Editorial

A Conversation with Katarina Juselius

Rocco Mosconi [1,*] and Paolo Paruolo [2]

1 Dipartimento di Ingegneria Gestionale, Politecnico di Milano, Via Lambruschini 4/B, 20156 Milano, Italy
2 European Commission, Joint Research Centre, Via E. Fermi 2749, 21027 Ispra, Italy;
  paolo.paruolo@ec.europa.eu
* Correspondence: rocco.mosconi@polimi.it

Abstract: This article was prepared for the Special Issue 'Celebrated Econometricians: Katarina Juselius and Søren Johansen' of *Econometrics*. It is based on material recorded on 30–31 October 2018 in Copenhagen. It explores Katarina Juselius' research, and discusses inter alia the following issues: equilibrium; short and long-run behaviour; common trends; adjustment; integral and proportional control mechanisms; model building and model comparison; breaks, crisis, learning; univariate versus multivariate modelling; mentoring and the gender gap in Econometrics.

Keywords: cointegration; CVAR; I(1); I(2); common trends; adjustment; breaks; model comparison; gender gap

JEL Classification: C32; B41; C01; C10; C30; C52

Introduction

On 30–31 October 2018 the authors sat down with Katarina Juselius in Copenhagen to discuss her contributions to Economics and Econometrics. Figure 1 shows photos of Katarina taken on that day; other recent photos are shown in Figure 2. The list of her publications can be found at https://www.Economics.ku.dk/staff/emeriti_kopi/?pure= en/persons/142900.[1]

In the following, frequent reference is made to Vector Autoregressive (VAR) models with Cointegration restrictions, labelled as CVAR, see Juselius (2006)–in particular Part II on the I(1) model and Part V on the I(2) model. In the rest of the article, questions are in bold and answers are in Roman. Text additions are reported between [ ] or in footnotes.

What do you think of micro-based versus macro-based Macroeconomic models?

I am sceptical to micro-based macro, partly because I have always been critical of the representative agent's approach. In the 1970s there appeared many excellent publications discussing its many unrealistic assumptions on aggregation. But for some reason, the criticism lost steam, the representative agent with rational expectations survived, and micro foundations of macro models became a "must" in academic work.

For me it is puzzling why the criticism did not have a greater impact on the mainstream theory considering that there are so many important aspects on the aggregate economy– such as unemployment, inflation, GDP growth, exchange rate, interest rate, inequality, speculation–that cannot be properly addressed in a representative agent's framework.

Considering all the major problems we face today both in the domestic and the international economy, it is obvious that more than ever we need an empirically well founded macro-based Macroeconomics. While Keynes already laid the foundations for such a theory, the old-fashioned Keynesianism needs of course to be modified to account for what we have learned about expectation formation based on imperfect knowledge/incomplete information, persistent equilibria, coordination failure and much more.

Figure 1. Katarina Juselius, 30 October 2018 in Copenhagen.

Joseph Stiglitz jointly with coauthors has proposed a new approach "disequilibrium Economics", which I think is a promising candidate for such a theory. Whether Stiglitz's disequilibrium Economics will change the direction of Economics is hard to predict, but based on previous experience perhaps one should not be too optimistic. When confronted with serious criticism, the Economics profession has too often responded with silence. For example, after the financial crisis, many methodologically oriented scholars, such as the editors and contributors of the Journal of Economic Methodology, were convinced the time for change had finally come, see Colander et al. (2009).

Numerous books were published addressing the mistakes and the misconceptions leading to the crisis, explaining why things went so wrong, and what could have been done instead. It might seem absurd, but the majority of the profession continued along the same path, modifying some assumptions here and there, but basically continuing as if the financial crisis was just a black swan.

I am often called "heterodox", "not an economist" or just ignored simply because I openly criticize mainstream models for relying on assumptions that do not describe the economic reality well enough. One of our prominent mathematical statisticians, Niels Keiding, once asked me "Why are economists not afraid of empirical data? Medical doctors sure are". While I had no really good answer, I know how hard it has been to raise a serious debate about the great divide between major empirical findings and standard mainstream assumptions, in spite of their important consequences for macroeconomic policy.

How can policy-makers learn about different policy options from a CVAR? Can CVAR answer policy questions?

I believe that policy-makers can primarily benefit from a CVAR analysis because it can improve our understanding of the dynamic transmission mechanisms of basic macroeconomic behaviour. For example, policy-makers facing a problem mostly think of one endogenous variable (the variable of interest) being pushed by a number of exogenous

variables. In practice, the assumed exogenous variables often exhibit strong feed-back effects from changes in the "endogenous" variables.

Figure 2. Katarina Juselius, 3 October 2016 in Milan.

The CVAR does not make use of the endogenous–exogenous dichotomy, but studies the economy as a system allowing for important feed-back effects in all equations of the system. Since policy-makers often are quite conservative in their economic beliefs, a well done CVAR analysis would highlight certain aspects of the model where such beliefs might be incorrect. Hence, a CVAR analysis could help policy makers avoid making bad decisions and, subsequently, to be criticized for them.

Another way a CVAR analysis can be useful is by learning from other countries' experience. For example, Finland, Sweden, and Japan experienced a housing bubble in the early nineties that resembled the more recent house price crisis in 2007. By applying a CVAR analysis to those countries–looking at the crisis mechanisms using the same perspective–policy-makers could have learned more about which policies are likely to work and which are not. They might even have been able to recognize the approaching crisis in time to prevent it.

The usefulness of addressing counter-factual policy questions with the CVAR might be more questionable. Judea Pearl would argue that a model like the CVAR is not appropriate, because the policy variables are set by the policy-maker and are not stochastically generated by the market as the VAR variables are assumed to be. Whether his–mostly theoretical–argument is empirically important is hard to say.

It is certainly the case that a policy variable like the federal funds rate is not behaving like a market determined stochastic variable. But at the same time, a CVAR analysis of the term structure of interest rates–inclusive the fed rate–seems to work reasonably well. But, perhaps one should be a little cautious with the conclusions in such as case.

What should we learn from crisis periods?

When the economy runs smoothly it doesn't matter much if you have a slightly wrong model, because things work anyway. When you are in a crisis period it matters a lot whether you correctly understand the economic mechanisms and how they work in the economy. The cost of wrong models can then be huge.

The question is, of course, whether it is at all possible to estimate economic mechanisms in a crisis period. For example in the official Danish macro model, the financial crisis is left out altogether with the motivation that it is too extreme to be analyzed econometrically.

I disagree. By experience I know it is possible to get plausible estimates over periods containing a serious crisis.

For example, I have used the CVAR model to address two very serious crisis periods: the house price crisis in Finland (Juselius and Juselius 2014) in the early nineties, and the more recent financial crisis in Greece (Juselius and Dimelis 2019). Both convinced me that it is possible to uncover the destructive forces that unfold during a crisis and that this would help policy-makers to mitigate the worst consequences of a crisis. So in principle I believe it would be a big mistake to leave out a crisis from the sample period.

People have sometimes asked me: "How can you use such periods, which are truly extraordinary, and then expect to find the mechanisms that apply in normal times". This is clearly a relevant question and I may not be able to provide more than a tentative answer: If the sample covers a crisis episode, then one usually needs to apply the I(2) model because it is explicitly specified to account for changes in equilibrium means and/or growth rates. In addition, it is specified to distinguish between levels, changes and acceleration rates, of which the latter is a key aspect of the crisis dynamics.

