presents the additional nice property that its asymptotic null distribution is standard normal because it is a weighted average of two standard normal statistics. Building on the work of BV05, HLT07 propose an ingenious test that mixes the two previous t-statistics, the one that is appropriate when $u_t \sim I(0)$ and the other that is appropriate when it is $I(1)$. The mixing is made with an auxiliary test statistic which consistently estimates the order of integration, $d \in \{0, 1\}$, and the mixing weights are designed to switch weight between those two HAC trend statistics.

Adopting a simple change in notation, denote the two previous HAC t-statistics of (5.4) and (5.5) with z_0 and z_1, respectively. Further, denote with U a unit root test statistic for testing the null that $\rho = 1$ (i.e., that $u_t \sim I(1)$), against the alternative that $\rho < 1$(i.e., that $u_t \sim I(0)$) and with S a stationarity test statistic reverting the roles of the hypotheses, i.e., the $I(0)$ null that $\rho < 1$ against the $I(1)$ alternative $\rho = 1$. Both statistics are computed with the detrended residuals of (5.2), $\widehat{u}_t$.

Considering a function $\lambda(U, S)$ on [0, 1], such that

- $\lambda(U, S) \xrightarrow{p} 0$ when $u_t \sim I(0)$, and
- $\lambda(U, S) \xrightarrow{p} 1$ when $u_t \sim I(1)$,

the statistic of HLT07 is the following weighted average

$$z_\lambda = [1 - \lambda(U, S)]z_0 + \lambda(U, S)z_1.$$

Therefore, provided U, S and $\lambda(U, S)$ satisfy a few unrestrictive conditions:

- when $u_t \sim I(0) \Longrightarrow z_\lambda \xrightarrow{p} z_0$, and
- when $u_t \sim I(1) \Longrightarrow z_\lambda \xrightarrow{p} z_1$.

Hence, at least asymptotically, the correct test will be selected in both situations. Moreover, the limiting null distribution of z_λ is always standard normal. As a by-product, HLT07 derive also an estimator of β that is consistent and asymptotically efficient regardless of whether u_t is $I(0)$ or $I(1)$. They derive also a tweaked version of the test statistic when there is a concern that the series might correspond to a near-unit root or near-$I(1)$ process, with $\rho = \rho_T - c/T$, $0 \leq c < \infty$. In this case they recommend using instead

$$z_\lambda^{m\delta} = [1 - \lambda(U,S)]z_0 + \lambda(U,S)\, z_1^{m\delta},$$

with $z_1^{m\delta} = \gamma_{\xi,\delta} R_\delta z_1$, $R_\delta = \left(\frac{\tilde{\sigma}_v^2}{T^{-1}\tilde{\sigma}_u^2}\right)^\delta$, where $\tilde{\sigma}_u^2 = (T-2)^{-1}\sum_{t=1}^T \widehat{u}_t^2$, $\delta = 1$ or 2, and the values for $\gamma_{\xi,1}$ and $\gamma_{\xi,2}$ are obtained from simulation and depend on the significance level; HLT07 present a table in p. 1312 for several significance levels.

As regards the statistics U and S and the function $\lambda(U,S)$, HLT07 recommend, respectively:

a) although standard unit root test statistics satisfy the conditions required, the DF-GLS$^\tau$ statistic of Elliot, Rothenberg and Stock (1996) for U;

b) for S, the standard KPSS (Kwiatkowski, Phillips, Schmidt and Shin, 1992) test denoted with $\widehat{\eta}_\tau$;

c) and based on some numerical experimentation,

$$\lambda(U,S) = \exp\left[-0.00025\left(\frac{\text{DF}-\text{GLS}^\tau}{\widehat{\eta}_\tau}\right)^2\right].$$

In terms of finite sample properties, according to HLT07, although the z_λ test presents some size distortion and is a bit worse than the z_λ^{m1} and z_λ^{m2} tests for pure unit root processes, its size-adjusted power behaviour is almost always superior to both these tests and to the $Dan - J$ test of BV05. It is therefore the most recommended for empirical work.

Yang and Wang (2017) propose a modified z_λ test statistic which is robust to time varying variance. As they show that this problem does not change the limiting null distribution of z_1, the modification needs to be concerned with z_0 only. A different extension is provided in Harvey, Leybourne and Taylor (2011), who propose a test for a quadratic (deterministic) trend robust to the order of integration as a by-product of their unit root test that allows a local quadratic trend.

The t_β^{RQF} test

Perron and Yabu (2009, PY09) also proposed a test for the linear trend when it is not known a priori whether the series is $I(0)$ or $I(1)$. Their approach is much different from those of BV05 and HLT07 and places a greater emphasis in the efficient estimation of β and ρ. It relies on a feasible GLS (FGLS) procedure, employing an estimate of the autoregressive coefficient which is truncated to unity when the usual estimate is in a certain neighborhood of that value. Using this estimate to build the FGLS transformed regression "*bridges the gap*" between the asymptotic distributions of the test statistic for the $I(0)$ and the $I(1)$ cases and the limiting null distribution is standard normal in both cases.

Although the assumptions about u_t are the same as in Canjels and Watson (1997) and BV05, PY09 prefer relying on its autoregressive representation

$$A(L)u_t = e_t, \quad \text{with} \quad A(L) = 1 - \sum_{i=1}^{\infty} a_i L^i,$$

so that

$$u_t = \rho u_{t-1} + A^*(L)\Delta u_{t-1} + e_t,$$

where $A^*(L) = \sum_{i=0}^{\infty} a_i^* L^i$, with $a_i^* = -\sum_{j=i+1}^{\infty} a_j$, and ρ now representing the sum of the autoregressive coefficients. Therefore, to estimate it, rather than relying on an autoregression of order one, PY09 employ

instead

$$\widehat{u}_t = \rho \widehat{u}_{t-1} + \sum_{i=1}^{k} \phi_i \Delta \widehat{u}_{t-1} + e_{tk}, \tag{5.6}$$

with $\widehat{u}_t$ still denoting the OLS residuals from equation (5.2), which is a truncated autoregresssion of order k selected with an information criterion. PY09 recommend using the modified AIC (MAIC) of Ng and Perron (2001), with $k_{MAX} = 12\left(\frac{T}{100}\right)^{1/4}$.

The t_β^{RQF} statistic of PY09, where RQF means "robust quasi feasible" (GLS), is an autocorrelation corrected t-statistic on β, obtained estimating the following quasi-GLS regression with OLS:

$$y_t - \tilde{\rho}_{MS}\, y_{t-1} = (1-\tilde{\rho}_{MS})\alpha + \beta\,[t - \tilde{\rho}_{MS}(t-1)] + (u_t - \tilde{\rho}_{MS}\, u_{t-1}), \quad t = 2, \ldots, T,$$

$$\text{and} \quad y_1 = \beta + u_1, \tag{5.7}$$

where $\tilde{\rho}_{MS}$ is obtained with the truncation

$$\tilde{\rho}_{MS} = \begin{cases} 1, & \text{if } |\tilde{\rho}_{WS} - 1| < T^{-1/2}, \\ \tilde{\rho}_{WS}, & \text{otherwise,} \end{cases}$$

where $\tilde{\rho}_{WS}$ is a weighted symmetric least squares (WSLS) estimate of ρ in equation (5.6), with one of two truncations explained in detail in p. 59 of PY07, $\tilde{\rho}_{MU}$ and $\tilde{\rho}_{UB}$.

The statistic can then be represented as

$$t_\beta^{RQF} = \frac{\tilde{\beta}}{\sqrt{\hat{h}(X'X)_{22}^{-1}}},$$

where $\tilde{\beta}$ is the estimate obtained from the quasi-GLS regression and $(X'X)_{22}^{-1}$ is the second diagonal element of the $(X'X)^{-1}$ matrix, with $X = [x_1, \ldots, x_T]'$, $x_t = [(1 - \tilde{\rho}_{MS}), t - \tilde{\rho}_{MS}(t-1)]$ for $t = 2, \ldots, T$, and $x_1' = (1, 1)$. $\hat{h}$ is a consistent estimator of the long-run variance of $\{u_t - \rho u_{t-1}\}$ (denoted with ϵ_t in section 5.3), for which, instead of a non-parametric estimator, PY09 suggest using an autoregressive spectral

density estimator (see their p. 61 for details).

While asymptotically this test is equivalent to the z_λ test of HLT07 because their asymptotic power functions are the same, in finite samples their behaviour can be much different. The simulation results in PY09 allow them to claim that their test, mainly the $t_\beta^{RQF}(MU)$ version, is preferable to the $Dan - J$ and z_λ tests, but the dominance is far from being uniform, both in terms of size and size-adjusted power properties.

Yang and Wang (2017) further extend the test to the case where, besides possibly strongly serially correlated, the error term in (5.2) is also affected by time-varying volatility.

A brief reference to the TR tests

In a similar vein to the two previous testing procedures and particularly to HLT07's, more recently Elliot (2020) proposed two new tests, based on two statistics that he names as TR_0 and TR_1. His main innovation lies in the component test statistic which is appropriate for the $I(1)$ and near-$I(1)$ cases, which he labels as WLR because it is a ratio of weighted likelihoods (under both hypotheses). Furthermore, to build the likelihood functions he considers two different assumptions regarding the initial value of the stochastic component: the more common that assumes that it is asymptotically irrelevant and the assumption that it is drawn from its unconditional distribution. A major problem with these procedures is that the calculation of the WLR statistics appears too much involved to be frequently adopted in empirical research.

5.6 Tests for linear segmented trends

Tests for linear segmented trends gained some popularity after the work of Perron (1989) on unit root tests which are robust to breaks but sometimes breaks are imposed on linear trends with only graphical evidence supporting them (which was the case with Perron's article). This contrasts with a growing and ever more refined theoretical approach to the

subject, which intersects the more general literature on the topic of structural breaks in time series. I have selected only a few approaches from this voluminous library: the one by Papell and Prodan (2014) and the tests by Hsu and Kuan (2001) and Wu (2016).

Characterizing long-run GDP growth

Papell and Prodan (2014, PP14) strongly defy the stochastic trend hypothesis concerning long-run economic growth. They propose a procedure to analyse the stability of long-run per capita GDP growth consisting of several tests for linear restrictions in the context of a regime-wise trend stationary model. They consider a sample of annual data from 1870 to 2008, taken from Maddison's database, for 19 OECD and 7 Asian countries.

PP14 seem to try to establish a relation between statistical and economic models but their work in this respect is somewhat unsuccessful. Rather, they obtain some success in the presentation of some stylized facts concerning long-run growth for many advanced economies and for some Asian countries. Statistically, after discarding the stochastic trend model for most of the countries on the basis of several unit root tests[7] and after estimating the number and the dates of the breaks in a linear trending model using Bai's (1999) procedure, the most general model they consider is

$$y_t = \alpha + \beta\, t + \gamma_1 DU_{1t} + \theta_1 DT_{1t} + \theta_2 DT_{2t} + \theta_3 DT_{3t} + \sum_{i=1}^{k} \rho_i y_{t-i} + u_t, \quad (5.8)$$

where y_t denotes logged per capita GDP, DU_{1t} is a step dummy variable defined by

$$DU_{1t} = \begin{cases} 1, & \text{if } \ t > Tb_1, \\ 0, & \text{otherwise,} \end{cases}$$

[7]Most of them allowing for breaks in the sample. PP14 do not control the size of their joint procedure, searching for a rejection of the unit root null with four distinct tests.

$DT_{it} = t - Tb_i = 1$ if $t > Tb_i$ and 0 otherwise, $i = 1, 2, 3$, are changing slope dummies, k is selected with the Schwarz information criterion (SIC or BIC), and u_t is (presumed to be) a well behaved error term, serially uncorrelated and homoskedastic.

For the countries that exhibit at least one rejection of the unit root tests, the following models are considered:

a) the linear trend model;

b) the level shift model;

c) the growth shift model;

d) the transition dynamics hypothesis.

As all the tests corresponding to these models are performed on a $I(0)$ context, i.e., for trend stationary series because the $I(1)$ environment was previously waived, resorting to standard inference methods, and particularly to the usual F-tests is a valid procedure[8].

Since PP14 find evidence of at least one break for all the retained countries, model a) is not really a pure, no-break linear trend model with a long-run average growth rate approximately constant throughout the sample. Instead, a "disturbed", non-aligned intermediate segment is allowed, the trend following the last break appearing as "*a linear projection of the trend preceding the first break*" (PP14, p. 469). Therefore, this model cannot be considered for the cases where only one break was found. The two sets of linear restrictions that allow obtaining this model as a particular case of equation (5.8) are

$$\sum_{l=1}^{i} \theta_l = 0 \quad \text{and} \quad \gamma_1 + \theta_1(Tb_i - Tb_1) = 0, \text{ for } i = 2, 3.$$

Two of the countries less affected by the world wars —- Canada and the U.S. — appear to be well characterized by this model, the first break

[8]Five countries were left out of this analysis because all unit root tests failed to reject the unit root null hypothesis.

located around the Great Depression and the linear trend resuming its previous trajectory at the beginning of the 1940s.

