

What can commonalities among time series teach us?

Citation for published version (APA):

Urbain, J. R. Y. J. (2007). *What can commonalities among time series teach us?* Maastricht University. <https://doi.org/10.26481/spe.20071011jpu>

Document status and date:

Published: 11/10/2007

DOI:

[10.26481/spe.20071011jpu](https://doi.org/10.26481/spe.20071011jpu)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

What can commonalities among time series teach us?

Colofon

Ontwerp en print: Océ Business Services, Maastricht

ISBN: 978-90-5681-268-3

NUR:780

Alle rechten voorbehouden. Niets uit deze uitgave mag worden veelevoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar gemaakt worden, zonder voorafgaande schriftelijke toestemming van de auteur of uitgever.

What can commonalities among time series teach us?

Oratie

In vrije vorm uitgesproken bij de aanvaarding van het ambt van hoogleraar Tijdreeksenconometrie aan de Faculteit der Economische Wetenschappen en Bedrijfskunde van de Universiteit Maastricht

op 11 oktober 2007

door

Jean-Pierre Urbain

Mijnheer de Rector Magnificus, Geachte Collega's, Dames en Heren,

In 2003, the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel was awarded to Professor Clive Granger and Professor Rob Engle for their contribution to the field of economics. The official announcement mentioned that Rob Engle was awarded half of the prize “for proposing methods of analyzing economic time series with time-varying volatility (ARCH)” and the other half was awarded to Clive Granger for having developed “methods of analyzing economic time series with common trends (cointegration)”.

For many people outside the academic world or without prior academic economics or statistics training, this was most likely the first time in their life that they heard these strange concepts of ARCH and cointegration. People like me, and I know this is true for many time series econometricians of my generation, viewed these awards like the recognition of the importance of the field that we, as young researcher in the late 80's, had been all modestly trying to understand and to which we secretly had hoped to contribute one day. I believe these developments now may be considered clearly as major revolutions in the analysis of economic data. Many of us were impatiently waiting for this to happen. I was fortunate enough to be a student, and later a young researcher, of a generation that grew up, academically speaking, in the late 80's very early 90's, during a period of an incredible intellectual excitement. Concepts, such as cointegration, were in their early development when I was a master student and PhD student. The original paper on cointegration had just been published and some of the most important developments were yet to come. The international conferences in economics and econometrics around the world had an uncountable number of sessions on these topics. Young researchers were discussing, collaborating and exchanging ideas about cointegration. To say that this revolution is one of the main reasons, in addition to meeting some extraordinary scientists, that motivated me to become an econometrician is certainly not an exaggeration.

20 years have passed since the publication of the seminal paper by Engle and Granger on cointegration¹ in the journal *Econometrica* as I am

¹ Engle, R.F. and C.W.J. Granger (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing, *Econometrica*, 55, 251-276

here today in front of you delivering this lecture for my appointment as Professor of Time Series Econometrics. During this lecture I will try to convince you that, even two decades later, what I will call searching for commonalities in economic data which is at the heart of cointegration analysis is a still fascinating research agenda.

I will definitely not cover all fields of time series econometrics and will deliberately remain silent on a number of research agendas that deserve to be discussed on their own, most notably all that has been done in the fields of volatility modelling, non-linear modelling, seasonality modelling to name just a few examples.

Econometrics, which makes use of tools from mathematics, probability theory and statistics to analyze economic reality has largely devoted its energy to building and developing techniques that are supposedly appropriate for economic data. The development of new techniques and methods are motivated by the continuous desire to provide tools that work in situation where existing tools, almost unanimously accepted in the profession and that were supposed to work well, are suddenly discovered to actually perform very poorly in situations that are likely to be important in reality. The history of econometrics is full of these revolutions.

It is certainly not my goal to review these here today, nor to rank any of these in terms of what I believe to be their importance. What I will attempt simply, besides hoping to convince you about the importance of time series econometrics as a sub-discipline of econometrics in itself, is to highlight the importance of a proper and deep understanding of these developments for empirical researchers in economics.

As the title of this lecture suggests, for the remainder of this lecture I will focus on the topic of commonalities, or equivalently similarities, between time series, illustrating with a few simple examples the concepts and ideas that have motivated my interest in this field. I will also try to convince you that the availability of rich new data sets offers new opportunities for time series econometricians to develop tools and techniques that may not only be of interest to economics but also to other sciences such as climatology. But before I go on, it is perhaps worthwhile to start at the beginning and explain what time series actually are.

A time series is simply an ordered sequence of observations on a variable of interest, usually observed at regular time intervals over a substantially period of time. The frequency of observation may be daily, weekly, monthly, quarterly, yearly or more recently even intra-daily if we think of financial data. Examples of such time series include national gross domestic product (GDP), the price level of a basket of goods, which is intended to represent an aggregate price level (consumer price index), stock market indices like the S&P 500 or the AEX index. Other noneconomic examples include the number violent crimes in a given city, the depth of the so-called eternal snow in the Alps measured daily, recordings of rainfall at a given location, chemical concentration in a river, the number of car accident in a given region, ... Figures 1 and 2 present some typical time series.

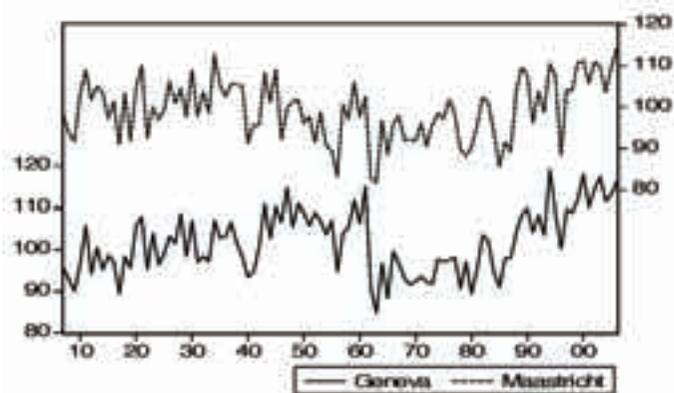


Figure 1: Average yearly temperature in Maastricht and Geneva

1907-2005

Why are time series data so different from other data? The answer is at first sight rather simple: unlike other types of data such as survey data or cross-section data, these data sets typically contain observations on a large number of individuals at a single point in time) there is a fundamental natural ordering in the time series data that we cannot change: *time*! Because there tends to be a relationship between the present, the past and the future reflected in these variables, the time ordering becomes an important piece of information for time series data.