In normal periods, however, you will observe that acceleration rates are essentially zero. Hence, the acceleration rates take the role of crisis dummies in the I(2) model. However, it should also be acknowledged that the crisis mechanisms of the model may no longer be relevant after the crisis. For example, in the Greek analysis, the crucial crisis mechanism–the strong self-reinforcing mechanism between the bond rate and the unemployment rate–is likely to disappear or at least to change somewhat when the crisis is finally over.

But based on my experience, the main CVAR results seems to hold both for the pre- and post-crisis period. Perhaps the great variability of the data during a crisis period is also a good thing as it is likely to improve the precision of the estimates.

The I(2) model is related to the notion of integral and proportional control developed in the 1960s and 1970s. Can the I(2) analysis be useful for understanding the pronounced persistence away from long-run equilibrium relations?

It is hard to come up with any argument to why the I(2) analysis would not be useful, as the I(2) model is basically designed for integral control. However, the I(2) model is useful not just in integral control situations, but in a more general setting. If growth rates exhibit persistence and the levels move persistently around long-run trends, then the I(2) model should naturally be the preferred choice.

Also, the I(2) model has also a richer structure than the I(1) model and it has frequently given me insights that I would not have gotten otherwise. This has, in particular, been the case with house and stock prices and the effect of their persistent movements out of equilibrium on the aggregate economy. Without using the I(2) model, I do not think it would be possible to capture the complex mix of error-increasing and error-correcting behaviour in house and stock prices that ultimately led to the financial crisis.

It is quite interesting that economic time-series seem to have become increasingly persistent in the period following financial deregulation. At least I have often found that the I(2) model cannot be rejected based on the trace test for this period. I might have been naive, but I thought this would lead to a greater interest for I(2) applications. However, when you make a search of "I(2)" in the Economics literature, you do not find many papers. It is almost as if this fabulously rich model does not exist. But, of course, the I(2) model is more complex than the I(1) model, albeit not more difficult than many other models.

Perhaps, people stay away from the I(2) model because they think that unit roots do not make sense in economic data. And, of course, economic series cannot drift away forever as unit root processes in theory can. If one considers a unit root to be a structural economic parameter, then I agree that neither I(2) nor I(1) would make much sense. But, if one thinks, like I do, that unit roots are useful approximations that measure the degree of persistence of economic time series, then it makes a lot of sense.

I believe macroeconomists could do much better by exploiting the richness of the I(2) model, rather than just ignoring it.

Do you think the notion of near unit root is crucial for measuring persistence?

I certainly do because near unit root Econometrics provide some powerful tools that help us to uncover important mechanisms that have generated persistence in key economic time-series.

Take for example the unemployment rate, defined as a ratio between zero and one. Because of this, many economists would argue that it is a stationary variable and, hence, should not be modelled as a unit root process. Nevertheless, it is a very persistent near unit root variable for which the largest inverse root [henceforth simply referred to as root] of the autoregressive characteristic polynomial is typically larger than 0.95.

If you have a sample size of say 80 quarterly observations, you would often not be able to reject the null of a unit root in this case. Many empirical econometricians would, therefore, argue that the unit root approximation is fine as long as the unit root hypothesis cannot be rejected based on the conventional 5% rule. But the economist would nonetheless (correctly) argue that it is not a structural unit root.

If, instead, we have a sample of 3000 daily observations and an empirical root of 0.99, then this empirically large root is likely to be rejected as a unit root, even though the degree of persistence is much higher in this case. The 5% rule has the consequence that the larger the sample size, the easier it is to reject a unit root and vice versa. Hence, sticking to this rule implies that an econometrician would treat a persistent (0.9 root) variable as nonstationary and a persistent (0.99 root) variable as stationary, whereas an economist would argue that both are stationary independent of the test outcome. Not exactly a situation of clarity!

I hold the pragmatic view that if persistence–for example a long movement away from equilibrium–is an important empirical property of a variable or relation, then we should try to model that property. And one way of doing it is by classifying one's data and relations as I(0), near I(1) and near I(2) and relate them to short run, medium run and long run structures in the data.

For example, a powerful way to uncover the puzzling persistence in unemployment rates is to collect the relevant data and estimate the I(2) model, then find out which other variable(s) are cointegrated with the unemployment rate and how the adjustment takes place in the long, medium and the short run, and which the exogenous forces are. If competently done such a model analysis would help us to understand much more about unemployment persistence and its causes than a conventional model analysis. But near-unit-root Econometrics would probably require a lot more research to offer well worked out procedures for empirical modelling.

I have used this idea to understand the Phillips curve, which has been declared dead numerous times, but still seems to be how policy makers think about unemployment and inflation. The former looks very much like an I(1) series with a small but persistent drift, whereas the latter is almost stationary with a small and persistent drift. That the two series have a different order of persistence explains the lack of empirical support for Phillips curve: inflation rate, being a near I(1) variable, cannot cointegrate with unemployment rate, being a near I(2) variable. To recover the Phillips curve we need to add at least one more previously omitted (ceteris paribus) variable.

Edmund Phelps argued in his "Structural Slumps" book, see Phelps (1994), that the natural rate of unemployment, rather than a constant, is a function of the real interest rate–possibly also the real exchange rate. I found that the long persistent swings in unemployment rate were cancelled by cointegration with the long-term interest rate implying that they shared a similar persistence and that the residual was cointegrated with the inflation rate. Thus, by exploiting the persistence in the data it was possible to recover the Phillips curve with a Phelpsian natural rate (Juselius and Dimelis 2019; Juselius and Juselius 2014) and, in addition, to learn a lot more about the internal system dynamics and the exogenous forces that had pushed the unemployment rate and the interest rate out of their long-run equilibria.

This way of exploiting the data, I have sometimes called the Sherlock Holmes approach to empirical modelling. By following it you will find results that either support or reject

your priors but you will also find new unexpected results. If you do not sweep the puzzling results under the carpet, but let them rest in your mind, you may very well later come across some new results that put the old puzzles in a new light. These are moments of pure happiness.

At one stage it struck me that I almost always needed to add one or two additional variables to my hypothetical economic relations to achieve stationarity. A systematic feature usually means a common cause. In retrospect, it took me embarrassingly long to realize that the common cause was related to expectations in financial markets formed by imperfect knowledge/incomplete information. Subsequently, I have learnt how crucial the impact of complex feedback dynamics from the financial sector is on the real economy.

Should inflation and interest rates be treated as stationary, or as I(1) even if they have a long-run equilibrium value?

As I already discussed above, my view of empirical modelling is rather pragmatic, as it has to be because every realistic application is immensely demanding. It is always a struggle to make sense of macroeconomic data relative to the theory supposed to explain it. In this struggle the "perfect" or the "true" easily becomes the enemy of the "good". This applies for sure to the modelling of inflation and interest rates: both of them are crucial for the economy and none of them obey mainstream economic theory.

Like unemployment rate, interest rates can be assumed to be bounded from below by zero (or that is what we previously thought) and from above by some upper limit. Inflation is not necessarily bounded but central banks usually do whatever they can to make it so. Whatever the case, both of them are persistent but differ in degree.

Inflation rates look more like a near I(1) process, whereas interest rates move in long persistent near I(2) swings around something that could possible be interpreted as a long-run equilibrium value. The question is of course what "long run equilibrium" means if economic relationships do not remain stable over long periods of time. For example, the period before and after financial deregulation describe two completely different regimes. Few equilibrium means remain constant across these two periods.

What is important in my view is that the inverse roots of the characteristic polynomial associated with nominal interest rates often contain a double (near) unit root–or rather one unit root and one near unit root. No theoretical prior would predict such an empirical finding and based on the conventional specific-to-general approach one would probably have swept this puzzling persistence under the carpet. But based on the general-to-specific approach it has been possible to suggest a coherent narrative in which a crucial element is financial market expectations based on imperfect knowledge (Juselius and Stillwagon 2018).