The level shift model leaves parameter γ_1 unrestricted and imposes $\sum_{l=1}^{i} \theta_l = 0$, $i = 2, 3$, that is, it allows a permanent change in the level of per capita income but imposes that changes in growth rates may only be temporary. Contrary to the previous one, this model may hold for the cases where only one break was previously identified; for this case, the only restriction is $\theta_1 = 0$. PP14 find that this model is appropriate for 6 countries.

The growth shift hypothesis requires the existence of permanent changes in long-run growth rates and therefore leaves the θ_i parameters unrestricted and imposes only that $\gamma_1 = 0$. Based on their preliminary break analysis, PP14 do not consider this to be a real possibility. Instead, they consider that a large group of countries belong to this class but allow that all the parameters of equation (5.8) are left unrestricted.

Finally, PP14 allow the possibility that after the last break there was not enough time until the end of the sample for countries to settle down to a new steady-state growth rate. This is the the transition dynamics hypothesis. They extend the sample with more 50 years of data generated artificially with autoregressive models estimated over samples starting in 1973 and consider the countries for which there was not evidence favoring the linear trend or the growth shift hypothesis and test again whether the restriction $\sum_{l=1}^{i} \theta_l = 0$ could be imposed. Only one more country now failed to reject these restrictions and therefore, in general, the results do not appear to be driven by transition dynamics.

Tests for partial parameter stability

Still in the context of the trend stationary, linear trend model of equation (5.2), sometimes interest centers on tests for partial parameter stability. When a structural break is identified, is it a crash-type of break (change in the intercept only), or a changing growth one (change in the slope only)? Obviously, in the case of a mixed break both parameters may change. Hsu and Kuan (2001, HC01) address this problem: testing each one of

the coefficients of the linear trend while keeping the other unchanged (and untested).

Allowing the error term to be serially correlated, with a similar framework to the one of section 5.3, HK01 firstly demonstrate that Wald-type tests are not appropriate for the purpose of partial parameter stability testing because they suffer from a serious size problem when the coefficient being tested is stable but the other coefficient changes. This is because although they are partial, these tests depend on all the coefficients of the model but the critical values are derived assuming the constancy of all the coefficients.

Instead, HK01 propose the following two-step procedure for the case where the date of the break is not known a priori. In the first step the constancy of the coefficients is tested jointly and the break date is estimated by least squares when the null hypothesis is rejected. Otherwise the procedure stops.

In practice, since minimizing the sum of squared residuals is equivalent to maximizing the LR statistic of joint parameter constancy, the least squares estimator of the change point observation, denoted with k as in HK01, may be obtained from the test statistic for the joint (J) test

$$LR_T(k, J) = (RSS_0(k, J) - RSS_1(k, J))/s_T^2(k, J),$$

where $RSS_0(k, J)$ is the constrained residual sum of squares (RSS) corresponding to the null hypothesis of parameter constancy, $RSS_1(k, J)$ is the unconstrained RSS allowing both parameters to change after observation k and $s_T^2(k, J) = RSS_1(k, J)/T$ is the variance estimator of the unconstrained regression, with

$$\hat{k} = \arg \max_{k \in [\underline{k}, \overline{k}]} LR_T(k, J),$$

where $\underline{k} = [T\underline{\tau}]$ and $\overline{k} = [T\overline{k}]$, $\underline{\tau}$ and $\overline{\tau}$ denoting the lower and upper limit for the trimmed sample interval adopted for the fraction of break dates, respectively; as is well known, usually $\underline{\tau}$ is a value close to zero, such as 0.05 or 0.10. Hence, the likelihood ratio test statistic plays two distinct

roles: testing joint constancy of the two coefficients and estimating the break date.

In the second step the constancy of each coefficient may be tested with

$$LT_T(\hat{k}, i) = (RSS_0(\hat{k}, i) - RSS_1(\hat{k}, i))/s_T^2(\hat{k}, i), \quad i = 1, 2,$$

with 1 denoting the intercept and 2 the slope coefficient. Under the null hypothesis

$$LT_T(\hat{k}, i) \xrightarrow{d} \chi^2_{(1)}, \quad i = 1, 2.$$

To control the overall size of these tests HK01 recommend performing each of them with half the usual size (say, 0.025), so that the desired overall size of the procedure is not exceeded

In one of the empirical illustrations they present, HK01 compare the results of their method over the Nelson-Plosser data with those of Perron (1989) and find that they agree for only five (of the 13) series. Different conclusions for the break types and dates are the dominant feature.

A test with monotonic power

Tests for a changing linear trend may possess the undesirable feature of non-monotonic power: as the change grows the power of the test decreases instead of increasing, as would be expected. The test proposed by Wu (2016) has monotonic power and allows for weak serial dependence, conditional heteroskedasticity and time-varying unconditional variance, and although it is derived mainly against smoothly changing alternatives, it has power to detect both abrupt and smooth structural changes. However, it does not provide any estimate of the breakpoint when the null of stability is rejected.

The basic idea is that under the null hypothesis the OLS residuals of equation (5.2), $\hat{u}_t$, are good estimates of a zero mean process. But under the alternative hypothesis, the instability of the coefficients will likely

induce that $\hat{u}_t$ will deviate from zero. Hence, a first statistic is

$$\tilde{\lambda}_T = \frac{1}{T^2 h_1} \sum_{t=1}^{T} \sum_{s \notin B(t)} k\left(\frac{t-s}{Th_1}\right) \hat{u}_t \hat{u}_s,$$

where $B(t) = [t - Th_2, t + Th_2]$ is a growing neighborhood around t, and h_1 and h_2 are bandwidth parameters such that as T grows, $h_1 \to 0, h_2 \to 0, h_2/h_1 \to 0$, and $k(\cdot)$ is a symmetric and bounded kernel function with support on $[-1, 1]$.

When the null hypothesis is true $\tilde{\lambda}_T$ should be close to zero, and when there is instability it will tend to be positive. However, it still needs to be standardized. Wu proposes the test statistic

$$\tilde{U}_T = Th_1^{1/2} \tilde{\lambda}_T / \tilde{\Sigma}^{1/2},$$

where the estimate of the variance is

$$\tilde{\Sigma} = \frac{2}{T^2 h_1} \sum_{t=1}^{T} \sum_{s \notin B(t)} k^2\left(\frac{t-s}{Th_1}\right) \hat{u}_t^2 \hat{u}_2^2,$$

which is asymptotically standard normal under the null of stability.

Based in some Monte Carlo evidence, Wu (2016) sets $h_2 = h_1 T^{-1/5}$ and recommends using $h_1 = 0.75T^{-1/5}$. This version of the test often exhibits better power performance in the simulations than such popular tests as CUSUM and *sup*Wald (or *max*Chow) tests.

5.7 Testing for a nonlinear trend

Several nonlinear models were presented in chapter 3. The purpose here is not to prescribe a test for each one of them but rather to deal with a general procedure to detect any nonlinear feature in the deterministic trend. The segmented linear model, containing at least one break or change in intercept and/or slope figures prominently among the alternatives but, of course, it is not the only alternative hypothesis. For instance, such a

procedure must be useful to detect smooth transition type behaviour in the trend. The idea is to search for any type of nonlinearity without any prior knowledge about its true functional form.

An approximation that has been servicing this purpose because it is known as capturing well many types of nonlinearities is the Fourier series expansion, where the approximating model is given by

$$y_t = \alpha + \beta t + \sum_{j=1}^{n} \gamma_{1j} \sin\left(\frac{2\pi k_j t}{T}\right) + \sum_{j=1}^{n} \gamma_{2j} \cos\left(\frac{2\pi k_j t}{T}\right) + u_t, \quad (5.9)$$

where the k_j coefficients are non-negative integers for $j = 1, \ldots, n$, which denote the particular frequency, and n represents the total number of frequencies. In the approach of Perron, Shintani and Yabu (2017, PSY17) the set of $k_j s$ can be a proper subset of all the integers between 1 and the maximum frequency, k_n, and hence k_n may not correspond to the n^{th} frequency. For instance, when $n = 2$ and $k_n = 3$, (k_1, k_2) can be either $(1, 3)$ or $(2, 3)$.

Note as well that the presence of the trigonometric terms does not imply that the series must be periodic. Rather, it is known that such terms can often capture the behaviour of any nonlinear function even when it is not periodic. Moreover, the approximation is particularly well suited to the case of trend stationary series affected by several structural breaks, particularly when they are smooth or gradual. It does not require any knowledge about the nature of those breaks, nor about their number and date. And since breaks distort the spectral density function of the series at low frequencies, values for k_j equal to 1 and 2 are usually recommended to deal with them. When used to test for a unit root in u_t, the usual recommendation is to use only one frequency so that there is no significant power loss resulting from over-fitting (see, e.g., Becker, Enders and Lee, 2006, hereafter BEL06, and Lopes, 2023).

In figure (5.1) the simulated series represented in figure (3.4) is presented again, together with the fitted values of the Fourier approximation

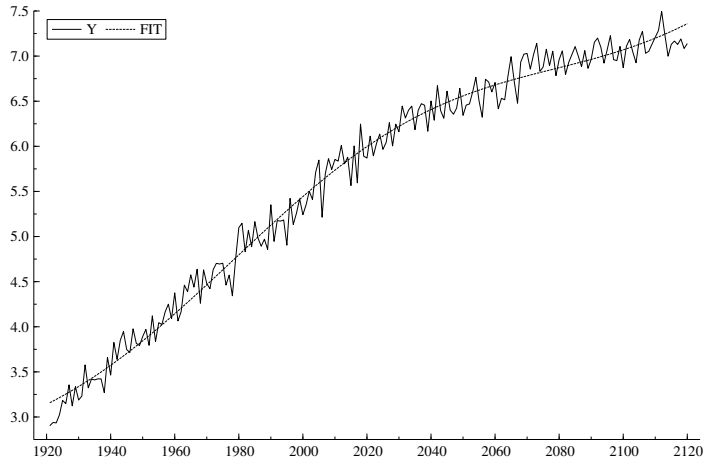

Figure 5.1: The simulated series of figure (3.4) and its fitted values with the Fourier approximation using a single frequency

using only one frequency, which is unity. The estimated model is

$$\widehat{y}_t = 3.51 + 0.021t - 0.103 \sin\left(\frac{2\pi t}{T}\right) - 0.373 \cos\left(\frac{2\pi t}{T}\right)$$

and the fit is particularly good as the trend changes and evolves smoothly after 2021.

As another example, the series of figure (5.2) was generated with three breaks, all affecting the trend negatively: a) the first, in 1980, affects both the intercept and the slope; b) the second, in 2020, affects only the slope (a growth shift); c) and the third, in 2080, affects only the intercept (a level shift). The model used for the approximation employed two frequencies, $(k_1, k_2) = (1, 2)$, and provided a much better fit than the approximation with a single frequency, all the estimated γ coefficients appearing highly statistically significant according to the standard criteria; moreover, a conventional F-statistic to test $H_0 : \gamma_{12} = \gamma_{22} = 0$ is also highly significant.

The test that was just mentioned is valid because the series is known to be trend stationary, $I(0)$, and the error term of its DGP is serially

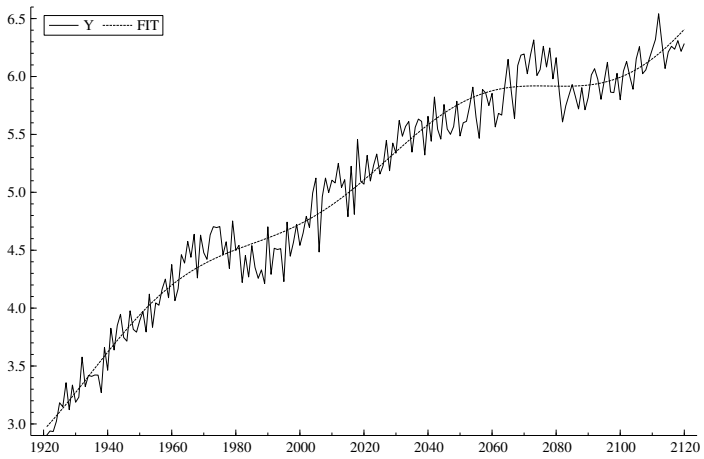

Figure 5.2: A series simulated with three breaks and its fitted values with the Fourier approximation using two frequencies

uncorrelated. But this information is almost never available. The typical situation is again subject to a circularity problem:

a) to perform valid tests to assess whether there is some nonlinearity in the data and to determine the correct order of the Fourier approximation, one needs to know in advance the order of integration of the series;

b) but the validity of unit root or stationarity tests that are necessary to determine the order of integration also depends critically on the adequate specification of the deterministic component, which again requires that the tests mentioned in a) are made beforehand.