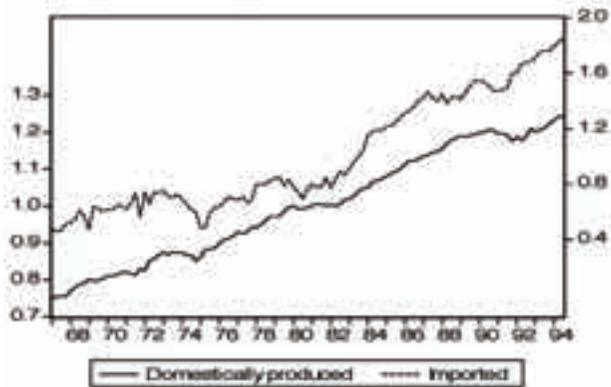


Figure 2: US Imported and domestically produced non-durable consumption goods

The second important characteristic of time series data is that these are almost always non-experimental and are made of records or measurements that represent the history of the phenomenon under study. In economics such a history is not reproducible in a laboratory, so we only have one observation of the series. Since we observe a single path, a single trajectory, the difficulties that we face in time series econometrics is that we need to develop and use tools that enable us to extract useful information from a single observation of a trajectory. Not surprisingly, this requires particular techniques and concepts.

The study of time series, either using graphics or using statistical methods is often aimed at trying to understand the past observed evolution through time with the potential goal of formulating hypotheses about the future evolution in the form of forecasts or policy predictions, or more generally confronting data evidence with predictions of economic theory.

Trends and cycles

As can be seen in Figure 2, economic time series data have a number of features which at first glance are easily visible graphically. One of the first features that comes visually to mind is the presence or absence of what we call a *trend*. Loosely speaking, a trend is the fact that the sequence of observations seems to go in a certain direction in the long run, over long periods. For example, the data may have a systematic tendency to increase, to decrease or more generally to show long smooth movements that need not be straight lines. Heuristically, the trend is thus what we identify as the direction of the *long run movements* in the data. The other features that are more or less easily observed graphically are abrupt changes, seasonal behaviour and periodic fluctuations as well as other movements around the trend. But visually, trends often seem to dominate the apparent behaviour of the data. The global movement in the data is most likely what the eyes immediately capture, relegating at a secondary level the other potentially informative movements of interest such as cycles that might be present in the data.

Identifying and isolating these features has always attracted a lot of interest in various fields of research, in economics most notably, but also outside economics. For example, the study of (average) temperature data over long periods for a given country is often used to assess the presence or absence of global warming. Researchers in climatology are struggling to find what is supposedly the best representation of the trend, and are struggling even more importantly on which techniques to best use to assess quantitatively the importance of the trend since differences of a few tenth of a degree Celsius might have dramatically different effects for some countries via the increase of the ocean and sea levels.

In economics, similar concerns have always attracted the interest of economists and motivated substantial research. In his review of the major developments in time series analysis, the Canadian econometrician Jean-Marie Dufour² notes that one can trace back to 1676 the first occurrence of the concept of economic cycle in William Petty's *Treatise of Taxes and Contributions*; while the first carefully drawn time series graphics in economics seems to have been proposed by the British economist

² All the dates and references are taken from Dufour, J.-M. (2006). *Historique de l'analyse des séries chronologiques*, miméo, Université de Montréal.

William Playfair in 1786 in his book *The Commercial and Political Atlas* which contained up to 44 figures of economic time series. Many further studies of cycles, in particular on the price of corn were conducted in the 19th century, while the opening of the 20th century saw a number of important concepts appearing, including the idea of decomposing a time series into a trend, seasonal variations, business cycle components and shocks. Interestingly enough, until 1934 and the work of Charles F. Roos, one of the founders of the Econometric Society, the standard practice of eliminating the trend components was never really questioned. This practice remained very popular until the early-mid 1980's.

From an economic point of view, isolating trend from cycles is also often considered to be of crucial importance in order to align what is being measured in the data at hand with what is supposedly being modelled by economic theory. A popular example in macroeconomics during the last few decades is the evaluation of models for business cycles, including the so-called real business cycle models that require the investigator to first isolate the cycle before being able to confront the theory with the data. From a purely economic point of view the decomposition, or more precisely the use of cyclical components alone, may nevertheless seem unsatisfactory since theories explaining only cycles cannot provide sufficient insights on the economic behaviour of how trend and cycles interact with one another.

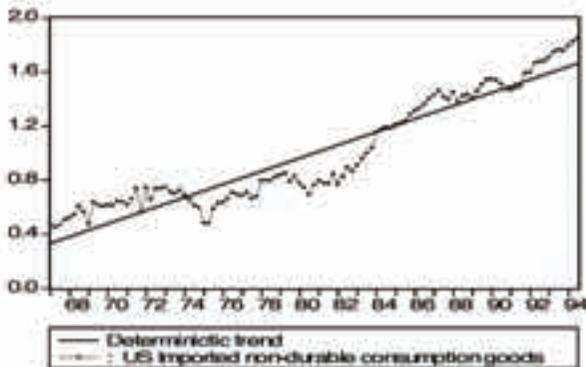


Figure 3

Historically, economists have focussed essentially on two competing representations of the trend. Until the 1980's, the standard in most empirical studies was to assume that the series had a constant growth rate. This is what we call a linear deterministic trend and is illustrated in Figure 3.

