# Measuring What Exactly? A Critique of Causal Modelling in Atheoretical Econometrics

**MSc Thesis** *(Afstudeerscriptie)*

written by

**Sebastian N. Køhlert**
(born October 22nd, 1993 in Esbjerg, Denmark)

under the supervision of **Dr. Federica Russo**, and submitted to the Examinations Board in partial fulfillment of the requirements for the degree of

**MSc in Logic**

at the *Universiteit van Amsterdam.*

| **Date of the public defense:** | **Members of the Thesis Committee:** |
|---|---|
| *June 9th, 2021* | Dr. Maria Aloni (Chair). |
| | Dr. Hein van den Berg. |
| | Dr. Dingmar van Eck. |
| | Dr. Federica Russo (Supervisor). |

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

## Acknowledgements

First, I would like to thank my supervisor, Federica Russo, for our fruitful and instructive conversations during my thesis project. The comments I received along the line helped me come a long way. Apart from that, I would like to thank the thesis committee for their flexibility in tough times. Furthermore, I would like to thank my family, girlfriend and my colleagues. I am sure it has not been easy to listen to me going on about my thesis all the time. However, without them, I would not be where I am now. Lastly, I want to thank my grandmother, who has always been there for me whenever I needed her. May you forever rest in peace.

# Abstract

An important part of econometrics is modelling causality. One way of getting causal predictions is to rely on data-driven models. This tradition is also known as atheoretical econometrics. Thus, atheoretical econometrics represents a range of methods that use models to infer causal relations directly from data. This is contrasted to theoretical econometrics that relies on economic theory. The main problem in getting causal knowledge from data in econometrics is that the investigator often faces large volumes of conflicting results from different models and that these models are highly sensitive, which conflict with one of the main goals of econometric modeling, obtaining stable outcomes. In this thesis, I strengthen the case against using atheoretical econometrics to infer causal relations from data, based on its inability to generate reliable evidence, due to its high sensitivity and lack of stable outcomes. I argue that we can understand econometrics models as measuring instruments not that different from thermometers and clocks, but what characterizes these measuring instruments are their high level of stability in outcomes. By relying on new literature in measurement theory, I show that the main problem in athereoteical econometrics occurs due to a misunderstanding of how measurement generates evidence and stable outcomes. In the end, I conclude that the evidence from Granger models is hardly strong enough to make any strong inferences based on it and argue that calibration may provide a way to bridge the atheoretical, and the theoretical view of econometrics.

# Contents

# 1

# Introduction: Causality in Econometrics

Vi veri veniversum vivus vici.

Faust

The famous motto of London School of Economics is *rerum causas cognoscere*, which refers to the critical importance of knowing the causes of things. Most actions in life are guided by one's beliefs about causes. For example, should I eat another piece of the pizza in front of me, knowing it might be poor for my health? Or should governments lower interest rates to stimulate consumer spending during economic turmoil? To better understand our environment and make better decisions, understanding the causes of things is all that matters. The same holds for economics. What should be done in light of the COVID-19 pandemic? Should there be stimulus spending or no stimulus? Getting economic decisions right is crucial. To ensure governments make the correct decisions, it is imperative that the causal relations in the economy are understood. It may seem like a trivial point that causality matters in decision-making. However, over the last century, this proposition was not always accepted in the sciences and philosophy. Traditions inspired by the influential works of David Hume were suspicious of the concept of causality, viewing all metaphysical notions as something that should be avoided. Hume's own project was to eliminate causal concepts by reducing them to regularities. For as Hume noted, one cannot observe causes; all one can observe is the constant conjunction of causes and effects. Hence, causality is wholly about regular associations between different events – in other words, regularities. The anti-metaphysical sentiment in 20th century philosophy was best captured by Russell's comparison between the monarchy and causality. One should use all means to avoid the use of causal notions. Following the works of Hume, Russell, and the logical empiricists, the main position regarding causality was that it is metaphysical and ambiguous, and as a consequence had no place in the sciences. The founding fathers of econometrics were not reluctant to engage in causal discussions, in spite of its development in the heyday of logical empiricism, as noted in M. S. Morgan (1990). The object of econometrics was defined in the first issue of Econometrica as:

> economic theory in its relation to statistics and mathematics (...) unification of the theoretical-quantitative and the empirical-quantitative approach to economic problems [(Frisch 1933), p. 1].

The main goals of those working in this early tradition of econometrics, including Jan Tinbergen, Tygve Haavelmo, and Tjalling Koopmans, was to develop a method that could identify causal relations from data with the help of theory [see Haavelmo (1944), T. C. Koopmans et al. (1950), and Tinbergen (1939)]. Other good examples are the works of H. A. Simon (1952) and H. Simon (1953), in which Simon makes use of the word cause multiple times. In H. Simon (1953), Simon presents a formal definition of 'causal order', in which causal order is equivalent to causal structure as that concept is used in the works of Tinbergen, Haavelmo, and Koopmans. It should be noted, however, that the works of Simon arguably mark the beginning of a downhill slope in discussions of causality in econometrics. Because Simon's work was not inconsistent with empiricist sentiments. However, because Simon might have argued that causality was a useful concept in the sciences, one of the main aims of Simon's work was to ensure that the concept was empirically respectful by operationalising it. Hoover (2004) presents a detailed graph that shows a decline in discussions of causality in econometric papers from this period and onward.

However, the taboo regarding discussions of causality has since disappeared in both the philosophical and econometric literature. Logical empiricism no longer dominates the econometric and philosophical literature. The works of Patrick Suppes, especially Suppes (1973), has noted the importance of the concept of causality. In econometrics, the development of Granger causality, gave rise to a new body of literature on causality in econometrics, see Granger (1969), Granger (1980), Granger (1988), Granger (1999), Sims (1972), and Sargent and Sims (1977). As I discuss later in this work, Granger's concept of causality is closely related to that of Suppes and is Humean in nature, since the temporal factor plays a crucial role in Granger's definition of causality. As I discuss at the beginning of this thesis, Granger's definition can be contrasted with that of Cowles Commission, which suggests that causes are supplied by economic theory. Apart from that, Cowles' position permits cause and effect to be simultaneous. The resurgence of discussions of causality can also be seen in the fact that a special issue of the Journal of Econometrics that focuses exclusively on causality in econometrics was released. The issue of the Journal of Econometrics provided several articles about how one should define causal relations and which methods should be applied to identify them. This shows that epistemological questions and questions regarding how to establish reliable grounds for the enterprise of econometrics are still important in econometrics. Thus, offering clear answers to the following questions is important:

1. How does one find out about causal relations?
2. When can one infer causality?

This thesis addresses these questions, given two treatments of causality, (i) a theoretical approach that is based on mechanisms and (ii) an atheoretical view, in which what is causal is determined by a set of instruments based on Granger's definition of causality. One (i) makes use of apriori theory and the other (ii) does not. Thus, the main goal of this thesis is twofold: first, I intend to contribute to the growing discussion of causality in econometrics; and second, I want to include the philosophy of measurement in this discussion, based on the idea that instruments in econometrics are not that different from thermometers, clocks, and other instruments humans use on a daily basis. The goal is to show that this offers another argument against atheoretical econometrics, in addition to a long list of other issues in the atheoretical approach, since theory play a vital role for these instruments to generate stable outcomes and reliable evidence. Further, it is important to remain cognisant of the practical effects of economic studies. For example, take a study like whether money causes income, which I will get back to in Chapter 4. How one interprets such a study depends heavily on the answers to the two questions presented above. Whether money does cause income might affect a large group of people through the policy that will be founded on it; however, as this thesis demonstrates, the answer one arrives at, as well as how one should interpret that answer, is highly conditional on the answer to question 1 and ones ontological and methodological commitments. It is not merely a neutral matter of letting the data speak or believing the facts. As Frisch also argued:

> The schools [empirical schools], however, had an unfortunate and rather naive belief in something like a theory-free observation. Let the facts speak for themselves. The impact of these schools on the development of economic thought was therefore not very great, at least not directly. Facts that speak for themselves talk in a very naive language [(Frisch 1970), p. 5].

This thesis starts in Chapter 2 by answering the first questions with two different approaches found in the literature. I transition to show the noticeable philosophical differences in the two views and provide a short introduction to both. After this, I question the atheoretical econometrics ability to provide a reliable foundation for inferring causality. I argue that it rests at a questionable idea that measurement is possible without theory. I argue in Chapter 3, with a sound basis in the philosophy of measurement, that this is not the case. Measurement relies on theory and is not reducible to relations among observables. In Chapter 4, I discuss how it is possible to infer causality by instruments. I defend the calibration view and note its resemblances to another view in the epistemology of evidence, evidential pluralism.

## 1.1 Aim

This thesis aims to provide a clear presentation of a particular kind of methodological approach to econometrics. I contrast this view to another methodological approach in econometrics to show the philosophical differences between the two. The goal, in the end, is to provide a well-founded criticism of the former, provide conceptual clarity in general, and link the study of measurement in econometrics to the literature in the philosophy of measurement. I need to state in the beginning that it's a particular kind of attitude to these models that I criticise; the perspective that these models provide a neutral way to knowledge. I find that mindset quite dangerous, especially in a discipline that gives policy advice; thus, I intend to show that there are nothing neutral about it, that background theory matters, and that philosophical inquiries matter too. I do not, however, intend to criticise every use of these models since they can be useful in their own right, just not for the purpose a lot of economists assign to them.

## 1.2 Structure of the Thesis

1. **Chapter 2:** This chapter provides a critical survey of causal models in econometrics, including a brief introduction to traditional textbook econometrics before transitioning to contemporary econometrics, mainly examining the works of C.J. Granger's, C. Sims, and T. Koopmans. This chapter begins by outlining the differences in the philosophical foundations ranging from textbook econometrics to contemporary econometrics. I argue that traditional econometrics is non-reductionist, which contrasts with contemporary econometrics, which I consider reductionist, based on the idea that modelling causes can be reduced to probabilities or probabilistic dependecies, see Moneta and Russo (2014a). From there, I continue chronologically by first introducing traditional econometrics, which I argue is a theoretical approach to causality in econometrics. Then, I transition to the main contemporary approach to causal modelling in econometrics, which I characterise as atheoretical.

2. **Chapter 3:** This chapter examines the idea behind numerical representation -that is, measurement- in economic models, mainly inspired by M. Boumans, L. Mari, and E. Tal. I pose two problems for econometrics, one of them being the problem of passive observation and the other how to generate evidence based on passive observation. The main question here will be, is measurement based on observation? Or is something else needed? I suggest that more is required, contrary to the idea that measurement without theory is possible, following E. Tal and K. Staley. However, one of the most popular analyses of measurement today, the representational theory of measurement, suggests that measurement is a homomorphic mapping of empirical relational systems to numerical relational systems; in other words, we can reduce measurement to relations between observables. The representational theory's relationship to econometrics and the problem of foundationalism in measurement is investigated. I emphasise the need to move beyond foundationalism, following the epistemological shift in measurement approaches, emphasizing the distinction between a reading and the measurement outcome. I argue that the measurement outcome is a range of acceptable values, consistent with both theoretical and statistical assumptions.

3. **Chapter 4:** In Chapter 3, I argued that as part of the shift from readings of instruments to outcomes, background theory is needed. Thus, the outcomes that measuring instruments provide, or what is often viewed as the evidence measuring instruments produce, is theory-laden. In this chapter, I investigate the implications of the theory- laden nature of evidence, including an investigation into what is needed to infer causality from econometric instruments. I first present a case study on whether money causes income to show the problems that follow without an acknowledgement that theory is needed. I note that the literature is inconclusive and point to the sensitivity of Granger tests. I then move on to argue that this means that acceptable empirical evidence in econometrics should be restricted by background theory. I further note that this means that one can conclude little from Granger tests without a sufficiently robust background theory, due to a problem of stability found in Granger tests. Lastly, I present the case for calibration in econometrics along the lines of Cooley (1997) and Kydland and Prescott (1996a), that makes use of both theory and data. I further note the resemblances to evidential pluralism noting (i) that neither probabilistic dependencies nor mechanisms are sufficient to establish causality in econometrics alone.

# 2

## Building Causal Models in Econometrics

*This chapter provides a critical survey of causal models in econometrics, including a brief introduction to traditional textbook econometrics before transitioning to contemporary econometrics, mainly examining the works of C.J. Granger's, C. Sims, and T. Koopmans. This chapter begins by outlining the differences in the philosophical foundations ranging from textbook econometrics to contemporary econometrics. I argue that traditional econometrics is non-reductionist, which contrasts with contemporary econometrics, which I consider reductionist, based on the idea that modelling causes can be reduced to probabilities or probabilistic dependecies, see Moneta and Russo (2014a). From there, I continue chronologically by first introducing traditional econometrics, which I argue is a theoretical approach to causality in econometrics. Then, I transition to the main contemporary approach to causal modelling in econometrics, which I characterise as atheoretical.*

> It is clearly a topic in which individual tastes predominate, and it would be improper to try to force research workers to accept a definition with which they feel uneasy. My own experience is that, unlike art, causality is a concept whose definition people know what they do not like but few know what they do like.
>
> C.W. Granger, 'Testing for Causality: A Personal Viewpoint'

Much of the development of how to properly model causality in econometrics has historically been linked to the problem of identification. The problem of identification is similar to the problem in the philosophy of science, known as the problem of underdetermination, or the Quine-Duhem Thesis [see for more S. Turner (1987), D. Turner (2005), Sawyer et al. (1997), and Yalçin (2001)]. For underdetermination appear precisely when the evidence is insufficient to identify which model to choose. Thus, both concern cases in which multiple hypotheses or relationships are compatible with the measurable statistical properties. Using economic terminology, I discuss the problem of identification throughout this thesis. The problem of identification has long been discussed in econometric literature, dating back to a Danish economist's publication at the beginning of the 20th Century. Although Macheprang did not use the precise term, he addressed a similar problem in Mackeprang (1906). Here, E. P. Mackeprang considered the problem of determining demand functions and demand elasticities. In calculating elasticities, Mackeprang considered a case involving the price of a given product at a certain time,$P_t$, and the demand of the same product, $D_t$. He calculated the price elasticities using a regression of $P_t$ on $D_t$ and vice versa. This yielded two different results: Mackeprang asked a question: which regression we should choose? He ultimately responded 'both' because he did not have a solution to the problem he had stumbled upon [For an English introduction to the works of Mackeprang, see Wold (1969a)]. Succinctly, the problem concerns how one isolates unobserved relationships in the variables of interest that have generated the data [for more on the problem of identification see T. Koopmans (1949)]; in other words, it asks how one chooses between competing possible relationships that are all compatible with the measurable statistical properties (e.g., correlation or covariation) of the data? The economists in the Cowles Commission traditionally chose to derive the solution from economic theory. Thus, causality was purely a theoretical relationship between variables postulated by economic theory. Economists in the later atheoretical tradition replaced the concept of causality from the Cowles Commission with a concept tied to statistical properties, Granger Causality, which could be tested with statistical tools. As Vining famously argued, 'statistical economics is too narrow in scope if it includes just the estimation of postulated relations' [(Vining 1949), p. 86]. Thus, the former is essentially a realist approach that relies on mechanistic evidence provided by economic theory and the latter a somewhat reductionist approach to the modelling of causality, that provide evidence of difference-making. I begin this chapter with a philosophical investigation into the causal models of econometrics. I begin by analysing the models themselves, distinguishing between a theoretical approach and an atheoretical approach to causal modelling, before turning my focus toward theory itself to examine what it is and how the two different traditions differ on theory. I then assess what this means for causal models in econometrics where I note that the theoretical approach uses mechanisms and the atheoretical approach and instrumental approach to discover causality. Lastly, I introduce both the theoretical and the atheoretical approaches to econometrics in greater depth.

## 2.1 Theory and Causal Models in Econometrics: Background

As James Heckman proclaimed, 'Just as the ancient Hebrews were "the people of the book", economists are "the people of the model"' [(Heckman 2000), p. 46]'. This emphasises how essential models are to economic science. This thesis' primary focus is the epistemological questions that arise in the epistemology of econometrics. For instance, how do we discover causal relationships in econometrics? How do we justify the discovery methods? When are such procedures correct? To establish reliable grounds for econometrics, it is ideal to focus on such questions alone. This section approaches econometrics from a philosophy of science perspectives in an attempt to uncover the philosophical commitments and foundations of modern econometrics. First, it should be noted that it is likely not possible to completely separate the epistemological question from the conceptual and ontological quibbles that underly it. To begin, we should therefore divide the problems concerning causality into three categories:

1. **Conceptual Analysis**. What does the term 'cause' mean?

2. **Ontological Analysis**. To which reality do causal relations refer? In other words, the ontological analysis of causality tries to answer the question, 'what is causality?'

3. **Epistemological Analysis**. How are beliefs about causal relations inferred? Or, in other words, the epistemological analysis of causality concerns how we learn about

causality.

The three questions are intertwined in multiple ways, assuming that there is no fixed meaning to the concept of causality. This does not solve the conceptual problem. Instead, a new one arises: how do we choose between multiple definitions of causality with a different meaning? The meaning of the term informs how we later approach its discovery. In order to make a discovery, we must have some awareness of what we are searching for. Furthermore, consider the ontological analysis, which includes questions such as whether causal relations exist independently of the observer, and, more importantly, whether we can reduce causal facts to non-causal facts. The latter has often been considered the central problem of the philosophy of causality [see Tooley (1990)]. A person who has an affirmative answer to this question is a reductionist. Denying the possibility of a reduction makes one a realist. I delineate two oppositional positions. One is the theoretical realist position to econometric model building. The second is the atheoretical reductionist approach. I do not suggest that these opposed positions divide neatly. However, I do find the delineation to leverage a better understanding of econometric analysis and the development of econometric thought in the 20th century. Even if the illustration provided in this section is somewhat problematic, these philosophical banners offer an improved understanding of the positions taken by the two sides. Here, it enhances understanding of the later methodological positions taken by the two opposed approaches to econometrics. Therefore, obtaining a better understanding of the philosophical foundation of theoretical and atheoretical econometrics provides the background necessary for a better understanding of why measurement, which will be discussed in Chapter 3, is essential and for understanding the evidence produced by such models, as chapter 4 discusses.

### 2.1.1 Models and Econometrics: Ontology and Epistemology

This section further expands on the metaphysical and epistemological properties of the two approaches I delineated in the introduction to this section. Following Granger (1999), I argue that there are two extremes in econometric literature on model building:

1. **Theoretical Econometrics**. The main view here is that theory should provide the structure of the empirical model. One may go so far as to claim that all residuals must have a theoretical explanation. This leaves little room for stochastics, uncertainty, and exogenous shocks in econometric models.

2. **Atheoretical Econometrics**. At the other extreme is the econometricians who claim that theory should play little or no role in the specification of an econometric model. Rather, we should build 'atheoretical models', which only analyse data by using the regularities found in it. The danger here is mining, 'particularly now that computing is both fast and cheap' [(Granger 1999), p. 18].

Following Lawson (1989), Moneta (2005b), Moneta (2005a), and Grabner (2016), I claim that the theoretical approach to econometrics is a realist one, and I show that this leads to a mechanistic approach to causality. Causal models are justified by pointing to the economic theory from which the mechanisms involved in the causal model are derived. On the contrary, I argue that what characterises the atheoretical approach to econometrics is metaphysical reductionism, or the idea that causal facts can be reduced to non-causal ones, in this case, statistical properties or probabilistic dependencies, see Moneta and Russo (2014a). This entails a certain epistemological reductionism, which helps to explain why most modern econometrics can be considered instrumentalist [for more see Giedymin (1976), Lawson (1989), Lagueux (1994), Moneta (2005b), Moneta (2005a), Reiss (2012), and Grabner (2016)]. What warrants the atheoretical nature of such an instrumentalist approach is the supposed neutrality of measurement itself, which I argue against in chapter 3. Later in this section, I examine the two approaches in greater detail. However, it should be noted here that, as argued in Moneta (2005b) and Moneta (2005a), the crucial question in econometrics is not the ontological question of whether causes can be reduced to regularities; instead, it is of which one is 'stable' and is best suited for the main objective of econometrics: to predict the future. By stable, I mean 'autonomous', denoting relations that are invariant to intervention [For more on the history of the concept of 'autonomy' in the history of econometrics, see Aldrich (1989)]. What macroeconomics desires is an autonomous relation between two parameters, say $A$ and $B$, meaning that manipulating A can enable the prediction of the outcome of $B$. This aligns closely with one of the critical goals of econometrics, policy intervention, and is closely connected to the reliability of some discovery procedures, $P$. If, for instance, a procedure $P$ chooses a certain autonomous relations $R$, the variability in outcomes becomes non-existent, making it easier to derive reliable conclusions and thereby infer true beliefs based on it. Thus, it is important that $P$ is not too sensitive, since that could prohibit $P$'s ability to pick out autonomous relations. As noted in the introduction to this chapter, the development on causal modelling depicted in the literature is closely connected to the debate of the problem of identification. This is primarily because the problem of stable relations is in turn closely connected to the problem of identification or, as argued in Moneta (2005b)[p. 298-99], 'The problem of identifying a structural model from a collection of economic time series is one that must be solved by anyone who claims the ability to give quantitative economic advice'. This subsection focuses on both approaches' philosophical foundations to provide a better understanding of the machinery behind the approaches to causal modelling in econometrics.

