

MACROECONOMETRICS: STATISTICAL ALCHEMY AT THE FINAL STAGE

Samuel Meng

Despite having faced a battery of criticisms, macroeconometrics has arisen and dominated for more than half a century. It is commonly used to model time-series data on the attempt to explain and predict macroeconomic variables such as GDP, investment, consumption, spending, exports, imports, employment and unemployment. By examining the foundations of macroeconometrics, this article demonstrates that all macroeconomic modelling work is based on shaky foundations because the conditions for the probability theory do not hold in time-series modelling. The article further shows that random experiment, random variable and random disturbance are a trinity, and that macroeconometricians have broken this trinity, detached the probability theory from its evidence, and thus embedded severe problems into their practice. The detrimental impact of macroeconomic modelling is such that it needs to be replaced with a more scientific way to conduct research in macroeconomics.

Macroeconomic modelling became steadily more prominent as a tool of mainstream economics during the second half of the twentieth century, despite diverse criticisms (*e.g.* Friedman 1948; Leontief 1971; Worswick 1972; Lucas 1976; Summers 1991; Keuzenkamp 1995; Solow 2010). By the end of the century, time-series analysis (*i.e.* macroeconometrics) had virtually become a synonym for macroeconomics. However, the failure of econometric models to predict the Global Financial Crisis (GFC) has subsequently cast doubt on the usefulness of macroeconometrics. Interestingly, the renewed criticisms have done little damage to its popularity, judging by the content of academic journals and university

Meng, S. (2017)

'Macroeconomics: Statistical Alchemy at the Final Stage'
Journal of Australian Political Economy
No. 80, pp. 155-77.

curricula. Macroeconometric modelling has become a major element in public policy debates too, with both proponents and critics of policy changes routinely citing the summary results of 'modelling' to support their propositions.

This article is organized as follows. The next section provides a review of the rise of macroeconometrics, and the criticisms levelled at it along the way, so as to provide the necessary background for the understanding of key problems in the field. The following section considers the shaky foundation of macroeconometrics, with particular reference to probability theory. Then comes a section which demonstrates the negative impact of macroeconometrics on the economy, as well as on economic research, and suggests a better way of bridging theory and data. The last section summarizes the arguments and provides concluding remarks.

The rise of macroeconometrics despite strong criticisms

In the late 1800's and early 1900's, statistical methods and probability theory developed significantly. Hence it was not surprising that statistical approaches were introduced into economics in the early 1900's. Because of his work in estimating demand curves (Moore 1911, 1917), Moore is considered the pioneer in applying statistical method to economic research, while Frisch and Tinbergen were pioneers in formalizing the statistical approach in economics. Frisch (1934) argued that most economic variables were simultaneously interconnected in 'confluent systems', which needed to be analyzed by means of multiple regressions. Tinbergen's pioneering work was contained in his report for the League of Nations, *Statistical Testing of Business-Cycles Theories* (Tinbergen 1939), in which he introduced multiple correlation analysis, built a parametrized mathematical model, and applied statistical testing procedures.

Influential economists such as Marshall, Edgeworth, and Keynes had been skeptical of the appropriateness of introducing statistical methods into economics. They argued that the *ceteris paribus* assumptions in economics made it difficult to apply statistical methods. Tinbergen's work was reviewed and heavily criticized by Keynes in *The Economic Journal* (Keynes 1939). Keynes first pointed out that Tinbergen did not adequately explain the conditions that the economic data must satisfy in

order to apply the statistical method. Then he pointed out the specific flaws in Tinbergen's work: (1) statistical tests cannot disprove a theory because multiple regression analysis relies on economic theory to provide a complete list of significant factors; (2) as not all significant factors are measurable, the model may not include all relevant variables; (3) different factors may not be independent of each other; (4) it is implausible to assume that the correlations under investigation are linear; (5) the treatment of lag and trends are arbitrary; and (6) historical curve-fitting and description are incapable of making an inductive claim.

Tinbergen's reply to Keynes was published in *The Economic Journal* the following year (Tinbergen 1940). He avoided the issue of logical conditions for applying multiple correlations, but answered the other questions. Regarding the complete list of relevant factors, Tinbergen claimed that it is not clear beforehand what factors are relevant, so they are determined by statistical testing. For independence of variables, he argued that the statistical requirement for explanatory factors was uncorrelated rather than independent. For the induction issue, he maintained that, if no structural changes take place, it is possible to reach conclusions for the near future by measuring the past. Tinbergen's reply did not answer Keynes' crucial questions such as the conditions for applying statistical method and the non-homogenous economic environment over time. However, these questions were subsequently answered by Haavelmo (1943, 1944). In the first of his two papers Haavelmo argued that the time-series can be viewed as data for random variables because random variables have to be introduced in order to confront theory with actual data. Once the random variables are introduced, the non-logical jump from data to theory is justified and thus a complete list of causes is not necessary for an econometric model. He considered that the data were produced by probability laws and that, if these laws were to persist, the model could predict the future. He also stated that statistical tests can prove or disprove a theory. These ideas were formalized in a second paper by Haavelmo (1944), which introduced the natural experiment concept to justify the foundation for applying the probability theory in macroeconomics. Haavelmo (1944) also distinguished observational, true, and theoretical variables, and demonstrated at length how to use a joint probability function to estimate simultaneous equations.

These arguments by Haavelmo somehow silenced Keynes. After Haavelmo's work, the probability approach was firmly established in

macroeconomics despite continued criticisms from different schools of economic thought. Representative criticisms and defenses are briefly reviewed here.