Because the stochastic trend in inflation is persistent of a lower degree than nominal interest rates, one would typically not find cointegration in a bivariate model of inflation and one interest rate and one would have to add at least one more variable. It turns out that by combining inflation with the spread between a short and a long interest rate one usually finds cointegration. This is because the long persistent swing in nominal interest rates are annihilated in the spread which then is cointegrated with the inflation rate. A plausible interpretation is that inflation is cointegrated with expected inflation measured by the spread.

The similarity to the Phillips curve model is quite striking: there we had first to combine unemployment rate with the (long term) interest rate to obtain a stationary cointegration relation for inflation. Thus, the long term interest rate needs to be cointegrated with either the unemployment rate or the short term interest rate to get rid of the persistent swings so that what is left can cointegrate with inflation rate.

Whatever the case, the real interest rate is generally too persistent to be considered stationary even though it is claimed to be so in many empirical papers. Such a claim is often based on a badly specified empirical model and, hence, a sizeable residual error variance that makes statistical testing inefficient. To me, such an analysis represents just a missed opportunity to learn something new.

The CVAR model has parameters related to long-run relations and to short-term adjustment. Other approaches in the literature focus only on the long-term relations, like for example Fully Modified Least Squares. What do you think is the advantage of CVAR over other approaches essentially focusing on the long-run relations?

Believing that a long run relation without the corresponding short-run dynamics is sufficient for understanding an economic problem is like thinking you only need your feet but not your eyes to get to your destination. In a certain sense, bivariate cointegration in a non-stationary world corresponds to correlation in a stationary world. It tells you there is a causal relationship but you need the short-run dynamics to understand the causal links.

There are numerous examples in my publications where it was the short-run dynamics that made me rethink the causal mechanisms of the economic problem. The application to the monetary transmission mechanisms in Denmark–discussed in every detail in my cointegration book (Juselius 2006)–is a good example. I found a stable, completely plausible money demand relation consistent with economic theory, but the estimated short-run dynamics contradicted that theory on essentially all counts. I then spent 10–20 years to understand why.

Another example of why the dynamic adjustment is so crucial is an empirical application to climate change which was done in collaboration with Robert Kaufmann [Robert hereafter], from Boston University (Kaufmann and Juselius 2016). I met Robert at a climate conference in Italy where he introduced me to a fascinating climate data set obtained from ice drilling in Vostok. The data base contained ten climate variables (among others: surface temperature, sea temperature, sea levels, CO_2, methane) measured over 400,000 years and based on a frequency of one observation per 1000 years. The dominant feature over this period is the regular occurrence of glacial cycles.

The well known Milankovitch theory was able to associate them with orbital variations such as precession, eccentricity and obliquity of the Earth relative to the Sun. However, the power to explain the glacial cycles with these measures was rather poor and most of the variation in temperature over the glacial cycles remained unexplained. Our purpose was to do better based on a CVAR for 10 climate variables conditional on the Milankovitch orbital variables.

The cointegration results showed that the long-run forces were important, but not as important as the dynamic feedback effects which were able to explain most of the variability. Our conclusion was that if you disregard the short-term adjustment, you will only explain a small part of the glacial cycle phenomena, whereas if you include the short-run feedback you can do much better.

Of course, this type of data and models are not meant to predict what will happen next year, but to learn something about the physics of climate change. For example, our results showed that CO_2 was the major determinant of surface temperature in this long period without anthropogenic effects on the climate. The results also showed that the CO_2 effect was strengthened by the feedback dynamics. This became strikingly evident when we estimated the effect on temperature from doubling CO_2 in the atmosphere.

While most climate models would predict that a doubling leads to an increase of roughly 3° or 4° Celsius, our CVAR model predicted an increase of almost 11° Celsius. That's a difference between something which is a disaster and the end of our civilization. But every time climate scientists update their models it is a little scary to learn that the new version shows that the previous one had again underestimated the effect of CO_2 on our climate.

Did you find I(2) behaviour in that case?

The data display long persistent cycles, between 80,000 and 100,000 years long, which showed up in the model as quite large complex pairs of inverse characteristic roots. The trace test, however, rejected I(2). I believe this was partly because the CVAR is not (yet) designed to handle large cyclical roots close to the unit circle, partly because of the compression of data into 1000 year averages. I would think that if instead we had access to

100 year observations there would be strong evidence of I(2). This is actually something I still would be keen on studying.

Your difficulties in publishing the result on CO_2 seem to suggest that journals and academia in general are somewhat conservative

"Somewhat conservative" is clearly an understatement. But, science is conservative and for good reasons. One should not jump away from an established path at every whim. What is harder to accept is the stubborn conservatism that is more about protecting one's theoretical stance. I always thought Economics was exceptionally conservative, partly because of its axiomatic foundation which makes it less prone to listen to empirical arguments.

The difficulties with getting our CVAR results published in climate journals suggest that it can also be difficult in physical sciences. I guess that the CVAR methodology may seem difficult and strange the first time you come across it. By now Climate Econometrics has become much more established and it is probably easier to publish papers today using cointegration techniques.

How can the CVAR methodology affect the learning process in economics?

If adequately done, the CVAR structures the data in economically relevant directions without imposing theory-consistent restrictions on the data prior to testing. By this you give the data the right to speak freely about the underlying mechanisms rather than to force them to speak your favourite story. Macro-data are quite fragile–one realization at each time *t* from the underlying process–and if you torture them enough they will usually confess. This, I believe, may partly explain the confirmation bias that seems quite prevalent in empirical Economics and which is not how to bring about innovation in learning.

The conventional specific-to-general approach starts with a theory model derived from some basic assumptions which are seldom tested. One example is the assumption of what is endogenous and what is exogenous in the model. Another is the assumption that omitted *ceteris paribus* variables do not significantly change the obtained results. Both of them tend to be rejected when tested within the CVAR model, and both of them tend to affect the conclusions in a very significant way. If the basic hypotheses are not correct, then the scientific value of the whole modelling analysis is of course questionable, because then it would be impossible to know which results are true empirical findings and which are just reflecting the incorrectly imposed restrictions.

Another example is the assumption of long-run price homogeneity which is an implicit assumption of most economic models. Central banks are mandated only to control CPI inflation, which makes sense under long-run price homogeneity. But over the last 30–40 years, long-run price homogeneity between CPI prices, house prices and stock prices has consistently been rejected due to the fact that stock prices and house prices have behaved completely differently from CPI prices. Central banks have focused primarily on CPI inflation, and by doing so, contributed to the devastating house and stock price bubbles and a steadily growing inequality in our societies.

I believe these problems could have been avoided if more attention had been paid to the signals in the data which were strong and clear after the financial deregulation in the eighties. But academic professors and policy makers were looking at the data through lenses colored by conventional theory, such as efficient markets, rational expectations and representative agents. Inconsistencies with data evidence were labeled theoretical puzzles and had no consequence for practical policy.

What can we do to change the status quo?

The question is of course if it is at all possible for empirical Econometrics to break the monopoly of theoretical Economics. While I do not have an answer to this question, I can at least refer to discussions I have had with other scholars.

One possibility is to make use of competitions like in other areas such as architecture. For example, if the government wants to build an opera house, they announce a

competition and whoever has the best project will win the competition. Similarly, if the government want to understand what the mechanisms are behind the soaring house and stock prices in order to avoid a new crisis, they could announce a competition. The team that most convincingly is able to explain past and present crisis mechanisms should win the competition. Of course, I can think of many relevant objections to such competitions, but in any case it might be an important step to bring Macroeconomics closer to empirical reality.