#### 2.1.1.1 Theoretical Econometrics

In econometrics' beginning in the previous century, its philosophy was that the quality of the data an economist had to work with was not high enough to stand alone. As noted in the editor's note to the first issue of Econometrica,

> Experience has shown that each of these three view-points, that of statistics, economic theory, and mathematics, is a necessary, but not by itself a sufficient, condition for a real understanding of the quantitative relations in modern economic life. It is the unification of all three that is powerful. And it is this unification that constitutes econometrics [(Frisch 1933), p. 2].

The first issue of Econometrica also provided one of the first definitions of an 'economic model' in the literature of econometrics, stating that a model is,

> A synthetic construction in which statistics, the assembly of observable facts, theory, the research of explanations of reality, and mathematics, the rigorous tool for the integration of facts and theory, are each constantly in service of the other [quoted from (Nell and Errouaki 2013), p. 158].

Thus, the job of the econometrician was to provide a bridge between theory and observable facts. Economic theory should postulate relationships between variables, and econometrics should measure the strength of these postulated relationships [see Moneta (2005b),Moneta (2005a)]. Hence, the model was a representation of a more general theory. The IS-LM model represented the economic theory presented by John Maynard Keynes in his General Theory. Paul Samuelson's models represented Ricardian economics, and the Cowles

Commission models represented Wallrasian general equilibrium theory. The most famous econometricians in this historic tradition were Mackeprang, Tinbergen, Klein, Haavelmoo, Koopmans, and Malinvaud. It is best summarised by the following passage from Klein: 'Without theory and other a priori information, we are lost', who also asked rhetorically 'I wonder why Sargent, Sims, and Geweke are trying to lead us away from the established path that was so long in being prepared?' [(Klein 1977) p. 208].

#### 2.1.1.1.0.1 Realism, Causality and Econometrics .

The theoretical approach to econometrics follows realism, as noted in the introduction to this subsection. It assumes that there are autonomous structures that are primary with respect to regularities, as argued in Moneta (2005b) and Grabner (2016). Thus, the theoretical approach is committed to the following ontological principle:

> **Realism (R)**: Causal claims exist independently of regularities.

Holding R does not exclude the possibility that statistical instruments can be useful in managing causality in econometrics. That said, most realists do claim that more is required, and statistical tools are insufficient. The most widely known approach to causality in this tradition and in econometrics was the Cowles Commission approach (CC). The CC approach argued that, as noted in Grabner (2016), Boumans (2010a), and Malinvaud (1988), mechanisms were that ingredient. For instance, Christ (1994a) felt that the theoretical approach to econometrics 'did not have much to say about the process of specifying models, rather taking it for granted that economic theory would do that, or had already done it'[1994a, p. 34], meaning that little attention was given to 'how to choose the variables and the form of the equations; it was thought that economic theory would provide this information in each case'[1994a, p. 33]. As argued by Koopmans,

> The analysis and explanation of economic fluctuations has been greatly advanced by the study of systems of equations connecting economic variables. The construction of such a system is a task in which economic theory and statistical method combine. Broadly speaking, considerations both of economic theory and of statistical availability determine the choice of the variables. [T. C. Koopmans et al. 1950, p 54].

It is exactly in the measurement of a system of equations that the problem of identification arises, as mentioned in the introduction. This is because the systems of equations can be written in multiple ways, thus 'Under no circumstances whatever will passive statistical observation permit [the econometrician] to distinguish between different mathematically equivalent ways of writing down that distribution' [(T. C. Koopmans et al. 1950), p. 64]. However, because the econometrician does not have any experimental control over the measured variables and instead observes them 'passively', 'the only way in which he can hope to identify and measure individual structural equations implied in that system is with the help of a priori specifications of the form of each structural equation'[(T. C. Koopmans et al. 1950), p. 64]. Historically, such a view is closely related to that of Keynes and was deemed an element of Keynesian macroeconomics by Lucas and Sargent in Lucas and Sargent (1981). There may be a certain 'irony in criticizing any econometrics as Keynesian, given Keynes's own scepticism of econometrics. (...) What is of course true is that most builders of large-scale macroeconometric models classified themselves as Keynesian', as noted in [(Hoover 1988b) , p. 270]. Despite this, these models were indeed Keynesian because they resembled Keynes' view of causal structures. Keynes was sceptical of econometrics, exemplified in his criticism of Tinbergen [see J. Keynes (1939)].

In view of this, it is ironic to refer to econometrics as 'Keynesian', but the CC approach nonetheless adopted the non- reductionist perspective found in Keynes' early criticism of econometrics. As Keynes argued, in order to apply statistical tools, what is needed is 'not merely a list of the significant causes, which is correct so far as it goes, but a complete list?' and 'it is necessary that all the significant factors should be measurable, this is very important' [(J. Keynes 1939), p. 560-561]. The former is not possible due to the problem of omitted variables, which may lead to an incorrect estimation of the quantitative importance of the included variables. The latter is problematic because, according to Keynes, economics includes multiple factors which are not measurable. This led Keynes to reject the econometric method applied to business cycle theory. Further, Keynes shared a criticism of econometrics often found in classical economic theory dating back to Mill [see Mill (1906), Mill (1836), and Hausman (1981)]. It posits that applying statistical tools to discover causal relationships is impossible because the underlying mechanism that produces the data, also known as the data generating process, is intertwined with other mechanisms and can therefore be difficult to isolate using statistical tools [see J. Keynes (1939)]. This led Keynes to conclude that,

> If so, this means that the method is only applicable where the economist is able to provide beforehand a correct and indubitably complete analysis of the significant factors. The method is one neither of discovery nor of criticism. It is a means of giving quantitative precision to what, in qualitative terms, we know already as the result of a complete theoretical analysis-provided always that it is a case where the other considerations to be given below are satisfied [(J. Keynes 1939), p. 560].

According to Keynes, if one subscribes to the theoretical view of econometrics, econometrics is not a method of testing or discovery but just one of 'measurement' in that it provides the qualitative relations already known about empirical content. The econometricians in the Cowles tradition did not disagree with such a view, as seen in T. Koopmans (1949) and T. Koopmans and Hood (1953). The main difference was that, while 'structural modelers accepted Mill's a priori approach to economics', they 'differed from Mill in their willingness to conduct empirical investigations' [(Hoover 2007), p. 4]. Further, Koopmans and Haavelmo agreed with Keynes that causal mechanisms existed, and that knowledge about them could be acquired. However, the way to acquire such knowledge was not through empirical means but by theoretical analysis. Thus, Keynes and the econometricians in the Cowles tradition were non-reductionists and realists, since he believed that causal facts were primary with respect to non-causal facts such as empirical regularities [(Moneta 2005a), p. 438]. Hence, beginning with instruments that measure such regularities would not derive any causes – not even if such instruments are assisted by economic theory.

As a result, the critique propagated by Keynes or even highly abstract classical economics are not inconsistent with the CC approach to econometrics. This is because the causal relationships were derived from economic theory and the specification of the model was not the concern of the econometrician according to CC. The job of econometrics was to give such causal relationships an empirical interpretation by measuring their strength. Haavelmo proposed the following tenets in his seminal publication, 'The Probability Approach to Econometrics' (1944)[for more see Moneta (2005a), p. 438]:

1. The economy can be characterised as a system, where 'everything depends upon everything else', but is built up from systems of relations of cause effect type [Haavelmo (1944), p. 22];

2. The structural parameters of such relations can be identified by, 'a theoretical relation, a design of experiments and a set of observations' [Haavelmo (1944), p. 14];

3. The relations are essentially stochastic, [Haavelmo (1944), p. 40].

The notion that there are more fundamental relations than just empirical regularities is also visible in Haavelmo, who argued that 'there are more fundamental relations than those that appear before us when we merely stand and look' Haavelmo (1944), [p. 38], and it is exactly those fundamental relations that are causal. What distinguishes autonomy from regularities is exactly that the former 'refers to a class of hypothetical variations in the structure, for which the relation would be invariant, while its actual persistence depends upon what variations actually occur' Haavelmo (1944), [p. 29]. As a consequence, causal connections should be viewed as autonomous relations, which are exactly those that exist independently of us and therefore cannot be reduced to empirical regularities. Haavelmo famously used an analogy to describe this:

> If we should make a series of speed tests with an automobile, driving on a flat, dry road, we might be able to establish a very accurate functional relationship between the pressure on the gas throttle (or the distance of the gas pedal from the bottom of the car) and the corresponding maximum speed of the car. And the knowledge of this relationship might be sufficient to operate the car at a prescribed speed. But if a man did not know anything about automobiles, and he wanted to understand how they work, we should not advise him to spend time and effort in measuring a relationship like that. Why? Because (1) such a relation leaves the whole inner mechanism of a car is a complete mystery, and (2) such a relation might break down at any time, as soon as there is some disorder or change in any working part of the car. (...) We say that such a relation has very little autonomy, because its existence depends upon the simultaneous fulfillment of a great many other relations, some of which are of transitory nature [(Haavelmo 1944), p. 27-28].

Thus, the distinguishing feature of an autonomous relation is its explanatory power and the fact that an autonomous relation is invariant under new conditions. In H. Simon (1953), a related concept is used for causality, that is 'invariance under intervention' [See Section 2.2.2]. This view is shared by other economists who take the CC approach, as seen in Haavelmo (1944), T. Koopmans (1947), Klein (1977), and Malinvaud (1988).

#### 2.1.1.2 Atheoretical Econometrics

The second group of econometricians includes contemporary time-series econometrics, especially VAR models, who mainly comprise Clive Granger and Christopher Sims's followers. However, a straight line runs through the econometric literature, as argued in Kaergaard (1984) from the Danish statistician J. Warming over W. Mitchell, A. F. Burns, R. Vining, and the work at the National Bureau [see for example Burns and Mitchell (1946)]] over C. Granger, to T.J. Sargent, and C. Sims. This is explained in the following quote from Sims (1980a), where he referred to 'identification claimed for existing large-scale models' as 'incredible'. Later in the same paper, Sims referred to 'a priori restrictions' as a 'genesis'. I will examine these two views more closely in sections 2 and 3. This section instead focuses on the philosophical foundations of the two views.

**2.1.1.2.0.1 Reductionism, Causality and Econometrics** As noted in the previous section, the theoretical approach begins from a sound and internally consistent economic theory that provides the basis and thereby a complete specification of the empirical model. According to the atheoretical view of econometrics, this is unhelpful. Sharing the Humean motto that 'I will seek relationships among events that seem always to hold in fact, and when it occurs that they do not hold, I will search for additional conditions and a broader model that will (until new exceptions are discovered) restore my power of prediction' [(H. Simon 1953), p. 53]. Thus, the Humean idea is shared on two fronts: one, the discovery methods will examine regularities in the data, and the second, the crucial point of modelling causality, is prediction. The principal problem with models formed on a theoretical basis is that, often, they do not provide a good fit for the data, as noted in Reziti and Ozanne (1997),

> a recurring problem in empirical studies of consumer and producer behavior is that the regularity properties implied by microeconomic theory have more often than not been rejected [(Granger 1999), p. 16].

As Granger argued, the main issue is that 'theory often fails to capture vital features of the data, such as trends or seasonal components or some of the structural breaks' [(Granger 1999), p. 16]. Hoover (2008) argued that what characterises the ontology of atheoretical econometrics in the tradition of Granger and Sims is its Humean roots. Given the standard interpretation of Hume, the main commitment of Hume was to the following principle,[1]

**Hume's Commitment (HC)** Causal relations are reducible to non-causal ones.

Thus, Hume's answer to the metaphysical question of whether we can reduce causality to regularities is affirmative. This makes Hume a reductionist. According to Hoover, Granger and Sims are also reductionists; the economists in the Cowles Commission, however, were anti- reductionists [see Moneta (2005a)]. In the philosophy of science, reductionists are often classified after the strength of their position, as noted in Silberstein (2012) and Moneta (2005a). The most common positions are,

1. **Eliminative Reductionism:** This position claim that there is an identity relation between regularity claims and causal claims. Thus causal claims are nothing but regularities. Hence, we can eliminate causal terminology. Since it does not add anything.

2. **Nomological Supervenience:** This position is weaker than eliminative reductionism. The main claim here is that 'causal relations are determined completely by the properties of regular conjunctions but not identical to them' [(Moneta 2005a), p. 435].

The reductionist project is not new to science. Most famously, Ernst Mach proposed eliminating the concept of causality from the scientific vocabulary at the beginning of the 20th century. Mach instead wanted to introduce the word 'function' because it did not have the same metaphysical baggage. Additionally, models should function as instruments for measuring and predicting rather than as tools for representing or mirroring an underlying theory, as proposed by the Cowles Commission. The main reason why we should avoid using economic theory for any purpose, according to Sims, was that,

> dynamic economic theories must inherently be incomplete, imprecise, and therefore subject to variation over time. One reason for this is that economic cause-effect relations involve a 'recognition delay' about which theory has little to say and may be expected to be variable . . . It is wrong, then, to expect economic theories to be complete, mechanical, and divorced from reference to specific historical circumstances [(Sims 1981), p. 579].

[1] What I take to be the standard interpretation of Hume here is the one found in Strawson (2014). For more see Beebee (2016). For more on the relation between Hume and Granger see Granger (1980), Hoover (2001), and Moneta (2005b). Although as we shall see, the inspiration was mainly through Suppes and his Probabilistic theory of causality, which we will see later [See Section 2.1.1.2.0.2].

Therefore, according to Sims, economic theory is subjective, and, as a consequence, the only benchmark for objectivity in macroeconomics is that of atheoretical, or uninterpreted, statistical models of aggregate data [(Sims 1987), p. 53.]. This is also the only basis any kind of consensus could have.

Sims view shows why eliminative reductionism entails a certain kind of epistemological reductionism [see Silberstein (2012)]. This becomes even clearer in the next section on discovery methods. For instance, the strongest version of epistemological reductionism argues that,

> **Epistemological Reductionsim (ER+)**: We can completely replace causal claims by regularities found in the data, or 'statistical claims'.

A view held by both Sims and Granger [see Granger (1969), Granger (1980), and Sims (1987)]. A weaker version of that principle is the following,

> **Epistemological Reductionsim (ER-)**: We can completely replace causal claims by regularities found in the data, or 'statistical claims'. But causal claims may have a certain 'pragmatic' power.

Combining HC with epistemological reductionism helps to understand why most contemporary econometrics are considered instrumentalist Boland (2014), Hoover and Dowell (2001), Moneta (2005b), and Moneta (2005a). I argue in subsequent sections that this metaphysical reductionism entails a certain reductionist view in measurement theory, which I contend is untenable.

**2.1.1.2.0.2 Reducing Causality to Statistics: Suppes' Probabilistic Causality** Patrick Suppes' probabilistic view has been at the centre of the philosophical debate on causality since the 1970s [for other central figures in the literature of probabilistic causality prior to Suppes see J. M. Keynes (1921), Good (1959), Good (1961), and Good (1962) and for good introductory works on these figures see Russo (2009) and Vercelli (1991)].[2] Suppes' objective was to reduce causality to mere probabilities or 'probabilistic depdencies', and this very idea is at the basis of the reductionism found in atheoretical econometrics. Granger did cite and discuss both Wiener, Good and Suppes in his own work, see Granger (1969) and Granger (1980), but failed to notice how similar his and Suppes' accounts really were, but I return to this in later sections.

Suppes' theory of causality was not meant to provide a correct definition of causality. Instead, Suppes began from what he saw as the least common denominator of the concept. For Suppes, this is a necessary premise for moving forward with a theory of causality. Further, it guarantees flexibility in that it only provides a lower bound of what many see as causality. Suppose that $C_t$ and $E_{t'}$ are events defined as subsets of all possible outcomes. Further assume that both $C_t$ and $E_{t'}$ are referred to at a well-defined instant

[2] The very notion of probabilistic causality is fairly new in the literature. The received view was that causality and determinism accompanied each other. The development of a probabilistic account of causality was also helped by developments elsewhere, in particular, Kolmogorov's axiomatization of probability. It should be noted that Suppes probabilistic account of causality can account for deterministic causality, as noted in Suppes (1973) and Vercelli (2017a),

**Definition 2.1.1.** *Deterministic Causality*(DC)

1. $C_t$ is a sufficient cause of $E_{t'}$;
2. $C_t$ is prima facie cause of $E_{t'}$;
3. P$(E_{t'} \cap C_t^c)$=1

of time; the starting point for Suppes' account is then the prima facie cause, see Vercelli (1991), Vercelli (2017a), Suppes (1973), and Russo (2009). Formally, we write [(Suppes 1973), p. 12-14]:

**Definition 2.1.2.** *Prima Facie Cause*(PMC)

1. $C_t$ is a prima facie cause of $E_{t'}$ iff;
2. $t'<t$;
3. $P(C_{t'}) > 0$;
4. $P(E|C) > P(E)$.

Here, a shift from ordinary definitions of causality is evident. Typically, a definition of causality states a sufficient and a necessary condition for identifying a causal relation. This definition by Suppes only states a necessary condition because a prima facie cause is only necessary for identifying a causal relation and not sufficient. That is, $C$ can prima facie cause $E$ without being an actual cause of $E$. In other words, a prima facie cause cannot discriminate between genuine and spurious cases of causality. Consider the following example:

**Example 1.** Assume we have a detection software (DS), which primary job is to detect a possible tsunami. Further suppose that DS works probabilistically. In this case $DS$ is clearly a prima facie cause of a tsunami since a shift in DS indicates an increased probability that a tsunami will happen immediately after. However, clearly, a shift in the DS does not cause the tsunami. Thus, neither condition (2) nor (3) can rule out the spurious cause in the case of DS and a tsunami.

The problem in example 1 is that the DS and the tsunami share a common cause. When we account for that cause, the effect is suddenly stochastically independent. We define a spurious cause in the following way [(Suppes 1973), p. 21-23]:

**Definition 2.1.3.** *Spurious Cause*(SC)

1. $C_t$ is a spurious cause of $E_{t'}$ iff;
2. $C_t$ is a prima facie cause of $E_{t'}$;
3. there is a $t'' < t$ and an event $E''_{t''}$;
4. $t'<t$;
5. $P(C_t \cap E''_{t''}) > 0$;
6. $P(E_{t'}|C_t \cap E''_{t''})$=P$(E_{t'}|E''_{t''})$;
7. $P(E_{t'}|C_t \cap E''_{t''}) \geq P(E_{t'}|C_t)$.

Hence, the lesson for Suppes is that, in order to arrive at causal relationships between events, frameworks are important. It is the only way we can identify a spurious cause. Suppes (1973) argued that such conceptual frameworks can be split into three [1973, section 2, p. 79-80], which are characterised by three main ingredients, according to Suppes,

1. **Conceptual framework**. Provided by some scientific theory $T$.

2. **Experimental framework**. Provided by the experimental setting.

3. **General framework**. Provided by the amount of information available to us at $t$ and our beliefs about them.

Thus, even if Granger causality is a subset of Suppes probabilistic causality, as we will see later, there are some obvious differences, that are important to keep in mind beforehand, as argued in Vercelli (2017a) and Vercelli (2017b). Suppes do think that causal claims are relative to a given conceptual framework, as argued in Williamson (2009). Apart from that Suppes also argue that causality is relative to ones conception of mechanisms [(Suppes 1973), p. 72]:

> the analysis of causes is always relative to a particular conception of mechanism, and it does not seem satisfactory to hold that the analysis of mechanism is ever complete or absolute in character.

This allows us to reformulate definition 2.3.2, given a background, following Vercelli (1991)[p. 108] to obtain the following, which will be useful later [see Section 2.3.2 and Section 4.2.3],

**Definition 2.1.4.** *Prima Facie Cause**(PMC*)

1. $C_t$ is a prima facie cause of $E_{t'}$ with respect to some background $B$ iff;

2. $t'<t$;

3. $P(C_t \cap B_t) > 0$;

4. $P(E_{t'}|C_t \cap B_t) > P(E_{t'}|B_t)$.