Interestingly, one of the early opposers of the econometric approach in macroeconomics was Milton Friedman (1940). His review of the work of Tinbergen (1939) concluded that Tinbergen's results could not be judged by the test of statistical significance because the variables were selected after an extensive process of trial and error. Friedman (1948) also pointed out that the econometric work of the Cowles Commission (an economic research institute where Tinbergen and Haavelmo stayed) was built on articles of faith, such as the possibility of constructing a comprehensive quantitative model from which one can predict the future course of economic activity. He argued that this kind of comprehensive model is possible only when decades of careful monographic work in constructing foundations had been done. Friedman (1951) proposed a naïve test to examine the predictive performance of econometric models. This test was frequently used by the Cowles Commission and, following the naïve-model approach, Sims (1980) developed the vector autoregression (VAR) model, which introduces multiple variables in a naïve model.

The poor forecasting performance of large structural macroeconomic models in the early 1970s discredited macroeconometrics and sparked a battery of criticisms. Leontief (1971) argued that econometrics was an attempt to use more and more sophisticated statistical techniques to compensate for the glaring weakness of the database available. Worswick (1972) claimed that econometricians are not engaged in forging tools to arrange and measure actual facts but are making a marvelous array of pretend-tools. Brown (1972) concluded that running regressions between time-series is only likely to deceive.

It was against this backdrop that Lucas (1976) published his famous paper '*Econometric Policy Evaluation: a Critique*'. By introducing 'rational expectation' into policy evaluation, Lucas (1976) demonstrated that any change in policy will systematically alter the structure of econometric models. Thus he claimed that there is a theoretical problem in structural macroeconomic models. However, the issue proved not to be fatal for macroeconometrics. Adding an expectation variable to the structural macroeconomic model, econometricians (*e.g.* Wallis 1977; Wickens 1982; and Pesaran 1987) sought to show that rational expectations could be modelled within a structural macroeconomic

framework. Along this line, the dynamic stochastic general equilibrium (DSGE) model was developed. The other consequence of Lucas' critique was that some econometricians (*e.g.* Sims, 1980) moved away from using a structural model to univariate time-series naïve models and subsequently to the vector autoregression (VAR) model. This approach has no backing from economic theory and thus also incurred heavy criticisms (*e.g.* Cooley and Leroy 1985).

The poor performance of macroeconomic models also led to the blaming of the modellers for their poor techniques. Hendry (1980) demonstrated how an inappropriate model can generate spurious regression and suggested the use of statistical tests to increase the robustness of econometric models. Leamer (1978) commented: 'the econometric modeling was done in the basement of the building and the econometric theory courses were taught on the top floor.' Leamer (1983) proposed the use of sensitivity analysis in order to take the 'con' out of econometrics and thus restore the credibility of econometrics. Spanos (2011) blamed the inappropriate statistical model specification and validation, such as the theory-driven approach, and advocated the error statistical approach and the use of general-to-specific procedures.

The claim that statistical testing can give econometrics a scientific status was not supported by other economists, notably, Summers (1991), Keuzenkamp (1995), and Ziliak and McCloskey (2008). Summers (1991) argued that formal econometric work had made little contribution to macroeconomics, while informal pragmatic empirical work had a more profound impact. He pointed out that the role of a decisive econometric test in falsifying an economic theory looked similar to, but was actually totally different from, the role of empirical observations in natural science. By examining the statistical tests in Hansen and Singleton (1982, 1983), Summers revealed that the test was just a confirmation of common sense because consumption of different consumers is not perfectly correlated with their wealth. In his review of the book by Hendry (1993), Keuzenkamp (1995) rejected Hendry's use of the 'data generation process', the general-to-specific approach, and the three golden rules of econometrics – 'test, test and test'. More recent criticism by Ziliak and McCloskey (2008) mainly focuses on the uselessness and deception of statistical significance and standard errors. Ziliak and McCloskey (2013) further demonstrate that statistical significance essentially proves nothing.

Freedman (1995) highlighted the crucial role of the assumption of an independent and identically distributed (IID) disturbance in computing statistical significance and in legitimate inferences. Freedman (1999) pointed out that econometricians tend to neglect the difficulties in establishing causal relations and that the mathematical complexities tend to obscure, rather than clarify, the assumptions for a statistical model. Freedman (2005) further identified many problems, including the faith in independent and identically-distributed disturbance, the use of linear functions, model selection problems, asymptotic assumption, and the inconsistency in Bayesian methods.

These twists and turns in the development, application and criticism of econometric technique did not create a disciplinary crisis. Indeed, they may be interpreted by macroeconometricians as consistent with the development of a 'normal science', but then in 2008-9 came the outbreak of the GFC. This stunned macroeconometricians (or should have) because no econometric model had shed light on the likelihood of this event.

This situation triggered widely-spread criticisms on econometrics both within and outside of the economics profession. In the *New York Times*, Krugman (2009) claimed that macroeconomics of the last 30 years was spectacularly useless at best and harmful at worst. In the *Financial Times*, Skidelski (2009) blamed banks for their blind faith in forecasting from mathematical models. Kling (2011) questioned the integrity of the macroeconomic models and claimed that macroeconomic models are unscientific because models using repeatable events are poorly suited to accurate prediction or historical explanation. In defense of the poor performance of econometric models, Hendry and Mizon (2014) attributed the problem to the shift in probability distributions during the economic recessions; and they proposed the need to recognize three types of unpredictability: intrinsic, instance and extrinsic unpredictability. Heterodox economists (*e.g.* Nell and Errouaki 2013; Birks 2015) blamed the assumptions used in econometric models. Rational expectations and optimization assumptions were at the centre of the critiques of Nell and Errouaki (2013), while Birks (2015) criticized economic research that focuses on theories, models and techniques with insufficient attention to their relevance to real-world issues.