I have discussed these issues many times with David Colander [Dave hereafter], one of the most innovative persons I have ever met. Some years ago he presented a proposal for how to reform university teaching of Economics based on a research oriented line and a more applied line. As the majority of students end up working for governments, research institutes, or institutions like the IMF, the ultimate aim was to offer a better training in how to solve real world problems.

On a practical level, one of Dave's suggestions was a big database into which the government as well as other public and private institutions could upload problems they wanted to be solved. University professors would then be allowed to pick problems related to their area of expertise, work out a proposal for how a research group of professors and students would address the problem, and submit the application to the relevant agency. This would have the advantage of bringing important problems closer to the university and would train students to solve real problems under qualified guidance. I should mention that the above is only a small part of his elaborate proposal which was then available as a written memo.

Is empirical research in Economics different from Physical Sciences? Do you think that changing the theories starting from evidence in the data is easier there?

Physical sciences tend to agree, to a larger extent than Economics, upon common rules based on which the profession is willing to accept results as being scientifically valid. But when this is said not everyone in physics agrees. For example, when I sometimes discuss the difficulties in social sciences with my son, who is a physicist, he argues that it's more or less the same in his field.

I believe there is a difference in grade in the sense that physical laws are laws in a much stricter sense. Once they have been established, after being suitably tested, they are hard to challenge, whereas economic laws are not "laws" in the same sense, they are much more mental inventions. Hence, one would think that the scientific community would be more willing to modify or change basic assumptions when they appear incompatible with reality.

In your applied research you address different problems: how do you select your research topics?

The short answer is that my research topics are forced on me by the many "why"s I stumble over in my CVAR analyses. This process started already with my first real economy application to the Danish money demand problem in the late eighties. I was fortunate to find empirical support for a stable, plausible money demand relation.

This was something I was really happy about, but there were other puzzling "why"s associated with the adjustment dynamics. So I decided to study German monetary transmission mechanisms hoping find an answer to my "why"s there. Some of the German results seemed to provide at least partial answers, but then they led to a whole bunch of new "why"s, which I subsequently tried to answer by studying monetary mechanisms in Italy and Spain.

As I was not able to satisfactorily solve the puzzling why's, I turned my attention on the international monetary transmission mechanisms where the purchasing power parity (PPP) and uncovered interest rate parity (UIP) provide the cornerstones. Again, some of the results made sense theoretically, but others raised new "why"s. The most important finding was that PPP needed the UIP to become stationary, indicating that they were inherently tied together.

Michael Goldberg stumbled over my first Journal of Econometrics paper discussing this and told me the results were exactly in accordance with the theory of imperfect knowledge based expectations he and Roman Frydman had worked out. It then dawned on me that many of my "why"s probably had to do with such expectations in financial markets and how they affected the real economy.

Two of the most important variables in the macro economy are the real interest rate and the real exchange rate and both of them exhibited this puzzling persistence. The idea that it was this persistence which had caused the puzzling persistence in unemployment rates suddenly struck me. This was a very important breakthrough in my research. From this stage onwards, I knew the direction.

Another example is a study of foreign aid effectiveness based on 36 African countries, which was commissioned by the UN-WIDER institute. Initially it involved one of my PhD students, but then the project grew and I also became actively involved. As it turned out, among those 36 countries, a few important ones, Tanzania and Ghana, were sticking out in a way that prompted many new "why"s. We picked them out for a much more detailed analysis which subsequently became another research publication. It is the trying to answer the "why"s of one paper that has often led to new papers.

Let's now discuss model building strategy. can you discuss the role of the deterministic components in the cointegrating vectors? how can structural breaks be distinguished from unit roots?

When I start a new project, I always spend a lot of time examining the graphical display of the relevant data. The first step is to examine the variables in levels and differences searching for features which stick out, such as a change in growth rate or a shift in the level of a variable. At this stage I also check the national economic calendar to identify the time points of major political reforms and interventions, because, in my view, an empirical analysis of a macroeconomic problem is always about combining economic theory with a institutional knowledge.

If I spot a sudden shift in the level of a variable followed by a blip in its difference and it coincides with a known political reform, I will add a shift dummy in the cointegration relations and an impulse dummy in the equations. In the final model I always check whether such a shift dummy is long-run excludable and whether the impulse dummy is statistically significant. The testing is important because a political reform often causes a shift in the equilibrium level of several variables so that the level shift may cancel in the cointegration relations.

While it is good scientific practice to test a prior hypothesis that a break has taken place at a certain known point in time, it is harder to defend a practice where step dummies are added only to be able to accept the stationarity of the variable. For example, as already discussed, unemployment is often found to be a very persistent process with a double near unit root. The trace test frequently concludes that it is not statistically different from an I(2) process, which can be a problem for a researcher believing it should be stationary.

By introducing sufficiently many deterministic level shifts so that stationarity around the level shifts can be accepted one might be able to solve the dilemma. But, whether you model the variable stochastically with the I(2) model or deterministically with many level shifts, you still need to address the puzzling persistence. I would clearly prefer to model it stochastically unless the breaks coincide with known policy reforms. To introduce breaks for the sole purpose of avoiding the I(1) or the I(2) model is not a good practice.

What about non-normality and dummies?

To assume Gaussian distributions is tempting, because then you have access to a very large tool box. And, because it is extremely demanding to adequately model macroeconomic time-series, you need as many tools as possible. This is because the series are often short, strongly autocorrelated, and subject to regime changes. In addition, macro models have to address path-dependencies, interrelated equations and aggregate behaviour that is typically different in the short, medium and long run. On top of all this inference is based

on a sample where you have just one observation at each time t from an underlying process which seldom is stable over extended periods of time. It is almost a miracle that the VAR model frequently is able to give a satisfactory summary of all this.

However, the assumption that the system is being hit by white noise shocks that cumulate via the dynamics of the process to generate the exogenous trends is a bold one, and an assumption that often needs to be modified.

Empirically, the VAR model is subject to many choices: we choose to study p variables among all the potentially relevant ones and we choose to cut the lag length at a not too large value k. In practice, normality is seldom accepted in the first unrestricted version of the VAR model. This is of course no surprise, as the residuals are not really estimates of white noise errors, but instead a summary of everything that has been left out of the model.

The effect of omitted variables can to some extent be accounted for by the VAR dynamics. But the effect of policy interventions and reforms are usually part of the residuals. Fortunately, policy events are numerous and their individual effect on the aggregated economy is mostly tiny. Hence, one can use the central limit theorem to justify the normality assumption.

The problem is that the effect of some of the policy events is far from small. For example, financial deregulation had an enormous effect on the economy, value added tax reforms exhibited also a very significant effect. The effect of other extraordinary events such as hurricanes, floods, fires, will often stick out as non-normal residuals. Such extraordinary effects have to be properly controlled for using dummies, or they will bias the VAR estimates. This is because the model will otherwise try to force these big effects onto the x variables.

I usually add dummies one at the time. First the ones I believe have to be there as they are a proxy for a real known event. Then I may add a few more if it is absolutely necessary to achieve residual normality or symmetry. Adding too many dummy variables to the model is generally not a good strategy as large effects are also very informative and dummying them out may destroy the explanatory power of your model.

The graphical display may also show transitory blips in the differenced series, that is a big blip followed by a blip of similar size but of opposite sign. They are typically the consequence of a mistake, sometimes a typing mistake, but mostly a reaction to a market misconception. For example, financial markets often bid up the price of an asset only to realize it was a mistake and the price drops back next period. But because they are symmetrical they affect excess kurtosis and not skewness, which is less serious. I often just leave them as they are. But if the jumps are huge, I usually control for them by a transitory impulse dummy $(\ldots 0, 0, +1, -1, 0, 0\ldots)$.