BEL06 tabulate the critical values of a $sup - F$ test to determine the order n which is valid only in a stationary environment. On the other hand, imposing a unit root, Enders and Lee (2012, EL12) derive the critical values for a F-test for a nonlinear trend. The tests of BEL06 will be oversized in case the data are $I(1)$; and the tests of EL12 will have reduced power in case the data are $I(0)$. Hence, as in the case of tests

for a linear trend, what is desirable is to have a test that is robust to the order of integration.

Harvey, Leybourne and Xiao (2010, HLX10) devised the first approach to the problem that is valid regardless of the order of integration (provided it is either $I(0)$ or $I(1)$). Considering equation (5.9), the test for a nonlinear trend is a test for

$$H_0 : \gamma_{1j} = \gamma_{2j} = 0,\ j = 1, \ldots, n, \quad vs.$$
$$H_1 : \text{at least one } \gamma_{1j}, \gamma_{2j} \neq 0, j = 1, \ldots, n.$$

Their method builds upon the work of Vogelsang (1998): a Wald statistic based on the partial sum or accumulation of equation (5.9) is considered and the limiting distribution under both assumptions, $I(0)$ and $I(1)$, is derived; the key feature is a modification that is applied to ensure that the critical values for each test coincide under both assumptions. To achieve this feature, the modification must depend on another, auxiliary statistic that converges in probability to zero in case the data is $I(0)$ but converges weakly to a pivotal limiting distribution when it is $I(1)$.

Moreover, HLX10 proposed also a robust algorithm to determine the order of the approximation, which again relies on a composite statistic. The complexity of the algorithm is not high but the process is somewhat cumbersome.

Astill, Harvey, Leybourne and Taylor (2015, AHLT15) further modify HLX10's procedure: instead of building a composite statistic, they use the auxiliary unit root test statistic to switch between the asymptotic critical values for the Wald statistic in the two contexts. The modification improves the small sample performance of the test, particularly in terms of power.

In both these cases, the limiting distributions of the test statistics are non-standard and depend on the choice of the frequencies. This is not the case when the approach presented in PSY17 is followed: the limiting distribution is standard χ^2 and depends only on the number of frequencies. And although it is also somewhat involved, its improved finite sample properties appear to justify the choice. It builds on the

work of PY09 addressed in subsection (5.5), sharing with it the following features:

a) it relies on a FGLS procedure which is supported by a estimator of the sum of the autoregressive coefficients (ρ) which is super-efficient when $\rho = 1$;

b) it builds the "bridge for the gap" between the limiting distributions of the test statistic for the I(0) and I(1) cases upon this property of the estimator for ρ, making the limiting distribution for the $I(1)$ case collapse into the one for the $I(0)$ case[9].

Differently from PY09, it needs to resort to the Prais-Wisten version of the FGLS estimation to ensure that the limiting distribution is not dominated by the initial condition. Indeed, as can be observed in equation (5.7), PY09 use the more popular Cochrane-Orcutt method, which implicitly assumes that the initial condition is zero; for further details about the distinction between these two versions of FGLS estimation see, e.g., Canjels and Watson (1998).

Equation (5.9) can be written in vector form as

$$y_t = x_t'\Psi + u_t, \tag{5.10}$$

where $x_t = (z_t', f_t')'$, with $z_t = (1, t)'$ and $f_t = (\sin(2\pi k_1 t/T), \cos(2\pi k_1 t/T), \ldots, \sin(2\pi k_n t/T), \cos(2\pi k_n t/T))'$, $\Psi = (\phi', \gamma')$, with $\phi = (\alpha, \beta)'$ and $\gamma = (\gamma_{11}, \gamma_{21}, \ldots, \gamma_{1n}, \gamma_{2n})'$. The assumptions for the error term u_t (and for Δu_t) are similar to those made in section (5.3):

a) in the $I(0)$ case, $u_t = C(L)e_t$, $C(L) = \sum_{i=1}^{\infty} c_i L^i$, $\sum_{i=1}^{\infty} i|c_i| < \infty$, $0 < |C(1)| < \infty$;

b) in the $I(1)$ case, $\Delta u_t = D(L)e_t$, $D(L) = \sum_{i=1}^{\infty} d_i L^i$, $\sum_{i=1}^{\infty} i|d_i| < \infty$, $0 < |D(1)| < \infty$,

[9] This becomes particularly easy to observe in the simple example offered by PSY17 for the case of a single trigonometric regressor.

with $e_t \sim (0, \sigma^2)$ a m.d.s. and $u_0 = O_p(1)$. As in PY09, the autoregressive representation is employed, $u_t = \sum_{i=1}^{\infty} a_i u_{t-i} + e_t$ or

$$u_t = \rho u_{t-1} + A^*(L)\Delta u_t + e_t,$$

ρ still representing the sum of the autoregressive coefficients, $\rho = A(1)$. As regards the null hypothesis for the absence of any nonlinear component, it may be written as

$$R\Psi = 0,$$

where $R = [0 : I_{2n}]$ is the $2n \times (2 + 2n)$ matrix of the restrictions. In the following paragraphs the steps of PSY17's procedure are described in detail.

i. Run the OLS regression (5.9) or (5.10) to obtain the residuals $\hat{u}_t$.

ii. Estimate ρ with the OLS regression

$$\hat{u}_t = \rho \hat{u}_{t-1} + \sum_{i=1}^{p} a_i^* \Delta \hat{u}_{t-i} + e_{pt}, \tag{5.11}$$

where p is the truncation lag order, that PSY17 recommend estimating with the modified AIC (MAIC) method proposed in Ng and Perron (2001); alternatively, resort to the usual AIC statistic or to the general-to-specific (GTS) popular method. PSY17 recommend starting with $p_{max} = [12 \left(\frac{T}{100}\right)^{1/4}]$, which is a generous lag order. Denote with $\hat{\tau}$ the usual unit root test statistic, $\hat{\tau} = (\hat{\rho} - 1)/se(\hat{\rho})$.

iii. Since the estimator of ρ based in equation (5.11) is known to be biased downward, particularly when ρ is close to 1, use instead the bias-corrected estimator:

$$\hat{\rho}_M = \hat{\rho} + C(\hat{\rho})\, se(\hat{\rho}),$$

where

$$C(\hat{\rho}) = \begin{cases} -\hat{\tau}, & \text{if } \hat{\tau} > \tau_{pct} \\ I_p T^{-1}\hat{\tau} - (1+r)[\hat{\tau} + c_2(\hat{\tau}+10)]^{-1}, & \text{if } -10 < \hat{\tau} \leq \tau_{pct} \\ I_p T^{-1}\hat{\tau} - (1+r)\hat{\tau}^{-1}, & \text{if } -c_1^{1/2} < \hat{\tau} \leq -10 \\ 0, & \text{if } \hat{\tau} \leq -c_1^{1/2} \end{cases}$$

with τ_{pct} denoting a percentile of the limiting distribution of $\hat{\tau}$ when $\rho = 1$ that PSY17 considered as $\tau_{0.50}$ and $\tau_{0.85}$, for the "median-unbiased" and "upper-biased" tests, respectively, $c_1 = (1+r)T$, $r = 2 + 2n$ is the total number of coefficients in (5.9) or (5.10), I_p is the integer part of $(p+2)/2$, $c_2 = [(1+r)T - \tau_{pct}^2(I_p+T)][\tau_{pct}(10+\tau_{pct})(I_p+T)]^{-1}$. The percentiles $\tau_{0.50}$ and $\tau_{0.85}$ are tabulated in PSY17, both when a singe frequency is used, with a value between 1 and 5, and for cases with multiple frequencies (see table 1 of PSY17, p. 830).

iv. With the previous estimator, construct the super-efficient estimator when $\rho = 1$ given by

$$\hat{\rho}_{MS} = \begin{cases} \hat{\rho}_M, & \text{if } |\hat{\rho}_M - 1| > T^{-1/2} \\ 1, & \text{otherwise.} \end{cases}$$

v. Use the previous estimate to build the Prais-Wisten FGLS regression, given by

$$(1 - \hat{\rho}_{MS}L)y_t = (1 - \hat{\rho}_{MS}L)x_t'\Psi + (1 - \hat{\rho}_{MS}L)u_t,\ t = 2, \ldots, T,$$

$$\text{and } (1 - \hat{\rho}_{MS}^2)^{1/2}y_1 = (1 - \hat{\rho}_{MS}^2)^{1/2}x_1'\Psi + (1 - \hat{\rho}_{MS}^2L)u_1,$$

to estimate Ψ and denote the residual of this regression with $\hat{v}_t$ since they estimate $v_t = (1 - \rho L)u_t$.

vi. Finally, the Wald statistic robust to serial correlation in v_t is

$$W_{\hat{\gamma}} = \hat{\Psi}'R'[\hat{\omega}^2 R(\tilde{X}'\tilde{X})^- R']^{-1}R\hat{\Psi},$$

where $\tilde{X}$ denotes the $T \times (2+2n)$ regressor matrix of the FGLS regression, i.e., its t^{th} row is $\tilde{x}_t' = (1-\hat{\rho}_{MS}L)x_t', t = 2, \ldots, T$, but $\tilde{x}_1' = (1-\hat{\rho}_{MS}^2)^{1/2}x_1'$,

$(\tilde{X}'\tilde{X})^-$ is the generalized inverse of $(\tilde{X}'\tilde{X})$, and $\hat{\omega}^2$ is a long-run variance estimator of v_t, which is given by

$$\hat{\omega}^2 = \begin{cases} (T-p)^{-1}\sum_{t=p+1}^{T}\hat{e}_{tp}^2, & \text{if } T^{1/2}|\hat{\rho}_M - 1| > 1 \\ T^{-1}\sum_{t=1}^{T}\hat{v}_t^2 + T^{-1}\sum_{j=1}^{T-1}\omega_j(j,m_T)\sum_{t=j+1}^{T}\hat{v}_t\hat{v}_{t-j}, & \text{otherwise} \end{cases}$$

with $\hat{e}_{pt}$ the residuals from equation (5.11) and $\omega_j(j,m_T)$ a weight function with bandwidth m_T; PSY17 recommend using Andrews's automatic selection method for m_T and the quadratic spectral window.

In both the $I(0)$ and $I(1)$ cases and under H_0,

$$W_{\hat{\gamma}} \xrightarrow{d} \chi^2_{(2n)}.$$

Regarding the choice between the median-unbiased and the upper-biased versions of the test, PSY17 recommend using this last one since its power performance is comparable to the first, only slightly inferior, but it is less affected by size distortions.

As regards the choice of frequencies, assuming that their set $(k_1,\ldots,k_n)$ is known, with $k_n = n$, PSY17 propose a simple GTS procedure which consists of the sequential application of the version of their test for subsets of coefficients, adapting the R matrix accordingly. For instance, starting with $n = 3$, the null hypothesis that the coefficients related to the maximum frequency are both equal to zero is tested. If this hypothesis is rejected, $n = 3$ is selected; otherwise n is set at $n = 2$ and the hypothesis that the coefficients related to $k = 2$ are zero is tested. The process stops when a decision for rejection is reached or when $n = 0$. A Monte Carlo study indicates that this method is more powerful than the one proposed in HLX10.

More recently, Perron, Shintani and Yabu (2020) returned to this problem, removing the assumption of a known set of frequencies and proposing a specific-to-general method which is again robust to the long-run properties of the data. Their procedure relies in two new tests, the mean and the sup-type tests. However, when the problem is confined to the nonlinearity in trend, not in the level, the assumption of a known set

of frequencies does not appear to be particularly restrictive. Actually, in unit root testing literature where the Fourier approximation is also used, a single frequency is usually employed and the only constraint is that its value must be low, not far from zero, to accommodate the effect of structural breaks. In the case at hand, considering $n = 2$, and so $k_n = 2$, must be sufficient for most cases. Abandoning the assumption of integer frequencies only, as has been done in that literature, appears to be a more promising avenue for research.

5.8 Two nonparametric tests

Tests for the detection of trends are not confined to economics, however. They can easily be found in research in earth science, and in particular in hydrology — as in Hamed and Rao (1998) and Shi et al. (2013) — and in climate change analysis. Actually, as recently as 2020, the *Journal of Econometrics* published an entire issue dedicated to this topic. As in the past in economics, where a linear time trend was often considered to provide a good approximation to the evolution of technology, a similar suggestion may be found in this area concerning the representation of anthropogenic influences.