The invalidity of macroeconometrics: a theoretical perspective

This brief survey indicates that macroeconometrics has been controversial from its origins to the present day. It is pertinent to ask why have the numerous criticisms not had more substantial adverse impact on the popularity of macroeconometrics? The author's view is that previous criticisms failed because they either tried to attack macroeconometrics from the outside without the use of a statistical framework or focused too much on methodological details. The former approach (*e.g.* Keynes 1940; Friedman 1948; Leontief 1971; and Solow 2010) attacked econometrics as a whole but failed to pinpoint where and why econometrics is wrong. The latter approach (*e.g.* Liu 1960; Lucas 1976; Freedman 2005) pointed out some technical flaws in econometrics but missed the main target – the framework of econometrics. In other words, the existing criticisms did not successfully attack the foundation of macroeconometrics. This section of the current article seeks to do so by focusing on the logical problems and foundational issues, rather than specific macroeconomic techniques (albeit some econometric techniques are mentioned). It postulates that, from a theoretical perspective, macroeconometrics is invalid because it violates the conditions of probability theory, is unable to include all relevant factors, and cannot reliably claim to have correct functional relationships.

Violating the condition for probability theory

Keynes (1939) pointed out that the econometricians must address the conditions for applying statistical method, not just apply it. Keynes' injunction has been addressed by several econometricians over time, but in an inadequate way. The foundation for macroeconometrics is probability theory. However, from the starting point, applying probability theory to macroeconomic data is questionable practice that involves a fundamental logical problem that is insolvable.

Haavelmo (1944: Ch.3) successfully argued that the probability theory should be applied to economics. He pointed out that: 'even in the case of an exact economic relationship, or when we say something is certain, we mean the probability nearly equals one'. However, the vital condition for probability theory is that of random experiment. Probability theory is

valid because it has been proven by experiments. For example, if you toss a coin, you cannot predict the outcome (heads-up or tails-up) of each toss. However, a large number of experiments reveal that there is a 50% chance of heads or of tails occurring. This is the nature of a 'random experiment'. Haavelmo (1944: 49) correctly stated that 'the notion of random experiments implies, usually, some hypothetical or actual possibility of repeating the experiment under approximately the same conditions'. Note that 'under approximately the same conditions' is actually very similar to the term 'other things being equal' (*ceteris paribus*) used in economics and similar to the 'same condition' used in scientific experiment. All sciences have something in common.

Based on the definition of 'random experiment', correctly implemented surveys and scientific experiments can be viewed as random experiments. However, it is obvious that the macroeconomic time-series data do not fit within the concept of 'experiments', let alone 'random experiments'. To justify the use of probability theory in time-series data, Haavelmo (1944) introduced a concept of 'natural experiments', namely, 'the experiments which, so to speak, are products of Nature, and by which the facts come into existence' (Haavelmo 1944: 50). However, even if we view macroeconomic time-series data as the result of such 'natural experiments', these experiments cannot be repeated under approximately the same conditions because conditions change remarkably over time. Since 'natural experiments' do not satisfy the conditions for 'random experiments', the former cannot be labelled as 'random experiments'. Consequently, it is invalid to apply probability theory to macroeconomic data.

Inability to include all factors in a model

There is a way to make natural experiments satisfy the condition for random experiments. That is, if all variables involved are included in the model, natural experiments can be viewed as being under approximately the same condition and thus can be treated as equivalent to a random experiment or a controlled experiment. While, to my knowledge, no econometrician has formally argued in this way, this is the underlying reasoning for the multiple-variable regression, especially the 'from general to specific' approach.

However, this approach is pragmatically impossible to implement because, no-one can claim to have included all significant variables in a model: there are an unknown number of (known or unknown) factors which may affect the time-series data. Because of the impossibility of being able to make such a claim, random experiments in statistics and controlled experiments in natural science are necessary.

Facing the countless numbers of factors involved, macroeconometricians claim to have a 'magic' tool – testing – to uncover important variables. According to macroeconometricians (*e.g.* Tinbergen 1940; Hendry 1980; and Leamer 1983), testing can discover important variables, transforming econometrics from alchemy to science. It is worth mentioning that here econometricians actually changed the task of including all important variables to the task of detecting important variables. Even if econometric tests are able to verify the importance of a variable, econometricians cannot claim that they have included all important variables because not all potentially important variables have been included in their testing list.

The least likelihood of selecting a correct function for a macroeconomic model

Even if econometricians could overcome the difficulty of including all important variables in a model, they face another difficulty – it is hard to select a correct function for a macroeconomic model because of the multiplicity of variables involved. A simple example can demonstrate this difficulty. Suppose there are 10 variables (which could be a very small number compared with the complete list of all significant variables) in a macroeconomic model and the influence of each variable may be expressed by 5 different functions, *e.g.* linear, quadruple, logarithm, exponential, and logarithm-quadruple. The number of possible functions for the model will be $5^{10}=9,765,625$. Only one function form is true in the almost 10 million possible function forms in our simplified case! It is obvious that the likelihood of the econometricians' obtaining a correct function is extremely small. In other words, function misspecification is very likely in a macroeconomic model.