How do you interpret the results of the trace test? How strictly do you use the 5% critical value in testing hypotheses?

Some people think that a "rigorous approach" to testing requires a strict adherence to standard rules (such as the 5% critical value). I have never been an advocate of the 5% rule, but have always based my choice on the whole range of empirical p-values. The 5% rule is reasonable when you strongly believe in the null hypothesis and, hence, are not willing to give it up unless there is massive evidence against it. Adhering to the 5% rule is particularly problematic in situations when the econometric null hypothesis does not coincide with the economic null.

The trace test of cointegration rank is a good example. The standard procedure relies on a sequence of tests where you start in the top testing the econometric null hypothesis "p unit roots, that is no cointegration". But this null seldom corresponds to the economic null as it would imply that your preferred economic model has no long run content. If the first null hypothesis is rejected, then you continue until the first time $p - r$ unit roots cannot be rejected. This means that the test procedure is essentially based on the principle of "no prior economic knowledge" regarding the the number of exogenous trends. This is often difficult to justify.

The econometric null is based on the number of unit roots (a simple hypothesis) and a 5% rule applied to a top-down series of tests will often favour the choice of too many common trends and, hence, too few cointegration relations. This is particularly problematic if your data contains a slowly adjusting economic long-run relation. Given the short samples usually available in Economics, a 5% trace test will often conclude that a slowly adjusting relation could possibly be a unit root process.

Hence, the top-down test procedure and a (blind) use of 5% critical values may lead to a rejection of a very plausible economic relation for the sole reason that it has a low mean reversion rate. As if this is not bad enough, treating a stationary relation as a common stochastic trend will also affect your model inference in unknown ways.

To circumvent this problem, I usually start the analysis by asking what is the number of exogenous trends consistent with the economic model in question. I usually test this number using the 5% rule, but I also check the plausibility of this choice against the closest alternatives, for example based on their trace test statistics and the characteristic roots. When deciding in favour or against adding one more cointegrating relation, I also look at the plausibility of the cointegration relation and the sign and the significance of the corresponding adjustment coefficients.

The most problematic situation is when there is no clear distinction between large and small canonical correlations and, hence, no distinct line between stationary and nonstationary directions. This is often a signal that your information set is not optimally chosen and that some important variables are missing. When in doubt about the right choice of rank I often try to enlarge the information set for example with a potentially important *ceteris paribus* variable such as the real exchange rate, a variable often ignored in the theory model but extremely important in practice. Surprisingly often this solves the problem.

Another illustration of the misuse of the 5% rule is the test of long-run exclusion in the CVAR. Here the econometric null is that a variable is not needed in the long run relations. In this case it is hard to argue that the econometric null coincides with the economic null as the variable was chosen precisely because it was considered an important determinant in the long-run relations. To throw it out only because we cannot reject that it might be long-run excludable on the 5% level seems a little foolish.

The main reason this problem arises is because the econometric null hypothesis is often chosen because of convenience, for example when the econometric null corresponds to a single value whereas the plausible economic null corresponds to a composite hypothesis. Whatever the case, whether you reject or accept a hypothesis, I think you have to openly argue why and then back up your choice with the *p*-value of the test.

How do you handle, in general, the problem of competing models? Do you like the idea of encompassing proposed by David Hendry?

Yes, I think it is a very useful idea. But I also think it is important to distinguish between encompassing in the econometric sense versus encompassing in the economic sense, even though the two concepts are clearly related. David introduced the concept of encompassing as a way of comparing empirical models. You may consider two models explaining Y, one as a function of a subset of X_1 variables and the other of X_2 variables. Then you estimate a model for Y as a function of X_1 and X_2 and ask which of the two models encompasses the big model.

David Hendry [David hereafter] and Grayham Mizon published the paper "Evaluating Econometric Models by Encompassing the VAR" (Hendry and Mizon 1993) which discussed the general-to-specific principle–which I am very much in favour of–applied to the VAR model as a baseline against which a more specific model should be evaluated. One may say the VAR model provides the econometrician with a set of broad confidence bands within which the empirically relevant model should fall. The advantage of encompassing is that it formalizes a principle for how to weed out models that do not describe the data sufficiently well.

However, the problem of competing models in Economics is even more important as there are many competing schools in Economics but no clear criterion for how to choose between them. Because there is one empirical reality–defined by the relevant data–but several models trying to explain it it seems obvious to discriminate between them by encompassing the CVAR.

I have tried to formalize this idea by the concept of a so called "theory-consistent CVAR scenario", which basically describes a set of testable hypotheses on the pulling and pushing forces in the CVAR model. In short, a scenario specifies a set of empirical regularities that one should find in a CVAR analysis, provided the theoretical assumptions of the economic model were empirically correct. Such a comprehensive testing often reveals a significant discrepancy between theory and empirical evidence.

The crucial question is why the reality differs so much from the theoretical model. It is a question that haunted me for many years until I begun to see a systematic pattern in the empirical results. They pointed to some theoretical assumptions associated with expectations in the financial markets that were clearly empirically incorrect but not questioned by the majority of the profession. The scenario analysis made it very explicit where the inconsistencies between theory and empirical evidence were and often helped me to understand why.

But, the formulation of a scenario is no easy task. While I was still actively teaching I used to ask my students to formulate a scenario prior to their econometric analysis, but in most cases it was too difficult without my help. This is a pity, because I am convinced it is a very powerful way to solve the dilemma of competing models in Macroeconomics and to bring macroeconomic models closer to reality.

Linearity is a common assumption. Do you think it might be important to consider non-linear adjustment?

I consider the CVAR model to be a first order linear approximation to a truly non-linear world. The question is of course how significant the second order or third order components are. If a first order approximation works reasonably well, then the second or third order components might not be so crucial. But, if the first order approximation works poorly, then it may of course be a good idea to consider for example non-linear adjustment. This could be the case in stock price models where adjustment behaviour is likely to be different in the bull and the bear market. Many people are risk averse and react differently when prices go up than when prices go down, so nonlinearity in the adjustment is likely to be useful in this case.

It is of course much easier to construct a linear model to start with. Take for example the smooth transition model as a very plausible nonlinear adjustment model describing adjustment from one equilibrium level to another. In the linear CVAR model, this can be approximated by a level shift (a step dummy) in the cointegration relations combined with a sufficiently flexible short-run dynamics. In many cases this linear approximation will work almost as well (sometimes better) than the nonlinear alternative.

Another example is the nonlinear model of shifts between stochastically evolving equilibria. These models have been proposed to describe the long-lasting swings we often see in the data. They are typical of variables strongly affected by financial market behaviour such as exchange rates, interest rates, and stock prices which tend to fluctuate between high and low levels. But these stochastically switching equilibrium models can in many cases be more precisely described by the I(2) CVAR model.

As a starting point, I think one could try to approximate potential non-linear effects with the linear I(1) or I(2) model with dummy variables and then exploit the CVAR estimates to develop a better nonlinear model. The difficulty is that the non-linear possibilities are almost infinite which makes it to hard know where to start, unless you have a very clear idea of where in the model the non-linear effects are.

Univariate models, small and large-scale multivariate macro models are all used in applied macroeconomics: how do you think they relate to each other?

Basically, a univariate time-series model of x_1 is a sub-model of a small-scale multivariate model of $x_1, \ldots, x_k$, which in turn is a sub-model of a large-scale multivariate model of $x_1, \ldots, x_k, \ldots, x_m$. Hence, one should be able to argue why the smaller model with less information is preferable to a larger model with more information.