The most well known trend test in hydrology is the Mann (1945)–Kendall (MK)[10] test, which is clearly presented in, e.g., Hamed and Rao (1998). Empirical applications vary much, ranging from rainfall data, to temporal pan evaporation series, to data on the streamflow of rivers or groundwater quality, to the occurrence of floods and droughts, etc. The MK test is a nonparametric test originated in the analysis of rank correlations between two groups of observations and particularized to the case where one of them is the order of the observations in time. Its focus is the correlation between the rank order of the observations and their order in time. The null hypothesis is that the data come from a random sample, i.e., that they are independently and identically distributed, and the commonly adopted alternative is that the data contain a trend.

[10]Kendall, M. K., 1955, *Rank Correlation Methods*, Griffin, London.

In the general case, the two sets of observations are denoted with $y_1, y_2, \ldots, y_T$ and $x_1, x_2, \ldots, x_T$, respectively. An original test statistic S is defined by

$$S = \sum_{i<j} a_{ij} b_{ij},$$

where

$$a_{ij} = \text{sign}(y_j - y_i) = \begin{cases} 1, & \text{if } y_i < y_j \\ 0, & \text{if } y_i = y_j \\ -1, & \text{if } y_i > y_j \end{cases},$$

and similarly for b_{ij} with the observations of X. Under the null hypothesis, S is asymptotically normal with $\text{E}(S) = 0$ and

$$\text{Var}(S) = \frac{T(T-1)(2T+5)}{18}.$$

When variable X is replaced by the order in time of the observations of Y, the test statistic adopts its trend form and it is given by

$$S = \sum_{i<j} a_{ij} = \sum_{i<j} \text{sign}(y_j - y_i).$$

It is frequently acknowledged that this test performs poorly, with serious size distortions when the data are serially correlated, i.e., the test often indicates the presence of a trend when none exists, simply because the data are serially correlated. And indeed the correct alternative hypothesis is much more general than the presence of a trend: any depart from randomness is a violation of the null. Several methods have been proposed to tackle this problem but I will not address them here. This test has not find its way into macroeconomics and, at least apparently, with good reason[11].

[11] An exception may be found in the work of Rogoff, Rossi and Schmelzing (2024), who refer to the test as providing evidence for a positive linear trend in global output; however, their strong claim for a downward linear trend in global long-maturity real interest rates is substantiated in visual inspection only, as they do not present the results for this or for any other formal test.

A simpler test, coming also mainly from hydrology, is the Cox-Stuart test. The sample is split into two parts, $y_1, \ldots, y_{T/2}$ and $y_{T/2+1}, \ldots, y_T$, and the sign of the difference between each pair $(y_t, y_{t+T/2})$ is retained[12]. The test statistic is given by the number of pairs in which $y_t < y_{t+T/2}$, that is

$$CS = \sum_{t=1}^{T/2} 1(y_t < y_{t+T/2}),$$

and its distribution under the null is the binomial with parameters $T/2$ and $1/2$ ($CS \sim \text{Bi}(T/2, 1/2)$). According to the Monte Carlo study in Rutkowska (2015), despite its simplicity the loss in power relatively to the MK test is small. However, in this study the problem of serial correlation is not addressed because it is assumed from the outset to be absent from the data, so that the test can be validly performed.

[12] T is assumed to be even; if it is not, the middle observation is removed from the series.

Appendix A

A brief introduction to filters

Although it is not claimed that economic time series display regular cycles, it may be useful to think that some of their variation results from the contribution made by periodic components. This is the subject of frequency analysis, which provides insights unattainable in the time domain and which is useful to understand linear filters. The brief overview that follows aims to provide a few tools to better understand this topic and starts by addressing a basic subject, that of complex numbers.

A.1 Complex numbers

Complex numbers as

$$z = a + bi,$$

are composed of a real part, a, and an imaginary one, $+bi$, where the complex unit is defined by $i = \sqrt{-1}$ ($i^2 = -1$). Complex numbers can be represented in the complex plane, with a plotted in the x axis and b in the y axis.

Euler's identity,

$$e^{\pm i\theta} = \cos\theta \pm i\,\sin\theta,$$

is often useful. It can be used to represent complex numbers in polar

coordinate form, as

$$\begin{aligned} z &= R\, e^{i\theta} \\ &= R\,[\cos\theta + i\,\sin\theta] \end{aligned}$$

where $R = (a^2+b^2)^{1/2}$ is the *modulus* of the complex number, i.e., $R = |z|$, and

$$\theta = \tan^{-1}(b/a)$$

is the angle, in radians ($\pi = 180^o$), that the point (a, b) makes with the real axis, which can be also characterized with

$$\cos\theta = a/R$$

$$\sin\theta = b/R.$$

Two useful identities that result from Euler's identity are

$$\cos\theta = (e^{i\theta} + e^{-i\theta})/2,$$

$$\sin\theta = (e^{i\theta} - e^{-i\theta}/2i.$$

Adding complex numbers is easy:

$$(a + b\,i) + (c + d\,i) = (a + c) + (b + d)i.$$

And multiplication is also easy:

$$(a + b\,i)(c + d\,i) = ac + ad\,i + bc\,i - bd = (ac - bd) + (ad + bd)i.$$

But it is easier in polar notation:

$$R_1 e^{i\theta_1} R_2 e^{i\theta_2} = R_1 R_2 e^{i(\theta_1+\theta_2)},$$

thus, the modulus of the product is the product of the two modulus and the angle of the product is the sum of the two angles.

The *complex conjugate*, denoted with an asterisk (some prefer to denote it with $\bar{z}$), is

$$z^* = (a + bi)^* = a - bi,$$

and, in polar coordinates,

$$(Re^{i\theta})^* = Re^{-i\theta}.$$

Note that the product of a complex with its conjugate is real:

$$(a + bi)(a - bi) = a^2 + b^2 = R^2.$$

Comparing the first equality with the expression for the modulus of a complex number, it can be deduced that the modulus of the complex number can be viewed as the square root of the product of the number with its complex conjugate:

$$R = |a + bi| = \sqrt{(a + bi)(a - bi)}.$$

A.2 Cyclical functions

Suppose one wishes to build a periodic deterministic model for a time series. Since a sinusoidal function is a possibility, one may consider, e.g.,

$$x_t = H \cos(\omega t + \phi),$$

where H is the *height* or *amplitude* of the sinusoidal variation, ω is its (angular) *frequency* in radians, that is, the number of radians per unit time, and ϕ is the angular amount by which the cosine wave is shifted, called the *phase*, which when is measured in time is ϕ/ω. Since the cosine function completes a full cycle in 2π radians, the *period* of this function, that is, the time that x takes to complete its complete sequence is given by

$$period = \frac{2\pi}{\omega},$$

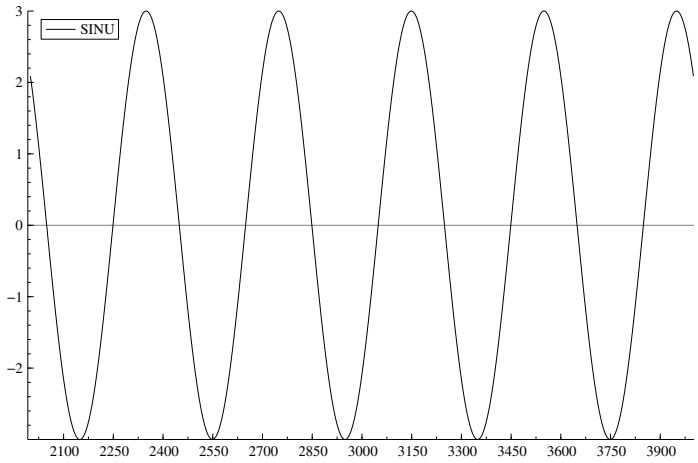

Figure A.1: An illustration of a sinusoidal function: the function $H\cos(\omega\, t + \phi)$ with $H = 3$, $\omega = \pi/2$ and $\phi = \pi/4$

which is sometimes called the *length of the period* or the *wavelength*[1]. This simple expression is very useful to economic reasoning.

The relation can obviously be inverted; for instance, if a trigonometric function repeats itself every 4 periods, its frequency is derived from $4 = 2\pi/\omega$, that is, $\omega = \pi/2$.

In figure (A.1) the sinusoidal function $3\cos(\frac{\pi}{2}t + \frac{\pi}{4})$ is represented. To approximate a continuous time variable, a time trend s was first generated for a sample of annual data from 2000 to 4000 and then t was made equal to $s/100$.

[1]Indeed, recalling that the sine and cosine functions are periodic with period 2π, and rewritting $\sin(\omega t)$ and $\cos(\omega t)$ as $\sin(2\pi\frac{\omega t}{2\pi})$ and $\cos(2\pi\frac{\omega t}{2\pi})$, repectively, it is easy to see that both these functions repeat whenever $\frac{\omega t}{2\pi}$ is an integer. Hence, setting $\frac{\omega t}{2\pi} = 1$ allows concluding that both repeat every $t = \frac{2\pi}{\omega}$ periods.

A.3 Fourier transforms

The (discrete) *Fourier transform* of a time series x_t is defined by

$$x(\omega) = \sum_{t=-\infty}^{\infty} e^{-i\omega t} x_t.$$

Thus, it transforms a function of t into a complex valued function of ω.

Given $x(\omega)$, the *inverse Fourier transform* is defined by

$$x_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega t} x(\omega) d\omega,$$

that is, it allows recovering the original series from its transform.

A.4 Spectral density

Suppose that the process $\{x_t\}$ is stationary, with autocovariances γ_j that are absolutely summable. Then its spectral density or spectrum is defined as the Fourier transform of its autocovariance function, i.e., as[2]

$$s_X(\omega) = \sum_{j=-\infty}^{\infty} \gamma_j \, e^{-i\omega j}.$$

But since the autocovariance function is an even function ($\gamma_{-j} = \gamma_j$) and using Euler's identity as well as the properties of the sine and cosine functions[3] it can be written as

$$s_X(\omega) = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j \, \cos(j\omega).$$

[2]Unfortunately there is no unanimity in the literature. For instance, in Hamilton (1994) and in Brockwell and Davies (2016), the (population) spectrum is this magnitude divided by 2π. Harvey (1993, p. 167) also adopts this convention but calls it the *power spectrum*. I will use the terms spectral density and spectrum interchangeably.

[3]Recall that $\cos(0) = 1$, $\sin(0) = 0$, $\sin(-\theta) = -\sin(\theta)$ and $\cos(-\theta) = \cos(\theta)$.

Hence, it is a real function, that can be shown to be nonnegative for all ω and symmetric around $\omega = 0$ because $\cos(\omega j) = \cos(-\omega j)$ for any ω and, since $\cos[(\omega + 2\pi k).j] = \cos(\omega j)$ for any integers k and j, $s_X(\omega + 2\pi k) = s_X(\omega)$ for any k, i.e., the spectral density is a periodic function of ω: it is sufficient to know the value of $s_X(\omega)$ for all $\omega \in [0, \pi]$ to know its value for any ω.

The autocovariance function can be recovered inverting the spectral density:

$$\gamma_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega j} s_X(\omega) d\omega.$$

In particular, with $j = 0$:

$$\gamma_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} s_X(\omega) d\omega,$$

that is, the variance of the process is the area under the spectral density between $\pm\pi$ divided by 2π. Cochrane (2005, p. 74) provides an interesting interpretation: "*this equation interprets the spectral density as a decomposition of the variance of the process into uncorrelated components at each frequency ω (if they weren' t uncorrelated, their variances would not sum without covariance terms)*".

This result can be generalized. Indeed,

$$\frac{1}{2\pi} \int_{-\omega_1}^{\omega_1} s_X(\omega) d\omega,$$

for any $\omega_1 \in [0,\ \pi]$ is a positive magnitude that can be interpreted as the proportion of the variance of x_t that is associated with frequencies ω that are smaller than ω_1 in absolute value. But since $s_X(\omega)$ is symmetric, twice the integral between 0 and ω_1, that is,

$$\frac{1}{\pi} \int_{0}^{\omega_1} s_X(\omega) d\omega,$$

represents the fraction of the variance of x_t that can be attributed to periodic random components with frequencies less than or equal to ω_1.

And of course we can convert these in terms of periodicities.

Moreover, if a spectrum exhibits a peak at a certain frequency, it means that the frequency and the frequencies close to it have an important contribution to the variance of the series.