Specifying Suppes' causality in this way by following Vercelli (1991) and Vercelli (2017a) allows us to utilise Suppes' account of causality in a theoretical framework to specify the necessity of a theoretical framework in understanding causality; that is, causality is relative to a set of information organised by a theoretical hypothesis. Reiss (2016) argues something similar and notes that he think that Suppes would have cited with the theoretical economists based on Suppes (1973) and Suppes (1966),

> Now, while I am not aware that Suppes ever commented on this debate between 'design-based' and 'structuralist' econometricians, it is probably safe to assume that he would side with the structuralists [The theoretical economists]. If anything, my guess would be that Suppes would urge economists not just to use economic theory but develop theories that are strong enough to have implications about all aspects of an empirical study that need to be addressed, including independence relations, functional form, error terms and so on, or at least implications that are strong enough so that we have a good reason to believe that tests of the statistical assumptions of lower-level empirical models yield informative results [(Reiss 2016), 298].

This is important to consider when I discuss Granger causality in the next section and when I discuss the importance of the information set in operationalizing Granger causality by different tests.

### 2.1.2 Discovering Causality: Methodology

The metaphysical and epistemological principles mentioned in the previous section entail a set of methodological doctrines. Theoretical econometrics and its commitment to realism [R] often means that true theories should be pursued. This is most often achieved by uncovering mechanisms that explain the variation in the underlying data. Such an approach is not new. As described in Hoover and Dowell (2001), using a mechanism to explain causality dates as far back as the days of Adam Smith. In A. Smith (1982)[1776], Smith sought to explain the causes of changes in the supply of silver [see A. Smith (1982)[1776] in book 1, Chapter 11]. The strategy employed here was inherently mechanism-based. Smith had a theoretical framework that provided the underlying structure and then measured the strength of the postulated relation [For more see Hoover and Dowell (2001), p. 142-143]. That said, atheoretical econometrics often takes an instrumental approach, as noted in Moneta (2005b), Moneta (2005a), Lawson (1989), Fullbrook (2008), Pheby (1991), and Grabner (2016). Such an approach is not new either, as Hoover and Dowell (2001), Reiss (2001), and M. S. Morgan (2012) noted. It is further argued that these instruments are no different than telescopes and thermometers, see Boumans (2004), Boumans (2015), and Hoover (2007). In the philosophy of science, instrumentalism typically refers to the idea that theories are instruments to pursue a certain set of prespecified goals [see Maki (2001)]. In this context, instrumentalism refers to the idea that models are instruments used to pursue a certain prespecified goal, predictivity. Since multiple definitions of causality in econometrics, like Granger causality, reduce the notion of causality to incremental predictivity. Thus, a causal model in the Granger tradition uses predictivity to explain underlying variations. This means that, if the prior values of some time series $X_{t-1}$ improve the prediction of $Y_t$, then $X_{t-1}$ explains the variation in the variables and therefore causes $Y_t$. It is important to note, however, that the CC approach is not a testing or discovery procedure. On the one hand, the only purpose of the CC approach is to give theories empirical content, as observed in the previous section. On the other hand, the contemporary atheoretical approach to econometrics is both a tool to test and discover causality. What I assume to be the common methodological basis of the two approaches to causality in econometrics is 'variation'. Both the theoretical econometrician and the atheoretical econometrician seek to explain exactly what produces a certain variation in the underlying variables. The disagreement is in what they add to the variation. The idea of variation as the most primitive notion of causality was especially well formulated in Russo (2009). As Russo (2009) noted, variation is where every causal analysis begins since there would be nothing for causality to explain in the absence of variation. The general intuition is formalised in the following way, following Wold (1969b) [p. 452]:

$$\text{x varies from x to x + } \Delta. \tag{2.1}$$

Indicating that the value of $x$ is changing with some unknown $\Delta$, which we denote in the following way:

$$x \uparrow \Delta. \tag{2.2}$$

Therefore, the variations we are interested in here are when we observe variations in $x$ at $t$,

$$x \uparrow \Delta. \tag{2.3}$$

That produces a change in some another variable at$t+1$

$$y_{t+1} \uparrow \Delta. \tag{2.4}$$

Thus, saying that $x$ causes $y$ is to say that we detect that $x_t \uparrow \Delta$, and we believe that $x$ is responsible for the variation in $y$ at $t+1$ [(Russo 2009), p. 94]. As noted in

Russo (2009), in the probabilistic theories that provide the philosophical basis of Granger's concept of causality, the underlying focus is $P(E|C)$ and the marginal probability $P(E)$. The comparison of these two is to 'analyse a statistical relevance relation' [(Russo 2009), p. 94]. That is, the purpose is to consider a change in $C$, or $C \uparrow \Delta$, given the effect $E$, if detected, is due to the difference in conditional and marginal probability, that is

$$P(E|C) > E. \tag{2.5}$$

Hence, causal claims in social science become variational claims. This is the same as saying that 'variation in the conditional probability of the effect is due to a variation in the marginal probability of the cause' [(Russo 2009), p. 95]. The question now becomes how we can specify the notion of variation. In other words, what should we add to variation to obtain causality? That is,

$$causality = variation + ?. \tag{2.6}$$

As the well-known maxim states, correlation or 'co-variation', is not causality. Neither of these two concepts provides a sufficient condition for causality. However, co-variation does provide a necessary condition for causality, as argued in Haynes and O'Brien (2000). Correlation does not provide a sufficient condition for causality either.[3]

Now, consider the standard reduced-form structural equation

$$Y = \beta X + \varepsilon. \tag{2.7}$$

We assume $Y$ to be some effect, $X$ to be some cause, $\beta$ a parameter, and $\varepsilon$ an error term. We then arrive at the following essential question: assume that there is some co-variation between the factors $X$ and $Y$; when is the particular co-variation chancy, and when is it causal? Co-variation here refers to when two variables vary with each other, often denoted $COV(X, Y)$.[4] This should not be confused with the correlation, which refers to when a change in one variable leads to a change in another. The population correlation is usually calculated in the following way and should be distinguished from the sample correlation [(Hoover 2007)]:[5]

$$Corr(X, Y) = \frac{COV(X, Y)}{\sigma_x \sigma_y}. \tag{2.8}$$

In other words, the population correlation is co-variation normalised by the standard deviation, denoted by $\sigma$.

In the instrumental approach suggested in atheoretical econometrics to modelling causality, the '?' in (2.6) would be predictivity. In the theoretical realist approach to modelling causality, the '?' in (2.6) would be a mechanism. Consequently, there are two different paths of discovery methods to causality. The atheoretical approach uses statistical instruments to test whether the prior values of one time series improve the prediction of another. Following Boumans (2015), I take such instruments to be triplets, that includes internal principles, bridge principles and calibration. Cartwright (1983) argued that internal prinpceples 'present the content of the theory, the laws that tell how the entities and process of the theory behave', bridge principles on ther other hand 'are supposed to tie the theory to aspects of reality more accessible to us' [p. 132]. Boumans add calibration that plays a

[3]Correlation does not provide a necessary condition since we can have a causal connection between uncorrelated variables, $A$ and $B$. This happens when there is a non-monotonic relationship between $A$ and $B$.

[4]This can be written in multiple ways: $COV(X,Y) = \sigma_{XY} = E[(x_i - \mu_X)(y_i - \mu_Y)]$.

[5]The sample correlation is an estimate based on a sample drawn from the underlying population. The population correlation is the 'true' correlation. In other words, the sample correlation is an estimate.

crucial role in transforming inexact relations into exact relationships [(Rodenburg 2004), p. 5]. I return to calibration in Chapter 3 and Chapter 4. The theoretical approach, though, uses economic theory to postulate relations a priori. I examine this more closely in the upcoming section.

### 2.1.3 Specifying Theory

Fifty years ago, Paul Samuelson quoted J. William Gibbs on the frontispiece of his Foundation of Economics: 'Mathematics is a language'. It may be more important to specify this now than ever before. Previous sections make great use of the concept of theory. This calls for an in-depth discussion of what theory means in this context.

As Leijonhufvud (1997) correctly noted, it is not unusual in economics to use theory and model interchangeably. This view reinforces the semantic interpretation of econometric practice, for which the structure of a given scientific theory $T$ is to be identified with a family $\mathcal{M}$ of models held in parts of the literature on the philosophy of economics [see Suppe (1991), Suppe (2000), and Halvorson (2012)].[6] However, I argue that there is an important distinction between models and theories which is often missing in a semantic interpretation wherein the two become one. Where I see 'theories', following Leijonhufvud (1997) and Boland (2014), as a set of beliefs about the economy and how it functions – which is (i) naturally prior to the model and (ii) about the world 'out there' – models are, in contrast, formal and partial representations of such theories, as noted in the Cowles approach [1997, p. 193]. This is also the main reason why a theory can be either 'true' or 'false', but a model is only said to be either 'correct' or 'incorrect'.

Therefore, my interpretation of theory in this thesis and in theoretical econometrics is also sometimes referred to as 'background knowledge', though the concept of 'background knowledge' extends further and includes the very context of the background knowledge, that might be institutional or environmental. Theory, then, includes every aspect, from political context to knowledge, of the population in question and theory. As a result, economic theory here has more restricted terms which only include the postulates of economic theory and take these as a given. However, reviewing the works of Koopmans and Haavelmo shows that that they were in favour of including contextual points – especially see Haavelmo (1944) and T. Koopmans (1953). That said, in most cases, it is a more restricted concept than 'background knowledge'. In this understanding, theory provides a list of crucial concepts and relations that are needed to perform empirical work. Concepts could be about individual markets, how households behave, and other areas. Relations could be the underlying mechanism of a given economic system. Thus, theoretical econometrics rejects the idea that 'pure facts' rather than theory-laden, is possible. Additionally, if that is needed for 'objectivity', then objectivity is simply impossible. It is exactly due to the non-experimental nature of economic data, the theory-laden quality of its data, that it is important to articulate theory clearly. Only here is it possible to ensure that no obvious mistakes are made in modelling. Further, theoretical econometrics should not be considered an attack on empirical analysis. Rather, the conception provided by the postulates of economics is precisely what allow us to understand economic data and history. The latter is even the clear goal of any economic inquiry. Economic theory is the servant of empirical work, and as argued by Austrian economist Ludwig von Mises, 'theory and the interpretation of historical phenomena are intertwined' [Von Mises (1996), p. 66].

[6] It is not unusual in the philosophy of econometrics literature, to view large parts of econometrics as being in accordance with the semantic view, see Chao (2005). Whether this claim is true is not for this thesis to decide. However, I will reject some parts of the semantic interpretation later in Chapter 3.

## 2.2 Theoretical Econometrics: Theory, Representation and Measurement

To better understand the atheoretical approach to econometrics, it is crucial to first comprehend the theoretical approach. My strategy in doing so is threefold. I present (i) the underlying principles, (ii) the important concepts of this approach, and (iii) an example that demonstrates how the Cowles approach works. Lastly, I provide arguments for why the Cowles approach lost traction in the literature. This serves as a bridge to the next section.

### 2.2.1 The Theoretical Approach as the Cowles Approach

Traditional econometrics began with the idea that econometric data is not strong enough to stand alone. This was the main motivation for applying theory in econometric analysis. Econometric practise took multiple a priori assumptions from economic theory, as noted in Kaergaard (1984):

1. Which variables should be a part of the analysis (this includes which variables should be given the coefficient zero?
2. What kind of function are we considering? For example, is it linear? Or perhaps non-linear?
3. How do we measure a certain latent variable? Perhaps inflation, which originally referred to an expansion of money supply, is now measured by either the CPI or the CPI deflator.

Haavelmo (1944)[p. 49] emphasised the importance of a complete stochastic model;[7] however, most of the restrictions placed on econometric models by economic theory are deterministic in nature. In the tradition of Haavelmo, Marschak, and Koopmans' modelling of causality, one important feature of structural models in the traditional econometric approach distinguishes it from Granger causality, as noted in Mouchart et al. (2010): that causality becomes relative to a given model. As argued in Christ 1994b[p. 6], we can reduce the Cowles approach to the following analysis on three levels:

1. **Methodology**: CC represents an attempt to bridge theory and empirical research. A way to do so is to explicate all assumptions made in the process. This would (i) facilitate discovery of problems and (ii) make it easier to adjust the assumptions themselves in light of new discoveries, as noted in Christ (1994b) and Gilbert and Qin (2007)[p. 253-255];
2. **Division of Labour**: According to the CC view, the job of the economist is to build theoretical models. The job of the econometrician is to estimate structural models based on those theoretical models. The best way to summarise the division of labour in the CC view between the economist and the econometrician was described by Dou Qin:

> Economic theory consists of the study of (...) relations which are supposed to describe the functioning of (...) an economic system. The task of econometric work is to estimate these relationships statistically [(Gilbert and Qin 2007), p. 254].

[7] Stochastic means that the model at least contain one random variable. A stochastic model on the other hand is a tool to estimate the probability distribution of potential outcomes.

3. **Technical**: The CC researcher begins from accepted theoretical models. The modelling procedure has been formulated as a simultaneous-equations model (SEM). These SEM models were seen as 'the most general (linear) theoretical model form since they encompass a dynamically extended Walrasian system' [(Gilbert and Qin 2007), p. 254-255]:
$$A_0 x_t = \sum_{i=1}^{p} A_i x_{t-i} + \varepsilon_t \tag{2.9}$$

In general, the main condition for causality in CC was exogeneity. In practise, the traditional method of econometrics, as seen in the Cowles approach, is summarised in the following way:

1. Theory provides a model specification (e.g., a linear Keynes model).

2. Data decides the parameter values by some estimation method (e.g., a least-squares method).

One of the most successful applications of this method is a macroeconomic model, most of which are Keynesian. Such a model is given in the structural form, as noted in Kaergaard (1984),
$$Y_t = \alpha Y_t + \beta X_t + \tau_t. \tag{2.10}$$
in which $Y_t$ denotes the vector that includes the endogenous variables, $X_t$ the exogenous variables, and $\tau_t$ the stochastic residual. The single equation in such a system of equations would normally describe the behaviour of a set of agents concerning a given variable. In this situation, $\alpha$ and $\beta$ are typically specified using many elements that have a zero coefficient. Often, a small group of determining factors are more than enough to describe the behaviour, while $Y_t$ and $X_t$ can include hundreds of variables.[8] If we solve equation (2.10), we obtain the reduced form,

$$Y_t = (I - \alpha)^{-1} \beta X_t + (I - \alpha)^{-1} \tau_t. \tag{2.11}$$

This equation directly shows the effect of the exogeneous variables on the endogenous variables. The main problem in (2.11) is the coefficient matrix, as noted by Kaergaard (1984), $[(I-\alpha)^{-1}\beta]$. Therefore, according to the traditional approach to econometrics, we instead estimate the relations in (2.10) due to the fact that most of the a priori information about $\alpha$ and $\beta$ is lost in the aggregation $[(I-\alpha)^{-1}\beta]$. This means that institutional changes are more easily identified. To better understand this, consider the following simultaneous equation models on the relationship between wage and status, in which wage depends on status and productivity as well as a range of other factors captured by $\tau_t$ and status to depend on wage and assets and other factors captured by $\tau_t$:

$$wage = \alpha_0 + \alpha_1 status + \alpha_2 prod + \tau_{t1}. \tag{2.12}$$

$$status = \beta 0 + \beta 1 wage + \alpha_2 assets + \tau_2. \tag{2.13}$$

This provides the structural form *derived from economic theory.* However, if we want to estimate these two equations using ordinary least squares (OLS), we find that wages are correlated with $\tau_2$ and status with $\tau_1$. Hence, our OLS estimation is biased. However, we could rewrite (2.12) and (2.13), thus obtaining what we refer to as the reduced form

[8] As a sidenote, the Danish Adam model includes about 3500 variables, among them 2500 endogeneous variables and 1000 exogeneous variables and close to 2500 equations.

by substituting equation (2.13) in (2.12) and (2.12) into (2.13), in which $\upsilon_i$ denotes a composite error term

$$wage = \gamma_0 + \gamma_1 prod + \gamma_2 assets + \upsilon_1. \quad (2.14)$$

$$status = \varkappa_0 + \varkappa_1 prod + \varkappa_2 assets + \upsilon_2. \quad (2.15)$$

Hence, what happens when we produce a reduce form becomes clear. The theoretical relationship disappears because we rearrange two postulated relationships algebraically. Since productivity and assets are considered exogenous, however, we can now estimate the two equations using OLS. That said, we would rather maintain the structural form that contains the theoretical relations derived from economic theory. Most importantly, though, the example here shows that using the structural approach in econometrics demands the use of prior theory. All the information we can extract from the equations is found in the structural form. Without that, it is impossible to derive the reduced form. Therefore, in both cases, the data provides at most an estimation of the parameters. To understand the entire picture, the structural form is required. There is no transition from the reduced form to the structural form. Subsequently, only the use of a priori theory restrictions on the model – for example, deciding the number of zero elements in $\alpha$ and $\beta$ – enables the estimation of data from a structural form, such as my choice of status and productivity as exogenous variables. As Kaergaard (1984) argued, one complication of the traditional approach is that structural equations are rarely linear. Even so, models are usually estimated from the structural (2.10) rather than the reduced form (2.11) even if the opposite would solve several statistical problems in estimating (2.11). One of these problems, the 'simultaneity bias', may appear because $Y_t$ on the right side of the equality sign, and is normally correlated with $\tau_t$.

### 2.2.2 The Causal Concept in the Cowles Approach: Simon on Identifiability and Exogeneity

One of the Cowles approach's principal influences in developing the causal concept was Herbert A. Simon in H. A. Simon (1952) and H. Simon (1953),[For more on Simon see Fennell (2005), Boumans (2010b), Hoover (2001), and Hoover (2008)] who noted in a monograph of the Cowles Commission that,

> In(...) scientific methodology, particularly those carried on within a positivist or operationalist framework, it is now customary to avoid any use of the notion of causation and to speak instead of 'functional relations' and 'interdependence' among variables [(H. Simon 1953), p. 49].

Hence, according to Simon, the sciences had become Humean, due to their acceptance of a basic Humean principle, which is that necessary connections cannot be perceived, and that, as a consequence, they do not have any empirical basis [(H. Simon 1953), p. 49]. However, Simon noted two main problems in the empiricist characterization of causality – in other words, the attempt to reduce causality to either 'interdependence', or 'functional relationships' [(H. Simon 1953), p. 50-51]:

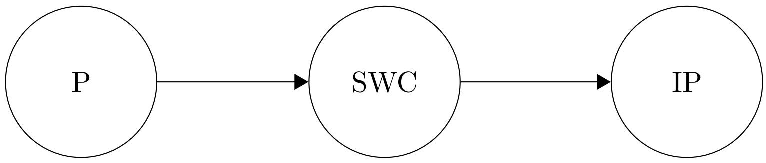
1. Replacing a deterministic viewpoint with a probabilistic one does not solve the problem at hand, since 'we can replace the causal ordering [structure] of the variables in a deterministic model by the assumption that the realised values of certain variables at one point or period in time determine the probability distribution of certain variables at later points or periods' [(H. Simon 1953), p. 50].

2. The problem of asymmetry. The main problem with ‘functional relationships’ and ‘interdependence’ is that their generally recognised symmetry. However, we acknowledge that $A$ causes $B$, we do not want to simultaneously determine that $B$ causes $A$.

In H. Simon (1953), Simon equated the problem of identification with the problem of finding the underlying causal structure.[9] The problem of causal structure that he adopted had been long discussed in the literature, dating back to the works of Tinbergen and Haavelmo. Tinbergen an account of causal structure in Tinbergen (1939), and the problem was discussed in Haavelmo (1944) as well, with him arguing that causal factors are of relative character depending on what one wants to explain:

> The causal factors (or the "independent variables") for one section of the economy may, themselves, be dependent variables in another section, while here the dependent variables from the first section enter as independent variables. [[(Haavelmo 1944), p. 22].