It is of course totally acceptable that people can choose between different perspectives when they approach a problem and it may be fully rational to focus on a smaller subset of the relevant information set. What I find to be problematic is the standard use of univariate Dickey–Fuller tests to pre-test the order of integration of each variable of a multivariate model. The absurdity of this becomes obvious when the result of the pre-tests is in conflict with the result of the more informative multivariate tests.

At one stage I became rather frustrated over this lack of coherence. To my great irritation I was often asked by referees to add univariate Dickey–Fuller tests to my papers, which I never did. Also, I consistently demanded any table with such tests to be removed if they had been added by a coauthor. They often reacted with puzzlement: why not calculate the univariate Dickey–Fuller tests? A simple thought experiment explains my concern.

Consider a paper which ultimately is analyzing a CVAR model but starts with a bunch of univariate Dickey–Fuller tests. Imagine now that the univariate pre-tests were placed at the end of the paper. Would this have any effect on the main conclusions of the paper?

I hired a student to find empirical CVAR analyses in papers published in a number of good-ranking journals over a period of 10 years that reported tables with pretesting. In most cases the pretests had no effect whatsoever on the final conclusions. In some cases the pre-tests led the researcher to make incorrect choices such as throwing out a relevant variable that was found to be stationary by the pretests.

To throw out variables as a result of pretesting is of course complete nonsense, because a multivariate model can easily handle a stationary variable but also because a pretested "stationary" variable may not be considered stationary in the multivariate model. This is because what matters is whether a variable corresponds to a unit vector in the cointegration space and this depends on the choice of cointegration rank. If this choice is too small–which is frequently the case–then the pretested "stationary" variable would often be rejected as a unit vector in β and the consequence would be a logical inconsistency in the analysis.

The perspective of large-scale macro models is usually different from small-scale models. This is in particular so if by large-scale you mean the large macro models used by finance ministries all over the world. They are typically characterized by a large set of behavioural (and definitional) relationships where the status of variables as endogenous, exogenous and ceteris paribus are assumed *a priori* and where little attention is given to dynamic feedback effects. As such, it is hard to argue that a small-scale multivariate model is a sub-model of these models as they represent two different approaches to macro-modelling. In my book (Juselius 2006) I have proposed a procedure to connect the two.

Common trends have been shown to be invariant to the extension of the information set. Can this be used to devise a progressive modeling strategy, where results from (one or more) small-scale CVAR models are inputs to a larger scale model?

The unsolved problem here is how to uniquely extract and identify individual common stochastic trends. While it is straightforward to determine the space spanned by the $p - r$ common stochastic trends in a p-dimensional CVAR model, it is much more difficult to economically identify these common stochastic trends. Kevin Hoover, Søren and I have worked on this difficult problem for many years with the purpose of solving the problem of long-run causality in economic models.

However, the potential of common trends analysis stretches far beyond this problem. It seems plausible that there are a limited number of common stochastic trends in the world. The invariance of common stochastic trends suggests that we should find linear combinations of these common stochastic trends in small-scale CVAR models. The set

of these extracted common stochastic trends could then be analyzed using cointegration techniques.

Let's say that we have extracted 10 common stochastic trends from a number of small-scale CVAR models and that we find the cointegration rank to be 7. This would be consistent with three fundamental stochastic trends in the economy, for example an inflation trend, a productivity trend, and a financial trend. The problem is, as already said, how to uniquely identify them so that we can put labels on them. I think it is a fantastic research problem.

What is the role of cross section versus panel data models?

Cross-section models can provide a different perspective on the economy than time-series models, because they add valuable information about individual characteristics at each point in time unavailable in aggregate time-series data. But the time perspective, such as feedback dynamics, is missing in cross section models.

In panel data models you have the possibility for both perspectives provided you have access to fairly long panel data sets. Personally I think reasonably long consumer panel data sets as we have in Denmark are extremely valuable as they combine the best of the two worlds. But, of course, they can not address all issues of macroeconomic relevance.

An interesting research project that has been on my wish list for a long time is to study the aggregated output from simulated agent-based models to learn more about the connection between micro and macro. For example, would the aggregate behaviour have similar properties in terms of cointegration, adjustment, and feedback dynamics as we usually find in our CVAR models?

You never used panel cointegration techniques. Is it because you are skeptical about them?

As I already said, I find panel cointegration models based on micro data to be potentially very valuable, but I am much more skeptical about such analyses based on a panel of countries. In my view, countries are individually too different to be merged into the same model structure. In most cases I have come across, the panel analysis is based on so many simplifying assumptions that in the end it is hard to know which of the results are true empirical results and which are due to the simplifying restrictions forced on the data.

For example, one can easily find examples of misuse of country panel data in Development Economics. This is because, for many of these countries, data are only available on an annual basis over a post-colonial period. The quality of data is often low, partly because data collection methods may not be very reliable, partly because observations are missing during periods of war and unrest. This has led many development economists to merge the countries in a panel to get more information out of the data. But as such this is no guarantee that the results to become more reliable; it can easily be the other way around.

To look into this problem, Finn Tarp, Niels Framroze Møller and I started a big project where we studied 36 Sub-Saharan countries regarding the effectiveness of their development aid on GDP growth, investment, private consumption and government expenditure. It was a huge data base and the computer output was almost killing.

Just the initial specification of an adequate VAR for each country was as a major task: first we had to identify the time points for extraordinary events such as wars, military coups, famines, droughts, and floods and then we had to control for them by appropriate dummy variables. Because the individual countries differed a lot, a major task was to classify them into more homogeneous groups.

Niels suggested a first coarse division according to whether aid had a significant long-run effect on GDP or investment, whether aid was exogenous to the system, whether it was purely adjusting to the macro-system, or none of the above. But also within these more homogeneous groups, individual countries differed a lot for example in terms of the magnitude of parameter estimates. We concluded that there were positive and significant effects of aid in almost all countries being studied. This was in stark contrast to panel data

studies published in high ranking journals which showed that foreign aid has had no, or even negative, effect on the growth of GDP and investment.

The lesson seems to be that unless you control properly for extraordinary events and other data problems before pushing the panel data button, you can get basically any result.

So your conclusion is that aid has been effective given some country-specific characteristics. Would it be possible to use these characteristics in a panel data set-up where you include all countries? Could they be a mediator of the effectiveness of aid?

No, I do not really think so. The countries are generally too diverse. As I already mentioned you might be able to use a smaller group as a panel, but not all of them. By studying each country separately, you can identify characteristic features which would be impossible to recognize if they are treated as a homogeneous group.

Take for example a country like Tanzania where Nyerere–the president up to mid-eighties–was a charismatic person with bold visions. The donor countries were generally favourable towards him and Tanzania received a lot of aid, substantially more than any other country in that same period. However, Nyerere believed in a strong currency and used the foreign aid to maintain a fixed exchange rate rather than to improve the development of the country.

Another example is Ghana where a military dictator took over the government in the early seventies. He declared that he had no intention to pay back previous development loans, the perfect recipe for not getting additional loans. As a consequence, the national currency was subject to extreme devaluations followed by hyperinflation.

These are just two examples of the type of heterogeneity you will come across in development countries and they are in no way "black swans". Just to understand the country-specific framework within which the aid is supposed to work requires a lot of work. If you pay attention to all the difficulties that must be solved before you do a panel analysis, the desire to do a panel analysis may evaporate altogether. And if you do all the necessary work, the need for a panel analysis may no longer be so great.

We put a lot of work into this project because of its importance. Rich donor countries give less than 1% of their GDP in foreign aid to improve the quality of life for very vulnerable people. Even though less than 1% is not a lot, there are many who would look for a good argument suggesting not to help those who are much worse off. It is unacceptable, in my view, if such an argument is based on too simplified and misleading econometrics.