This way of looking at the world was challenged in the late 1970's when it was observed that the tendency to revert back to the trend is far from obvious and in fact sometimes never occurs and certainly does not happen at any regular frequency. The proposed alternative was a so-called stochastic trend assumption, whereby the apparent trend may be constantly changing and the growth rates may be rather unpredictable. At each point in time, the trend may increase by some fixed amount on average, but in any given period the change in the trend will deviate from its average by some unpredictable amount. This type of evolution is well known under the label of a random walk with drift, easily related to the way a drunken man walks. Each of his steps are unpredictable (up to the drift term), and the best guess we can make of where he will be at the next step is where he is now plus the constant drift. The fundamental difference between the deterministic and the stochastic trend assumption is that in the latter there is no tendency to return to a linear deterministic trend function.

It should be noted that the assumption of a stochastic trend is not merely a pure statistical assumption or simplification. A number of theoretical models in economics and finance imply under some hypotheses that a stochastic trend should exist. For example, some forms of the efficient market hypothesis or the so-called permanent income consumption model³ predict the presence of a stochastic trend.

³ Hall, R.E. (1978). Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence, *Journal of Political Economy*, 86, 971-987.

Recognizing that most macroeconomic variables seem to be of the stochastic trend type instead of the deterministic trend type has been at the source of a major revolution in econometrics over the last 25 years. There are different reasons for this. It was shown, first by simulations, later analytically, that the nature of the trend may have important implications for the validity or invalidity of several standard tools that were, and sometime still are, routinely used by empirical researchers. Also, it became apparent that the way cycles are calculated is potentially very sensitive to the definition of the trend. As there are a number of situations where it may be desirable to disentangle the trend from the cycles in the observed data, it is clear that a better understanding at least of the statistical nature of the trending behaviour in the data leads to a better understanding of what remains after the trend has been removed.

Commonalities among trending time series

If all series seem to be reasonable well described by the stochastic trend model, why do we have any interest in finding commonalities?

For time series analysis to be of any value for economics, it must be the case that the phenomena behind the time series that we observe are on the one hand sufficiently different so that each additional time series provide us with more information, but on the other hand sufficiently similar so that combining these can help us to understand the fundamental economic structure behind the movements. An outsider looking at economic time series data could almost legitimately question the need for the development of complicated statistical techniques and it seems reasonable to ask whether simply looking graphically at various data series through time is not sufficient to discover common behaviour. Unfortunately, we know since many years that one should be very careful in comparing different time series, as already pointed out by Yule in 1926. Why is this?

To try to explain without relying on any formal analysis, let me come back to the simple example of the drunken guy whose walking is perfectly represented by a random walk. As drinking alone is not the most exciting thing to do, let us imagine that he has been drinking in a pub where other customers have been engaging in similar activities. At the end of the night, two clients decide to go home and co-incidentally

move out of the pub at the same time. On their way back, to an outside observer, their manner of walking may look very similar, leading one to infer an apparent relationship between them. Nothing however will tie them together after a few hundred meters. They are both drunk, driven by alcohol, but nothing relates them to each other in any way beyond this.

The situation is different if in addition to being drunk they share a true and long-lasting friendship and decide to leave the pub together for a walk. Their individual trajectories may still wander apart from each other during some steps, but real friendship being ultimately always stronger than the effect of alcohol; they will never wander completely apart from each other. Either one of the two friends observes that he is deviating from his friend trajectory and he will (try to) adjust his own trajectory. Either both become concerned when seeing their friend getting too far away and consequently adjust their own trajectories to come back in the neighbourhood of their friend. In the former case the drunken man that is not adjusting may be viewed as the driving force of the observed movement; in the latter case both in one way or another contribute to the driving force. The common driving force, here the alcohol, is unobservable, but its effect is observable if we follow our two friends over a sufficiently long distance.

This unobserved force would be the *common random walk component*, or more precisely what we call *common stochastic trend* in the time series econometrics jargon. It represents the fundamental force behind the movement of the two series. As economic time series with a random walk component are often called integrated processes (of order one), the situation we face here is referred to as *cointegration*, highlighting that while individually the processes show a random walk behaviour, they will never wander apart from each other over a long period in an unbounded way due to the existence of a cointegration relationship that plays the role of an attractor, here their friendship.

Transposed into economics, our two wine or beer fanatics may become economic variables of interest and the friendship that ties them together is nothing but a long run equilibrium that may, or may not be predicted by economic theory. Macroeconomics, finance or international economic theory in particular are rich with theoretical results that imply that if the economic series under study have random walk like behaviour, then they should be tied together in the long run by equilibrium relationships, e.g.

they should be *cointegrated*. Demand for Money, private consumption, exchange rate (purchasing power parity), or present value models in finance are among the most frequently studied phenomena, but numerous other examples exist such as import demand, or labour market dynamics, where dynamic structural macroeconomic models may also imply cointegration as we have shown with David de la Croix and Franz Palm⁴.

The above mentioned example also allows a behavioural interpretation of the evolution of our two friends. Since at least one of two corrects his trajectory when observing that he is deviating from his buddy, we call this an error correction behaviour (or equilibrium correcting behaviour) and accordingly this can be modelled using linear (or possibly non-linear) error correction models, which have become extremely popular over the last 25 years.

At the mathematical and statistical level, the implications of such commonalities, in the form of common trends, are by now well understood. One of the most important implications is that cointegration ensured immunity against the spurious regression problem and therefore became a fundamental concept in econometrics.

Spurious regression or spurious correlation among time series is a phenomenon that was first mentioned⁵ by G. Udny Yule in 1926 when he observed that correlations were often useless when applied to time series. Transposed to our case of the two unrelated drunken guys, the idea is that being both driven by a random walk behaviour, to an outside observer they may spuriously give the impression of being related.

Modelling cointegration is now part of the standard toolkit of any empirical researcher dealing with time series data, thanks to the important theoretical work that was developed mainly by Clive Granger and Rob Engle when they were both in San Diego, by Søren Johansen from Copenhagen and Peter Phillips from Yale. In this world, concepts such as *causality* and *exogeneity* have been reconsidered as the

⁴ de la Croix, D. and J.P. Urbain (1998). Intertemporal Substitution in Import Demand and Habit Formation, *Journal of Applied Econometrics*, 13, 589-613. de la Croix, D., Palm, F.C. and J.-P. Urbain (2000). Labour Market Dynamics when Effort Depends on Wages Growth Comparison, *Empirical Economics*, 3, 393-419.