In practice, linear functions (or transformed linear functions) are used in the vast majority of macroeconomic models because of the convenience of linear regression. Macroeconometricians try to downplay

the importance of having the correct function by arguing that there are many mechanisms (*i.e.* probability laws according to Haavelmo 1944, or data generation processes according to Hendry 1993) which govern the economic data and that, if the linear model can fit the data well, the model uncovers one of the mechanisms. This reasoning is flawed and deceptive. Haavelmo (1944) regarded probability laws as mechanisms of macroeconometrics, but the truth is that a probability law is anything but a mechanism, because mechanisms are not the concern of probability theory. For example, probability theory can tell you, in the experiment of tossing a coin, the chance of heads-up and tails-up is 50/50, but the theory tells you nothing about what factor or mechanism causes this probability. Regarding probability law as a mechanism is equivalent to saying that there is no mechanism and thus there is no truth.

The importance of using a correct function which reflects the true mechanism is highlighted by the GFC. Macroeconometricians may show how good their within-sample prediction is, but this prediction means nothing because it results from their fitting-the-data strategy. The failure of econometricians on the GFC (or 'the profession's blindness to the very possibility of catastrophic failure', according to Krugman 2009) indicates that econometric models did not reveal the true mechanism. However, macroeconometricians still have not understood that there is only one true mechanism for one reality which their models have failed to uncover. Instead, they continue to harbor their statistical illusions.

Attempts to make macroeconometrics scientific: problems in practice

The previous section shows that, because the theoretical foundation for macroeconometrics is flawed due to its inability to satisfy the condition for random experiment, macroeconometrics should be dead at its birth. On the contrary, it has thrived and dominated macroeconomic studies. Why has the invalidity issue been ignored? One reason is that, apparently, no one is able to make the time-series data satisfy the condition for random experiment. Otherwise, a new method would have replaced or upgraded macroeconometrics. However, the inability to do random experiments is no excuse for macroeconometricians to assume that data can be viewed as if it came from random experiments and to generate invalid results from invalid assumptions. The other reason for

the thriving of macroeconometrics is that most economists have been convinced that econometric methods are able to fix the problems arising from data and, as such, they believe econometrics can qualify as a rigorous science. This section will show that what macroeconometricians have produced is a mirage.

The macroeconometricians' approach is to make strict assumptions first, then to relax these assumptions and use econometric tools to bring the conditions of real data close to strict assumptions. An econometric textbook might start with the five or six conditions for simple or multiple regression. The important ones are: random error term assumption, constant covariance assumption, zero multiple collinearity assumption, zero autocorrelation assumption, and stationary time-series assumption. These assumptions have not been true for macroeconomic data, so the econometric estimation suffers from a number of problems, *e.g.* the endogeneity problem, the heteroscedasticity problem, the multicollinearity problem, and the autocorrelation problem. Any standard econometric textbook explores these problems, generally leading to the view that they are no unresolvable challenges to econometricians because a number of tools or models can be used to address these problems 'thoroughly'. For example, a number of macroeconomic techniques are devoted to addressing the stationarity and spurious correlation issues.

This approach indeed looks rigorous, but it is only a game confined within the box of statistics theory. By using random variable/disturbance as an example, we examine the possibility that the strict assumptions proposed by macroeconometricians can hold. Then, we reveal the nature of using econometric method to bring the estimation conditions close to the strict condition. Finally, we illustrate a number of problems related to econometric practice.

Breaking the trinity of random experiment, random variable and random disturbance

The assumption of random variables and random disturbance is vital for valid model estimation and testing, and thus this assumption is crucial for macroeconometrics. Although macroeconometricians have tried to downplay the importance of the validity of this assumption, we need to investigate this thoroughly because, as Friedman (1953) noted, incorrect assumption will lead to incorrect theory.

Haavelmo (1944) did rigorously define 'random variables' but did not use the term according to his own definition. He gave two types of systems of random variables. One refers to 'random sampling' (Haavelmo, 1944:46). This type is clearly unrelated to macroeconomics because time-series are not survey data. The other type refers to 'stochastically independent' variables, which obey the joint elementary probability law: the probability of concurrently occurrence of all events x_1, x_2, \dots, x_n is the multiplication of probability of occurrence of each events. Mathematically:

$$p(x_1, x_2, \dots, x_r) = p_1(x_1) * p_2(x_2) \dots * p_r(x_r)$$

where p, p_1, p_2, \dots, p_n are the probability functions.

To satisfy the above equation, the probability function for each of the r variables, i.e. $p_1(x_1), p_2(x_2), \dots, p_r(x_r)$, must be independent. In econometric jargon, each variable must have an independent dimension in space R , or space R is of r -dimension. In non-statistical language, the probability of each event does not affect the other. However, this definition has never been used as the criterion for the assumption of a random variable in a macroeconomic model. The assumption is made simply because of the need for a macroeconomic model rather than its being based on evidence.

Trying to justify that the probability theory can be applied to time-series and that a complete list of significant factors is not necessary, Haavelmo (1943) stated that 'there can be no harm in *considering* economic variables as stochastic (random) variables having certain distribution properties' and 'only through the introduction of such notions are we able to formulate hypotheses that have a meaning in relation to facts' (Haavelmo 1943:13, italics added). Apparently, Haavelmo was simply 'considering' that economic variables are random variables because he *needed* this assumption. When illustrating the way to estimate a linear consumption-income function, Haavelmo (1943) added two random disturbances after stating 'to make it a real hypothesis to be tested against facts, we *have to* introduce some random variables' (Haavelmo 1943:17, italics added). Nowadays, the assumption of random disturbance is simply used by theoretic or applied econometricians even if they know the assumption is not true in reality. The reason is that they *need* this assumption so that estimation and testing of the macroeconomic model is valid. Once the economic variables are considered random, it is fairly easy to dismiss Keynes' request for a complete list of variables because

the probability of one variable does not affect that of another. The reality is that most economic variables cannot satisfy the definition of the random variable because of multicollinearity. Thus, Haavelmo's assumption is untenable.