A.5 Some spectral densities

For a white noise process, $x_t = \epsilon_t \sim iid(0, \sigma_\epsilon^2)$, that is, with $\gamma_0 = \sigma_\epsilon^2$ and $\gamma_j = 0$ for $j > 0$,

$$s_X(\omega) = \sigma_\epsilon^2 = \sigma_x^2.$$

Thus, the spectral density is flat; all the frequencies are equally represented and actually the name white noise derives from the analogy with white light, which contains all the frequencies in the colour spectrum.

For the MA(1) process, $x_t = \epsilon_t + \theta\epsilon_{t-1}$, with $\gamma_0 = (1+\theta^2)\sigma_\epsilon^2$, $\gamma_1 = \theta\sigma_\epsilon^2$ and $\gamma_j = 0$ for $j > 1$, it is

$$\begin{aligned} s_X(\omega) &= (1+\theta^2)\sigma_\epsilon^2 + 2\theta\sigma_\epsilon^2 \cos\omega \\ &= (1+\theta^2+2\theta\cos\omega)\sigma_\epsilon^2. \end{aligned}$$

MA(1) processes with $\theta > 0$ have spectral densities with low frequencies emphasized, while those with $\theta < 0$ have spectral densities with high frequencies emphasized.

For a stationary AR(1) process, $x_t = \phi x_{t-1} + \epsilon_t$, the spectral density is

$$s_X(\omega) = \frac{\sigma_\epsilon^2}{1+\phi^2-2\phi\cos(\omega)}.$$

When $\phi > 0$ the denominator is monotonically increasing in ω over $[0, \pi]$, and hence $s_X(\omega)$ is monotonically decreasing. When $\phi < 0$, $s_X(\omega)$ is a monotonically increasing function of the frequency.

A.6 Linear filters

It is well known that several characteristics of a time series can be modified by means of a linear transformation. In particular, linear filters modify the spectral characteristics of a series in a predictable way, and hence are useful to highlight the behaviour of the series at certain frequencies, i.e., to enhance certain parts of its spectrum.

Denote with x_t the stationary input series and consider that the series y_t is constructed as a weighted average of x_t:

$$y_t = \sum_{j=-r}^{s} a_j x_{t-j},$$

with weights $a_{-r}, \ldots, a_0, \ldots, a_s$ real and fixed. This transformation is a linear time-invariant filter because the weights do not vary with time. When the weights sum unity the filter is usually called a moving average.

Using the common notation of polynomials in the lag operator (L), the filter may be represented as

$$y_t = A(L)x_t,$$

with

$$A(L) = a_{-r}L^{-r} + \ldots + a_{-1}L^{-1} + a_0 + a_1 L + \ldots + a_s L^s.$$

The weights or coefficients are called the *impulse response function*. A very simple example is the case of the moving average defined by

$$y_t = \frac{1}{3}(x_{t-1} + x_t + x_{t+1}),$$

which has $A(L) = (1/3)(L^{-1}+1+L)$ and whose impulse response function is

$$a_j = \begin{cases} 1/3, & j = -1, 0, 1 \\ 0, & \text{otherwise.} \end{cases}$$

In the case that the filter is infinite,

$$y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j},$$

its coefficients are required to satisfy an absolute summability condition, $\sum_{t=-\infty}^{\infty} |a_t| < \infty$, so that the Fourier transform

$$A(e^{-i\omega}) = \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j},$$

which is called the *frequency* response function, is well defined.

A.7 Relation between the spectral densities

A very important result concerns the relation between the spectral densities of the input and output series. This relation is given by

$$s_Y(\omega) = |A(e^{-i\omega})|^2 \, s_X(\omega),$$

where the term $|A(e^{-i\omega})|^2$ is called the *power transfer function*. Thus the filter changes the spectrum of the input series through a frequency-by-frequency multiplication by the squared modulus of the frequency response function.

Recalling the well known property of the variance which states that if x denotes a random variable with variance σ_x^2 then, $y = bx$ has variance $\sigma_y^2 = b^2\sigma_x^2$, there is therefore a close analogy between this property and the effect of filtering on the spectrum of a series.

A.8 Some examples

In the first example probably the most popular filter is addressed, the one of first differencing:

$$y_t = \Delta x_t = (1 - L)x_t.$$

It is an example of a *high-pass* filter because it passes the higher frequencies while reducing or attenuating the low frequency variation of the series. Baxter and King (1999) strongly criticize this filter as a band-pass (business cycle) filter, isolating fluctuations in the data associated with a certain range of periodicities, because it induces a "*a dramatic reweighting of frequencies*" (p. 585) towards the higher frequencies while downweighting lower frequencies.

The weights of the impulse response are simply $a_0 = 1, a_1 = -1$ and $a_r = 0$ otherwise. Therefore,

$$A(e^{-i\omega}) = 1 - e^{-i\omega},$$

and

$$\begin{aligned} |A(e^{-i\omega})|^2 &= (1 - e^{-i\omega})(1 - e^{i\omega}) \\ &= 2(1 - \cos\omega). \end{aligned}$$

Therefore,

$$s_Y(\omega) = 2(1 - \cos\omega)\, s_X(\omega).$$

This power function is represented in figure A.2. The filter attenuates the lower frequencies — and it is even equal to zero at $\omega = 0$ — and enhances the higher frequencies.

A simple application of the filtering formula concerns the MA(∞) representation, which expresses any stationary process as an infinite linear filter of a white noise process. Indeed, from the Wold representation of the series,

$$y_t = \Psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j},$$

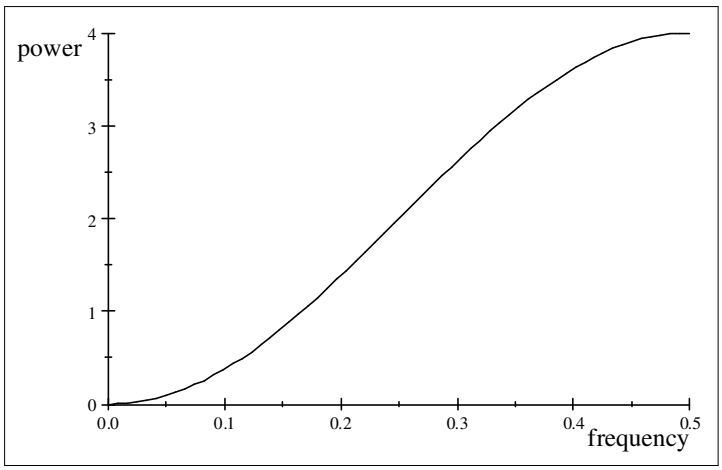

Figure A.2: The power or squared frequency response function of the first difference filter

one can easily derive that

$$s_Y(\omega) = \Psi(e^{-i\omega})\Psi(e^{i\omega}).\sigma_\epsilon^2,$$

and hence the spectral density of any stationary ARMA process can be derived.

For instance, for the MA(1) process

$$y_t = \epsilon_t + \theta\epsilon_{t-1} = (1 + \theta L)\epsilon_t,$$

$$\begin{aligned} s_Y(\omega) &= (1 + \theta e^{-i\omega})(1 + \theta e^{i\omega})\,\sigma_\epsilon^2 \\ &= [1 + \theta(e^{i\omega} + e^{-i\omega}) + \theta^2]\,\sigma_\epsilon^2 \\ &= [1 + 2\theta\cos(\omega) + \theta^2]\,\sigma_\epsilon^2, \end{aligned}$$

as previously indicated.

A.9 Gain and phase

Applying a filter to a series produces two different types of effects:

a) it changes the relative importance of the various cyclical or periodic components;

b) it may shift the position of the series in time.

The first effect is usually the one that is sought with the filter: to increase or enhance and to reduce or attenuate the amplitude of certain cyclical components. This effect is called the *gain*, $G(\omega)$, and it is given by the modulus of the frequency response function:

$$G(\omega) = |A(e^{-i\omega})|.$$

Hence the power transfer function is the gain squared. At any given frequency, it indicates the extent to which the filter augments or reduces the variance of the filtered series in relation to the original series (Baxter and King, 1999).

I take an example from Harvey (1993, p. 193) to illustrate this effect. It is the case of a symmetric moving average, targeted to smooth the series, i.e., to reduce or even eliminate the irregular fluctuations and to emphasize the trend, i.e., the low frequency component. Suppose that the moving average has length $m = 2r + 1$, with equal weights summing unity:

$$y_t = \frac{1}{m} \sum_{j=-r}^{r} x_{t-j}.$$

Then it can be shown that

$$A(e^{-i\omega}) = \frac{\sin(m\,\omega/2)}{m \sin(\omega/2)},$$

which is real because the filter is symmetric.

Figure A.3 illustrates the gain function for the case of a five-period moving average filter. Since the gain is small at high frequencies, irregular movements of the series are dampened, smoothing the series and making the trend appear more clearly. Notice also that the gain is zero at the frequency corresponding to a five-period cycle, $\omega = 2\pi/5$.

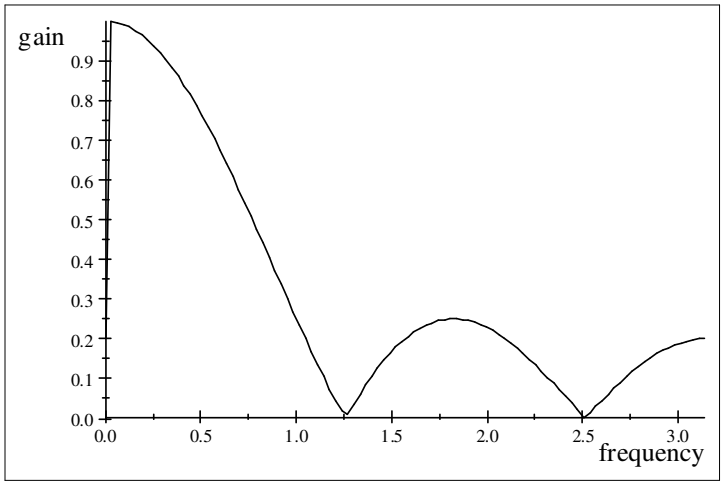

Figure A.3: The gain function for a five-period moving average

The second effect is called *phase shift* and it is represented with $Ph(\omega)$. As a simple example consider that the filter simply shifts the series four periods back in time:

$$y_t = x_{t-4}.$$

If the original series, x_t, were generated by a cyclical process as, for instance

$$x_t = \cos \omega t,$$

the filtered series would be

$$y_t = \cos \omega(t-4) = \cos(\omega t - 4\omega),$$

and hence the phase shift is $Ph(\omega) = 4\omega$, i.e., a backward shift of four time periods is transformed in a phase shift of 4ω in the frequency domain.

The analysis of phase shifts appears to be both less interesting and more involved than that of the gain. However, a few results are available that may be helpful. First, a sinusoidal input of a filter originates a

sinusoidal output at the same frequency. Intuitively, as pointed out by Cochrane (2005, p. 82), "*all you can do to a sine wave is stretch it or shift it*". Harvey (1993) provides a brief discussion about the calculation of phase shifts.

Second, symmetric filers — as, for instance, most moving averages — have real response functions and do not exhibit any phase shift.

A.10 Ideal filters

Suppose now that one wishes to find an *optimal* or *ideal filter*, such that when it is applied to a series it retains or extracts only those components of the series whose frequencies lie (in absolute value) in a certain subset Ω^* of $[-\pi, \pi]$. The ideal *low pass filter* is the filter that leaves the low frequencies, long-run or slow movements unchanged and "wipes out" the high frequency components of the series. It is denoted with $A_L(\omega)$ and it must pass only frequencies in the interval $\Omega^* =]-\omega_c, \omega_c[$, where ω_c is called the *cut-off frequency*. Its frequency response function must be

$$A_L(\omega) = \begin{cases} 1, & \text{if } \omega \in \Omega^* \\ 0, & \text{if } \omega \notin \Omega^* \end{cases}.$$

It should be also phase neutral, that is, not inducing any temporal shift, and this requires that the filter is symmetric, $A_L(\omega) = A_L(-\omega)$. Resorting to the inverse Fourier transform, Mills (2003, pp. 85-86) derives the weights of this filter:

$$a_{L,0} = \frac{\omega_c}{\pi}, \quad \text{and} \quad a_{L,j} = \frac{1}{\pi j}\sin(\omega_c j),\ j \neq 0.$$

This is an infinite moving average and, although the weights $a_{L,j}$ tend to zero as j increases, an infinite sample of observations is required. Therefore, the filter is unfeasible.

Conversely, if the gain of the filter is small (or even zero) for low values of ω but large for high values of ω, the filter is a *high-pass filter*. Finally, a *band-pass* filter can be viewed as a combination of a low pass

and a high pass filter: "wiping out" very high and very low frequencies and allowing only the intermediate frequencies to "get through". That is, components of the series corresponding to cycles with lengths inside a band, defined by a lower and a upper bound, are passed through or extracted, and the remaining are filtered out. This is the type of filter typically used in business cycle analysis; Canova (2007, pp. pp. 94-99) provides an overview of these filters.