⁵ Yule, G.U. (1926). Why do we Sometimes get Nonsense Correlations between Time Series – A Study in Sampling and the Nature of Time Series, *Journal of the Royal Statistical Society*, 89, 1-69.

existence of cointegration implies fundamental links that tie the variables together. My early research and the research of my PhD thesis dealt with these issues. All this is now part of the standard econometric practice and has been extensively used by many international research institutes, National banks, ...

There might be a life besides cointegration

As I mentioned earlier on, while trends are the dominating feature of many time series, there is no reason why we should limit ourselves to the trending behaviour, and in particular to the commonalities regarding the trending behaviour of the data, even if these are representing the most fundamental forces that push the data in the long run. Observing common short run movement such as what we call *common cycles* may also lead to a better understanding of the evolution of the underlying economic phenomena and are in some cases also implied by economic theories such as the real business cycle model of King, Plosser and Rebello. In work with Franz Palm and Alain Hecq, that received the Christian Huygens prize in 2003 for his PhD on common cycles, we have been developing a number of tools and models⁶ that enable the analysis of such cyclical commonalities, leading in some cases to a better understanding of the dynamic interrelationships between economic time series.

Other commonalities could also be considered and have been in fact been considered by many colleagues here in The Netherlands and abroad, including common seasonality, common volatility -of particular importance in finance- , non-linearity, structural changes, ...

⁶ See for example Hecq, A., Palm, F.C; and J.-P. Urbain (2002). Separation, Weak Exogeneity and P-T Decomposition in Cointegrated VAR Systems with Common Features”, *Econometric Reviews*, 21, 273-307

“Similar” Commonalities in panel time series

Until now this lecture has focused on the importance of studying carefully commonalities (common-cycles or common trends) among a relatively small number of variables, 5 or 6, for let us say a single country, of a single phenomenon of economic interest.

During the last 10 to 15 year, we have seen the emergence and a substantial increase in the availability of a new type of data set that combines the standard time series dimension (observation of long periods) with what is usually called the cross sectional dimension. A typical example would be a data set with observations of private consumption and real disposable income over the period 1953-2006, not just for one country like The Netherlands but for a large number of countries including all OECD countries. One of the first historically important data sets of this type is the Summers and Heston data set. More recently, our colleagues from the Groningen Growth and Development Center have built a similar rich collection of data that covers many macro-economic aggregates as well as demographic data for up to 125 countries, observed over periods as long as 50 years. Similar data sets are encountered in finance where the number of units (firms or stocks) are even much larger. Outside of economics, environmental sciences and climatology are also rich with similar large data sets. These are what we generically call panel data sets. However, to distinguish from data sets where only a very few observations over time are available while the number of units (individuals, household, firms) are in the hundreds or even thousands, we call this panel time series, time series panels, or even macro-panels if the data are representing macroeconomic aggregates.

In practice, a key feature of techniques developed for macro panels is the emphasis on the dynamic properties of the data. This is in part due to the fact that macro panels tend to be longer in the time dimension, and one therefore has better ability to accommodate the dynamic features of the data. But it is also no doubt in part due to the fact that macroeconomic questions are often more explicitly oriented toward the dynamic properties of the data.

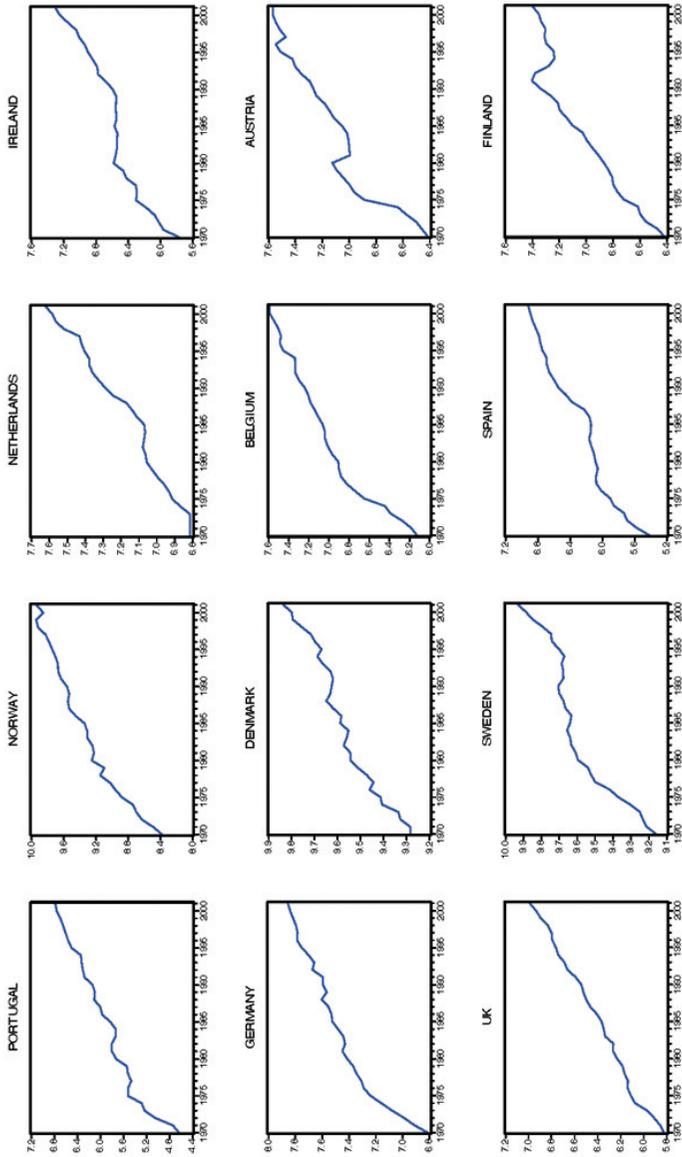


Figure 4: Health care expenditure in some OECD countries

Figure 4 illustrates the type of data set that underlies for example a number of panel studies on health care expenditures over the period 1970-2001. The first things that pops up for the observer is that many of the series seem to have again similar behaviour over time and a natural question is whether there is anything to be gained by studying all these countries together. Stated differently, the question is whether observations on more than one country are helpful to assess the validity of economic models or to understand the mechanics behind the data such as the nature of the trends, or the potential commonalities in the form of cointegration or common cycles.