Our paper was finally published in the Oxford Bulletin (Juselius et al. 2014), after first having been rejected by those Development Economics journals that had published the studies we criticized. One of the Oxford Bulletin referees wrote that he had been of the firm opinion that foreign aid was not contributing to development, but had changed his mind, because–as he wrote–the analysis of our paper was so carefully done that he could not find anything to criticize. I felt very proud.

Can you provide more details on the quality of the data in this research and the consistency across countries of the variables you analyzed?

We used annual data starting from the 60s, which is when most African countries became independent. But because the 60s was a very volatile transition period we decided to leave out this decade for most countries. The data–consisting of total foreign aid and five key macro variables–were collected from the official data bases, the Penn World Tables and World Development Indicators where data are reported in a reasonably consistent way across countries.

A few countries were excluded due to many missing data points and for two countries we had to add variables to be able to make sense of the results. We were able to keep 36 countries for a detailed CVAR analysis based on roughly 45 annual observations. Every step was carefully reported, but with six variables and only 45 data points it was more or less pointless to apply the recursive tests to check for parameters stability.

Therefore, the parameter estimates should be thought of as representing average effects over the sample period. Despite the shortness of our time series, the data were surprisingly informative, possibly due to their large variation. Still, I think it is plausible that the transmission mechanisms of foreign aid have undergone changes over the last decades, similarly as macroeconomic mechanisms have changed in the industrialized part of the world.

It could, therefore, be quite interesting to extend our study with a more recent data set based on quarterly macro data. For many of the countries such data are available starting from the 90s. But on the whole I believe the results we obtained from our annual sample were completely plausible, often telling an interesting story about vulnerable economies struggling to find their way out of poverty.

You said that when the sample is short it is not easy to analyze the stability of the parameters and the possibility of structural breaks: can you elaborate on this?

An interesting example of long-run stability is the Danish money demand relation which is thoroughly analyzed in my book (Juselius 2006). It was my first illustration of Maximum Likelihood cointegration and was published in 1990 in the Oxford Bulletin of Statistics and Economics based on fifteen years of quarterly data from 1972 to 1987 (Johansen and Juselius 1990).

It was a volatile period covering two oil crises, several devaluations of the Danish krona and a far-reaching political decision to deregulate financial movements. A priori there was good reason to suspect that the parameters of the estimated money demand relation would not be totally stable. Even though the recursive stability tests did not signal any problem, the rather short sample of fifteen years made these tests rather uninformative.

At a later stage I updated the data by adding data from 1988 to 1994. To my relief I got essentially the same parameter estimates for the money demand relation as before (Juselius 1998). Based on the extended data, the recursive stability tests were now more informative and they confirmed that the money demand relation was stable.

However, the recursive tests also showed that this was not the case with the first cointegration relation–a partially specified IS relation–which exhibited a complete structural break around mid eighties due to Denmark's financial deregulation. Ironically, the 5% rule would have selected a non-constant, meaningless, relation and left out the stable and meaningful one, a good illustration of the hazards of blindly using the 5% rule.

When I started writing my book I decided to use the Danish money demand data as an empirical illustration of the CVAR methodology. A first version of my book was based on the 1994 data set, but in 2004 when the text was more or less finished, I could no longer ignore the fact that the data were rather old. So I updated it once more with additionally 10 years of quarterly observations, now up to 2004.

The first time I run the CVAR with the new data, a whole flock of butterflies fluttered in my stomach. If the empirical results had changed significantly, then I would have had to rewrite large parts of my book. But, fortunately, all major conclusions remained remarkably stable, albeit the estimates changed to some minor extent.

After my retirement in 2014 somebody asked me if the Danish money demand relation was still going strong. So out of curiosity I updated my data once more and found out that the money demand relation was no longer in the data! Adding the period of unprecedented credit expansion that led to the overheated economy ending with the financial crisis, seemed to have destroyed the stability of the money demand relation.

As such it is an interesting finding that prompts the question "why?". Is it because of the exceptionally high house and stock prices in the more extended period compared to the almost zero CPI inflation rate and historically low interest rates have changed the determinants of money demand? Would we be able to recover the old relationship by extending the data with house price and stock price inflation? I would not be too surprised if this was the case.

All this raises an important discussion about the stability of economic mechanisms. The underlying rationale is of course that social norms and behaviour tend to change over time as a consequence of political reforms but also of political views or propaganda, as there nowadays is too much evidence for around the world. However, also economic norms and dogmas are likely to influence behaviour. If economic models show that competition is good and greed is even better then some politician will use it in their propaganda as evidence in favour of their policy.

Of course it would be absolutely fantastic if we had access to powerful econometric tools which could tell us exactly when a structural change has occurred, but I doubt very much this will ever be the case, not even remotely so. Structural change is seldom a black or white event; things change in a much more blurred way. Take for example the overheated economy at the beginning of this century that ended with the financial crisis–the so called long "moderation" period.

If this turns out to be a transitory event, albeit very long-lasting, then the breakdown of the money–demand relation may not represent a structural change. Updating the money–demand data to the present date might give us back the old parameter estimates. Even though I doubt it very much, it is nonetheless a possibility. In most of my professional life I have struggled with questions like this.

Econometric analysis is fantastic when it helps to make complex structures more transparent, when it forces you to understand puzzling features you would otherwise not have thought about, and when it teaches you to see the world in a new light. But it does not let you escape the fact that it is you who are in charge, it is your judgement and expertise that is a guarantee for the scientific quality of the results.

You have been mentoring many Ph.d. students and young researchers, like the younger version of the two of us. What did you like or dislike about this? Any forward-looking lessons for other econometricians?

In all these years I have immensely enjoyed guiding students both at the Economics Department in Copenhagen, but also at other departments during our many travels. But to be both a good teacher, a good supervisor and a good researcher at the same time is basically "mission impossible" as long as 24 h a day is a binding restriction. Even though I spent all my time (including late evenings, weekends, and holidays) on these activities, I nevertheless always felt I should have done more.

On top of all this I also had the ambition to engage in the public debate, not to mention obligations to family and friends. So, time was always in short supply and every day was a struggle to meet deadlines and a compromise between everything that needed to be done. It took surprisingly long until my body begun to protest increasingly loudly. In the end it forced me to slow down a little. This is the not-so-good aspect of being an (over)active researcher.

My best teaching memories are without comparison from our many Summer Schools of the Methodology of the Cointegrated VAR. To experience highly motivated, hard-working students willing to give up all other temptations in beautiful Copenhagen only to learn a little more econometrics was a very precious experience and I feel enormously privileged to have had it.

The secret behind this success was that we offered the students a firm theoretical base, a well-worked out guidance for how to apply the theory to realistic problems and a personal guidance of their own individual problems, often a chapter of their PhD thesis. A typical day started with Søren [Johansen] discussing a theoretical aspect of the CVAR (and students came out looking happy and devastated at the same time), then I illustrated the same aspect using the Danish money demand data (students begun to look somewhat more relaxed), and in the early afternoon a teaching assistant explained the same aspect once more based on a new application (students begun to say now they had grasped it).

Finally in the late afternoon, early evening, they had to apply the theory on their own data (students were totally lost, but after competent guidance happiness returned). It was a tough experience, but many students learned immensely in three weeks. One of them said

he had learned more than during three years of full time studies at home. If I should give any lesson for other econometricians, this is the one.