All these filters are unfeasible and only approximate solutions are feasible. The graphical representation of their gain functions is box-shaped but those of their feasible, finite sample, approximations is bell-shaped. Thus, relatively to their ideal versions, feasible filters imply some loss of power at the edges of the band; this is called the *leakage* of the filter. It also implies an increase in importance of the frequencies in the middle of the band, and this phenomenon is called *compression*. The opposite phenomenon is called *exacerbation*.

For the case of low pass filters, Mills (2003, p. 87) shows that the weights of the ideal approximation filter sets $a_j = a_{L,j}$ for $j = 0, 1, \ldots, m$ and $a_j = 0$ for $j > m$, that is, the ideal approximation simply truncates the set of the (infinite) unfeasible ideal filter weights. According to Baxter and King (1999), this result can be generalized to any approximating filter, provided the loss function is quadratic and attaches the same weights to the (squared) approximation errors at different frequencies.

Besides providing a very pedagogical introduction to linear filters, Baxter and King (1999) derive the ideal band-pass filter for business cycles analysis as well as its optimal approximation, which they represent with $BP_K(p, q)$: it passes cycles between p and q periods in length, with a truncation at point $K(= m$ in Mills'2003 notation). Since Burns and Mitchell (1946) specified that business cycles components were those with periodic fluctuations between 6 and 32 quarters, for quarterly data the ideal business cycles band-pass filter is denoted with $BP_\infty(6, 32)$ and its optimal approximation is represented with $BP_{12}(6, 32)$. For annual data Baxter and King recommend $BP_3(2, 4)$. A brief presentation of Baxter and King's filter is provided in Álvarez and Gómez-Loscos (2018) and a

more extensive, though brief presentation is provided in the main text.

Murray (2003) shows that the Baxter and King filter, and more generally any bandpass filter, is incapable of isolating the cyclical component of the UC modeling approach when the trend is an integrated process because the first difference of the trend passes through the filter and joins the cyclical component. That is, the definition of the cyclical component according to the frequency domain approach does not appear to be compatible with the UC modeling one.

A.11 Spurious effects

Besides the previously mentioned drawbacks, filters may also distort fluctuations in such an extension as to provoke spurious effects. A well known case may occur when applying summing operations over previously differenced data, a spurious cycle likely emerging from the combined effect. This phenomenon is known as the Yule-Slutsky effect, and is described in Harvey (1993, pp. 195-6). While differencing attenuates low frequencies, summing attenuates high frequencies and hence certain intermediate frequencies will tend to appear reinforced.

The most well known example of this phenomenon is provided by a work from Kuznets published in 1961[4]. While searching for evidence on long swings in economic data, Kuznets found it with a periodicity of around 20 years. A few years later it came out that this was an artifact of the two filters that he had employed, a five-year moving average followed by a five-year differencing (i.e., $y_t = x_t - x_{t-5}$). The distortion emerged from the combined effect of these two filters, whose transfer function is given by

$$|A(e^{-i\omega})|^2 = \left[\frac{2\sin 5\omega \sin(5\omega/2)}{5\sin(\omega/2)}\right]^2,$$

which has a very high peak at a frequency corresponding to cycles with length of 20.3 years.

[4]Simon S. Kuznets (1961), *Capital and the American economy : its Formation and Financing*, NBER, New York.

References

[1] Álvarez, Luis J. and Ana Gómez-Loscos (2018). A menu on output gap estimation methods, *Journal of Policy Modeling*, 40, pp. 827–850.

[2] Arata, Linda, Enrico Fabrizi and Paolo Sckokai (2020). A worldwide analysis of trend in crop yields and yield variability: Evidence from FAO data, *Economic Modelling*, 90 pp. 190–208.

[3] Astill, Sam, David I. Harvey, Stephen J. Leybourne and A.M. Robert Taylor (2014). Robust tests for a linear trend with an application to equity indices, *Journal of Empirical Finance*, 29, pp. 168–185.

[4] Astill, Sam, David I. Harvey, Stephen J. Leybourne and A.M. Robert Taylor (2015). Robust and powerful tests for nonlinear deterministic components, *Oxford Bulletin of Economics and Statistics*, 77, pp. 780-99.

[5] Ayat, Leila and Peter Burridge (2000). Unit root tests in the presence of uncertainty about the non-stochastic trend, *Journal of Econometrics*, 95, pp. 71-96.

[6] Ball, Michael and Wood, Andrew (1996). Trend growth in post-1850 British economic history: the Kalman filter and historical judgement, *Journal of the Royal Statistical Society: series D — The Statistician*, 45 (2), pp. 143-152.

[7] Baxter, Marianne and Robert G. King (1999). Measuring business cycles: approximate band-pass filters for economic time series, *The Review of Economics and Statistics*, 81(4), pp. 575-93.

[8] Becker, Ralph, Walter Enders and Junsoo Lee (2006). A stationarity test in the presence of an unknown number of smooth breaks, *Journal of Time Series Analysis*, 27, pp. 381-409.

[9] Bernard, Andrew, B. and Steven N. Durlauf (1996). Interpreting tests of the convergence hypothesis, *Journal of Econometrics*, 71, 161-73.

[10] Beveridge, Stephen and Charles R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory

components with particular attention to measurement of the business cycle, *Journal of Monetary Economics*, 7, pp. 151-174.

[11] Blanchard, Olivier J. and Stanley Fischer (1989). *Lectures on Macroeconomics*, the MIT Press, Cambridge.

[12] Bolt, Jutta, and Jan L. van Zanden (2020). Maddison style estimates of the evolution of the world economy. A new 2020 update. *Maddison-Project Working Paper* WP-15.

[13] Brockwell, Peter J. and Richard A. Davis (1987). *Time Series: Theory and Methods*, Springer-Verlag, New York.

[14] Brockwell, Peter J. and Richard A. Davis (2016). *Introduction to Time Series and Forecasting*, 3rd ed., Springer.

[15] Bunzel, Helle and Timothy J. Vogelsang (2005). Powerful trend function tests that are robust to strong serial correlation, with an application to the Prebisch–Singer hypothesis, *Journal of Business & Economic Statistics*, 23 (4), pp. 381-394.

[16] Burns, Arthur F. (1934). *Production Trends in the United States Since 1870*, NBER, New York.

[17] Burns, Arthur F. (1944). Frickey on the Decomposition of Time Series, *The Review of Economics and Statistics*, 26, pp. 136-47.

[18] Burns, Arthur F. and Wesley C. Mitchell (1946). *Measuring Business cycles*, NBER, New York.

[19] Burnside, Craig (1998). Detrending and business cycle facts: A comment, *Journal of Monetary Economics* 41, pp. 513-532.

[20] Campbell, John Y. and N. Gregory Mankiw (1987). Are output fluctuations transitory? *The Quarterly Journal of Economics*, 102, pp. 857-880.

[21] Campbell, J. Y. and Pierre Perron (1991). Pitfalls and opportunities: what macroeconomists should know about unit roots, *NBER Macroeconomics Annual*, pp. 141-219.

[22] Canjels, Eugene and Mark W. Watson (1997). Estimating deterministic trends in the presence of serially correlated errors, *The Review of Economics and Statistics*, 79 (2), pp. 184-200.

[23] Canova, Fabio (1998). Detrending and business cycle facts, *Journal of Monetary Economics* 41, pp. 475-512

[24] Canova, Fabio (2007). *Methods for Applied Macroeconomic Research*, Princeton university Press, Princeton.

[25] Cerra, Valerie, Antonio Fatás and Sweta C. Saxena (2023). Hysteresis and business cycles, *Journal of Economic Literature*, 61(1), pp. 181–225.

[26] Chalmovianský, Jakub and Daniel Němec (2022). Assessing uncertainty of output gap estimates: evidence from Visegrad countries, *Economic Modelling*, 116, pp. 1-25.

[27] Chan, K. Hung, Jack C. Hayya and J. Keith Ord (1977). A note on trend removal methods: the case of polynomial regression versus variate differencing, *Econometrica*, 45 (3), pp. 737-744.

[28] Chatfield, Chris (2004). The Analysis of Time Series – An Introduction, 6th ed., Chapman & Hall/CRC.

[29] Choi, I. (2015). *Almost All About Unit Roots*, Cambridge University Press, Cambridge.

[30] Christiano, Lawrence, J. and Terry J. Fitzgerald (2003). The band pass filter, *International Economic Review*, 44 (2), pp. 435-65.

[31] Chudik, Alexander, M. Hashem Pesaran and Ron Smith (2022). Revisiting the Great Ratios Hypothesis, *Cambridge Working Papers in economics*.

[32] Clark, Peter K. (1987). The cyclical component of U. S. economic activity, *The Quarterly Journal of Economics*, 102 (4), pp. 797-814.

[33] Clements, Michael P. and David F. Hendry (1999). *Forecasting Non-stationary Economic Time Series*, the MIT press, Cambridge.

[34] Cochrane, John H. (2005). *Time Series for Macroeconomics and Finance*, manuscript, available at `https://www.johnhcochrane.com/research-all/time-series-\\for-macroeconomics-and-finance`.

[35] Cournot, Augustin (1838). *Researches into the Mathematical Principles of the Theory of Wealth*, english edition published in 1897, MacMillan, New York.

[36] Crafts, Nicholas and Terence Mills (2017). Six centuries of British economic growth: a time-series perspective, *European Review of Economic History*, 21, pp. 141-58.

[37] Cuestas, Juan C. and Dean Garratt (2011). Is real GDP per capita a stationary process? Smooth transitions, nonlinear trends and unit root testing, *Empirical Economics*, 41, pp. 555-63.

[38] Dakos, Vasilis, Marten Scheffer, Egbert H. van Nes, Victor Brovkin, Vladimir Petoukhov, and Hermann Held (2008). Slowing down as an early warning signal for abrupt climate change, *The Proceedings of the National Academy of Sciences*, 105, no. 38, pp. 14308—14312.

[39] De Bonis, Riccardo and Andrea Silvestrini (2014). The Italian financial cycle: 1861–2011, *Cliometrica*, 8, pp. 301-334.

[40] de Jong, Robert M. and Neslihan Sakarya (2016). The econometrics of the Hodrick-Prescott filter, *The Review of Economics and Statistics*, 98(2), pp. 310–317.

[41] De Long, J. Bradford and Lawrence H. Summers (1988). How does macroeconomic policy affect output?, *Brookings Papers on Economic Activity*, vol 2, pp. 433-494.

[42] Dessavre, Adjani G., Emma Southall, Michael J. Tildesley, and Louise Dyson (2019). The problem of detrending when analysing potential indicators of disease elimination, *Journal of Theoretical Biology*, 481, pp. 183—193.

[43] 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, pp. 427-31.

[44] Dickey, David A. and Wayne A. Fuller (1979). Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica*, 49, pp. 1057-72.

[45] Drake, Leigh and Terence C. Mills (2010). Trends and cycles in Euro area real GDP, *Applied Economics*, 42, pp. 1397–1401.

[46] Durlauf, Steven N. and Peter C. B. Phillips (1988). Trends versus random walks in time series, *Econometrica*, 56 (6), pp. 1333-1354.

[47] Enders, W. (2010). *Applied Econometric Time Series*, 3rd ed., John Wiley & Sons, New Jersey.

[48] Enders, W. and Junsoo Lee (2012). A unit root test using a Fourier series to approximate smooth breaks, *Oxford Bulletin of Economics and Statistics*, 74 (4), pp. 574-599.

[49] Evans, Lewis T. and Neil C. Quigley (1995). What can univariate models tell us about Canadian economic growth 1870-1985, *Explorations in Economic History*, 32, pp. 236-52.

[50] Fernald, John, Robert Inklaar and Dimitrije Ruzic (2024). The productivity slowdown in advanced economies: common shocks or common trends? *Review of Income and Wealth*, forthcoming.

[51] Franses, Philip H. (1998). *Time Series Models for Business and Economic Forecasting*, Cambridge University Press, Cambridge.

[52] Frickey, Edwin (1934). The problem of secular trend, *The Review of Economics and Statistics*, 16 (10), pp. 199-206.

[53] Frickey, Edwin (1942). *Economic Fluctuations in the United States: a Systematic Analysis of Long- Run Trends and Business Cycles, 1866-1914*, Harvard University Press, Cambridge.

[54] Fritz, Marlon, Thomas Gries and Yuanhua Feng (2019). Growth trends and systematic patterns of booms and busts - testing 200 years of business cycle dynamics, *Oxford Bulletin of Economics and Statistics*, 81 (1), pp. 62-78.