It is arguable that the currently popular econometric estimation method – the Bayesian method (for an introduction to Bayesian method, see Bolstad 2007) – may be free of the problem because there is no random disturbance in it. However, the assumption of random disturbance is embedded in the Bayesians' assumption of random coefficients. Since the coefficients are random, Bayesians must calculate the mean of the coefficients – an equivalent to the point estimates in traditional econometric estimation. Random disturbance has to be assumed when Bayesian econometricians apply the law of large numbers to calculate the mean and variance of random coefficients. In this calculation, the random disturbance assumption is converted or equivalent to the Bayesians' assumption of random coefficients.

When can the assumption of random variable/disturbance be true? For this, it is necessary to trace back to the origin of statistical theory. The random disturbance/variable assumption originally used in statistics is not a mysterious assumption resulting from the imagination of a genius statistician. The assumption comes directly from and can be proven by random experiments: random disturbance is the difference between the outcome of each random experiment (*e.g.* coin tossing) and the statistical mean of a large number of random experiments. Random disturbance is caused by random variables (*e.g.* the movement of molecules in the air in the case of coin-tossing experiment). In other words, random variable, random disturbance, and random experiment are a trinity: when we use the concepts of random variable and random disturbance we imply that random experiments are conducted. If the conditions for random experiments do not hold, the assumption of random disturbance/variable is invalidated and, thus, all econometric estimations based on this assumption are invalid. This is exactly the case for time-series modelling.

In practice, the assumption of random variable/disturbance is not only simply assumed by macroeconometricians but is also used as the black/magic boxes or trash bins in macroeconometrics – any excluded (omitted, unwanted or unexplained) factors being thrown into it. The treatment of uncertainty is a good example. Macroeconometricians are very interested in modeling 'uncertainty' and use it as an explanation for

treating time-series data as experimental data (e.g. the GDP next year is random because it can be affected by uncertainties). Since an uncertain event (e.g. cyclones) may 'randomly' happen at any time in the future, so macroeconometricians think 'uncertainty' is random and, hence, it can be accommodated by random disturbance. However, in order to include the future uncertainty in the disturbance, one must judge if uncertainty is random according to the 'random' concept in the statistical theory. That is, one must do random experiments. There is no way to do random experiments regarding uncertainty, however, so the claim of 'random uncertainty' is purely based on macroeconometricians' imagination. Since 'uncertainty' is not a random variable with certain distribution, there is no way macroeconometricians can model 'uncertainty' successfully.

Detaching probability theory from its evidence

Although the strict conditions for a macroeconomic model are never met, this never seems to bother macroeconometricians because they claim that they can use a battery of econometric tools to bring the conditions close to the strict conditions (e.g. bring a non-random error term close to a random one). If this claim is true, does this make macroeconomics valid? On the surface, it might appear this approach is logical and rigorous. Thinking one step further, however, we can find that this is not the case.

The key lies in the source and implication of the strict assumptions made in econometrics. Where do the strict assumptions for regression analysis come from? They come from statistical theory. However, any theory originates from reality and will be tested by reality. Statistical theory is no exception, as it comes from and can be examined by random experiments. The strict assumptions proposed by econometricians are actually the features of random experiments. If these strict conditions are not satisfied, then this indicates that the experiments are not random, *i.e.* the condition for random experiments is not satisfied. In this case, the statistical theory is invalidated, and so is the regression. Macroeconometricians can use econometric tools to perform a 'cosmetic surgery' to bring the data conditions close to the strict assumptions, but this practice does not bring a non-random experiment close to a random experiment. Using an analogy, covering a wolf with a sheep skin does not

make the wolf a sheep. The ‘cosmetic surgery’ by macroeconometricians reminds the author of a scene in the movie ‘Cinderella’: Lady Tremaine tries to fit her daughter’s foot into the glass slipper by cutting off her toes!

Scientific illusion of statistical tests

In the effort to recover confidence in macroeconometrics after the prediction failure of many macroeconomic models, rigorous statistical testing is highly recommended. For example, Hendry (1993) suggested that rigorously tested models would greatly enhance any claim to be scientific. Are the statistical tests scientific enough to save macroeconometrics?

One big problem about statistical tests is that the test criteria (i.e. critical values) are established based on artificially generated random series. Since these artificial data come from random experiments (*e.g.* Monte Carlo simulation), the tests and the criteria are proven valid for artificial data. From this point of view, macroeconometrics can be said to be scientific on artificially generated data. However, because the real macro time-series do not satisfy the definition of random experiment, the tests and criteria are invalidated when they are applied to macroeconomic data. This is why Keuzenkamp (1995) claimed that statistical tests are not genuine.

In practice, statistical tests in macroeconomics appear to be objective but are often quite subjective. What will an econometrician do if the indication from the test contradicts common sense or economic theories? Most people will choose common sense over tests. Moreover, there are a number of tests for the same problem, *e.g.* many kinds of tests for unit root, for autocorrelation, for heteroscedasticity, for endogeneity, and for cointegration. It is not surprising that the different tests do not agree with each other. What would an econometrician do when faced with this situation? He/she has to make a decision. Most likely, he/she will select the model that can produce desirable results.