Contrary to some initial belief, the answer is not a trivial “yes” and will depend on how “common” the commonalities are. The use of some hypothesized commonalities across the individual members of the panel is indeed a key feature of panel time series techniques that seek to provide an advantage relative to conventional individual time series techniques. Broadly speaking, in the context of this lecture, I will emphasize two basic minimal types of commonalities.

One corresponds to common properties (such as common cycles, cointegration, stochastic nature of the trend, ...) of the data for different members of the panel. The other corresponds to common properties of the question that is being asked for different members of the panel.

- *Common property, common commonality.* The most obvious approach is to consider whether different members of the panel share a common characteristic that describe the underlying unknown process that generates the observable data. For example, one might feel confident that the parameters describing the (steady state) relationship between aggregate prices and nominal exchange rates are the same among different countries, while the parameters describing the dynamics of adjustment to these steady states differ across countries
- *Common properties of the question that is being ask to the data.* Another way in which to think about the type of commonality that we often minimally require in the context of hypothesis testing pertains to the properties of the question that is being asked. For example, we may wish to exploit the fact that if a certain state

of the world (under the null hypothesis) is true for one member of the panel, then it must be true for all members of the panel. Here again, however, it is possible to exploit weaker restrictions on commonalities pertaining to the question of interest that does not require the answer to the question to be the same for all members of the panel. Intuitively, one can think of this as a condition that requires the answer for one member to have some bearing on the likelihood of the answer for another member of panel. Provided one of these two types of commonalities is considered to be present, then the additional dimension of the panel data adds new opportunities both theoretically and empirically.

It was for example shown in the mid-late 90's independently by Peter Pedroni in his PhD thesis at Columbia University and later by Peter Phillips and Roger Moon in a seminal theoretical⁷ paper, that under some conditions, that will turn out later to be of little relevance for many applications, the danger of spurious regression is as such not an issue anymore. Secondly, in an ideal world, the procedure based on these panels are also supposedly displaying much better properties in terms of being able to distinguish between alternative representation of the world (what is called power of tests statistics). If we add to these two first points the very rapid dissemination of some computer programmes that enabled applied researcher to implement recently developed techniques; then I believe that we have the main reasons of the rapidly growing popularity of these techniques.

In an ideal world, these “nice” results hold, but rely on assumptions that turned out to be of limited use for practical analysis. I'll come back to this issue in a few minutes. At the more fundamental level, one may regret that insufficient attention has sometimes been paid to the exact nature the questions that are being asked to the data, or at least to what the existing techniques imply. Indeed, the nature of the commonalities that are being exploited have an important bearing not only on the way in which information is pooled, but also on how one formulates hypotheses.

⁷ Phillips, P.C.B. and H.R. Moon (1999). Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica*, 67, 1057-1111.

Formulating hypotheses about similar commonalities

There are different options when formulating questions about the state of the world using a panel. To illustrate this point, let me consider the following situation. Time series analysis seems to suggest that many countries have a certain property *A* that is of economic interest. It seems hence logical to investigate if this property *A* is common to the considered set of countries. This is typically what is assumed in most applications of panel data econometrics. The question is thus whether the units (here the countries) in a panel *all* satisfy the property *A*. This property *A* may take numerous forms that are encountered in economic time series data or in times series data more generally. These may include the nature of the trending character (stochastic versus deterministic), the presence of cycles, the presence of abrupt changes (structural breaks), ... or any other features deemed important by the researcher for either an economic reason or more simply a statistical reason. While my own recent research has mainly focussed on the nature of the trend⁸ (deterministic versus stochastic) or existence of co-movements⁹ (cointegration, common cycles), the idea is more general and is potentially equally valid for other features of interest.

As usually the existence of this property is assessed by comparing two models of the world: one where all countries share properties, and another one where *none* of the countries share that property. In terms of what we call hypotheses to be tested, we have several solutions for formulating precisely the hypothesis that we want to test and each of these may lead to a different understanding of the world. Ideally this should depend on the degree of homogeneity or commonality that we want to assume under both hypothesized state of the world. Let me consider three different cases:

- Case 1. The hypothesis we want to test is the existence of a representation of the world (in statistics we call this our null hypothesis) where *all* of the countries have property *A* against another representation of the world (the alternative hypothesis) where *none* the countries share property *A*.

⁸ Gengenbach, C., Palm, F.C. and J.-P. Urbain (2007). Panel Unit Root Tests in the Presence of cross-sectional Dependencies: Comparison and Implications for Modelling, *METEOR Research Memorandum*, Universiteit Maastricht.

⁹ Hecq, A., Palm, F.C. and J.-P. Urbain (2000). Testing for Common Cyclical Features in Nonstationary Panel Data Models, *Advances in Econometrics*, 15, 131-160. Gengenbach, C., Palm, F.C. and J.-P. Urbain (2006). Cointegration Testing in Panels with Common Factors", *Oxford Bulletin of Economics and Statistics*, 6, 683-719

- Case 2. The hypothesis we want to test is the existence of a representation of the world where *all* of the countries have property A against another representation of the world where only *some* of the countries share property A.
- Case 3. The hypothesis we want to test is the existence of a representation of the world where *some* of the countries have property A against another representation of the world where *none* the countries share property A.

If the property A is for example the presence of common cyclical movements among the series for each countries (let us say output, investment and private consumption) then in Case 3 the null hypothesis tested will hold as long as some countries have common cycles and is tested against the alternative that none display common cycles. In Cases 2 and 3, the alternative world we consider is the negation of the word that we assume under the null. This is not the case for the first situation. Different approaches (e.g. statistical tests, techniques) use *different* formulations, but what we should be bear in the mind is the correct interpretation of the exercise, as these different approaches are testing different states of the world. Formulating and interpreting correctly the hypotheses that are being tested seems a minimal requirement for valid inference in practical empirical exercise if something is to be learned from the data.