[55] Fuller, Wayne A. (1996). *Introduction to Statistical Time Series*, 2nd. ed., John Wiley & Sons, Inc., New York.

[56] Gao, Jiti and Kim Hawthorne (2006). Semiparametric estimation and testing of the trend of temperature series, *Econometrics Journal*, 9, pp. 332-55.

[57] Gómez, Victor (1999). Three equivalent methods for filtering finite nonstationary time series, *Journal of Business & Economic Statistics*, 47 (1), pp. 106-16.

[58] Gómez, Victor (2001). The use of Butterworth filters for trend and cycle estimation in economic time series, *Journal of Business & Economic Statistics*, 19 (3), pp. 365-373.

[59] Granger, Clive W. J., Tomoo Inoue and Norman Morin (1997). Nonlinear stochastic trends, *Journal of Econometrics*, 8, pp. 65-92.

[60] Granger, Clive W. J. and Paul Newbold (1974). Spurious regressions in Econometrics, *Journal of Econometrics*, 2, pp. 111-120.

[61] Granger, Clive W. J. and Timo Teräsvirta (1993). *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.

[62] Grant, Angelia L. and Joshua C.C. Chan (2017). Reconciling output gaps: unobserved components model and Hodrick–Prescott filter, *Journal of Economic Dynamics & Control*, 75, pp. 114–121.

[63] Hafner, Sasha (2003). Trends in maize, rice, and wheat yields for 188 nations over the past 40 years: a prevalence of linear growth, *Agriculture, Ecosystems and Environment*, 97, pp. 275–283.

[64] Haldrup, Niels, Robinson Kruse, Timo Teräsvirta and Rasmus T. Varneskov (2013). Unit roots, non-linearities and structural breaks, *in*, Hashimzade, Nigar and Michael A. Thornton (eds), *Handbook of Research Methods and Applications in Empirical Macroeconomics*, Edward Elgar, Cheltenham, pp. 61-94.

[65] Hamed, Khaled H. and A. Ramachandra Rao (1998). A modified Mann-Kendall trend test for autocorrelated data, *Journal of Hydrology*, 204, pp. 182-196.

[66] Hamilton, James D. (1994). *Time Series Analysis*, Princeton University Press, Princeton.

[67] Hamilton, James D. (2018). Why you should never use the Hodrick-Prescott filer, *The Review of Economics and Statistics*, 100 (5), pp. 831-43.

[68] Han, Yang, Zehao Liu and Jun Ma (2020). Growth cycles and business cycles of the Chinese economy through the lens of the unobserved components model, *China Economic Review*, 63, pp. 1-28.

[69] Hansen, Bruce E. (2022). *Econometrics*, Princeton University Press, Princeton.

[70] Harvey, Andrew C. (1985). Trends and cycles in macroeconomic time series, *Journal of Business & Economic Statistics*, 3 (3), pp. 216-227.

[71] Harvey, Andrew C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge.

[72] Harvey, Andrew C. (1993). *Time Series Models*, 2nd ed., Harvester Wheatsheaf.

[73] Harvey, Andrew C. and Albert Jaeger (1993). Detrending, stylized facts and the business cycle, *Journal of Applied Econometrics*, 8 (3), pp. 231-247.

[74] Harvey, Andrew C. and Thomas M. Trimbur (2003). General model-based filters for extracting cycles and trends in economic time series, *The Review of Economics and Statistics*, 85(2), pp. 244-255.

[75] Harvey, David I., Stephen J. Leybourne, and A. M. Robert Taylor (2007). A simple, robust and powerful test of the trend hypothesis, *Journal of Econometrics*, 141, pp. 1302–1330.

[76] Harvey, David I., Stephen J. Leybourne, and A. M. Robert Taylor (2011). Testing for unit roots and the impact of quadratic trends, with an application to relative primary commodity prices, *Econometric Reviews*, 30 (5), pp. 514-547.

[77] Harvey, David I., Stephen J. Leybourne and Lisa Xiao (2010). Testing for nonlinear deterministic components when the order of integration is unknown, *Journal of Time Series Analysis*, 31, pp. 379-91.

[78] Hassler, Uwe (2000). Simple regressions with linear time trends, *Journal of Time Series Analysis*, 21 (1), pp. 27-32.

[79] Hatanaka, Michio (1996). *Time Series based Econometrics: Unit Roots and Cointegration*, Oxford University Press, Oxford.

[80] Hayashi, Fumio (2000). *Econometrics*, Princeton University Press, Princeton.

[81] He, Changli and Rickard Sandberg (2006). Dickey–Fuller type of tests against nonlinear dynamic models, *Oxford Bulletin of Economics and Statistics*, 68, Supplement, pp. 835-61.

[82] Higgins, Benjamim (1955). Interactions of Cycles and Trends, *The Economic Journal*, Vol. 65 (260), pp. 594-614.

[83] Hodrick, Robert J. (2020). An exploration of trend-cycle decomposition methodologies in simulated data, *NBER Working Paper 26750*.

[84] Hodrick, Robert J. and Edward C. Prescott (1981). Post-war U.S. business cycles: an empirical investigation, *Discussion Paper* 451, Southwestern University.

[85] Hodrick, Robert J. and Edward C. Prescott (1997). Post-war U.S. business cycles: an empirical investigation, *Journal of Money, Credit and Banking*, 29 (1), pp. 1-16.

[86] Hooker, Reginald H. (1901). Correlation of the Marriage-Rate with Trade, *Journal of the Royal Statistical Society* , Vol. 64, No. 3, pp. 485-492.

[87] Hooker, Reginald H. (1905). On the Correlation of Successive Observations, *Journal of the Royal Statistical Society*, Vol. 68, No. 4, pp. 696-703.

[88] Hsu, Chih-Chiang and Chung-Ming Kuan (2001). Distinguishing between trend-break models: method and empirical evidence, *Econometrics Journal*, 4, pp. 171–190.

[89] Hyndman, Rob J. (2009). Moving averages, manuscript.

[90] Inwood, Kris and Thanasis Stengos (1995). Segmented trend models of Canadian economic growth: rejoinder, *Explorations in Economic History*, vol. 32, pp. 253-61.

[91] Òscar, Jordà, Moritz Schularick and Alan M. Taylor (2017). Macrofinancial History and the New Business Cycle Facts, *in NBER Macroeconomics Annual 2016*, vol. 31, edited by Martin Eichenbaum and Jonathan A. Parker. Chicago: University of Chicago Press.

[92] Jevons, William S. (1878). *The Principles of Science*, vol. IV, MacMillan, London.

[93] Jönsson, Kristian (2020). Real-time US GDP gap properties using Hamilton's regression-based filter, *Empirical Economics*, 59, pp. 307–314.

[94] Kaldor, Nicholas (1957). A Model of Economic Growth, *The Economic Journal*, Vol. 67 (268), pp. 591-624.

[95] Kaldor, Nicholas (1961). Capital accumulation and economic growth, in Lutz, F. A. and Hague, D. C., *The Theory of Capital*, New York, St. Martin's Press.

[96] Kapetanios, George (2003). Threshold models for trended time series, *Empirical Economics*, 28, pp. 687–707.

[97] Kiefer, Nicholas M. and Timothy J. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests, *Econometric Theory*, 21, pp. 1130–1164.

[98] Kitchin, Joseph (1923). Cycles and Trends in Economic Factors, *The Review of Economics and Statistics* , 5 (1), pp. 10-16.

[99] Klein, Judy L. (1997). *Statistical Visions in Time: a history of time series analysis 1662-1938*, Cambridge University Press, Cambridge.

[100] Klein, Lawrence R. and Richard F. Kosobud (1961). Some Econometrics of Growth: Great Ratios of Economics, *The Quarterly Journal of Economics*, vol. 75 (2), pp. 173-198.

[101] Koopmans, Tjalling C. (1947). Measurement Without Theory, *The Review of Economics and Statistics*, 29 (3), pp. 161-172.

[102] Kuznets, Simon S. (1930). *Secular Movements in Production and Prices*, Houghton Mifflin, Boston.

[103] Kydland, Finn E. and Edward C. Prescott (1990). Business cycles: real facts and a monetary myth. *Federal Reserve Bank of Minneapolis Quarterly Review*, Spring 1990.

[104] Ladiray, Dominique, Gian Luigi Mazzi, and Fabio Sartori (2003). *Statistical methods for potential output estimation and cycle extraction*, Eurostat (European Comission).

[105] Lanne, Markku, Helmut Lütkepohl and Penti Saikkonen (2002). Comparison of unit root tests for time series with level shifts, *Journal of Time Series Analysis*, 23 (6), pp. 667-85.

[106] Lenton, T. M., V. N. Livina, V. Dakos, E. H. Van Nes and M. Scheffer (2012). Early warning of climate tipping points from critical slowing down: comparing methods to improve robustness, *Philosofical Transactions of the Royal Society A*, 370, pp. 1185—1204.

[107] Leser, Conrad E. (1961). A Simple Method of Trend Construction, *Journal of the Royal Statistical Society, Series B*, Vol. 23 (1), pp. 91-107.

[108] Leybourne, Sthephen, Paul Newbold and Dimitrios Vougas (1998). Unit roots and smooth transitions, *Journal of Time Series Analysis*, 19 (1), pp. 83-97.

[109] Li, Yi, Ning Yao and Henry W, Chau (2017). Influences of removing linear and nonlinear trends from climatic variables on temporal variations of annual reference crop evapotranspiration in Xinjiang, China, *Science of the Total Environment*, 592, 680–92.

[110] Lima, Luciano B., Francisco Cribari-Neto and Dilermando P. Lima-Junior (2022). Dynamic quantile regression for trend analysis of

streamflow time series, *River Research and Applications*, 38, pp. 1051-60.

[111] Lin, Chien-Fu, J. and Timo Teräsvirta (1994). Testing the constancy of regression parameters against continuous structural change, *Journal of Econometrics*, 62, pp. 211-28.

[112] Lopes, Artur S. (2015). *Raízes Unitárias — uma introdução*, Almedina, Coimbra.

[113] Lopes, Artur S. (2023). Non-convergent incomes with a new DF-Fourier test: most likely you go your way (and I'll go mine), *mimeo*, ULisboa.

[114] Lu, Ye and Adrian Pagan (2023). To Boost or Not to Boost? That is the Question, *CAMA Working Paper 12/2023*, Australian National University.

[115] Lucas, Robert E. (1980). Two illustrations of the quantity theory of money, *The American Economic Review* , vol. 70 (5), pp. 1005-1014.

[116] Maddala, G. S. and In-Moo Kim (1998). *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press, Cambridge.

[117] Mann, Henry B. (1945). Nonparametric tests against trend, *Econometrica*, 13 (3), pp. 245-259.

[118] March, Lucien (1905). Comparaison numérique de courbes statistiques, *Journal de la Société Statistique de Paris*, tome 46, pp. 255-277.

[119] Martin, Vance, Stan Hurn and David Harris (2013). *Econometric Modelling with Time Series*, Cambridge University Press, Cambridge.

[120] Miller, John P. and Paul Newbold (1995). Uncertainty about the persistence of economic shocks, *Journal of Business & Economic Statistics*, 13, pp. 435-440.

[121] Mills, Terence C. (2003). *Modelling trends and cycles in economic time series*, Palgrave MacMillan.

[122] Mills, Terence C. (2009). Modelling trends and cycles, *Cliometrica*, 3, pp. 221-244.

[123] Mills, Terence C. and Nicholas F. R. Crafts (1996). Trend growth in British industrial output, 1700–1913: a reappraisal, *Explorations in Economic History*, pp. 277-95.

[124] Mills, Terry C. and Patterson, Kerry D. (2015), Modelling the trend: the historical origins of some modern methods and ideas, *Journal of Economic Surveys* 29 (3), pp. 527–548.

[125] Mitchell, Wesley C. (1913). *Business Cycles*, University of California Press, Berkeley.

[126] Mitchell, Wesley C. (1927). *Business Cycles: The Problem and Its Setting*, NBER, New York.

[127] Moore, Henry L. (1914). *Economic Cycles: Their Law and Cause*, MacMillan, New York.

[128] Morgan, Mary S. (1990). *The History of Econometric Ideas*, Cambridge University Press, Cambridge.

[129] Morley, James and Jeremy Piger (2012). The asymmetric business cycle, *The Review of Economics and Statistics*, 94(1), pp. 208–221.

[130] Morley, James C., Charles R. Nelson and Eric Zivot (2003). Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different?, *The Review of Economics & Statistics*, 85 (2), pp. 235-43.

[131] Müller, Ulrich K., and Mark W. Watson (2008). Testing models of low-frequency variability, *Econometrica*, 76, pp. 979 - 1016.