The GFC destroyed the macroeconometricians’ illusion of the power of statistical tests along with the ability of their models to predict the future. Consequently, even the macroeconomic models built by econometric leaders who advocated statistical tests were rejected. These econometric leaders were unwilling to admit the failure of their statistical tests as well

as their econometric models; instead, they rendered an explanation that the underlying distribution of the shock changed during a recession. Based on this explanation, the econometric leaders were effectively stating that even a correct econometric model with robust statistical tests is unable to predict the future.

The inability to solve the spurious correlation problem within a statistical framework

In a broad background, the econometricians' untenable confidence in statistical tests results from their conviction that they can use the statistical framework to solve reasoning problems in economics. In the statistical framework, the variable name for a particular time-series (e.g. real GDP data) is of no relevance, so a statistical theory can identify if two time-series A and B are correlated but the theory is unable to determine whether a correlation makes sense or not. By spurious correlation, we mean that the correlation does not indicate any causality. This is a logical judgement which cannot be solved within a statistical framework but needs to be solved by experiments and logic reasoning. Thus, any attempt to overcome the problem of a spurious correlation by developing statistical tests is misguided and futile. This is demonstrated in the macroeconometricians' long journey of fighting against spurious correlation.

The problem of nonsense correlation or spurious regression has a long history. Yule (1926) found high nonsense correlations between the standardized mortality and the proportion of Church of England marriages to all marriages during 1866-1911. Hendry and von Ungern-Sternberg (1980) demonstrated an almost perfectly fitted curve between UK inflation and the cumulated rainfall. Macroeconometricians always try to solve this problem within a statistical framework. Yule (1926) found that, in the case of nonsense correlation, there are strong serial correlations in the estimated residuals, so he thought the autocorrelation in the estimated residuals must be the reason for a nonsense correlation. This reasoning is fatally flawed because a feature (e.g. deflation) associated with something (e.g. economic recession) is not necessarily the cause of something.

Yule's invalid reasoning was followed by his successors and a large body of literature has been devoted to testing and modelling non-stationary

time-series. Is this practice able to overcome the problem of spurious regression? As stated previously, non-stationarity and autocorrelation are the symptoms (or indicators) that the macroeconomic time-series data do not satisfy the requirement of the probability theory. It is impossible to make a regression valid simply by artificially fixing the symptoms without addressing the cause. Moreover, even if a cointegration or first-differencing model could overcome the non-stationary issue and also could provide valid estimation, the modeller cannot guarantee that the revealed correlation indicates causality and thus the modeller cannot rule out that the correlation is not spurious. The logical mistake that macroeconometricians have made here is that they changed the task of testing and avoiding nonsense correlation to the task of detecting and avoiding autocorrelations.

The detrimental impact of macroeconometrics and the way forward

Nowadays, a popular expression regarding the performance of econometrics is: 'all models are wrong but some are useful', or simply 'harmless econometrics'. But is macroeconometrics useful and harmless? At first sight it may appear so. Getting some quantitative prediction, even if only approximate, could help on the adjustments to economic policy. Moreover, whether the modelling results/projections are right or wrong, the modelling prediction has little impact on the real-world economy. Contrary to these comforting claims, it can be demonstrated that macroeconometrics can cause harm in a number of ways. First, it results in a great waste of resources. Many institutions are engaged in macroeconomic modelling; time-series analysis is taught in most universities; and macroeconomic models proliferate in academic journals. All these activities waste time and money because the outcome of macroeconomic modelling is of little use due to the serious flaws in this approach. Secondly, the projection from macroeconomic models about the economy can be misleading and result in complacency. This is harmful because people are vulnerable when they are unprepared for disasters. To make things even worse, when economic recessions (such as the GFC) strike, econometricians provide no way for the public to react to the economic shock. Last but not most importantly, macroeconometrics delays or even suppresses more useful research in

economics. If macroeconometrics had not dominated macroeconomics for over sixty years, there was a possibility that the cause of and remedy for economic recessions might have been found and the GFC might have been avoided.

Compared with its indirect damage on the economy, the impact of macroeconometrics on scientific research work is more devastating because macroeconometricians have hijacked the well-established scientific research goal and method. The goal of scientific research is to find causality, truth, or true mechanism, which can be used to explain phenomena, predict future events, and provide solutions to existing problems. The established method to achieve the goal of scientific research is an ongoing process of 'data-hypothesis-theory'. From experimental or empirical data we can derive a hypothesis which can be further developed into a theory. The theory will be examined by newly available data and this will lead to a new hypothesis and an improved theory. In principle, this procedure goes on continuously and leads us closer and closer to truth. However, macroeconometrics involves a totally different goal and different means to achieve the goal – to predict the future by producing a close fit for historical data. In this approach, truth is unimportant and dispensable. As previously argued, this research method is fundamentally flawed. A prediction is reliable only when it is based on a true mechanism.

What then is the way forward for macroeconomic study? No silver bullet can be offered because the essential problem is that macroeconomic data are aggregate non-experimental time-series data which cannot satisfy the 'other things being equal' requirement in economic theory or 'under approximately the same conditions' requirement in probability theory. It is acknowledged that relying on time-series data is not unique to the economics discipline. Time-series data are also used for many macro-level natural and social science studies, *e.g.* studies of the environment, earthquakes, volcanos, wild animals, and astronomy. To the author's knowledge, however, no study uses the approach advocated by macroeconometricians: using tools to satisfy the statistical requirement artificially. Scientific prediction (*e.g.* weather forecasting) exists but it is based on proven theories or laws, *e.g.* the inertia law in physics. By contrast, economic predictions based on correlation between two variables or multiple correlations contain no true mechanism and thus are not scientific.