“Common” commonalities in panels and cross-sectional dependence

As discussed above, the use of panel time series data requires a minimal amount of commonality. A limitation however of the standard techniques in panel data econometrics is the assumption that the members of the panels that are under study, while being assumed *similar* in order to satisfy the commonality requirement, are at the same time assumed to be independent.

For example, when we consider a cross-country panel for analyzing the nature of economic growth, standard techniques developed until a few years ago were assuming GDP growth in Germany and GDP growth in the Netherlands were almost unrelated. This rules out what we call *cross-sectional dependence*. You don't need to live in Maastricht or in

Aachen to see why this assumption is hard to maintain. While cross-section dependence can often be expected for household or firms from a given country; it is always present in cross-regional, cross-country and financial panel data sets. Recognizing the empirical relevance of the dependence has been at the source of important developments over the last three years. Things are not that easy though.

Let me take the following simple example where I'm analyzing yearly data on health care expenditure as a share of Gross Domestic Product for a set of 25 countries including the EU countries over a long period of 30 years. The main interest is the nature of the trend in health care expenditure. If countries are independent and satisfy the same type of trend model then the panel data provide us with an enriched information set that could be used to improve our knowledge of the trend. Schematically speaking we may then view the panel data set as carrying *25 times more* information on the nature of the trend. Health care expenditure of EU countries is however tied together as a consequence of geographical proximity, EU policy or more generally as a consequence of the complex interrelationships. It is consequently unlikely for the health care expenditure levels in the Netherlands to be totally independent of those in Germany and in Belgium, just to take one example that seems particularly valid here in Maastricht. In the face of such cross-sectional dependence the panel does not really carry 25 times more information; and the "amount" of extra additional information will depend on how many truly *independent* trends there are, something usually unknown and unobservable to the researcher but fundamental, as these trends are the fundamental sources that drive the data. From a theoretical point of view the implications of cross-sectional dependence have been studied extensively over the last few years. Many of the tools that had become so popular in the recent empirical literature lose most of their nice attractive properties as has been shown in recent studies, including some of my own recent work in this field with Christian Gengenbach and Franz Palm or with Joakim Westerlund.¹⁰

The solution is to model this dependence between the countries, but this turns out to be complicated. Indeed, in contrast to the time series

¹⁰ Gengenbach, C., Palm, F.C. and J.-P. Urbain (2006). Cointegration Testing in Panels with Common Factors", *Oxford Bulletin of Economics and Statistics*, 2006, 683-719
Urbain, J.-P. And J. Westerlund (2007). Spurious Regression in Nonstationary Panels with Cross-Unit Cointegration, *METEOR Research Memorandum*, Universiteit Maastricht.

dimension where the observations have a natural ordering (the time), there is no such ordering between the countries or more generally units in a panel, and more importantly economic theory is of almost no help in this case for assuming a priori a specific form for the dependence. One solution is to use geographical closeness or distance and assume that relationships between Belgium and The Netherlands are more intense than between The Netherlands and a remotely located country. This is what spatial econometrics tries to do and often leads to rather restrictive dependence structures.

Fortunately enough, recent work¹¹ by Jushan Bai and Serena Ng showed that it might be very convenient, given our total ignorance of the mechanics behind cross-sectional dependence, to adopt a so-called approximate factor model to address these issues. Factor models have a long history in economics and other social sciences. The idea is simple and elegant: it is to assume that a feature or a property observed at the level of a member of the panel (country, region, ...) may be due to a common source (a global common factor similar to the common trends that I mentioned before); to an idiosyncratic (country specific, region specific) source or to both. It has become one of the standards for studying large time panels and has been adopted in some of my own methodological research with Christian Gengenbach and Franz Palm where we propose to adopt this idea to study cointegration in panels with common factors. The advantage over time series cointegration analysis is here clear as this framework allows the researcher to have both common and idiosyncratic trends, or more general features, which in some applications may be particularly appealing.

Perhaps the best will be if I illustrate this idea using an example outside economics, where this framework could lead to insightful new results. This is the study of global climate change over a long period and a large number of meteorological stations. For this purpose I used a fairly rich data set on yearly average temperature in Europe.¹² The data covers

¹¹ Bai, J. and S. Ng (2004). A PANIC Attack on Unit Roots and Cointegration, *Econometrica*, 72, 1127-1177.

¹² These are preliminary illustrative results from an on-going research project on climate change that is conducted in collaboration with Joachim Freyberger who is currently a master student in econometrics in Maastricht. The data are taken from an update of Klein Tank, A.M.G. et al. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, *International Journal of Climatology*, 22, 1441-1453.

more than 40 meteorological stations spread all around Europe for which reliable data were available over the period from 1906 to 2005. A simple graphical analysis (confirmed by statistical calculations) shows that many of the series are slightly positively trending, certainly over the last decades. The crucial question is to quantify this trend and more interestingly to ask what part of the trend can be attributed to one or more “global” warming factors, a phenomena that is common to all locations, and what part can be labelled idiosyncratic and hence essentially due to local conditions and characteristics such as geographical location, such as closeness to the sea , altitude,