[132] Müller, Ulrich K., and Mark W. Watson (2017). Low Frequency Econometrics, *in Advances in Economics and Econometrics, Eleven World Congress*, vol. II, Honoré, Bo, Ariel Pakos, Mnika Piazzezi and Larry Samuelson (eds.), Cambridge University Press.

[133] Müller, Ulrich, and Mark W. Watson (2024). Low-Frequency Analysis of Economic Time Series, draft chapter for *Handbook of Econometrics*, Volume 7, edited by S. Durlauf, L.P. Hansen, J.J. Heckman, and R. Matzkin.

[134] Murray, Christian J. (2003), Cyclical properties of Baxter-King filtered time series, *The Review of Economics and Statistics*, 2003, 85(2), pp. 472-76.

[135] Nachane, Dilip and Aditi Chaubal (2022). A comparative evaluation of some DSP filters vis-à-vis commonly used economic filters, *Journal of Quantitative Economics*, 20 (Suppl 1), pp. S161–S190

[136] Nakajima, Jouchi (2023). Estimating trend inflation in a regime-switching Phillips curve, Discussion Paper Series A No.750, Institute of Economic Research, Hitotsubashi University.

[137] Nelson, Charles R. (2008). The Beveridge-Nelson decomposition in retrospect and prospect, *Journal of Econometrics*, 146, pp. 202-206.

[138] Nelson, Charles R. and Charles I. Plosser (1982). Trends and random walks in macroeconomic time series: evidence and implications, *Journal of Monetary Economics*, 10, pp. 139-162.

[139] Nelson, Charles R. and Heejoon, Kang (1984). Pitfalls in the use of time as an explanatory variable in regression, *Journal of Business & Economic Statistics*, 2 (1), pp. 73-82.

[140] Nerlove, Marc, David M. Grether and José L. Carvallho (1997). *Analysis of Economic Time Series*, rev. ed., Academic Press, London.

[141] Neusser, Klaus (1999). An investigation into a non-linear stochastic trend model, *Empirical Economics*, 24, pp. 135-153.

[142] Ng, S. and Pierre Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power, *Econometrica*, 69, pp. 1519-54.

[143] Norton, John P. (1902). *Statistical Studies in the New York Money-Market*, MacMillan, New York.

[144] Oh, Kum H. and Eric Zivot (2006). The Clark model with correlated components. Unpublished manuscript, Department of Economics, University of Washington.

[145] Oh, Kum H., Eric Zivot and Drew Creal (2008). The relationship between the Beveridge-Nelson decomposition and other permanent-transitory decompositions that are popular in economics, *Journal of Econometrics*, 146, pp. 207-219.

[146] Osborn, Denise R. (1995). Moving Average Detrending and the Analysis of Business Cycles, *Oxford Bulletin of Economics and Statistics*, 57 (4), pp. 547-58.

[147] Papell, David H. and Ruxandra Prodan (2014). Long run time series tests of constant steady-state growth, *Economic Modelling*, 42, pp. 464-74.

[148] Pedregal, Diego J. (2022). Automatic identification and forecasting of structural unobserved components models with `UComp`, *Journal of Statistical Software*, Vol. 103 (9), pp. 1-33.

[149] Perron, Pierre (1989). The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, *Econometrica*, 57 (6), pp. 1361-1401.

[150] Perron, Pierre (2006). Dealing with structural breaks, *in*, Mills, Terence C. and Kerry Patterson, eds., *Palgrave Handbook of Econometrics*, vol. 1, Palgrave MacMillan, pp. 277-352.

[151] Perron, Pierre, Mototsugu Shintani and Tomoyoshi Yabu (2017). Testing for flexible nonlinear trends with an integrated or stationary noise component, *Oxford Bulletin of Economics and Statistics*, 79 (5), pp. 822-50.

[152] Perron, Pierre, Mototsugu Shintani and Tomoyoshi Yabu (2020). Trigonometric trend regressions of unknown frequencies with stationary or integrated noise, *mimeo*, Boston University.

[153] Perron, Pierre and Tomoyoshi Yabu (2009). Estimating deterministic trends with an integrated or stationary noise component, *Journal of Econometrics*, 151, pp. 56-69.

[154] Perron, Pierre and Tatsuma Wada (2009). Let's take a break: trends and cycles in US real GDP, *Journal of Monetary Economics*, 56, pp. 749–765.

[155] Persons, Warren M. (1919). Indices of business conditions, *The Review of Economics and Statistics*, vol. 1 (1), pp. 5-107.

[156] Phillips, P. C. B. (2005). Challenges of trending time series econometrics, *Mathematics and Computers in Simulation*, 68, pp. 401—416.

[157] Phillips, P. C. B. (2010a). Unit roots, *in* Durlauf, S. S. and Blume, L. E. (eds.), *Macroeconometrics and Time Series Analysis*, Palgrave MacMillan, pp. 347–68.

[158] Phillips, P. C. B. (2010b). The mysteries of trend, *Cowles Foundation Discussion Paper* no. 1771.

[159] Phillips, Peter C. B. and Sainan Jin (2021). Business cycles, trend elimination, and the HP filter, *International Economic Review*, Vol. 62 (2), pp. 469-520.

[160] Phillips, Peter C. B., Xiaohu Wang and Yonghui Zhang (2019). HAR testing for spurious regression in trend, *Econometrics*, 7 (50), pp 1-28.

[161] Phillips, Peter C. B. and Zhentao Shi (2021). Boosting: why you can use the HP filter, *International Economic Review*, Vol. 62 (2), pp. 521-570.

[162] Pollock, Stephen G. (1998). Trends in economic time series, unpublished manuscript.

[163] Pollock, Stephen G. (2011). *Topics in Econometrics*, unpublished manuscript.

[164] Pollock, Stephen G. (2013). Filtering macroeconomic data, *in* Handbook of Research Methods and Applications in Empirical Macroeconomics, Nigar Hashimzade and Michael A. Thornton (eds.), pp. 95-136, Edward Elgar, Cheltenham.

[165] Pollock, Stephen G. (2016). Econometric filters, *Computational Economics*, 48, pp. 669-91.

[166] Poynting, John H. (1884). A Comparison of the Fluctuations in the Price of Wheat and in the Cotton and Silk Imports into Great Britain, *Journal of the Statistical Society of London*, 47, 1, pp. 34-74.

[167] Quast, Josefine and Maik H. Wolters (2022). Reliable real-time output gap estimates based on a modified Hamilton filter, *Journal of Business & Economic Statistics*, 40 (1), pp. 152-168.

[168] Rappoport, Peter and Lucrezia Reichlin (1989). Segmented Trends and Non-Stationary Time Series, *The Economic Journal*, 99 (395), Supplement: Conference Papers, pp. 168-177.

[169] Rivas, María Dolores G. and Jesús Gonzalo (2020). Trends in distributional characteristics: existence of global warming, *Journal of Econometrics*, 214, pp. 153–174.

[170] Rogoff, Kenneth S., Barbara Rossi and Paul Schmelzing (2024). Long-run trends in long-maturity real rates 1311-2021, *American Economic Review*, 114, pp. 2271–2307.

[171] Rutkowska, Agnieszka (2015). Properties of the Cox–Stuart test for trend in application to hydrological series: the simulation study, *Communications in Statistics —– Simulation and Computation*, 44, pp. 565–579.

[172] Rybski, Diego and Armin Bunde (2009). On the detection of trends in long-term correlated records, *Physica A* 388, pp. 1687-695.

[173] Salamaliki, Paraskevi K. and Ioannis A. Venetis (2014). Smooth transition trends and labor force participation rates in the United States, *Empirical Economics*, 46, pp. 629–652.

[174] Schüler, Yves S. (2021). On the cyclical properties of Hamilton's regression filter, manuscript.

[175] Seater, John J. (1993). World temperature-trend uncertainties and their implications for economic policy, *Journal of Business & Economic Statistics*, 11 (3), pp. 265-277.

[176] Shapiro, Matthew D. and Mark W. Watson (1988). Sources of business cycles fluctuations, *NBER Macroeconomics Annual*, Vol. 3, pp. 111-156.

[177] Shi, Wanli, Xuezhong Yu, Wengen Liao, Ying Wang and Baozhen Jia (2013). Spatial and temporal variability of daily precipitation concentration in the Lancang river basin, China, *Journal of Hydrology*, pp. 197–207.

[178] Shook, John (2020). The Meaning of 'Secular' as a Scientific Concept. *Secularism and Nonreligion*, 9 (1), pp. 1-11.

[179] Smith, Bradford B. (1925). The Error in Eliminating Secular Trend and Seasonal Variation Before Correlating Time Series, *Journal of the American Statistical Association*, Vol. 20 (152), pp.543-545.

[180] Smith, Ron P. (2019). The Great Ratios in Economics: A retrospective, *Working Paper*.

[181] Sollis, Robert (2005). Evidence on purchasing power parity from univariate models: the case of smooth transition trend-stationarity, *Journal of Applied Econometrics*, 20, pp. 79-98.

[182] Stewart, Jon and Len Gill (1998). *Econometrics*, 2nd ed., Prentice Hall Europe, Hemel Hempstead.

[183] Stock, James H. (1994). Unit roots, structural breaks and trends, in *Handbook of Econometrics*, Volume IV, edited by R.F. Engle and D.L. McFadden, pp. 2739-2841.

[184] Stock, James H. and Mark W. Watson (1988). Variable trends in economic time series, *Journal of Economic Perspectives*, Vol. 2 (3), pp. 147–174.

[185] Stock, James H. and Mark W. Watson (2015). *Introduction to Econometrics*, updated third ed., Pearson Education Limited, Harlow.

[186] Teräsvirta, Timo (1994). Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association*, 89 (425), pp. 208-18.

[187] Teräsvirta, Timo (1998). Modeling economic relationships with smooth transition regressions, *in* A. Ullah and D. E. A. Giles (eds.), *Handbook of of Applied Economic Statistics*, Marcell Dekker, New York, pp. 507-52.

[188] Teräsvirta, Timo, Dag Tjøstheim and Clive W. J. Granger (2010). *Modelling Nonlinear Economic Time Series*, Oxford University Press, Oxford.

[189] Tödter, Karl-Heinz (2002). Exponential smoothing as an alternative to the Hodrick- Prescott filter?, *in* I. Klein and St. Mittnik (eds.), *Contributions to Modern Econometrics — From Data Analysis to Economic Policy*, pp. 223-237, Dordrecht, Boston, London, Kluwer Academic Publishers.

[190] van Dijk, Dick, Timo Teräsvirta and Philip Hans Franses (2002). Smooth transition autoregressive models – a survey of recent developments, *Econometric Reviews*, 21(1), pp. 1-47.

[191] Ventosa-Santaulária, Daniel and Manuel Gómez-Zaldívar (2010). Testing for a deterministic trend when there is evidence of a unit root, *Journal of Time Series Econometrics*, 2 (2), article 3.

[192] Vining, Rutlidge (1949). Koopmans on the choice of variables to be studied and the methods of measurement, *The Review of Economics and Statistics*, 31 (2), pp. 77-86.

[193] Vogelsang, Timothy J. (1998). Trend function hypothesis testing in the presence of serial correlation, *Econometrica*, 66 (1), pp. 123-48.

[194] Vougas, Dimitrios V. (2007). Is the trend in post-WW II US real GDP uncertain or non-linear? *Economics Letters*, 94, pp. 348-55.

[195] White, Halbert and Clive W. J. Granger (2011). Consideration of Trends in Time Series, *Journal of Time Series Econometrics*, article 2.

[196] Wu, Jilin (2016). A test for changing trends with monotonic power, *Economics Letters*, 141, pp. 15-19.

[197] Working, Holbrook (1934). A random-difference series for use in the analysis of time series, *Journal of the American Statistical Association*, 29 (185), pp. 11-24.

[198] Yamada, Hiroshi (2018). Why does the trend extracted by the Hodrick-Prescott filtering seems to be more plausible than the linear trend?, *Applied Economics Letters*, 25 (2), pp. 102-105.

[199] Yamada, Hiroshi (2020). A smoothing method that looks like the Hodrick-Prescott filter, *Econometric Theory*, 36, pp. 961–981.

[200] Yang, Yanga and Shaoping Wang (2017). Two simple tests of the trend hypothesis under time-varying variance, *Economics Letters*, 156, pp. 123–128.

[201] Yule, Udny G. (1921). On the time-correlation problem, with especial reference to the variate- difference correlation method, *Journal of the Royal Statistical Society*, 84 (4), pp. 497-537.

[202] Zarnowitz, Victor (1992). *Business Cycles: Theory, History, Indicators, and Forecasting*, University of Chicago Press.