This practice in macro-level natural science may shed some light on research in macroeconomics. Correlation and simple multiple regressions can still be useful auxiliary tools to examine a theory, but caution must be taken in interpreting the results. When these tools are used, it is assumed that the effects of other variables are negligible. In fact, many other factors may play significant roles in data, so the empirical results can only be indicative. As a result, the correlations or estimation results cannot prove or disprove a theory or predict the future with confidence, but they can indicate how far away the theory is from the data. This kind of 'informal' approach is not the anti-theory approach suggested by Summers (1991). On the contrary, statistical theory supports this approach. Since the time-series data do not satisfy the condition of random experiments, what we can do is to be aware of the impact of other factors in the interpretation of modelling results.

After obtaining the empirical results of correlation or simple multi-regression, researchers need to investigate the difference between the theory and empirical results in order to discover the causes of the difference. In this way, the difference between theory and data can help the researcher to reform and refine the theory. The refined theory is then subject to data testing again. Through multiple procedures, both from theory to data and from data to theory and by using the combined induction and deduction of logical reasoning, the gap between data and theory can be reduced and, more importantly, our understanding will become closer to the truth.

This approach can be illustrated by the evolution of consumption theory. When Keynes' consumption function was confronted by data, the results were mixed. The aggregated time-series estimation showed a marginal propensity to consume (MPC) around 0.90, which implies a unit income elasticity of consumption and a constant saving rate in the long run. However, the studies based on household survey data showed a MPC in the range of 0.60 to 0.80 (Bunting, 1989). This inconsistency in empirical results led to the new theories on income and consumption, namely the relative income hypothesis posited by Duesenberry (1949), the permanent income hypothesis posited by Friedman (1957) and the life-cycle hypothesis posited by Modigliani (1986). Although these theories were supported by cross-sectional studies, they could not explain the high MPC in the time-series study. Bunting (1989) argued that the comparison between the results from the aggregate time-series study and those from the cross-sectional study was not valid. By dividing the

aggregate data by the number of households each year and using the ungrouped dataset of cross-section data, Bunting substantially reduced the gap of MPCs from time-series data and from survey data. This example also shows that, in order to bridge the gap between theory and data, we need use our logical reasoning power to do a rigorous job on both theory and data. The emphasis on research design suggested by Angrist and Pischke (2010) and the broad-based analysis with both quantitative and qualitative research suggested by Babones (2014) and Birks (2015) are steps in this direction.

Conclusions

This article has briefly reviewed the history of the rise of macroeconometrics and identified fundamental flaws in its approach. Based on the way time-series data are used in natural science, it has suggested a better way to conduct research in macroeconomics. That is, multiple procedures both from theory to data and from data to theory with combined induction and deduction of logical reasoning. While correlation and multiple-variable regression can still be used, the researcher must be conscious that these empirical results are only indicative and need to be interpreted carefully because the variables outside of the model may also have impacts on the results.

To answer Hendry's (1980) self-addressed question 'is econometrics science or alchemy?', the author would say macroeconomic theory is a rigorous 'science' on artificially-generated series, so it is irrelevant to the real world. 'Modern' macroeconomic practice, which aims to predict the future by fitting the historical economic time-series, is statistical alchemy at the last stage. Even after the GFC delivered a loud and clear message that macroeconomic models were unable to predict the future, macroeconomicians have not admitted their failure and continue a practice that is equivalent to 'transforming lead into gold'. By pointing out the baseless nature and fundamental problems of macroeconometrics, the author hopes to see an end to the seemingly harmless but potentially disastrous practice of statistical alchemy. Practical economists would be better engaged in utilizing their valuable resources to solve real economic problems.

Samuel Meng is a senior research fellow at the University of New England, Australia. He would like to thank the editor and three anonymous referees for their useful comments, which sharpen the arguments and make the article more concise and rigorous.

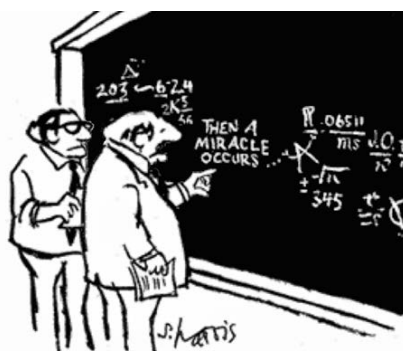
xmeng4@une.edu.au

References

- Angrist, J., and Pischke, J., (2010) The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24(2): 3-30.
- Babones, S., (2014) *Methods for Quantitative Macro-Comparative Research*, Los Angeles: Sage Publications.
- Birks, S. (2015) *Rethinking economics: from analogies to the real world*. Singapore: Springer.
- Bolstad, W., (2007) *Introduction to Bayesian Statistics* (2nd edition), Wiley publishing.
- Brown, P., (1972) The Underdevelopment of Economics, *Economic Journal*, 82:1-10.
- Bunting, D., (1989) The Consumption Function "Paradox", *Journal of Post Keynesian Economics*, 11(3):347-359.
- Cooley, T., and Leroy, S., (1985) Atheoretical Macroeconometrics: a Critique, *Journal of Monetary Economics*, 16: 283-368.
- Duesenberry, J., (1949) *Income, Saving, and the Theory of Consumer Behaviour*, Cambridge, MA: Harvard University Press.
- Freedman, D., (1995) Some Issues in the Foundation of Statistics, *Foundations of Science*, 1: 19-83.
- Freedman, D., (1999) From Association to Causation: Some Remarks on the History of Statistics, *Statistical Science*, 14: 243-258.
- Freedman, D., (2005) *Statistical Models: Theory and Practice*, Cambridge University Press.
- Friedman, M., (1948) *Memorandum about the Possible Value of the CC's Approach toward the Study of Economic Fluctuations*, Rockefeller Archive.
- Friedman, M., (1951) Comment, in *Conference on Business Cycles*, p. 107-114, New York: National Bureau of Economic Research.
- Friedman, M., (1957) *A Theory of Consumption Function*. Princeton, NJ: Princeton University Press.
- Frisch, R., (1934) *Statistical Confluence Analysis By Means of Complete Regression Systems*, Oslo: Institute of Economics.
- Haavelmo, T., (1944) The Probability Approach in Econometrics, *Econometrica* 12(Supplement), iii-115.