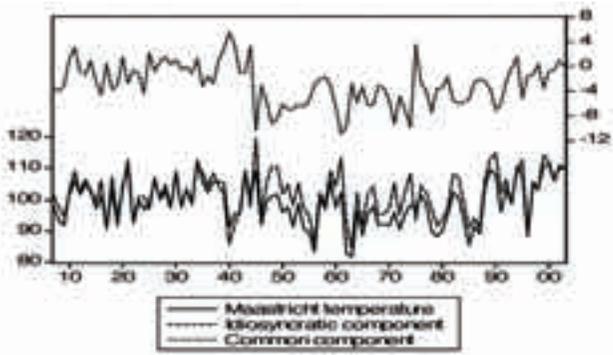


Figure 4: Decomposition of average yearly temperature in Maastricht

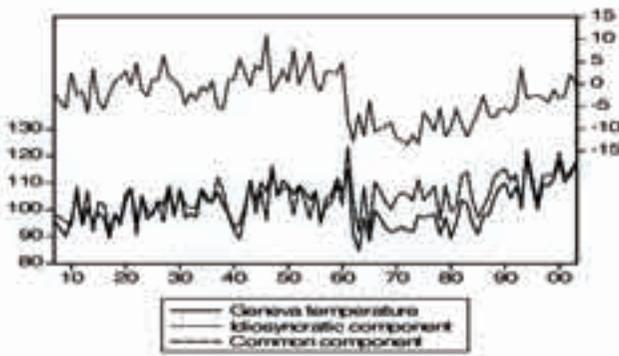


Figure 5: Decomposition of average yearly temperature in Geneva

Applying these statistical tools to the panel of temperature data clearly shows the existence of a single common factor over the period 1950-2006 that capture some, if not most, of the common trend in the data. Using this global common factor, it is then possible to decompose the individual temperature data into their idiosyncratic part, or component, and a part that is due to the common factor. Figures 4 and 5 present such decompositions for two particular locations namely Maastricht and Geneva, the unit being tenth of Celsius degrees.

A detailed discussion of the results is out of place here, but it is interesting to note that the analysis shows that for most of the locations considered in the panel, the decomposition highlights the common factor as the major source of the movement in the temperature series, reinforcing the idea of a *common* global warming factor. For many series the observed trends in the temperature are effectively captured by this global factor as can be seen even graphically from the Geneva temperature decomposition, where for example the idiosyncratic components only show a few tenths of a degree deviation from the common factor. Interestingly enough, further analyses also shows a significant change in the trend of the global common factor during the last part of the sample. I will however resist discussing these features any further at this time, since this examples is simply intended to illustrate the usefulness of these recently developed panel time series techniques for fields other than economics.

Future prospects

Looking back at the last 20 years of research in time series and dynamic econometrics I would like to draw a number of observations that might be useful for the future.

One first observes that the number of studies focussing on trends and cycles in economic data are almost “countably infinite”, and continuing to trend upward, to use a terminology familiar to some of you. But this does not imply that we know what trends actually are. Econometricians and economists should be modest enough and admit that while we know what they statistically mean, we still ignore what the fundamental trends really are, and how they could be explained by economic theory in a convincing manner. There are plenty of statistical models out there

for the observed “trend” in the data: linear or non-linear, deterministic and stochastic, slowly changing trends, trends with structural breaks, whether all these are rich enough to satisfactorily model and understand a number of changes observed in the data is an open question for the future, and an important challenge for the future in econometrics given the continuous rise of productivity and of economic growth observed in some parts of the world. This is true for univariate time series analyses, but also for panel time series where the economic interpretation of the common factor remains an open issue for many empirical applications.

It has become standard since some years in the profession to argue and claim, mistakenly I believe, that studying trends and the nature of trends in economics time series or time series panel data has become a non-interesting, even dead-end field. I would like to argue that on the contrary we probably need more than ever to understand what trends actually are. A few years ago Peter Phillips, from Yale, who has been instrumental in the development of this field, made a statement that I truly think is still of application nowadays. Commenting on the usual critique against unit root and related fields of research, he mentioned¹³: “we have only begun to mine this enormous subject and that the veins of most important development are yet ahead of us”.

The development of computer capabilities also offers new opportunities to econometricians. The increase in computer power has opened new opportunities using computer intensive techniques such as what we call bootstrapping and subsampling techniques, indirect inference, ... Provided these are not blindly used and applied, which seems unfortunately frequently the case in empirical research, these techniques may enable us to focus on some central issues in a way that is more robust to, or less affected by, the potential invalidity of the model chosen. Unfortunately, there is always another side to the coin, as these new computer intensive techniques can also easily be used to “elegantly disguise our ignorance”. There’s nothing fundamentally wrong with that. But more frequent careful attention to the theoretical foundations and validity of these tools for their applicability to practical empirical problems would be something welcome. Some of my recent

¹³ Phillips, P. (2001), Trending Time Series and Macroeconomic Activity : Some Present and Future Challenges, *Journal of Econometrics*, 100, 21-27.

more theoretical work¹⁴ with Franz Palm and Stephan Smeekes goes in this direction and seems promising.

Compared to the natural sciences, research in time series econometrics and in particular empirical research in economics using time series data gives substantial importance to probability and statistics based formal inference. Economic data are essentially non-experimental, measured with errors and often do not correspond to the concept that they are supposed to measure. Empirical research also often faces the lack of a single accepted theory describing economic phenomena in a dynamic and possibly time varying world. Other sciences such as atmospheric sciences, climatology, environmental sciences and to some extent epidemiology also share some of these characteristics. For example studies of weather and climate change, as I've tried to illustrate with simple examples, have very limited experimental capacity and also have to rely on non-experimental data. But compared to empirical research in economics these fields tend to have more data. As pointed out already in 1996 by Christopher Sims¹⁵, with the increasing interest in studying in depth the effects of gas emissions on global warming, climatology begins to come close to empirical research in economics, with on the one hand empirical studies of observed data and on the other hand an increasing number competing models that give potentially very different policy relevant predictions. I expect and hope in the future that more and more theoretical advances in time series econometrics will find their way in other fields where nonexperimental time series are studied, most notably climatology. We most likely all have something to gain.

Given all these circumstance, I'm truly convinced that the theoretical development of sound new methods, new concepts and new tools in time series econometrics are fundamental if we want empirical research to be helpful in understanding, or more modestly, describing the evolution of the world, and in evaluating the socio-economic and also environmental consequences of human economic behaviour. This doesn't make the life of a time series econometrician easier, on the contrary, but it surely makes it fascinating and exciting.

¹⁴ Palm, F.C., Smeekes, S. and J.-P. Urbain (2007). Bootstrap Unit Root Tests: comparison and extensions", *Journal of Time Series Analysis*, forthcoming.

¹⁵ Sims, C.A. (1996). Macroeconomics and Methodology, *Journal of Economic Perspectives*, 10, 105-120.