This idea was later adopted in T. C. Koopmans et al. (1950), where Koopmans discussed which variables should be taken as ‘exogeneous’ and which as ‘endogenous’. ‘Exogenous’ variables refer to those outside the model that explains them, and ‘endogenous’ variables refer to those variables being explained. An example like the following from H. Simon (1953) p. explains this [(H. Simon 1953), p. 52]:




Here P denotes poor weather; SWC denotes small wheat crops and lastly IP denotes increase in the price of wheat. In this example, we cannot intervene in the price of wheat to change the weather since ‘The weather’ is an ’exogenous’ variable, making the price of wheat an ’endogenous’ variable.’ These two concepts represent properties of the variables and play a key role in the characterization of a causal structure. Koopmans ultimately distinguished the ‘main principles’ from a ‘departmental principle’, and the ‘causal principle’ when characterizing endogenous and exogenous variables. The departmental principle regarded variables outside the realm of economics as exogenous, which could include a natural disaster. The causal principle, though, took those variables as exogenous; they influence but are not influenced (endogenous) [see T. C. Koopmans et al. (1950), 393-395]. As noted in Hoover (2008)[p. 5], ‘the Cowles Commission, related causality to the invariance properties of the structural econometric model. This approach emphasised the distinction between endogenous and exogenous variables and the identification and estimation of structural parameters’. Further, as argued by Boumans, it was ‘Herbert Simon’s paper "Causal Ordering and Identifiability" (1953) that linked causal structure with the problem of identification, thereby cutting off definitively the problem of causality from any empirical approach’ [(Boumans 2010b), p. 103]. This further shows the connection between identification and causality.

Simon’s main goal was to specify the algebraic properties that should be satisfied by a model to be causal. However, Simon’s discussion in H. Simon (1953) may have indicated another focus than is usually ascribed to him. His focus on these aspects may indicate that

[9] As noted in Fennell (2005), Simon uses the term ‘causal ordering’ to mean the same as ‘causal structure’. In order to be consistent with the rest of the thesis, I will use structure in the entire thesis.

he was ultimately interested in mapping causal relations that exist independently in the world. This would also suggest a realist view in line with the rest of Cowles. The tools for measuring and testing the postulates of economic theory were the simultaneous equation models (SEM), an approach to causal modelling that implicitly assumes the existence of an underlying mechanism that generates the data in the world [Also known as the data generating process]. Further, it emphasises that, without restricting the model a priori, it is impossible to identify causality, as I will discuss in [Section 2.2.3].

Simon began from,

> self-contained linear structure can be decomposed into a number of distinct "minimal" self-contained subsets and a "remainder." Minimal self-contained subsets were defined as self-contained subsets of a linear structure that do not themselves contain self-contained (proper) subsets [(Boumans 2010b), p. 103].

We say further, that when there is one or more self-contained subsets in the linear structure, the remainder is not empty; instead we say that the structure is causally ordered. The second step according to Boumans is that we simply 'repeat the partitioning as follows. The equations of the minimal subsets are solved, and the values of their variables are substituted in the equations of the remainder' [(Boumans 2010b), p. 103-104]. This yields a new self-contained structure derived from the first, which in turn produces a 'derived structure of first order'. If this is also causally ordered, the process can be repeated. The consequence of such an structure is that variables appear endogenous in one complete subset, and that they appear in a structure of higher order as exogenous, making the separation of variables into endogenous and exogenous variables a central pillar of Simons' account of causal structure [(Boumans 2010b), p. 103-105].

The operational meaning of causal structure was understood in the following way:

> We suppose a group of persons whom we shall call "experimenters." If we like, we may consider "nature" to be a member of the group. The experimenters, severally or separately, are able to choose the nonzero elements of the coefficient matrix of the linear structure, but they may not replace zero elements by nonzero elements or vice versa (i.e., they are restricted to a specified linear model). We may say that they control directly the values of the nonzero coefficients [(H. Simon 1953), p. 69].

As a result, causal structure would only have an operational meaning if one condition was satisfied, which is that 'We must have a priori knowledge of the limits imposed on the 'experimenters'—in this case, knowledge that certain coefficients of the matrix are zeros' (1953: p. 65) — expressing the importance of distinguishing between zero and nonzero elements and the importance causality theory in Cowles. There was no causal account without a priori theory.

Simon then 'showed that the conditions for a well-defined causal order are equivalent to the well-known conditions for identification' [(Hoover 2008), p. 7]. To understand how we could define causality between endogenous variables in an SEM model, though, consider the following bivariate system, remembering that Simon abandoned the idea of a temporal basis for asymmetry found in the Humean analysis in favour of a recursive structure [The example is from (Hoover 2008), p. 9-10]:

$$Y_t = \varphi X_t + \varepsilon_{1t}, \tag{2.16}$$

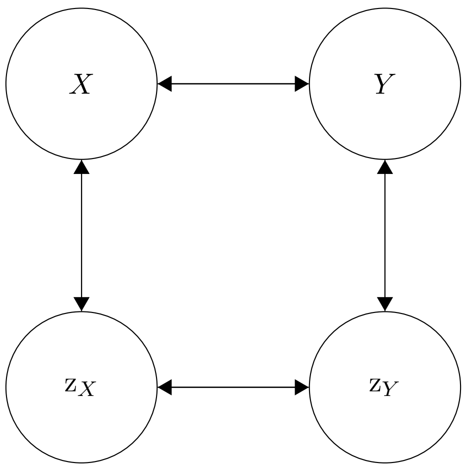
$$X_t = \varepsilon_{2t}. \tag{2.17}$$

The random error $\varepsilon_{1t}$ is independent, identically distributed, and $\varphi$ is a parameter. According to Simon, $X_t$ causes $Y_t$ because $X_t$ is recursively ordered ahead of $Y_t$. Thus, one knows all about $X_t$ without knowing about $Y_t$, but one must first know the value of $X_t$ to determine $Y_t$. This indicates that a change in the variance of $\varepsilon_{2t}$ in (2.17) would impact (2.16), while it does not hold in the reverse. This means that we could use Xt to control Yt. Hoover (2008) argued that this recursive approach does not work because it does not guarantee asymmetry of information or control. Thus, we remain in possession of a structure in lists of exogenous and endogenous variables with no way to order the endogenous variables.

However, the endogenous/exogenous approach to causal structure was highly criticised in the 1970s and 1980s, since one can show that model exogeneity of a variable is neither sufficient nor necessary for treating a certain variable as fixed [For more see Geweke (2017)]. However, the connection between model exogeneity and the exogeneity of Cowles was questioned in Cooley and Leroy (1985). I return to model exogeneity in the next section on atheoretical instruments because Sims also utilised the concept in VAR models. He transformed the concept into a testable one given the Humean assumption that the cause must precede the effect.

### 2.2.3 The Cowles Approach: A Textbook Example

Consider the following example, which is a classic textbook example of the Cowles approach. We have observed a possible connection between two variables, $X$ and $Y$. In this case, we do not know which of these two factors, if either, is the cause. However, we have also observed that another factor varies with $X$, perhaps $z_X$, and yet another factor varies with $Y$, perhaps $z_y$. This yields the following figure. Including all variables, we obtain the following case:




At stake here is identification of the causal relationship between $X$ and $Y$ . The probability of $Y$ given $X$ is indeed higher, but we simply do not know whether this is to $X$ or $z_X$. Hence, removing the arrow at the top and the arrow at the bottom provides the updated understanding. Moving to the right of the system as a system of two equations, we obtain the causes on the right-hand side and the effects on the left-hand side, allowing $c_X$ to denote $X$ as an effect of $Y$ and $c_Y$ to denote $Y$ as an effect of $X$. Finally, allowing $e_X$ and $e_Y$ to denote all other variables, we obtain the following:

$$X = c_Y x + e_X z_X. \tag{2.18}$$

$$Y = c_X y + e_Y z_Y. \tag{2.19}$$

The two equations here show $X$ and $Y$ respectively as a function of all other variables and each other. We can observe $X$ and $Y$ directly. However, there are still four relations that we cannot measure:

1. What is the effect of $X$ on $y(c_Y)$?
2. What is the effect of $Y$ on $x(c_X)$?
3. What is the correlation between $X$ and $e_X$?
4. What is the correlation between $Y$ and $e_Y$?

This system is clearly not identifiable. As a result, we encounter the identification problem: we simply cannot identify the effect of $X$ on $Y$. What is the solution? As observed in the previous section, applying theory and adding restrictions so the number of equations equals the number of unknowns is the next step. This allows us to identify the model [see Section 2.2.1-Section 2.2.2]. For instance, perhaps we learn from theory that y is an exogenous variable and further that there is no correlation between $z_X$ and $z_Y$ , that is,

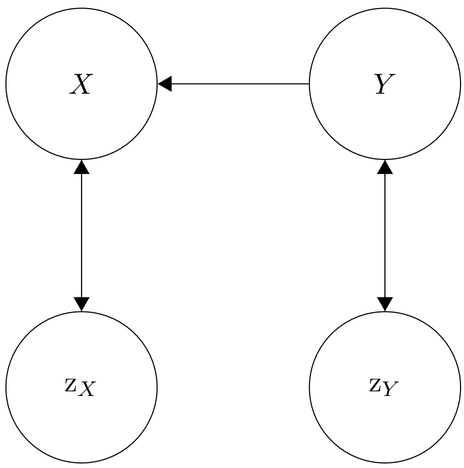
1. $c_Y = 0$.
2. $E[z_X z_Y]]$=0.

The second condition is important as well, since it is not enough that the dependent and independent variables are not affected by the same variable. Indeed, if $X$ and $Y$ correlate with different variables that might be connected, the second condition is breached. This identifies the original variables, from which we obtain the following:

$$X = c_X y + e_X z_X. \tag{2.20}$$

$$Y = e_Y z_Y. \tag{2.21}$$

Here, we obtain the causal structure in which $Y$ is the cause of $X$ and $c_X$ provides the size of the causal effect. $Y$ is decided in (2.21), where $X$ is no longer present. Only thereafter is $X$ decided, in (2.20). However, since $Y$ is decided first, it has causal priority. Therefore, it is exactly the exogeneity that enables us to identify the causal connection. As a result, we obtain the following figure:




In this case, it was economic theory which allowed us to observe that $Y$ causes $X$. Without economic theory and with it the knowledge that $Y$ is exogenous, we would have been unable to identify the model, and, consequently, we would not know that $Y$ causes $X$. This shows how vital economic theory is in the Cowles approach to econometrics. In contrast, an atheoretical approach would instead address the same problem by running a statistical causality test, as I discuss in section 2.3.

### 2.2.4 The Downfall of Theoretical Econometrics

In the group that propagated the standard approach to econometrics, there is a straight line from the early works of Danish econometrician Macklesprang to Tinbergen, Haavelmoo, and Koopmans, to Malinvaud in 1981, and to the structural movement prominent in contemporary literature [For more on this, see Keane (2010) and Keane (2013)]. The first criticism against what would later be the traditional approach to econometrics was already voiced before Tinbergen wrote his well-known book Statistical Testing of Business Cycle Theories (1939); in Nationaløkonomisk tidsskrift [National Economic Journal], by Danish statistician Jens Warming who asked the following question:

> How can it even be allowed, to choose one or the other arbitrary formula, and then – here you go – let it represent one of the living laws of turnover?.[10]

Here, Warming questioned why we should believe the postulates from economic theory, asking why we should trust that a list of exogenous variables are indeed exogenous. This is similar to the famous remark by Sims in Sims (1980b) where he referred to the restrictions placed on macroeconomic models as 'incredible' throughout his article. As Epstein argued,

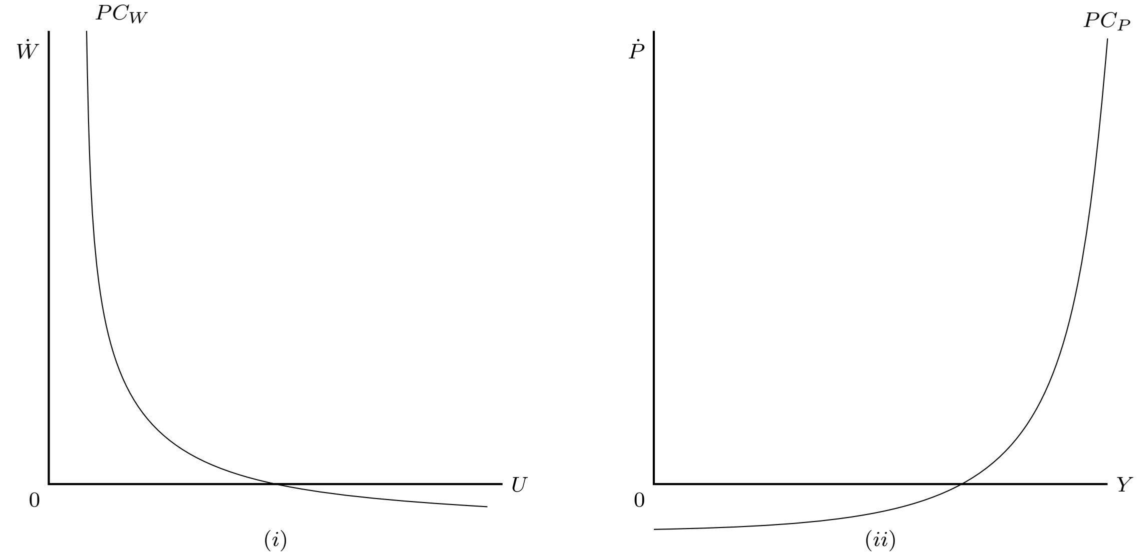
> Sims (1980) doubts that identification of simultaneous behavioural equations in macroeconomics is practicable. Granger (1969) denies that economic relations are really governed by simultaneity (see also Wold, 1954). Both authors refuse to allow the concept of an exogenous variable into their work. Their models mimic time series methods without pretending to have too much prior economic theory [(Epstein 2014), p. 6].

However, Sims simply resurrected, [see Sims (1980b)].

> an old article by Liu (1960) which insisted that it was incredible to regard 'B' and 'C ' of the system of structural equations. The argument touches a chord with anyone involved in the construction of computable general equilibrium models. If decisions on consumption, labor supply, portfolio allocations, etc. are all determined by the same set of variables, consequently, theoretical considerations would predict no difference in the menu of variables entering different equations, although the quantitative importance of individual variables is most likely to vary with the type of decision. Prescription of the zero elements in B and C therefore involves excluding variables with coefficients close to zero. In this respect, the action is little different to what is done in any attempt to model reality by capturing the major influences at work [(Nell and Errouaki 2013), p. 211-212].

It is important to note, however, that Sims presented little evidence supporting the idea that the identifying restrictions posed on macroeconomic models are invalid, as Epstein (2014) observed [see, p. 206]. That said, as noted in Pesaran and R. Smith (1992) and Pesaran and R. Smith (1995), there were other reasons behind the departure as well. For instance, an increasing amount of data had shown that models were not a good fit with the data, especially the problems associated with the Phillips curve. Consider the following figure:

[10] The quote from Warming is based on my own translation. The original quote in Danish is: 'Hvordan kan det overhovedet være tilladt at udvælge en eller anden vilkaarlig Formel, og så - værs'go - lade den forestille en af den levende omsætnings love? [(Kaergaard 1984), p. 8]'




(i) (ii)

The Phillips curve present an inverse relationship between unemployment and rises in wages. The relationship between employment and inflation is then easily deduced from his findings. Friedman famously argued that this was false and was proven correct with the stagflation of the 1970s.[11] Another crucial setback to the Cowles approach was the Lucas Critique. Lucas (1976) showed that, due to dynamic optimisation with rational expectation, the idea of estimating structural relationships was likely impossible; as a result, econometric models were not useful for policy purposes.

This caused a shift from the idea that models should be a representation of a more general theory and that the role of econometrics is to estimate economic theory, which had been a central pillar of the Cowles approach to econometrics. Instead, the focus was directed toward 'developing measures of model adequacy, a proliferation of diagnostic and misspecification tests' [(Pesaran and R. Smith 1995), p. 58]. All of these were intended to contain as little theory as possible and, in cases like that of Granger causality, no theory at all. As Smith noted, however, the Lucas critique was not the final critique empiricists would make against econometrics. Parts of the economic literature saw the problems in the 1970s and 1980s as problems of econometrics itself. One notable example is Summers, who argued that,

> (...) formal empirical work which, to use Sargent's (1987, p. 7) phrase, tries to "take models seriously econometrically" has had almost no influence on serious thinking about substantive as opposed to methodological question (...) [(Summers 1991), p. 129].

Summers continued by contrasting economics with the natural sciences. Where theoretical physicists would often await empirical physicists in the hopes that their theory would prove correct, this never occurred in economics. Theorists tended toward ambivalence when faced with empirical results, and that was if they were even aware of such attempts to estimate and test their theories. The connection between the theoretical enterprise of economics and the empirical work conducted in econometrics never appeared as early econometricians Frisch, Haavelmoo, and Tjalling Koopmans had hoped they would. As Summers (1991) noted, there are few references to econometrics in journals on economic theory, and, when econometrics does have an impact, it is mostly in relation to qualitative

[11] For a good examination of the Phillips curve see Qin (1993).

propositions rather than estimations of structural parameters. To conclude Epstein noted that

> The principal difference with the Cowles approach is that it does not seem likely to develop a reliable theoretical base for the future. Perhaps as a corollary, these investigators do not emphasise that statistical inference in their work is highly contingent on the adequacy of asymptotic approximations to the true finite sample distributions of estimators in models with lagged dependent variables. They retain the use of linear difference equations, similar to Tinbergen's final forms, but seem less concerned with the problem of model selection – or even hypothesis selection – in this framework than many other schools of econometric thought. The approach tends to stress forecasting and prediction with little regard for changes in underlying economic structure [(Epstein 2014), p. 6].

## 2.3 Atheoretical Econometrics: Measurement Without Theory

The atheoretical approach to econometrics was a reaction to the idea that theory was responsible for specifying economic models. Instead, the time-series analysts proposed another way, which would push data to the forefront of the econometric enterprise. As this section shows, it had a few consequences for econometric research, one of which is that the emphasis on theory in the specification of models was abandoned. In the first part of this section, I further show how large a role it plays and the extent to which it affects the result of the model. Furthermore, I introduce two causality tests in the time-series literature: the direct Granger test and the Sims test. Lastly, I discuss how the atheoretical turn in econometrics led to empirical treatment of important concepts in econometrics. I focus on the concept of exogeneity due to its importance in causal modelling.

### 2.3.1 The Atheoretical Approach as the Time-Series Approach

The approach I present in this section is an attempt to move in a new direction. The idea that theory should not decide the specification of the model characterised the time-series econometricians. This is conducted using statistical tests on data to determine the direction of causality and the way to specify the relation in question. This section introduces the basic terminology and formalism used in econometrics [This section is based on Hamilton (2020), mainly Chapter 1].

What we consider in time-series analysis are assumed to be available in the form of economic time-series ($X_t$,....,$X_{tn}$), in which t denotes a certain time. Now, consider the following linear first- order difference equation,

$$X_t = \alpha x_{t-1} + y_t. \tag{2.22}$$

Equation (2.22) relates that, at an unspecified point in time, the value $X$ takes t to another variable, $Y_t$, and to the value $X$ had at an earlier point in time. A difference equation is an expression that relates a variable $X_t$ to its previous values. Equation (2.22) is a first-order difference equation because the first lag of the variable appears in the equation. With lag, I refer to the past period values of a given explanatory variable. Now, consider the following equation,

$$X_t = \alpha_1 x_{t-1} + \beta_1 y_{t-1} \beta_2 y_{t-2} + \varepsilon_t. \tag{2.23}$$

in which $\varepsilon$ denotes the independent and normally distributed error term. In this model, $y_{t-1}$ and $y_{t-2}$ is a cause of $X_t$, if $\beta_1$ and $\beta_2$ is different from zero, and the information set is given by $x_{t-1}$. It isnoteworthy that there is no guarantee of this holding if we were to expand the information set with another variable to, for example,

$$X_t = \alpha_1 x_{t-1} + \beta_1 y_{t-1} \beta_2 y_{t-2} + \vartheta_1 z_{t-1} \upsilon_t \tag{2.24}$$

This model could present a more optimal fit. Perhaps $\beta_1$ and $\beta_2$ are zero, in which case $y_{t-1}$ and $y_{t-2}$ are not a cause of $X_t$ in the new and extended information set that also includes $z_{t-1}$. Whether something is causal is tested by examining whether the regression parameters equal zero. This is often conducted by comparing residual variances, which is in turn done by performing an F- Test, as this section later shows.

There is, however, no doubt that the use of different diagnostic assessments like causality tests are uncontroversial. As future sections discuss, the flowering of these methods has engendered a revolution of the meaning of essential economic concepts, which has rendered the concept empirically testable. The intention was to replace theoretical economic arguments with empirical arguments in model specification. Some of these tests were even developed to settle the quarrel between the Monetarists and the Keynesians, as we will see in a case-study in Chapter 4.