- Haavelmo, T., (1943) Statistical Testing of Business Cycle Theories, *Review of Economics and Statistics*, 25:13-18.
- Hansen, Lars Peter and Singleton, Kenneth J. (1982) Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50 (5), 1269-86, Sept.
- Hansen, Lars Peter and Singleton, Kenneth J. (1983) Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy* 91 (2), 249-65, March.
- Hendry, D., (1980) Econometrics: Alchemy or Science? *Economica*, 47: 387-406.
- Hendry, D., (1993) *Econometrics: Alchemy or Science? Essays in Econometric Methodology*. Oxford: Blackwell.
- Hendry, D., and Mizon, G., (2014) Unpredictability in Economic Analysis, Econometric Modelling and Forecasting, *Journal of Econometrics*, 182: 186-195.,
- Hendry and von Ungern-Sternberg, T., (1980) Liquidity and Inflation Effects on Consumer's Expenditure, In *Essays in the Theory and Measurement of Demand* (Deaton, ed.), Cambridge University Press.
- Keuzenkamp, H., (1995) The Econometrics of the Holy Grail, *Journal of Economic Surveys*, 9:233-248.
- Keynes, J., (1939) Professor Tinbergen's Method, *Economic Journal*, 49: 558-568.
- Keynes, J., (1940) Comment, *Economic Journal*, 50: 154-156.
- Kling, A., (2011) Macroeconometrics: the Science of Hubris, *Critical Review*, 23:123-133.
- Krugman, P., (2009) How Did Economists Get It So Wrong? *The New York Times*, Sept. 6.
- Leamer, E., (1978) *Specification Searches: Ad Hoc Inference with Non-experimental Data*. New York: John Wiley.
- Leamer, E., (1983) Let's Take the Con out of Econometrics, *American Economic Review*, 73:31-43.
- Leontief, W., (1971) Theoretical Assumptions and Non-observed Facts, *American Economic Review*, 61: 1-7.
- Liu, T., (1960) Under-identification, Structural Estimation, and Forecasting, *Econometrica*, 28: 855-865.
- Lucas, R., (1976) Econometric Policy Evaluation: A Critique. In, *Carnegie Rochester Conference Series on Public Policy* (Karl Brunner and Alan Meltzer 9eds.), Vol.1, 19-46.
- Modigliani, F., (1986) Life Cycle, Individual Thrift, and the Wealth of Nations, *American Economic review*, 76:297-313.
- Moore, (1917) *Forecasting the Yield and Price of Cotton*. The Macmillan Company: New York.
- Moore, H., (1911) *Laws of Wages: An Essay in Statistical Economics*, The Macmillan Company: New York.
- Nell, E., and Errouaki, K., (2013) *Rational Econometric Man, Transforming Structural Econometrics*, Edward Elgar, Aldershot.
- Pesaran, M., (1987) *The Limits to Rational Expectations*, Oxford: Basil Blackwell.
- Sims, C., (1980) Macroeconomics and Reality, *Econometrica*, 48: 1-48.

- Skidelski, R., (2009) How to Rebuild a Shamed Subject. *Financial Times*, Aug. 6.
- Solow, R., (2010) Statement of Robert M. Solow, in *Building a Science of Economics for the Real World*, U.S Government Printing Office, p.12-15, <http://www.gpo.gov/fdsys/pkg/CHRG-111hhr57604/pdf/CHRG-111hhr57604.pdf>
- Spanos, A., (2011) Foundational Issues in Statistical Modelling: Statistical Model Specification and Validation, *Rationality, Markets and Morals*, 2: 146-178.
- Summers, L., (1991) The Science Illusion in Empirical Macroeconomics, *Scandinavian Journal of Economics*, 93: 19-148.
- Tinbergen, J., (1940) On a Method of Statistical Business Cycle Research, A reply. *Economic Journal*, 50:141-154.
- Tinbergen, J., (1939) *Statistical Testing of Business Cycle Theories: Part I: A Method and Its Application to Investment Activity*, Agaton Press, New Work.
- Wallis, K., (1977) Multiple Time Series Analysis and the Final Form of Econometric Models, *Econometrica*, 45: 1481-1497.
- Wickens, M., (1982) The Efficient Estimation of Econometric Models with Rational Expectations, *Review of Economic Studies*, 49: 55-68.
- Worswick, G., (1972) Is Progress in Economic Science Possible? *Economic Journal*, 82:73-86.
- 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): 1-64.
- Ziliak, S.T., and D.N. McCloskey. (2008) *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor (MI): The University of Michigan Press.
- Ziliak, S.T., and D.N. McCloskey. (2013) We Agree That Statistical Significance Proves Essentially Nothing: A Rejoinder to Thomas Mayer, *Econ.*



"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."