Another extremely good experience was a series of Nordic–later also European–workshops from 1989 to 2000 where we met two, three times a year to discuss ongoing research in the cointegrated VAR model. This was a different way of guiding young researchers–like the two of you–by offering a direct involvement in the research process. It was truly learning-by-doing research. Most of the cointegration results in Søren's book and in my own were developed and intensely discussed in this period. A workshop usually lasted for 3–5 days and we were engaged in discussions every single minute. When the workshop closed I think we were all practically dead.

But I believe we found it enormously exciting. It was a once-in-the-lifetime experience. This is also a lesson I would happily give to other econometricians.

Is there a "gender gap" in Econometrics?[2]

When I started my academic career as a young econometrician, the gender gap was very large indeed. There were only a few female colleagues at the department and very, very few female professors altogether in Economics. But, even though the gap has become smaller, it has not disappeared.

To some extent, I believe it is a question of a male contra a female culture. The traditional language/jargon in Economics is a male-dominated language foreign to many women. For example, theoretical ideas are formulated in the abstract terms of a "representative agent" who maximizes a well-defined utility function derived from a preference function that often reflects greed.

This way of thinking is not very attractive to many women, who would chose Economics because they are concerned about the huge income gap between industrialized and developing countries, the well-being of their parents, children, friends (not an "agent") and would like to understand why a good friend became unemployed and what to do about it. I think it is quite telling that the two most popular fields among female economists are Labour Economics and Development Economics.

The question is whether the abstract way of formulating Economics is absolutely necessary from a scientific point of view. I find it probematic that trivialities or common sense results are often presented in an almost opaque language which tends to make economic reasoning inaccessible to laymen.

In the book, Chang (2014) "Economics: a user's guide", the well-known Cambridge economist Ha-Joon Chang argues that 95% of Economics is just common sense, but made more complex by abstract mathematics. His accessible and highly qualified text illustrates this point. On the whole I believe more women would be attracted to research in Economics if one would allow more common sense reasoning and pluralism into the teaching of Economics.

Many times in my teaching, I noticed the cultural difference between male and female students. My male students were often fascinated by the technical aspects, whereas my female students were more excited by the applied aspects. For example, when I demonstrated the derivation of the trace test, the guys were flocking around me after class ended asking questions to the technical aspects. When I illustrated how one could use the technical stuff to ask relevant empirical questions, the female students did the same. They were willing to learn the technical stuff, but mostly because it was necessary for the empirical applications.

The gender gap in publications also reflects a similar difference in attitudes. Many top journals tend to favour "technical" work, partly because it's easier to assess whether a mathematical result is right or wrong than an applied empirical result. But the fact that the editorial boards of top journals are mostly populated by men might also contribute to the gender gap. Notwithstanding today's strong emphasis on empirical work, top journals tend to favour rigorously applied theories and mathematical models which are illustrated with simple examples or alternatively applied to simple problems.

Since many real life problems are much more difficult to formulate using rigorous mathematics they are, therefore, much harder to publish in top journals. When I started teaching the CVAR methodology–which is based on rigorous mathematical statistical principles–I thought it would help female (and male) students to overcome this problem. But it did not work out as I had hoped. The main problem was that the empirical reality seldom supported the rigorously derived economic model.

As I have learnt over and over again, journal editors are not happy to accept a paper reporting results which contradict previously published ones. The consequence was that my PhD students often tried to "sit on two chairs": on one hand they wanted to use the CVAR method in a rigorous way, on the other hand they wished the results to support mainstream economic models. I believe it was still another "mission impossible" and I sometimes regret that I had put them in this situation.

Nowadays there are more female students in Economics than in the past, so things are slowly changing. Many of them are still interested in empirical work often in Labor and Development Economics, but their research is much more related to Microeconometrics than to what I would call disequilibrium Macroeconometrics.

Did you feel kind of alone in this mainly male environment?

If I feel alone, then it is because I am rather alone in my view about what is important in empirical macroeconomic modelling and how it should be done. Considering all disasters in the world around us, it is obvious to me that we desperately need a much better economic understanding of real world problems, rather than still another model of a toy economy.

I am also aware of the dilemma between rigour and empirical relevance in Economics. I have seen numerous examples of really bad empirical CVAR applications where the data have been read in, the CVAR button has been pushed and meaningless results have been printed out that say nothing useful about our economic reality. While there should be no shortcuts in science and empirical results should be derived in a transparent way obeying accepted scientific rules, I strongly believe there should also be room for informed judgement, what Dave Colander would call "the art of Economics". I also believe this is what a rigorously done CVAR analysis can do for you.

Since I have always been outspoken with my views both in academic forums and in the public debate, I have also got my part of male anger, more nowadays than when I was younger. Perhaps I was more diplomatic then or just more good-looking. Whatever the case, being one of the very few female economists was not just negative, I probably did not have to fight as hard for attention as a comparable male econometrician.

But the fact that my research has received a lot of interest among econometricians, economic methodologists and the public is something I value very highly. Many of my absolutely best and most valued colleagues and friends are male economists or econometricians, as for example Søren, the guest editors and the contributors of this wonderful Special Issue. So, on the whole I have been very fortunate in my professional life.

Notes

[1]   Accessed on 8 April 2022.

[2]   See Card et al. (2021).

References

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri. 2021. *Gender Differences in Peer Recognition by Economists*. Technical Report 28942. Cambridge: National Bureau of Economic Research. [CrossRef]

Chang, Ha-Joon. 2014. *Economics: A User's Guide*. Torrance: Pelican.

Colander, David, Michael Goldberg, Armin Haas, Katarina Juselius, Alan Kirman, Thomas Lux, and Birgitte Sloth. 2009. The financial crisis and the systemic failure of academic economics. *Critical Review* 21: 249–267. [CrossRef]

Hendry, David F., and Graham E. Mizon. 1993. Evaluating dynamic econometric models by encompassing the VAR. In *Models, Methods and Applications of Econometrics*. Edited by Peter C.B. Phillips. Oxford: Blackwell, pp. 272–300.

Johansen, Søren, and Katarina Juselius. 1990. Maximum likelihood estimation and inference on cointegration—With applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52: 169–210. [CrossRef]

Juselius, Katarina. 1998. A structured VAR for Denmark under changing monetary regimes. *Journal of Business & Economic Statistics* 16: 400–11. [CrossRef]

Juselius, Katarina. 2006. *The Cointegrated VAR Model: Methodology and Applications*. Oxford: Oxford University Press.

Juselius, Katarina, and Sophia Dimelis. 2019. The Greek crisis: A story of self-reinforcing feedback mechanisms. *Economics: The Open-Access, Open-Assessment E-Journal* 13: 1–22. [CrossRef]

Juselius, Katarina, and Mikael Juselius. 2014. Balance sheet recessions and time-varying coefficients in a Phillips curve relationship: An application to Finnish data. In *Essays in Nonlinear Time Series Econometrics*. Edited by Niels Haldrup, Mika Meitz and Pentti Saikkonen. Oxford: Oxford University Press.

Juselius, Katarina, Niels Framroze Møller, and Finn Tarp. 2014. The long-run impact of foreign aid in 36 African countries: Insights from multivariate time series analysis. *Oxford Bulletin of Economics and Statistics* 76: 153–84. [CrossRef]

Juselius, Katarina, and Josh R. Stillwagon. 2018. Are outcomes driving expectations or the other way around? An I(2) CVAR analysis of interest rate expectations in the dollar/pound market. *Journal of International Money and Finance* 83: 93–105. [CrossRef]

Kaufmann, Robert K., and Katarina Juselius. 2016. Testing competing forms of the Milankovitch hypothesis: A multivariate approach. *Paleoceanography* 31: 286–97. [CrossRef]

Phelps, Edmund. 1994. *Structural Slumps*. Princeton: Princeton University Press.