### 2.3.2 The Causal Concept in the Time Series Approach as Granger Causality

To paraphrase a famous quote by Clive W.J. Granger, in terms of causality, most people know better what they do not like than what they do like. When working with Granger causality, it is not unusual to specify that we manage Granger causality by noting that $x$ Granger-causes $y$ instead of merely causing it. Zellner (1979) and Cooley and Leroy (1985) both raised doubts about the philosophical legitimacy of Granger causality, but, as noted in Spohn (1983), Granger causality conveys a specific subset of Suppes' probabilistic account of causality presented in section 2.1.1.2 as the philosophical basis of the atheoretical reductionist position. Thus, the philosophical legitimacy of Granger causality is difficult to question, since Granger causality is part of a larger philosophical tradition. The criticism cannot be made in isolation, which is why focusing on the causal concept in modern econometrics is important. It is a part of a broader philosophical tradition. The general idea in Granger causality, formulated in Granger (1969), builds on the definition of causality provided earlier [See Section 2.1.1.2.0.2., Definition 2.1.2]: if two variables, $X$ and $Y$, are observed over time, $t$, then $X_t$ is taken to be the cause of $Y_t$ if we can explain $Y_t$ better by including past observations of $X_t$, or $X_t, X_{t-1}, .., X_{t-k}$, than if we only included past observations of $Y_t$, or $Y_t, Y_{t-1}, .., Y_{t-k}$. Consequently, Granger causality is best understood in terms of 'predictability' and understands causality as mainly difference-making. This stands in contrast to the theoretical approach to causality, which mainly concerns mechanisms. Granger causality understands a connection as causal if including the past values of the cause makes a difference in our ability to predict the effect. As argued by Granger himself,

> We say that $Y_t$ is causing $X_t$ if we are better able to predict $X_t$ using all available information than if the information apart from $Y_t$ had been used [(Granger 1969), p. 428].

Thus, what we measure across time is predictability. This yield the following formal definition of Granger causality Granger (1969) and Granger (1980):

**Definition 2.3.1.** Granger Causality

1. $X_t$ causes $Y_{t+1}$ iff $\mathrm{P}(Y_{t+1} \in A|\Omega_t) \neq P(Y_{t+1} \in A|\Omega_t - X_t)$ for some $A$.

The definition of Granger causality, then, is not that the occurrence of a given cause $X$ necessitates the effect, $Y$; rather, variations in the values of $X$ at $t_{x-1}, ..t_{x-k}$ produce a change in the probability or predictability of $Y$ at $t+1$.[12] One can define Granger causality in Suppes' terminology, first argued in Spohn (1983), in the following way, let $E_{t'}$=$Y_{t+1}$, $C_t$=$X_t$, $B_t$=$\Omega_t$, $C_t \cap B_t$=$\Omega'_t$. then, we obtain that:

**Definition 2.3.2.** *Granger Causality*($GC*$)

1. $C_t$ is a prima facie Granger cause of $E_{t'}$ with respect to some background $B$ iff;
2. $t'<t$;
3. Both $C_t$ and $E_{t'}$ occurs;
4. $\mathcal{F}(E_{t'}|C_t \cap B_t \neq \mathcal{F}(E_{t'}|B_t)$.

The definition provided here is based on three basic axioms specified by Granger in which we the first axiom is known as the temporal axiom, *the temporal axiom* [(Granger 1980), p. 330]:

> **Axiom** $T$: The past and present may cause the future, but the future cannot cause the past.

This axiom is clearly inspired by Hume's account of causality and primarily excludes the possibility of backward causation and contemporaneous causation. The second axiom is named The nonredundancy axiom *The nonredundancy axiom* [(Granger 1980), p. 330]:

> **Axiom R**: The knowledge set $\Omega$ contains all the information in the universe on a given time $t$. However, $\Omega$ does not contain redundant information.

This axiom is important and rather restrictive. The main reason for having Axiom R for Granger is that redundant information could generate the false idea that a $PMC$ is spurious, especially in economics—as Varcelli noted—where time series often follow similar fluctuating patterns. The last axiom is called the constancy axiom [(Granger 1980), p. 335]:

> **Axiom** $C$: All causal relationships remain constant in direction in time.

Axiom C has been criticised by economists as overly strict [see (Zellner 1979; Zellner 1988)]. However, as noted by Granger, this is common in causal analyses and scientific inferences. More problematic is that this axiom often yields the assumption in many Granger tests that the series in question are covariance stationary. According to Granger, this is highly important for practical purposes − for developing the tools to measure causality − but

[12] Due to the symmetry of conditional probability presented in Chapter 1, it follows that $X_t$ causes $Y_{t+1}$ just in case that $Y_{t+1}$ causes $X_t$ (for more see Salmon (2003)). Simply because:

$$P(Y_{t+1}|\Omega, X_t) = \frac{P(Y_{t+1} \cap X_t|\Omega_t)}{P(X_t|\Omega)} \tag{2.25}$$

$$= \frac{P(X_t|\Omega_t, Y_{t+1})P(Y_{t+1}|\Omega_t)}{P(X_t|\Omega_t)} \tag{2.26}$$

$$\neq P(Y_{t+1}|\Omega_t) \tag{2.27}$$

As one can see time series solve the problem of symmetry. But in cases of instantaneous causality the problem arise ones again. For if we substitute $X_t$ by $X_{t+1}$ in equation (2.25)-(2.27) and it becomes impossible by empirical means to decide the direction of causality.

highly problematic for several reasons because time series are rarely stationary. Perhaps most important, however, is that Axiom R and Axiom C would actually warrant a need for theory. In practise, the information set $\Omega$ could never contain all information in the universe. Rather, the axiom is an idealization. In practise, the content of the information set is restricted. This leads to two questions: how is it restricted, and what restriction procedure is justified? This issue warrants the need for theory.

One of the main issues with the Granger method is its inability to distinguish between genuine and spurious cases of causality. This leads to a lack of agreement between different Granger tests and the problem of how to choose evidence (my second research question). As Vercelli (1991) and Vercelli (2017a) noted, it may appear that the supporters of Granger causality have been unaware of the limits of inductive methods and, as a consequence, have accepted one of the following three questionable axioms:

1. That measurement is possible without theory - and actually that it is possible to measure causality without theory. I take this to mean, that measurement is reducible to relation between observations. I argue against this point following Klein,

    > To some extent vector autoregressions are associated in my mind with the concept that Koopmans introduced, 'Measurement without Theory.' I think that they are eventually going to be misleading from that point of view. I look at the problem in the following way: When we first put our models together, people said that the relevant test should be the random walk, or today equals yesterday. Then, after that became a not very severe test – after it was shown that that was not a good standard – people went on to the next more sophisticated criterion, today's changes equal yesterday's changes. Then they went to autoregression, then they went to ARIMA models; and now they have gone to vector autoregression. So I regard vector autoregression as being in this sequence of moving from the most simplistic model of testing, which we call the naive model, to a semi-naive model which is, in the present state, a vector autoregression. In all these tests we have noticed that the systems that represent 'measurement with theory' break down at turning points; they break down under unusual circumstances and they cumulate error fast. The vector autoregression is the first of such systems that doesn't seem to cumulate error very fast, a least at this stage of the process [(Nell and Errouaki 2013), p. 212-123].

2. Correlation implies causation.

3. post hoc ergo propter hoc.[13]

This thesis is principally concerned with showing that the first of these axioms is incorrect, as Chapter 3 discusses. Following that, Chapter 4 discusses the third questionable axiom and shows that the lack of theory is the primary reason for the sensitivity in atheoretical instruments.

### 2.3.3 Time Series Econometrics: Different Tests

As noted in Maziarz (2015), Granger causality is the most commonly applied definition of causality in economics and econometrics. Indeed, the literature has produced multiple tests

[13]Sims (1972), [p. 543] has the following remark: ' Finally, we ought to consider whether the bivariate model underlying this paper could be mimicking a more complicated model with a different causal structure. The method of identifying causal direction employed here does rest on a sophisticated version of the post hoc ergo propter hoc princip'. Here Sims turns around the famous criticism levied on these atheoretical models by Tobin (1970).

to discover causality understood in the Granger sense. As noted in the previous section, a lack of a theory does mean that no background theory is a part of these tests, and, as observed in the first section of this chapter does hold that causal claims can be replaced by statistical properties such as incremental predictivity, and that such models can capture causality [Section 2.1.1]. I will mainly be focusing on the following instruments,

1. **Sims Test**. Proposed in Sims (1972) , the Sims test was the first procedure created to test for Granger causality. The Sims test demands more observations than other tests in that it includes both past and future observations of X [(Sims 1972), p. 545]. As noted in Kirchgassner et al. (2012), this test is rarely applied today.

2. **The Direct Granger Test**. This test was proposed in Sargent and Wallace (1976) and was derived directly from the Granger definition proposed in Granger (1969). It is important to note that the direct Granger test is actually a test of Granger non-causality and not exactly a test of Granger causality [for more, see Kirchgassner et al. (2012), 3.3.].

3. **Pierce Test**. The Pierce is different in the way that it applies estimated residuals of the univariate models for $X$ and $Y$ [(Kirchgassner et al. 2012), p. 104].

To illustrate the Direct Granger Test, consider two simple vector autoregression models, $X_t$ and $Y_t$:

$$X_t = \sum_{i=1}^{n} \alpha_i X_{t-1} + \sum_{j=1}^{n} \beta_j X_{t-j} + \varepsilon_{1t}. \tag{2.28}$$

$$Y_t = \sum_{j=1}^{n} \vartheta_j Y_{t-j} + \sum_{i=1}^{n} \varphi_i X_{t-i} + \varepsilon_{2t}. \tag{2.29}$$

From these two models, there are four possible causal relations between the time series, as noted in Granger (1969), Kirchgassner et al. (2012) suggest 8,

1. $X_{t-i}$ improve the prediction of $Y_t$. In this case $\mathrm{X}_{t-i}$ *granger causes* $Y_t$ (one-sided causal relation).

2. $Y_{t-j}$ improve the prediction of $X_t$. In this case $\mathrm{Y}_{t-j}$ *granger causes* $X_t$ (one-sided causal relation).

3. $X_{t-i}$ improve the prediction of $Y_t$ and $Y_{t-j}$ improve the prediction of $X_t$ (two-sided causal relation).

4. Neither improve the prediction of the other (no correlation).

Granger also mentioned instantaneous causality, although this appears to contradict his axioms. It is notable that instantaneous causality is inconsistent with the axioms since the effect will always occur after the cause in time. In other words, there will always be a minor time lag.[14]

[14]It should be noted that this is highly disputed in the philosophical literature, since some parts of physics seem to contradict it. One of these is the second law of classical Newtonian mechanics:

$$\overrightarrow{F} = m\overrightarrow{a}. \tag{2.30}$$

or

$$\overrightarrow{F} = m\frac{d^2\overrightarrow{x}}{dt^2}. \tag{2.31}$$

. Another example mentioned is the Lorentz equation, which states, as noted in Huemer and Kovitz (2003), p. 560 that a body, $B$, with some charge $q$ moving at some velocity, $v$, through electric and magnetic fields

In a direct Granger test, an ordinary least-squares regression analysis of the following model, in which $\tau$ denotes the residual at t, is typically performed:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_i Y_{t-i} + \tau_t. \tag{2.33}$$

After performing the first ordinary least-squares regression analysis, another ordinary least- squares regression analysis of an expanded model including lagged values is performed. In this test, $\varepsilon$ to denotes the residual at $t$:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_i Y_{t-i} + \vartheta X_{t-1} + ... + \vartheta_j X_{t-j} + \varepsilon_t. \tag{2.34}$$

Lastly, we test the null hypothesis: $H_0$

$$H_0 : \vartheta_1 = ... = \vartheta_j = 0 \tag{2.35}$$

Testing the null hypothesis implies that lagged values of $X$ do not contribute to explanations of $Y$ beyond what the lagged values of $Y$ explain. To test the null hypothesis, a simple F-test is performed. We obtain the F-ratio using the following formula:

$$F = \frac{RSS_0 - RSS_1/i}{RSS_1/N - i - j - 1} \tag{2.36}$$

in which $SS$ represents the residual sum of squares of the equation, and RSS1 represents the residual sum of squares of equation. Further, $N$ represents the number of observations. If the F-test shows statistical significance, we reject the null hypothesis. This is interpreted as $X$ Granger-causes $Y$ ([for more see Freeman (1983), Granger (1969), and Cuddington (1980)]. This is easily generalised to the multivariate case, as argued in Cuddington (1980). It is crucial to recall at this point that, contrary to the Cowles approach outlined in the previous section, there is no theory involved here – neither in the set-up nor in the interpretation of these tests – and this is by design. These models were intended to work as an algorithm. One provides a problem as an input, and the model delivers a verdict that should settle theoretical disputes. An example is whether money causes income [I will return to this case in Chapter 4]. The monetarists claimed that money does cause income, or GNP, also known as 'the monetarist hypothesis'. The Keynesians, however, claimed that the causal chain runs from money trough interest rates and then to income. Thus, the Keynesians the importance of interest rates. In the test I present next, though, Sims was able to show that money does 'granger cause' income. That should have provided a final verdict in that case, but, as we observe in Chapter 4, the empirical literature on the monetarist hypothesis varies widely.

The Sims test mentioned in Sims (1972) is similarly worthy of consideration of historical reasons, since it played a crucial role in the debate over whether money causes income. The basic assumptions of the Sims test are as follows [(Sims 1972), p. 544]:

**Assumption 1** Time series are jointly covariance stationary, by only (i) considering linear predictors.

**Assumption 2** Our criterion of predictive accuracy is given by taking expected squared forecast error.

---

experiences a force given by the following equation:

$$\overrightarrow{F} = q\overrightarrow{E} + q\overrightarrow{v} * \overrightarrow{B}. \tag{2.32}$$

. Therefore, the idea that causality is necessarily sequential is problematic and can be questioned. For more on these examples see Huemer and Kovitz (2003).

Now, consider two stochastic processes $X$ and $Y$, which satisfy the assumptions provided above; then, if $X$ and $Y$ are jointly, purely, linearly indeterministic, we write [(Sims 1972), p. 544]:

$$X_t = \sum_{i=0}^{\infty} a * u_{t-i} + \sum_{i=o}^{\infty} b_i v_{t-i}. \tag{2.37}$$

$$Y_t = \sum_{i=0}^{\infty} c_j u_{t-i} + \sum_{i=o}^{\infty} d_i v_{t-i}. \tag{2.38}$$

in which $a_i, b_i, c_i, d_i$ are constants and $u_t$ and $v_t$ are white noise processes. Sims then showed the following theorem, which was not proven in Granger (1969):

**Theorem 1.** *$Y$ does not cause $X$ in Granger's definition if and only if $a$ or $b$ can be chosen identically 0.*

*Proof.* See Sims (1972). □

In other words, a test of Granger causality is equivalent to $a_i$ and $b_i$ chosen identically to zero for all $i$.

Now, if the bivariate system $X'_t(X_t, Y_t)$ has an autoregressive representation,

$$X_t = \sum_{i=0}^{\infty} \varphi_i * y_{t-i} + \varepsilon_t. \tag{2.39}$$

In this, $\varphi_{ii}$ denotes parameters, and $\varepsilon$ is an unobservable innovation. Sims used this equation to show that the residuals from a projection of $X_t$ onto $Y_{t-i}$, $i \geq 0$ are uncorrelated with all past and future $Y_t$ if $X_t$ does not Granger cause $Y_{t+1}$ [(Sims 1972), p. 545]. Thus, that $X_t$ does not Granger cause $Y_{t+1}$ is equivalent to the condition that $Y_t$ is strictly exogenous [(Kursteiner 2016), p. 10291].[15] The main problem with these tests is that they are extremely sensitive to the number of variables included in the analysis, as argued in Dunne and R. P. Smith (2010). Results are also highly sensitive to the test used; see Nelson (1981), Geweke et al. (1983), Freeman (1983), and Dunne and R. P. Smith (2010). As we will see in chapter 3, this is mainly because there is a non-trivial inference from readings of an instrument to the outcome of a measurement procedure. A non-trivial inference that relies heavily on theory.

### 2.3.4 How Time-Series Econometrics Reshaped Exogeneity

The development of different empirical causality tests led to a revision of the exogeneity concept in economics. The idea that exogeneity was equal to predeterminedness in the Cowles Commission led to the separation of exogeneity and predeterminedness in Sims (1972) and Sargent and Sims (1977). Sims also showed the connection between Granger causality and what he referred to as strict causality. The results proven in Sims (1972) and Sargent and Sims (1977) yielded the idea that, as noted in Vercelli (2017a) [p. 413],

1. Granger causality is a necessary and sufficient condition for exogeneity (which is a necessary condition for efficient estimation).

[15] I should be noted that even if this is the official narrative, Jacobs et al. (1979), p. 409 argued that, 'is actually a test of the informativeness hypothesis and is not a test for exogeneity or causality as is generally believed.'

In 1983, Econometrica published an influential paper about exogeneity, see Engle, Hendry, et al. (1983a).[16] This article distinguished three different exogeneity concepts: (i) weak exogeneity, (ii) strong exogeneity, and (iii) super exogeneity. It was published primarily because precise definitions of "exogeneity" had been elusive. Furthermore, it is unclear precisely what the discovery that a certain variable is "exogenous" entails in a given definition [(Engle, Hendry, et al. 1983a), p 277].

The starting point is a probability distribution for the variables in question. For instance, assume we have the probability distribution for two variables, x and y; then, we can separate these two into the probability of one of them multiplied by the probability of the other conditional on the value of the first, [see Engle, Hendry, et al. (1983a), [p. 276-282], Milhoj (1985)[ p. 24-27]]

$$P(x, v|\Phi) = P(x|y\Phi_x) * P(y|x\Phi_y). \tag{2.40}$$

In which $\Phi$,$\Phi_x$ and $\Phi_y$ y are the probability distributions of the variables in question. Say we are only interested in $\Phi_x$ and assume there is no connection between $\Phi_x$ and $\Phi_y$. Then one can obtain $\Phi_x$ by applying a maximum likelihood estimation solely by maximizing $P(x|\Phi_x)$ and ignoring $P(y|y, \Phi_y)$, which only would be presented as a proportionality factor. The variable $y$ can then be seen as irrelevant and be kept outside the analysis. Say we are instead interested in $\Phi(y)$ and assume further that no connection exists between $\Phi(y)$ and $\Phi(x)$. Then, it is possible to estimate by $P(y|x, \Phi_y)$ and ignore $P(x|\Phi_x)$, since $x$ is determined outside the model. This is understood as $x$ being weakly exogeneus. To better understand this idea, consider a model $\mathcal{M}$, and a collection of variables $R_t$. $\mathcal{M}$ is assumed to be given by the density function, , which relies on the parameters $\Phi$ and the past values prior to $t$ [that is $X_{t-i}$, for which we have that $X_{t-i}$=$(X_{t-1}, X_{t-2}, ..)$]. Thus we have that:

$$\mathcal{M} = \mathcal{F}(R_t|X_{t-1}, \Phi) \tag{2.41}$$

We now spilt the collection of variables $R$ into two $Y_t$ and $Z_t$ and substitute $R_t$ for $Y_t$ and $Z_t$, obtaining:

$$\mathcal{M} = \mathcal{F}(Y_t, Z_t|X_{t-1}, \Phi) \tag{2.42}$$

The density function is then factorised, as seen in (4.25). This results in

$$\mathcal{M} = \mathcal{F}(Y_t|X_{t-1}, Z_t, \Phi_Y) * \mathcal{F}(Z_t|X_{t-1}, \Phi_Y) \tag{2.43}$$

If we are only interested in, for instance, $\Phi_Y$ , and there is no connection between $\Phi_Y$ and $\Phi_Z$, then, as previously discussed, we can simply ignore everything but $\mathcal{F}(Y_t|X_{t-1}, Z_t)$ and consider $Z_t$ as exogeneous or weakly exogeneous in relation to all the parameters $\varphi$ that only depend on $\Phi_Y$ . In this way, exogeneity becomes relative to the variables considered but not to any background theory. Thus, the idea of estimating the values as a true model is gone [as seen in Section 4.1.1]. We become solely interested in the distribution of $\varphi$ rather than in how $Z$ came to be. Consequently, we do not place any restriction on $Z_t$.