Words of thanks

Having succeeding in reaching the end of this lecture I would like to use this opportunity to thank all the people that have had a direct or indirect role in my appointment as full professor of Time Series Econometrics here at Maastricht University and that have participated, voluntarily or not, in my personal development that lead to this happening today.

Allereerst wil ik het *College van Bestuur* van de Universiteit Maastricht en iedereen die heeft bijgedragen aan mijn benoeming als Hoogleraar Tijdreeksenometrie bedanken voor het in mij gestelde vertrouwen. Ik ben u daarvoor zeer erkentelijk.

Hooggeleerde van Dijk, beste Herman,

Zonder jou zou ik hier vandaag nooit aanwezig zijn geweest. We hebben elkaar ontmoet in 1987 toen je “visiting professor” was in CORE, in Louvain La Neuve. Je hebt me toen aangenomen als scriptant. Een jaar later ben je mijn promotor geworden; en ik jouw allereerste AIO. Ik reisde toen regelmatig gedurende drie jaar van Luik naar Rotterdam met de trein om met jou een dag of een middag te kunnen praten, soms een beetje over vanalles, over allerlei ideeën die je in een of ander tijdschrift of op een conferentie was tegengekomen, maar meestal ging het toch wel over mijn eigen werk. Je hebt een fundamentele rol gespeeld in mijn ontwikkeling als onderzoeker: je hebt me mijn eigen weg laten volgen zonder van mij per se een Bayesiaan econometrist te willen maken. Van jouw motivatie voor onderzoek, jouw nieuwsgierigheid, creativiteit en fascinatie voor serieus theoretisch en empirisch onderzoek binnen de econometrie heb ik veel geleerd. Een wereldberoemde econometrist, ik zal zijn naam even niet noemen, heeft me een keer op een conferentie gezegd “*Herman, he is the nicest human being in the econometric business*”, dat had ik al lang gemerkt. Ik zal je nooit genoeg kunnen bedanken.

Beste collega's van de vakgroep Kwantitatieve Economie,

Het is nu al 15 jaar geleden dat ik hier als post-doc in Maastricht kwam. Als ik er nog steeds ben, met het plezier nu ook voorzitter van onze vakgroep te zijn, moeten er zeker goede redenen zijn. Jullie

allemaal zijn een van de belangrijkste redenen. De vakgroep KE is in mijn ogen een prachtig voorbeeld van een dynamische groep waar mensen van verschillende afkomst samen kunnen werken, praten, en nog belangrijker van elkaar genieten. Een speciale dank aan Stan die mij zo goed heeft voorbereid voor mijn huidige taak binnen de vakgroep. Ik wil ook een woord van dank zeggen aan Haydeé, Karin en Yolanda voor de bijzonder efficiënte en altijd plezierige secretariële ondersteuning en samenwerking. Merci beaucoup!

Als ik praat over goede redenen waarom ik nog steeds hier in Maastricht ben en nog lang zal blijven, dan ben ik natuurlijk de belangrijkste reden vergeten, die heeft een naam: Franz Palm.

Cher Franz,

Franz permets moi de t'adresser ces quelques mots en français, langue que nous pratiquons si souvent ensemble dans le cadre de nos très nombreuses discussions. Il m'est difficile d'exprimer à quel point je te suis reconnaissant, sur le plan intellectuel et professionnel naturellement, mais également sur le plan personnel. Tu m'as invité à rejoindre le département d'économie quantitative en 1992 après avoir été membre de mon jury de thèse de doctorat, et tu as depuis toujours été là à chacune des étapes de ma carrière académique, en tant que collègue, co-auteur et ami. Nous avons à ce jour collaboré à plus d'une dizaine d'articles, nous encadrons ensemble de nombreux étudiants, avons énormément de projets, j'espère que ce n'est toujours là qu'un début. Je voudrais te remercier pour toute la confiance que tu as placé en moi. Te fréquenter au quotidien m'a appris et m'apprends toujours énormément de choses.

Dear co-authors,

Doing research is everything but a lonely man journey and without you life at the university wouldn't be exciting and motivating as it is. I want to thank Luc Bauwens, Peter Boswijk, Bertrand Candelon, Albert Corhay, David de la Croix, Denis de Crombrughe, Herman van Dijk, Alain Hecq, Frank Kleibergen, Sébastien Laurent, Franz Palm, Joakim Westerlund, current co-authors and PhD students, Stephan Smeekes, Christian Gengenbach and Jeroen van den Berg: thanks to you all, joint work and PhD supervision are some of the most enjoyable aspects of

research that I can think of. A special thanks to Peter Pedroni for his comments on this lecture, for the endless discussions on methodological issues that we had over the years and for his incredible cellar.

Dear students,

Research is one part of my life here at the faculty, but as you know better than anyone else, teaching is another one. It is a constant pleasure to be engaged by all of you, to face your challenging and sometimes difficult questions. I thank you for all the motivation you show and for the incredible amount of time and energy you put in my courses, most notably the time series course that I teach at the graduate level.

Chers amis, chère famille,

Il y a heureusement une vie après l'économétrie, et la mienne a la grande chance d'être incroyablement riche grâce à vous avec qui il est si agréable de partager des passions bien éloignées de l'économétrie, mais ô combien indispensables pour mon équilibre. Merci à tous. Je voudrais également dire un mot spécial de remerciement à toute ma famille, ma mère, Bernadette, Joseph, ainsi que mon père qui nous a quittés il y a quelques temps. Sans vous tous je ne serais pas ce que je suis, je ne serais pas qui je suis.

Marina, Céline et Emilie,

Je ne sais pas si j'ai réussi aujourd'hui à vous faire comprendre ma passion pour cette drôle de chose qu'est l'économétrie des séries chronologiques, elle n'atteindra de toute façon jamais l'intensité de ce que vous représentez pour moi. Merci d'être comme vous êtes car c'est ce dont je suis le plus fier au monde.

Ik wil deze rede opdragen aan mijn moeder en aan de nagedachtenis van mijn vader.

Ik heb gezegd.