If interested in knowing or forecasting multiple periods ahead, learning how $Z_t$ came to be is of the highest importance. We want to know whether $Z_t$ is affected by earlier periods $Y_t$. There might be a feedback mechanism from $Y_t$ that is analysed in later periods of $Zs$. In this case, it is primarily important to model these feedback mechanisms. Two conditions must then be met for a variable to be considered strongly exogenous: (a) $Z_t$ must be weakly exogenous for $Y_t$ , and (b) $Z_t$ must not be granger caused by $Y$.

[16]The three economists that wrote the paper comes from different camps. Two of them from views I do not consider here, Jean-François Richard working mostly on Bayesianism, and David Forbes Hendry, who is one of the most famous economists in the LSE camp. Robert F. Engle is the only one of them who mainly does time-series econometrics.

Super exogenous is strongly inspired by Lucas' famous problem of policy. The basic idea of this is that one cannot use econometric models as a condition for economic policy because the change of policy alters the models in question. For example, even if taxes might be exogenous variables in a model $\mathcal{M}$, you cannot use $\mathcal{M}$ to calculate the effects of tax reform, $P$, since the changed policy alters the consumption patterns calculated in $\mathcal{M}$, yielding a new model $\mathcal{M}_1$. Consider a case in which the Department of Commerce has calculated that, given an $\mathcal{M}$, a sugar tax would earn a certain amount of revenue. This revenue is based on the old model's micro patterns, $\mathcal{M}$. However, the tax is likely to alter the old patterns, making people consume less. This renders $\mathcal{M}$ incapable of calculating the effects of $P$, since the consequence of $P$ is a movement from $\mathcal{M}$ to $\mathcal{M}_1$. To avoid this situation, $Z$ must be weakly exogenous, and the parameters of interest must not be affected by a change in the underlying mechanisms that generate $Z$. This is what is referred to as super exogeneity. This notion is extremely close to the autonomous relations mentioned in the beginning of this chapter [see Section 2.1.1-Section 2.1.2].

Other influential papers on exogeneity were also published in the 1980s, including Sargent and Sims (1977), Cooley and Leroy (1985), and Leamer (1983). The terminology in Cooley and Leroy (1985) follows that from Sargent and Sims (1977) and notes that, for linear models, weak exogeneity corresponds to predeterminedness, and strong exogeneity corresponds to strict exogeneity. However, as noted in Cooley and Leroy (1985), if we consider predeterminedness [weak exogeneity] to be the right exogeneity concept for causality, then Granger causality is neither necessary nor sufficient for exogeneity since Granger causality is neither necessary nor sufficient for predeterminedness. As Cooley and Leroy (1985) p. [289-290] elaborated,

> Since predeterminedness is the exogeneity concept relevant for the analysis of interventions, it follows that the Granger and Sims tests are irrelevant to whether a causal interpretation of a conditional correlation is justified.

Cooley continues this line of thinking, stating that,

> Further, predeterminedness is also the exogeneity concept relevant for econometric estimation, implying that the Granger and Sims tests are equally irrelevant to the question of whether a model is consistently estimated.

If this is the case, then the idea that Granger causality is necessary and sufficient for exogeneity and efficient estimation is incorrect. To emphasise this point, consider the following model:

$$\mathcal{M} = \mathcal{F}(X_{1t}, ..., X_{nt}, ... = u_t. \tag{2.44}$$

$u_t$ here is a stochastic residual with the mean 0. Predetermined variables are variables in which the values of the residuals are uncorrelated to the present and the future, meaning that $X_{nt}$ is predetermined if

$$E(X_{nt} \cdot u_{t+1})) = 0, \quad \text{for i} \geq 0. \tag{2.45}$$

in which '$\cdot$' refers to the dot product. At the same time, exogenous variables are understood as variables independent of residual in infinity, yielding the same result:

$$E(X_{nt} \cdot u_{t+1})) = 0. \tag{2.46}$$

It is notable that strict exogeneity [strong exogeneity] does, in contrast, imply Granger noncausality. However, the acceptance of Granger non-causality does not imply strict exogeneity but is rather consistent with it. Furthermore, strong exogeneity is the concept closest to exogeneity's role in econometrics. However, two requirements are needed, as previously mentioned, leading to the next point. If we assume that a given variable $Z$ is weakly exogenous, then, as noted in Cooley, Granger non-causality and strict causality become the same thing, given predeterminedness. If for instance, $Z$ is established as weakly exogenous, a Granger or Sims test can establish whether the $Z$ is strongly exogenous.

However, this is only the case if the fact that $Z$ is weakly exogenous – that $Z$ is predetermined – has already been established. In order to justify that $Z$ is weakly exogeneous, Sims and Granger tests are irrelevant. Thus, if strong, strict exogeneity is the concept of exogeneity, Granger causality may be necessary but insufficient for establishing causality. Sims failed to provide an atheoretical presentation of the concept, still relying on Z's weakly exogenous status, which is guided by a priori theory as established in Cooley and Leroy (1985)). Thus, we conclude from Engle, Hendry, et al. (1983b), Cooley and Leroy (1985), and Leamer (1983) that, if we take strong exogeneity to be the exogeneity concept of modern econometrics, then Granger causality is necessary. However, it is not sufficient and therefore neither a necessary nor a sufficient condition for correct estimation.

## 2.4 Concluding Remarks

This chapter had two intentions. One was to outline the philosophical foundations of two different approaches to causal modelling in econometrics, which provided the following characterization:

| | Theoretical | Atheoretical |
|---|---|---|
| Metaphysics | Realist | Reductionist |
| Methodology | Mechanistic (Given deductively) | Instrumentalist (Statistical Induction) |
| Background (a priori) | Economic Theory | None (Only variable selection) |
| Assumptions | Theoretical | Statistical; Probabilistic |

Some remarks on this table are necessary. First, in the Cowles approach causal claims are provided as postulates from economic theory. Koopmans specifically argued that econometric models measure the strength of a postulated relation rather than correctness. Further, Koopmans rejected his early falsificationist approach to econometric modelling, [see T. Koopmans (1937)]. I have left the background empty for atheoretical approach, although variables could be mentioned here. The main reason for this move is due to the fact that they do not have any background theory. In addition, I introduced the two approaches in chronological order, beginning with the theoretical approach. I outlined the causal concept in both and showed how crucial economic concepts like exogeneity expanded from possessing only a theoretical interpretation to a statistical one as well. I also provided examples of both approaches, listing three questionable axioms underlying the atheoretical approach – one of them which was that measurement is possible without theory. This is the main topic of chapter 3, in which I show why this is not the case. I argue that the main problem in atheoretical econometrics is exactly a failure to recognize the need for theory in measurement. Another questionable underlying axiom, the 'post hoc ergo propter hoc', is addressed in Chapter 4 in a discussion of an important case in the history of econometrics: whether money causes income.

# 3

## Instrumentalism: Measuring Causality in Atheoretical Econometrics

*This chapter examines the idea behind numerical representation -that is, measurement- in economic models, mainly inspired by M. Boumans, L. Mari, and E. Tal. I pose two problems for econometrics, one of them being the problem of passive observation and the other how to generate evidence based on passive observation. The main question here will be, is measurement based on observation? Or is something else needed? I suggest that more is required, contrary to the idea that measurement without theory is possible, following E. Tal and K. Staley. However, one of the most popular analyses of measurement today, the representational theory of measurement, suggests that measurement is a homomorphic mapping of empirical relational systems to numerical relational systems; in other words, we can reduce measurement to relations between observables. The representational theory's relationship to econometrics and the problem of foundationalism in measurement is investigated. I emphasise the need to move beyond foundationalism, following the epistemological shift in measurement approaches, emphasizing the distinction between a reading and the measurement outcome. I argue that the measurement outcome is a range of acceptable values, consistent with both theoretical and statistical assumptions.*

> If you cannot measure, measure anyhow.
>
> Frank Knight, 'A General Method in Proofs of Undecidability'

Chapter 2 provided an introduction to causal models in econometrics and its philosophical foundation; therein, I argued that causal economic models in atheoretical time-series econometrics and its instrumentalist approach to econometrics show that macroeconomic models should be considered measuring instruments, not much different from thermometers and clocks. I also argued that the time-series approach is committed to what I refer to as HC, which is that causal relations are reducible to non-causal ones. Further, the atheoretical view advocates implicitly for an epistemological doctrine that I referred to as ER+, which states that we can replace causal claims with regularities found in the data. ER+ follows trivially since HC claims the identity between regularities and causal claims. I located this approach in the tradition of probabilistic causality and further note that Granger causality is a subset of the very same probabilistic approach. Further, I argued that this instrumentalist view is also based on a reductionist view of measurement in which measurement is reducible to relations between observations. In this chapter, I contend that one problem remains: that measurement is not possible without theory, and that measurement is not reducible to relations between observations; for in order to obtain the stability that characterize measurement, theory is necessary. Thus, I maintain that there is no purely empirical way to transform qualitative observation into numerical representation. I further claim that this is precisely what the atheoretical approach in econometrics takes for granted. In short, I dismiss that measurement outcomes can be reduced to relations between qualitative observations alone. There is a non-trivial inference from readings of instruments to outcomes and it is exactly this non-trivial inference based on theory that generates stability in outcome. Thus, I begin by highlighting the general problems of economic measurement and rejecting the idea that observations play a special role in measurement. Instead, I move on to assert that it is exactly the relative high stability of measurment compared to other forms of quantitative estimation that makes measurement unique and it is eaxtly this stability that atheoretical econometric instruments lack. In the two last sections in the chapter, I introduce the empiricist theory of measurement, which I reject, before examining the model-based theory of measurement, which stresses the importance of theory in generating stability in measurement outcome.

## 3.1 Is Measurement Observational? Establishing a Theoretical Basis of Measurement

In the previous chapter, I note that the atheoretical position rests on the idea that measurement is possible without theory – that it is indeed possible to measure causality without theory. This means that measurement is reducible to relations between observations, as represented in the commitment to HC. To better understand what this means, I analyse some of the definitions in this section. I begin, though, by discussing some of the literature on measurement in econometrics. After this, I discuss what I mean by 'observation'. Then, I examine a problem that appears between conflicting intuitions about measurement: one defending the notion that measurement can be reduced to relations among observations and the other rejecting that this is the case.

### 3.1.1 The Problem of Economic Measurement: Passive Observation and Accuracy

As noted by Hood and Koopmans,

> the measurements being made by or for the scientist, but their exactness not entirely controlled by him. In some fields, the experimental control exercised by the scientist is a more significant factor than in others. In economics, it is of much less significance than in most. But in all cases the choice and the usefulness of a method (...) of testing a hypotheses depend on the character (assumed or known) of the process generating the observations [T. Koopmans and Hood (1953), 114].

The main point here is that the scientists have far less experimental control of the measurements being made in economics than in the natural sciences. This is also known as the problem of passive observation [See Boumans (2010b)]. Instead, econometrics relies on prior knowledge of the process that generated the observations in question [The DGP]. Most modern theories of statistical inference describe this process in the following way, assume that, for any number of observation, N, an unknown probability distribution function $\mathcal{F}_n(x)$, exists of as many variables, $x = (x_1, x_2, ..., x_n)$ as there are observations, and that the observations made are random drawings from this distribution [example from T. Koopmans and Hood (1953), p. 114]. The way to obtain knowledge about the underlying distribution from the observations, which are random drawings from the distribution, is to specify a set $\mathfrak{F}_n$ of distribution functions, of which the distribution $\mathcal{F}_n(x)$ generating the data is assumed to be an element. The set $\mathfrak{F}_n$ now represents the a priori knowledge that the statistician or econometrician feels justified in specifying the underlying DGP. This is driven by either confidence in the correctness of $\mathfrak{F}_n$ or a wish to explore its implications. For Haavelmo, the problem of passive observation is close to that of identification, as I noted in the previous chapter. This specific subproblem of identification can be referred to as the problem of simultaneity, as argued in Boumans (2010b). Further, Haavelmo's paper, 'The Problem of Testing Economic Theories by Means of Passive Observations', was mainly about this problem, which he defined in the following way:

> Can we measure economic structure relations (e.g., individual indifference surfaces or other "behavioristic" relations) by means of data which satisfy simultaneously a whole network of such relations, i.e., data obtained by a "passive watching of the game" and not by planned experiments? [Quoted from (Boumans 2010b),p. 77.]

This problem was also highlighted in Frisch:

> This is the nature of passive observations, where the investigator is restricted to observing what happens when all equations in a large determinate system are actually fulfilled simultaneously [Quoted from (Boumans 2010b),p. 77.].

Boumans (2010b) continue, 'this problem came to be known as the problem of identifying structural relationships when all we have are passive observations, so it is closely related to the problem of identification'. It clarifies the need for theory in measurement, which is more important than the need for theory in laboratory sciences since the ability to isolate causes is far more difficult in sciences outside the laboratory [P. 77]. The only way out of this cycle, according to Haavelmo, is economic theory:

> Our hope in economic theory and research is that it may be possible to establish constant and relatively simple relations between dependent variables, y (of the type described above), and a relatively small number of independent variables, x [(Haavelmo 1944), p. 22-23].

Morgenstern, another economist extremely critical of the accuracy of measurement in economics, also noticed the problem of passive observation, even considering it the main problem in mathematical economics. As Morgenstern argued,

> most frequently made of unique phenomena. Sometimes the same event is observed simultaneously by different observers who are, however, seldom scientific observers [quoted from (Boumans 2012a), p. 114].

Here, Morgenstern problematises the lack of experimental data in economics. He even argued that the problem of measurement in economics can be reduced to passive observation. For instance, consider that a scientist, $S$, witnesses an event $E$. In justifying $E$, $S$ cannot appeal to the reliable laboratory standard of reproducibility or what Morgenstern considered the ideal standard. Instead, the reliability of the justification of $E$ is partly based on the reliability of $S$ and her ability of interpretation. These, according to Morgenstern, will always be highly dependent on the theoretical framework in which they are observed. Thus, Morgenstern argued, as the economists in the Cowles Commission, affirming a theoretical framework is necessary for knowledge in economics – most explicitly argued in T. Koopmans (1953). Hence, the reliability of that same knowledge depends on the theoretical framework affirmed by the interpreter. For Morgenstern, the problem of passive observation and the problem of error meant that the problem of measurement is severe and important in economics and is worthy of greater focus than the economic profession had given it. Morgenstern noted that the lack of focus on the problem of measurement had led to a rather uncertain development in the field of economics, see Morgenstern (1963) . I follow this line of thought in this chapter.

### 3.1.2 The Problem of Ecomomic Measurement: Generating Evidence on the Basis of Passive Observation

As observed in chapter two, measurement is also characterised by two views of how measurement procedures generate evidence. One is a reductionist position claiming that observations play a key role in justifying measurement outcomes, and the other claims that observations do not play a special role in the justification of measurement [(Tal 2016b), p. 3],

> **Intuition 1**: Measurement is the most reliable source of evidence and is more reliable than other forms of quantitative estimation, including computer simulation and theoretical prediction.
>
> **Intuition 2**: If Intuition 1 is correct, it is because observation plays a crucial role in justifying measurement outcomes.
>
> **Intuition 3**: If Intuition 1 is correct, it is not due to the role of observation. On the contrary, observation does not play a special role in measurement.

The conflicting intuitions are intuitions two and three, and the most pressing question at stake is a reformulation of the HC:

> **Measurement Reductionism (MR):** Is measurement outcomes reducible to qualitative relations between observables?

Choosing to accept either intuition 2 or 3 is not a new discussion in econometrics, as illustrated by the famous measurement without theory controversy between the two camps introduced in the previous chapter. Unsurprisingly, atheoretical econometricians claim that it is possible to measure without theory, as outlined in Burns and Mitchell (1946). However, Tjalling Koopmans, a Cowles Commission econometrician, argued the opposite.

The controversy itself began with a review of A. Burns and W. Mitchell by Tjalling Koopmans in T. Koopmans (1947). This review led to a series of articles being exchanged between Vining and Koopmans. In the paper T. Koopmans (1947), Koopmans distinguished between two stages of the development of a determinate science:

1. **The Kepler Stage**. In the Kepler stage of scientific development, researchers are still at the logical level of abstraction, trying to figure out the basic relations between the examined variables: causal laws or a set of causal laws.

2. **The Newton Stage**. In the Newton stage of scientific development, a general theory has been discovered: we have causal laws, or a set of causal laws, which means that predictability is possible.

According to Koopmans, the Cowles Commission was working on the Newtonian level based on a Walrasian foundation. They were working from the idea of the Walrasian general equilibrium theory, which was essentially built upon the aggregation of individual agents as units. The Walrasian theory narrowed the parameter space for Cowles and presented a range of acceptable values for different variables to take. That said,

> It appears to be the intention of Burns and Mitchell – in any case it is the opinion of the present reviewer – that their book represents an important contribution to the 'Kepler stage' of inquiry in the field of economics [(T. Koopmans 1947), p. 162].

Thus, the Cowles commission had a general theory, but Burns and Mitchell did not. As Koopmans argued, that is the only way to achieve a

> Fuller utilization of the concepts and hypotheses of economic theory . . . as a part of the practices of observation and measurement promises to be a shorter road, perhaps even the only road, to an understanding of cyclical fluctuations [(T. Koopmans 1947), p. 162].

Although the Cowles commission had a clear theoretical basis from which they worked, Qin, a well-known historian of econometrics, provided a different story:

> Since the central task of the Cowles Commission was to formalise the statistical methods (...) for econometric analyses, given economic theory, they consciously left open the issue of how to put particular economic theory into a particular structural model [(Qin 1993), p. 63].

This story does not seem entirely correct, however, and this is because the econometrics researchers at the Cowles Commission did not '[leave] open the issue'. The answer was the Walrasian general-equilibrium theory. There was a clear underlying economic theory that prohibited that model selection from becoming an issue at Cowles, and this was, according to Koopmans, exactly where the people working there differed from those working being at the NBER. Burns and Mitchell did not have any economic theory in *Measuring Business Cycles*, as Koopmans observed:

> The examples given illustrate the authors' scientific "strategy," in which measurement and observation precede, and are largely independent of, any attempts toward the explanation of economic fluctuations [(T. Koopmans 1947), p. 166].

Instead, Burns argued that such theoretical restrictions were not necessary for measuring economic cycles. According to Koopmans, though, this lack of theory was a hindrance similar to that of pilots flying blindfolded without autopilot. Walrasian general equilibrium theory provided an idealisation and distribution of systematic errors with an awareness that measurement was partly a theoretical endeavour. None of this was present in Burns and Mitchell's book. Consequently, Koopmans presented three criticisms against the apparent 'measurement without theory' found in Burns and Mitchell's publication [(T. Koopmans 1947), p. 164-169]:

1. The choice of time-series and variables are not well defended.
2. No awareness of the problem of identification and no apparent willingness to solve it.
3. The choice of statistics, or in other words, that the result is too sensitive to the measurement procedure used and thus incapable of producing stable causal claims. As Koopmans argued, 'The question should therefore now be raised whether the authors' finding of strong domination of random variation over possible traces of systematic change in cyclical "behaviour" is not at least partly due to the choice of the particular "statistics" studied'.

Vining responded on behalf of NBER, accusing the researchers working at Cowles of being guilty of the same sin. They also undertook measurement without the appropriate theory, as argued in Vining (1949). Vining even called Walrasian general equilibrium theory 'a pretty skinny fellow', since the theory of how agents behave had not been full conceptualised – at least not enough for it to be the foundation of high-powered statistical estimation. The criticism is summarised here:

> Some of his [Koopmans] discussion suggests that we have already at hand a theoretical model (...) Koopmans doesn't give his hypotheses specific economic content. He discusses the mathematical form that the model should (or must) take; and suggests the kind of content it should have in very general terms. . . . But apparently all he has to insist upon at present is the mathematical form, and from his discussion it appears not unfair to regard the formal economic theory underlying his approach as being in the main available front works not later than those of Walras [Vining quoted from (Qin 1993), p. 177].

Highlighting that the Walrasian model of the economy was still awaiting confirmation and that it was a 'pretty skinny fellow' of untested capacity shows that Vining did not understand the argument proposed by Koopmans on two fronts.

Firstly, Koopmans' project was to defend a dependence on economic theory in general rather than a particular theory. His main intention was to suggest that measurement is not reducible to relations between qualitative observations. Instead, those measurements relied on theory. There is no doubt that Cowles worked from a particular theory, but this was public knowledge. However, Koopmans was uninterested in defending Walrasian theory. At the same time, Vining defended the relationship between measurement and observation against the attacks launched by Koopmans but gave an unclear response to the problems raised by Koopmans. Furthermore, demanding confirmation of the underlying theory on which one bases choices leads to circularity. Eventually, choices must be made, and they must based on some measure not endorsed by the data. Here, though, I favour articulated theory over personal judgements, in which the personal judgements are not argued for or defended. The debate does clearly show, though, that the NBER economists defended a reductionist and partly instrumental approach – that is, intuition 2 to econometrics – which was exceedingly close to the MR principle in the beginning. In contrast, Koopmans, a Cowles economist, defended intuition 3. In this chapter, I defend intuition 3 along the line of Koopmans.

### 3.1.3 A Closer Look at Observation

Hoover (2007) argued that econometrics is a kind of observation and he argued in favour of what he referred to as 'weak a priorism'. In short, Hoover argued in part for the idea that measurement is reducible to qualitative relations among observations. Since the measuring instrument is just another way to observe, this position contrasts to that of Koopman's 'strong aprioism', and Hoover is closer to the atheoretical than the theoretical position. According to Hoover, one can define the two positions in the following ways [(Hoover 2007), p. 70-72]:

> **Strong a priorism**. Strong a priorists like Koopmans have considered the objective of econometrics to be the direct measurement of a structure suggested by economic theories. In other words, economic theory supplies replicas of economic reality that the econometrician then estimates.
>
> **Weak a priorism**. Weak a priorists like Hoover have rejected the main objective of strong a priorism. The task is not to supply replicas of economic reality, for such a

> task is not practicable. We do not have a background that sufficiently imposes such constraints on the structure since we do not know which constraints reality places on economic structure. Consequently, 'measurement requires prior theory; equally, theory requires prior measurement.'

The main problem is that econometrics and other measuring instruments are not solely 'observations'. Econometric results are not 'first and foremost, calculations, summaries of observable data' because measurements are not 'epistemically or semantically reducible to relations among qualitative observations' [(Tal 2016a), p. 5]. Rather, the measuring instruments used in econometrics and the sciences are highly sophisticated theoretical devices. Indeed, the characterisation of Koopmans' position is at best imprecise. Koopmans correctly believed that economic theory should provide structure, postulating that the main point of a model is to represent such. However, this is not because Koopmans thought it possible to provide a replica of the economy. T. Koopmans (1937) was actually an early proponent of the idea that exact relations are not possible in econometrics, thereby rejecting that it is possible to measure what Hendry referred to as the 'data generating process' [DGP] or what Hoover referred to as 'replicas of the economy'. Koopmans also rejected the idea that we have sufficient background to be confident that a given restriction is the right one. But it is all we have. To solve the problem of identification, we must provide background theory that eliminates observationally equivalent alternatives. So, all Koopmans contended was that some restrictions are necessary to make sense of econometric models T. Koopmans (1937), T. Koopmans (1941), and T. Koopmans (1947) Hoover 2007 [p. 73 further argues is already noted that 'Measurement requires prior theory; equally, theory requires prior measurement.']. However, as Mari (2005) astutely argued, this circularity is a direct consequence of a foundationalist approach to measurement, for, as Hanson (1965) explained, data is always theory-laden.

To better understand whether measuring instruments in econometrics are observations, it is useful to unpack what is meant by 'observation'. As noted in Israel-Jost (2011), the literature has identified three different views on observations [(Israel-Jost 2011), p. 29-31]:

> **The Empiricist View**(EV). Observations are mere perceptions given in a nontheoretical vocabulary – in other words, when the senses are applied to obtain information.
>
> **The Empiricist View***(EV*). Observations are mere perceptions but include observations made by sophisticated instruments and theoretical terms.
>
> **Theoretical View**(TV). There is no such thing as a theory-free observation.

As Israel-Jost (2011) noted, EV has been widely rejected. Thus, most either share EV* or TV. It is true that measurement is observation in the TV view, as noted in Tal (2016a). In this context, observations are used in a much broader sense than what is usually the case in science. However, Tal (2016a) noted, observation is interesting in cases where we understand it as unaided sense perceptions. In the EV* view, one should be careful about what the user observes. The person using a measuring instrument does not observe what the instrument is measuring; that is the data gathered. One observes the instrument's reading, such as a line of numbers on a display. In this sense, I partly agree with Hoover that measurement is observation. That said, the reading of the instrument is not the interesting part of measurement. That is the outcome, Inferring anything from a reading of an instrument includes idealisation, inference, abstraction, statistics, and other parameters. It is a technical process that transcends observation in any empirical sense. The main reason why the empiricist explanation is so tempting is precisely that the justification of the reading is immediately present, for,

> If measurement outcomes are no more than observational reports cast in mathematical language, the empirical content of any measurement outcome is reducible in principle to some set of observations [(Tal 2016a), p. 2].

The problem, though, is that measurement outcomes are not. This is because there is a nontrivial inference to be made from an instrument to a measurement outcome. Consider, for example, two thermometers, $T_1$ and $T_2$, that are meant to measure the same temperature. The problem now is that different thermometers tend to provide different readings. Consequently, they will not provide consistent readings of temperature. The conclusion from this phenomenon is not that the patient, the temperature being measured, is exhibiting different temperatures or different kinds of temperature. Rather, they infer that $T_1$ and $T_2$ approximate the temperature that the patient exhibits. Hence, we idealise. Secondly, the approximation is not uniquely determined by observing the reading of $T_1$ and $T_2$. As a result, the outcome we infer from the reading is not entirely determined by data and is therefore underdetermined by it. Hence, theory is needed.

### 3.1.4 What Makes Measurement Unique: The Stability of Evidence

We now have a question to answer: What makes measurement unique if not observation? Earlier in this section, I denied the notion that observation is the unique characteristic of measurement. Instead, following Tal (2016a), I argue that stability is the unique characteristic of measurement, or what Staley (2004) and Staley (2012) refer to as 'security'.[1] Staley define the concept in the following way:

**Definition 3.1.1.** Security of evidence Suppose that $\mathcal{W}_0$ is the set of all epistemically possible worlds relative to epistemic situation $s$, and $\mathcal{W}_0 \subseteq \mathcal{W}_1$. A proposition $P$ is secure throughout $\mathcal{W}_1$ relative to $s$ if for any world $w \in \mathcal{W}_1$, $P$ is true. If $P$ is secure throughout $\mathcal{W}_0$, then $P$ is fully secure [(Staley 2012), p. 30].

In short, we assume that an evidence claim, $P$, is stable, and that the stability is higher if an agent, $A$, possesses fewer possible scenarios for which the claim is incorrect. Consequently, when we use a thermometer, an audiometer, a compass, a measuring spoon, a pH meter, among other tools, we consider very few scenarios in which these measuring devices exhibit incorrect readings. There are no fully secure instruments, but we often view measurement outcomes as unconditional factual claims simply because we have become accustomed to them being correct. We rarely think about the background theory and assumptions necessary for the instruments to function. The question then becomes, why are measuring instruments so stable? This is due to the hard work of metrologists who continually improve the instruments and make them less sensitive to multiple factors. They accomplish this by applying strategies; the more sensitive $P$ is, the more probable it is that there is some possible scenario, in which $P$ is not true, as noted in Tal (2016a). These factors could be

1. The Procedure in Question.

2. Changes in parameter values.

3. Correcting for Systematic Errors.

4. Environmental circumstances.

[1] To stay consistent, I will after the definition use 'stability' and not 'security', although, both concepts can be used interchangeably.

5. Changes in contested underlying assumptions. In evaluating the stability here, Staley offers two questions [(Staley 2004), p. 471]:
   (a) How strong is the evidence for the applied assumption?
   (b) How sensitive is the evidence claim, $P$, to the failure of the applied assumption?

This is the very problem that atheoretical econometrics faces: that evidence claims provided by the models are overly unstable. They are too sensitive to the items on this list, and there is a direct relationship between robustness and stability. As Staley (2004)[p. 467]) argued, robustness, which occurs 'when multiple independent tests yield the same (convergent) positive result,' enhances the stability of an evidence claim, $P$.[2] Thus, the more robustness tests an evidence claim $P$ survives, the more stable it is. In metrology, it is common to test for the robustness of evidence claims produced by measuring instruments through the application of different strategies:

1. Repeatability and reproducibility tests [Hacking (1983) and Culp (1994).]
2. Calibration and traceability [See Boumans (2004),Boumans (2012b),Cooley (1997)].
3. Revising uncertainties of incompatible outcomes.
4. Avoid contested assumptions, [Tal (2016a)].

It is essential to distinguish between the stability of an evidence claim, $P$, and the strength of $P$. As Staley (2004) noted, the strength $P$ is indicative of the correctness of a hypothesis, $H$. Despite this, stability navigates $P$'s sensitivity to failing due to an auxiliary assumption. Emphasizing stability to assumptions in econometrics is not a new concept. For example, Leamer (1983) argued that, in cases where the inference is sensitive to the assumptions one makes, one should suspend judgement as Leamer did in his proposed case study of the death penalty [(Leamer 1983), p. 42]. Leamer (1983) focused on '(...) a more complete, but still an economical to report the mapping of assumptions into inferences'. In so doing, it is vital that 'all assumptions in a specific set lead to essentially the same inference' [(Leamer 1983), p. 38]. In statistics, these ideas were propagated by Danish statistician Georg Rasch, who referred to this characteristic as 'specific objectivity' [For more see Kaergaard (1984)]. As he noted, a model that possesses this characteristic is insensitive to unknown parameters in the overall system [(Kaergaard 1984), p. 14-15].

The ideal way to increase stability is by applying theory. I will later argue that it is in the form of calibration. I have argued so far that measurement is not unique because it stands in a special relationship with observation, that is by preserving qualitative observation structure. Rather, it is unique because it constrain the parameter space for possible values, by calibrating instrument to a given standard. Consider a medical example: a fever. It is usually defined as a temperature above 38 degrees Celsius. This is not the result of purely data outputs but a complex process that includes knowledge of the body, theory, data analysis tools, simplification of assumptions, corrections for systematic errors, and other parameters. This process teaches that, if we want to discover fever, we should restrict our parameter space of possible values to above 38 degrees Celsius and we calibrate our instruments according to this standard. Why? Because this is where the best evidence of observed indications most likely lie. Hence, the stability that makes measuring instruments unique, are not due to its proximity to observation, but a much more complex process. The same is true for econometric models. If we want models that produce more stable

[2]Multiple philosophers have argued for the importance of robustness in experimental results, for more see Hacking (1983) and Culp (1994).

evidence claims, we must abandon the idea of 'letting the data speak'. Since we know that some regressions are nonsense correlations, the knowledge of which dates back to Yule (1926), we need background knowledge to restrict the parameter space to exclude those very correlations as possible explanations of any real-world phenomena. This is obtained here, since these nonsense correlations are not likely to be in the range of acceptable evidence.

## 3.2 Rejecting Theory in Measurement: The Empiricist View of Measurement

To demonstrate why measurement cannot be reduced to relations among observables, I examine the most prominent effort to do just that: the representational theory of measurement (RTM). I then show that the representational view is common in econometrics. To help identify the conditions for better measurements, I discuss some of the flaws of the RTM, particularly its inability to deal with systematic errors, and the main problem of operationalism and underdetermination.

### 3.2.1 The Representational Theory of Measurement

One of the main 20th century developments in the realm of measurement theory was *The Representational Theory of Measurement $RTM$*, a breakthrough made by Patrick Suppes (among others), set out in series of articles and books Scott and Suppes (1958), Suppes and D. H. Krantz (2007), Suppes and Zinnes (1962), and E. a. Krantz D. (1971). This literature provides a foundationalist theory of measurement. The breakthrough provided an empiricist answer to the paradox raised in the introduction of this paper, thereby offering an answer to a century-old question on how to move from qualitative observations to quantitative representation. $RTM$ is usually seen as one of most important account of measurement in the literature, as noted in Swistak (1990), Cartwright (2008), and Boumans (2016). The $RTM$ does not rely on any metaphysical intuitions about the character of the world or whether quantities are in fact properties of the objects themselves. Instead, the RTM holds that [(Mari 2005), p. 262]:

1. The world is non-quantitative.

2. We impose quantities on the world when we assign numbers to it.

As noted by E. a. Krantz D. (1971) RTM provides: 'basic procedures for assigning numbers to objects or events on the basis of qualitative observations of attributes' [(E. a. Krantz D. 1971), 2-3,]. Hence, RTM offers a way for researchers to pass directly from sense data to numerical presentation in a non-arbitrary way, in the language of mathematics. The main problems RTM sets out to solve is the following, as noted in Chao (2020):

1. **The problem of Representation:** How do we justify assigning numbers to phenomena or objects? This is done by proving a representational theorem showing that some empirical structure ($S_E$) are isomorphic to some arithmetical structure ($S_A$). This is significant because it is possible to apply different computational methods to $S_A$ in order to obtain facts about the isomorphic $S_E$.

2. **The problem of Uniqueness:** How can one determine the scale type of a given procedure? Usually, this is done by proving a uniqueness theorem.

This approach to measurement is based on the idea that there exists a foundational scale of measurement: counting. All other measurement scales were derived from this foundational scale. For example, compare the members of a set with mass. In the former case, the number of members is unique, but the same does not hold for the latter. This is because the empirical procedure does not determine the unit of mass [(Suppes 2002)].

The contrast has often been criticised for being too *abstract* as noted in Boumans (2016) and Mari (2005), which is at odds with Suppes more practical approach to Psychology. However one may see this as a way to bridge the semantic view of theories with empirical experiments. Heilmann (2015) argued that 'RTM is simply viewed as a library of theorems. That is to say, in what follows, the term RTM will refer to the theorems in the three books that contain the authoritative statement of RTM' [(Heilmann 2015), p. 791]. Further it should be noted that the $RTM$ does not account for 'measurement procedures, devices, and methods; and it applies only to error-free data' [(Boumans 2016), p. 2], thus it cannot help to settle the epistemological problem that this thesis deals with. I will return to these problem in Section 3.1.1.

One example of fundamental measurement, as Suppes (2002) discusses, is the ordinal theory of measurement. Models of this theory are called 'weak orders.' We understand weak orders and structure in the following way[(Suppes 2002), p. 62]:

**Definition 3.2.1.** Simple Relational Structure
Let $S$ be a non-empty set and $R$ a binary relation on $S$. Then the pair $\langle S,R\rangle$ is a *simple relation structure.*

**Definition 3.2.2.** Weak order
A *simple relation structure* $\langle S,R\rangle$ is a weak ordering iff ∀xyz (i) $xRy \vee yRx$ and (ii) $xRy \wedge yRz \rightarrow xRz$

Another important definition here is that of an isomorphism to a *simple relation structure* (Suppes 2002):

**Definition 3.2.3.** Isomorphism
Let $\langle S,R\rangle$ and $\langle S',R'\rangle$ be pairs, then $\langle S,R\rangle$ is isomorphic to $\langle S',R'\rangle$ iff there is a function $\mathcal{F}$ such that:

1. The domain of $\mathcal{F}$ is $S$ and the range of $\mathcal{F}$ is $S'$.
2. $\mathcal{F}$ is a one-one function.
3. If x and y are in $S$ then $xRy$ iff $\mathcal{F}(x)R'\mathcal{F}(y)$

As Suppes (2002) notes, there are cases in which homomorphisms, the weaker notion, are more interesting than isomorphisms. One of these cases is the theory of measurement. The main problem relating to isomorphisms is a tendency to assign the same amount/number to different objects, thus violating the one-one function condition of F. However, according to Suppes, weakening condition 2 in definition 3.1.2. is the only weakening necessary to obtain an adequate account for theories of measurement. The concept of homomorphism can accommodate this change, since this is the only condition that changes in a homomorphism:[3]

**Definition 3.2.4.** Homomorphism
Let $\langle S,R\rangle$ and $\langle S',R'\rangle$ be pairs, then $\langle S,R\rangle$ is isomorphic to $\langle S',R'\rangle$ iff there is a function $\mathcal{F}$ such that:

[3]As noted in Suppes (2002) the third condition is often weakened in algebra to *if* and not *iff*. In measurement theory, *iff* is used.

1. The domain of $\mathcal{F}$ is $S$ and the range of $\mathcal{F}$ is $S'$.

2. $\mathcal{F}$ is a many-one function.

3. If x and y are in $S$ then $xRy$ iff $\mathcal{F}(x)R'\mathcal{F}(y)$

Suppes begins with a set of empirical entities, objects, and possible outcomes. For which he distinguishes between extensive and intensive properties, treating extensive properties as properties that can be added and intensive properties as properties that cannot. The RTM exploits the formal construct of simple relational structures by interpreting it in terms of empirical and numerical relational structures [(Frigerio2010), p. 126-127]:

**Definition 3.2.5.** Empirical Relational Structures
An empirical relational structure is a relational structure for which the domain is a set of empirically accessible objects (phenomena, physical objects, etc.).

**Definition 3.2.6.** Numerical Relational System
A numerical relational structure is a relational structure for which the domain is a set of mathematical objects).

We then understand these in terms of scales,

**Definition 3.2.7.** Scale
A scale is triple $\langle S_E, S_N, h\rangle$, where:

1. $S_E$ is an empirical relational structure;

2. $S_N$ is a numerical relational structure;

3. The relational structure type of $S_E$, and $S_N$ is the same;

4. h is a homomorphism from $S_E$ into $S_N$.

Providing us with the following definition of measurmenet [(Frigerio 2010), p. 126-127]:

**Definition 3.2.8.** Measurement
Measurement consists of the construction of a measurement scale –that is, defining a morphism—such that a numerical relational structure represents an empirical structure

To better understand the above definitions, I present a common examples from the literature, a comparative relational structure. Typically, there are two different types of comparison relations [(Domotor and Batitsky 2008), p. 130]:

1. **Comparison relation ($\geq$):** The determination of whether some object $o \in D$, has some quality, $q \in Q$, *compared to some degree*($\geq$), to another object $o_1 \in D$. Hence, one may say that *o is greater than* $o_1$, thus creating a comparison relation between $o$ and $o_1$.

2. **Addition ($\Upsilon$):** The creation of some new object $\mathrm{o}_2$ out of two given ones $o$ and $o_1$, by some additive operation $\Upsilon$ on $o$ and $o_1$.

In this thesis, the intended interpretations can be either objective or subjective probabilities. Thus, one can see the following [(Suppes2002), p. 63-64]:

1. $S$ is set of possible outcomes in a given empirical situation.

2. $\mathcal{F}$ refers events in the ordinary sense.

3. $A \geq B$ iff A is judged at least as probable as B.

One can then obtain the following definition of a comparative relational structure between subjective probability [(Frigerio 2010), p. 127-128]:

**Definition 3.2.9.** Comparative relational structure Let $S_E = \langle S, \succ \rangle$ be an empirical relational system. Then, $S_E$ is a comparative relational system iff, we can map $S_E$ into $S_N = \langle N, \geq \rangle$, that is, iff there exist a homomorphism (def. 3.1.3.) such that:

1. $\mathcal{F} : S_E \mapsto S_N$
2. $\forall x, y \in S_E (x \succ y \rightarrow \mathcal{F}(x) \geq \mathcal{F}(y))$

An ordinal scale measurement theory based on the algebraric interpretation of measurement furthermore identify sufficient conditions (and in this case necessary conditions too) for $S_E$ to be represented by $S_N$, in this case, these conditions are just that $\geq$ is a weak order, hence:

**Theorem 2.** *A simple relational system $S_E = \langle S, \succ \rangle$, such that $\succ$ is a weak-order is a comparative relational system.*

*Proof.* See [(E. a. Krantz D. 1971), ch. 1] □

We refer to these as representation theorem, showing us that a given empirical relational structure can be represented by a specific numerical relational structure.

### 3.2.2 RTM in Econometrics

Examining measurement in the RTM setting is not a rare occurrence, nor is it in economics in general. For example, in utility theory, one might wish to measure the utility of doing a certain choice $C$. Proving a representation theorem here is essential to build a utility function and representation theorems in general play a crucial role in econometrics Chao (2002), Chao (2007), Chao (2005), Chao et al. (2013), and Backhouse (2007). However, one should note that, the axiomatisation of utility theory does not proceed in the usual set-theoretic way. Moreover, it is important to note, that RTM is not a precise model of economic practice, but it is a close enough approximation. As Backhouse argues,

> (...) when economists use the word they refer to systems that can be presented using mathematical notation – using algebra or geometry. This approach to the subject is the method that is best articulated, and as a result it is the most visible. It has much in common with the assumptions underlying the representational theory of measurement [(Backhouse 2007), p. 135].

Hence, one can view RTM as a useful way of modelling econometric practice. Furthermore, as Backhouse and Boumans argue, general econometric practise often views models as representations of the underlying structure. As Chang further argues, representation theorems are also proved in econometrics to ensure that a model actually represents the underlying economic structure. One of the most famous representation theorems in the literature is the granger representation theorem, also known as the Engle-Granger-Johansen thoerem, established in Johansen (1992) and Engle and Granger (1987). It shows that cointegrated variables may be represented as an error correction model. However, representation is not the only problem in measurement approaches. Once more, the main problem in proving a uniqueness theorem is the problem of identification [Chao (2007), p. 288-289]. In this particular situation, the issue is that of observationally equivalent structures. Imagine we have three variables, $o_1, o_2, ..., o_3 \in O$, that generate the same data. In other words, they have the same probability density function. Thus, the question is how to determine which one to choose. This is the main reason why Backhouse includes metaphysical assumptions in his characterisation of important background knowledge in the modelling process:

> Metaphysical assumptions are included as a separate category as a reminder that economists appear, much of the time, to be committed to some of many of the assumptions made in their models for reasons that appear to have little to do with evidence [(Backhouse 2007), p. 136].

However, the kind of econometrics that I surveyed in Chapter 2 cannot appeal to any metaphysical assumptions in the evaluation process. As Chao (2007), argue, the problem in the atheoretical operationalisation of causality in Granger models in Granger (1969), Sims (1972), and Sargent and Sims (1977) is that vector autoregression models (VAR) are not identified. Here we may understand the identification problem in the following way [(Christ 1966), p. 298]:

> It is a truism that any given observed fact, or any set of observed facts, can be explained in many ways. That is, a large number of hypotheses can be framed, each of which if true would account for the observance of the given fact or facts [(Christ 1966), p. 298].

Hence, it then follows, as noted in Chao (2007) that one 'can easily find observationally equivalent VARs that generate the same probability distribution for the data', which gives rise to an identification problem: which VAR should one choose? Particularly, one might ask which VAR he or she should choose without appealing to theory. When theory is applied econometricians hold strong priors, as noted in Hoover (1988a). It is also possible to use other principles to identify the model, such as *simplicity*, Christ (1966):

> The purpose of a model, embodying a priori information (sometimes called the maintained hypothesis), is to rule out most of the hypotheses that are consistent with the observed facts. The ideal situation is one in which, after appeal has been made both to the facts and to the model, only one hypothesis remains acceptable (i.e., is consistent with both). If the "facts" have been correctly observed and the model is correct, the single hypothesis that is consistent with both facts and model must be correct; . . . . In a typical econometrics problem the hypothesis we accept or reject is a statement about the relevant structure or a part of it or a transformation of it [quote from (Chao 2007), p. 289].

Appealing to additional data is of no assistance if the choice is underdetermined by empirical evidence. Thus, VAR models do not satisfy the uniqueness requirement. As noted in Chao (2007), representation theorems in econometrics is important, if the main goal of econometrics is to represent. These are not often stated explicitly but usually defined in a set theoretical way. Later in chapter, I will argue against the idea that the main goal of econometrics is to represent.

### 3.2.3 Problems for Atheoretical Measurement Theory

I argue here that there are two main problems for atheoretical measurement theory. Firstly, I mention the problem of underdetermination, before I move on to the problem of systematic error. Both problems show how difficult it is to obtain stability in measurement outcomes without theory.

#### 3.2.3.1 Underdetermination

One of the main problems raised in Tal (2016a) is the problem of idealization and underdetermination. Tal (2016a) uses the example of a clock and ask the reader to consider two clocks $C_1$ and $C_2$ that that are meant to instantiate the same frequency. The problem here is that 'Even the most stable atomic clocks exhibit a systematic frequency drift when compared accurately enough' [(Tal 2016a), p. 2]. In other words, it is possible to show that $C_1$ and $C_2$ will not provide consistent estimates of duration. But scientists do not deduce that there are multiple clocks as a result. Rather they 'ideal ordering of time intervals and assume that concrete clocks merely approximate this ideal [(Tal 2016a), p. 2]'. This leads to a second problem for the empiricist, for such an approximation is not uniquely determined by observation. In other words, the outcome we infer from the reading is not fully determined by data, and as a consequence, theory is needed.

#### 3.2.3.2 Problems for Atheoretical Measurement Theory: Systematic Error

Uncertainty plays a crucial role in the reliability of measurement results. In this section, I introduce another problem with the atheoretical approach to measurement theory: that of systematic error. It is accurate to state that the 'true value' is never reached due to error. This is the case even when

> all of the known or suspected components of error have been evaluated and the appropriate corrections have been applied, there still remains an uncertainty about the correctness of the stated result, that is, a doubt about how well the result of the measurement represents the value of the quantity being measured [(Boumans 2016), p. 26].

To begin, it is important to distinguish between error and uncertainty. Both the social and natural sciences often conflate these terms; indeed, that there is any distinction between them is not always clear. Metrology, however, clearly differentiates the two concepts. The main authority on these matters, as noted in Boumans (2016), is the Joint Committee for Guides in Metrology (JCGM) who published the Guide to the Expression of Uncertainty in Measurement (GUM; JCGM Working Group I [2008]) and the International Vocabulary of Metrology (VIM) (JCGM Working Group II [2012]).

However, the traditional 'error-approach' understands accuracy as agreement between the reading and true value of the measurand [for more see Willink (2013)]. Precision, though, is not committed to the existence of a true value. Instead, precision examines the agreement between two different measurements of a measurand. We also know precision to be an indicator of a given measuring instrument's ability to reproduce [for more, see Barlow (2002) and JCGM (2008), p. 36].

Advancing toward a definition of error on the error-approach, it is important to understand the very purpose of measurement, which is to arrive as close to the 'true value' as possible. As a result, the very idea of measurement must presuppose that there is an ideal result or at least an ideal range of permissible results. Following this reasoning leads to a definition of error in terms of true value (which is in turn given in terms of accuracy; [(Willink 2013), p. 7]:

$$\varepsilon = e - \Phi \tag{3.1}$$

In (3.1), $\varepsilon$ denotes the measurement error and e refers to the reading of a given instrument, I. The issue with this is that, by definition, $\varepsilon$ is a signed quantity [(Willink 2013), p. 8], meaning that the value is unknown. If one knew the true value of the measurand, no measurement would be required. The scientist in question would simply correct for the error in e, and the error, $\varepsilon$=0, would vanish. The literature distinguishes between two different kinds of error: random error and systematic error. Following GUM, I define random error in the following way [(JCGM 2008), p. 37)]:

**Definition 3.2.10.** Random error
The difference between the measurement result, $e$, and the mean that would result from an infinite number of measurements of the same measurand performed under repeatability conditions.

However, if the measurement procedure is biased, then the reading, $e$, produced by mean measurement, would not equal $\Phi$. This leads to the definition of systematic error as the:

**Definition 3.2.11.** Systematic error (GUM)
mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand [(JCGM 2008), p. 37].

This means that the systematic error does not disappear over time by continually repeating the same measurement. Furthermore, as noted in GUM, the definition of systematic error is closely related to the notion of a 'true value'. Systematic errors can manifest in multiple ways, including as follows:

1. Omitted Variables (confounding variables).
2. Wrongly calibrated instrument.
3. Selection Bias.

It is important to understand that the notion of error does not always carry the same meaning in science as in ordinary language. As Barlow (2002) noted, in statistics, error can refer to both a mistake and a discrepancy.[4] To better understand the difference between mistake and discrepancy (uncertainty), consider the case in which we have applied an instrument $I$ and obtained the following results [Barlow (2002) p. 1]:

$$1{,}40, 1{,}42, 1{,}48, 1{,}44, 1{,}47, 1{,}88.$$

Ignoring one reading of the instrument $I$ is exhibiting high precision, with some uncertainty in the third decimal. We identified the reading 1,88 as a mistake, since it is an outlier. Statistical instruments can help to identify such outliers but cannot determine the reason behind them and therefore cannot help to fix them. Further, statistical instruments can provide information about the precision of $I$ but not about the accuracy of $I$. The readings provided by $I$ do not state whether the most accurate reading is the outlier or the bundle of readings with the same first decimal. Simply viewing a list of readings of an instrument is not sufficient for determining whether the readings in question are accurate.

Error as a discrepancy, can be written in two ways [Barlow (2002) p. 1]:

$$y_i = mx_i + c + \varepsilon_i. \tag{3.2}$$

One can either use an error term to refer to the difference between the measured value and the ideal value, also known as the 'error in equation' method, or they can instead write

$$y_i = mx_i + c. \tag{3.3}$$

[4]By mistake here I mean either a Type I mistake, that is mistakenly rejecting a true hypothesis or Type II mistake, that is mistakenly accepting a false hypothesis.

Simply dropping the error term $\varepsilon$ and allowing the equality sign to signify agreement relative to some uncertainty $\sigma$. This is known as the 'uncertainty approach.' Since the notion of error is omitted, it no longer plays a crucial role. Therefore, as previously stated, it is important to distinguish between error and uncertainty. While the former signifies the difference in value between the measured value and the true value, uncertainty signifies the effect of these errors. A basic premise of the uncertainty approach, according to Boumans (2004), is that

> it is not possible to state how well the essentially unique true value of the measurand is known, but only how well it is believed to be known. Measurement uncertainty can therefore be described as a measure of how well one believes one knows the essentially unique true value of the measurand. This uncertainty reflects the incomplete knowledge of the measurand [Quoted from (Boumans 2016), p. 27].

We subsequently transition from assessing the reliability of measurement in terms of error to focusing on uncertainty. Metrology recommends two ways to examine uncertainty, per Farrance and Frenkel (2012), in which we understand standard uncertainty as 'the result of a measurement expressed as a standard deviation',

1. Type A: 'method of evaluation of uncertainty by the statistical analysis of series of observations' [(Farrance and Frenkel 2012), p. 52].
2. Type B: 'method of evaluation of uncertainty by means other than the statistical analysis of series of observations' [(Farrance and Frenkel 2012), p. 52].

It is important to note that GUM procedures rely on the fact that all systematic errors have been corrected. The uncertainty only exists toward the content of the chosen correction rather than in the uncertainty that one has failed to correct for all systematic errors. The uncertainty approach, then, refers to the uncertainty of random errors and the corrections for systematic error, as noted in Farrance and Frenkel (2012); however, it does not entertain the possibility that one has not corrected for all systematic errors, which is a prime concern in econometrics – especially the effect of a possible omitted variable. The complication is that there are multiple ways to distribute errors. This must be done relative to the chosen idealised structure. In the case of Cowles, it would be conducted according to Walrasian general- equilibrium theory, for the underlying selection of variables is not uniquely determined by observation. To illustrate, consider three different econometric models, $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$, that are built on different theories. What is considered a systematic error relative to $\mathcal{M}_1$ may not be considered a systematic error relative to $\mathcal{M}_2$. An example I return to in chapter 4 is whether money causes income. If interest rates are included in the set of variables, this changes the strength of the cause-effect relationship; so, is the omission a systematic error? According to the Keynesians, it is. According to the monetarists, though, it is not. However, no amount of data gathered will be able to tell whether $\mathcal{M}_1$ or $\mathcal{M}_2$ is the best fit. There are multiple ways to distribute errors, all of which are compatible with observation. Thus, the only way to close the gap between the reading of the instrument and the measurement outcome is through theoretical considerations, and through that, economic theory.

## 3.3 Defending the Need for Theory in Measurement: The Model-Based View

In this section I set out to show that the only way to get around the problems raised in the previous section, is to apply theory. I present the model-based approach, which at its core is meant to navigate epistemological issues. I note that they share a common thought: the idea that outcomes are underdetermined by observation. I then define outcomes as the set of acceptable values generated by an idealised model of the economics process, following Tal (2016a). This idea is important for what comes next in Chapter 4, where I defend the need for theory in the context of calibration, which is a way to decide the value range of a parameter (i.e. outcome).

### 3.3.1 The Model Based Account of Measurement

Before the epistemological turn in measurement theory, the literature on the topic predominantly engaged in the metaphysics of quantity – in other words, the mathematical theories of measuring scales, as the RTM, as Chang (2004), Tal (2012), and Mari (2005), and others have discussed. The main problem with this is that RTM completely disregards the epistemological aspect of measurement by claiming that measurement theory does little more than examine the relationship between empirical structures and numerical structures. Thus, RTM provides the ideal pathway for the empiricist; it offers a direct path from qualitative observation to numerical representation. However, for the empiricist, the primary issue is that RTM does not provide any method for understanding epistemological issues in measurement. For example, consider a measurement procedure, $\mathcal{P}$:

1. What does $\mathcal{P}$ measure?

2. How accurately does $\mathcal{P}$ measure?

3. Say we have another procedure $\mathcal{P}$', how can we tell whether $\mathcal{P}$ and $\mathcal{P}$' measure the same thing? This problem can be seen as a variation of uniqueness introduced in the previous section.

The model-based approach (MA) on the other hand provide an answer to these epistemological challenges [for more, see Boumans (2004), Boumans (2015), Chang (2004), Mari (2003), Tal (2016a), Tal (2017b), Tal (2017a), and Tal (2019)]. At the very foundation of the MA approach one finds two important distinctions:

1. A distinction between instrument 'readings' and instrument 'outcomes'.

2. A distinction between measurement 'processes' and measurement 'procedures'.

The first distinction is based on the idea that the inference from 'indications' to 'outcome' is non-trivial. The indications are readings of the final state of the instrument being applied. When one applies a thermometer to take a person's temperature, they will have different thermometers to choose between. One thermometer takes the person's temperature from their ear, while another thermometer takes the temperature in the armpit and yet another reads temperature orally. They all diverge in approach and measurement and cannot all be correct. As a result, one should not confuse the reading with the state of the measurand – the object being measured. The instrument's indications do not assume any notion of reliability [see Tal (2017a)]; instead, it solely assumes the intention to measure the measurand reliably with a certain instrument. Furthermore, the indications are not in themselves numerical. Applying numerical representation is simply convenient for a mathematical model. Hence, readings are not claims of knowledge. Instead, instrument outcomes provide knowledge claims by abstracting away from a concrete method. These outcomes thus objectively pertain to the relation in question by attributing the knowledge claim to a measured object instead of the measuring instrument. The outcome often includes a specification of uncertainty, a unit, and a particular scale. The next distinction is between (i) measurement processes and (ii) measurement procedures, in which the latter is a subset of the former. Furthermore, the model-based accounts' principal claim is that any knowledge claim about points 1–3, as outlined above, are based on the latter rather than the former. Here, we take a measurement process to include two aspects [(Tal 2017a), p. 25-29]:

1. **Physical**: Interaction between instruments, samples, human operators and the environment.
2. **Symbolic**: Data processing operations, such as error correction, and data reduction.

Unsurprisingly, a model of measurement processes functions to represent the final indication of a process in the value of the measured quantity. That said, a measurement procedure is a process represented under a specific set of assumptions. As a consequence, one may have multiple procedures within a given process that are all specified by a set of assumptions. In order to say something about the epistemological challenges provided in points 1–3, one must subsume a given process under a set of specific idealised assumptions. It then follows from the model-based approach that any measurement outcome is relative to a model and, ultimately, relative to a background theoretical framework. Without it, (i) no objective claim is possible and (ii) it is impossible to ascribe a certain outcome to that which is under measurement. The idealised assumptions play a crucial role in correcting for systematic error, which the last part of this section demonstrates.

Further, as noted in Boumans (2005)[p. 275] 'In economics, there exist two different and separate traditions of measurement. Ignoring one of these traditions would mean understanding only half of how economics proceeds as a science'. The problem is that regression to the mean, contrary to what Chao (2002) and Chao (2020) argued, does not provide an adequate model of understanding measurement practises in parts of the econometric and macroeconomic literature. The problem with the view that is commonly expressed in representational theory – that measures are an operation that preserves observed relationships among things – is that a problem of 'obtaining empirical significance' arises. It should be noted, however, that this was not the goal of those developing the RTM, as argued in F. Roberts (1979). They were not interested 'in the interaction between the apparatus and the objects being measured.' Instead, the main goal was to provide a foundation for measurement theory. However, the problem with this is that this foundation was mainly developed with logical considerations in mind, rather than trying to satisfy empirical requirements, as Boumans (2005) argued. As Granger noted, 'a theory may be required to be internally consistent, although consistency says little or nothing about relevance' [quote from Boumans (2005), p. 277]. M. Morgan (1988) also argued that the idea in the early history of econometrics was that a satisfactory model is to be defined based on the subject matter. These models were matched both with theory and data in order to provide a bridge between them.

### 3.3.2 What Measurement Outcomes Really Are: The Role of Theory in Determining Outcomes and Macroeconomic Measurement

A useful way to defend Intuition 3, the problem addressed in the introduction to this section, is to examine the history of measurement. Chang (2004) provided a helpful case study in which the author focused on temperature measurements. Other applicable cases include Tal (2011), Chang (2001), and Boumans (2004). Chang's case study showed that measurement of a quantity and the theory behind are closely connected and progress through a process of mutual refinement [see Reiss (2001)]. Reiss summarised several items of importance from an earlier article by Chang [(Reiss 2001), p. 295-298]:

1. Justifying measurement requires knowledge of the mechanisms underlying the data. As Reiss pointed out, the important comparability criterion found in the works of Regnault states that (i) under the same conditions, the measuring instruments must give the same readings if they are functioning correctly, and (ii) if a particular type of measuring instrument is measuring a measurand correctly, all tokens should agree in readings. One can see this as a stability criterion.

2. Background theory is involved in justification in the way that it guarantees stability. Background theory may appear in different forms. In the explanation provided by Pierre-Simon Laplace, an explicit theory was involved in justification: caloric theory. In Regnault's explanation, an assumption about an unobservable item was present. One could see the fact that something exists called 'temperature', which can be measured by a particular measuring instrument as a metaphysical principle.

3. At last, the very justification of measurement outcomes are contextualised both by (i) theoretical formulations and (ii) scientific knowledge, both of which affect the choice of measurement procedure and ensure stability.

I argued earlier in this chapter that stability is the characterising property that makes measurement unique, which is enforced by the points just mentioned. I further noted a number of strategies that might be used to secure stable measurement results. One such strategy is calibration, which is well-discussed in economics literature, see Press (1981), Cooley (1997), Hansen and Heckman (1996), and Sims (1996a). As I discussed in Chapter 2, Boumans argued, that one should view measuring instruments as triplets [See Section 2.1.2]. They include internal principles, bridge principles, and calibration. Calibration refers to a specific method of obtaining parameter values and can be defined in the following way see Boumans (2015),

> **Calibration**. Calibration is a process, that 'fine tune' an instrument to provide a result withinga set of acceptable values.

The crucial thing here is that background knowledge which one's instruments should be calibrated against, including statistical and theoretical assumptions, are as stable as possible. Choosing what to calibrate one's instruments according to is, again, not something data can be used to decide. This highlights the importance of economic theory. Calibration makes use of economic theory in a specific way. It allows a researcher to map a framework onto the measured data, which is something Suppes argued was necessary for the appropriate use of the probabilistic theory of causality [See Section 2.1.1.2.0.2 and (Suppes 1973), section 2]. As noted earlier, the theoretical approach in econometrics argued that measurement without theory is essentially impossible. The very principle of identification is about extracting as much information as possible from the data. Calibration captures this idea, but it also recognises that the relationship goes in both directions. For a given calibration approach, measurements give theory empirical content. In turn, theory allows one to take measurements that are more focused in two ways: (i) how one measures and (ii) what to measure. Additionally, it allows one to recognise when something is wrong. It is this relationship between theory and measurement that ought to characterise econometrics [for more see Cooley (1997)].

Calibration supplement the model-based approach greatly, since it allow us to to estimate a range of acceptable evidence, i.e. measurement outcomes, by providing an idealized model of the measurement process. For according to Tal (2012), to measure is exactly to estimate the value of a parameter in an idealised model, given theoretical and statistical assumptions. Moreover, the goal of calibration is 'selecting the best range of predictors of each possible set of observed indications'; in other words, it is about determining the set of acceptable values [(Tal 2016a), p. 4]. Further, this acceptable range of values must be 'coherent with background theoretical and statistical assumptions, maximally accurate, and "invariant" under changes to the measurement procedure and environment' [(Tal 2016a), p. 4]. It is worth noting that this set of acceptable values must also be refutable, meaning that the values should only predict a small subset of all possible values. Thus, the set of acceptable values must be a small subset in the total space of values. Cooley argues that there are four factors that support best practises when calibrating a measurement [(Cooley 1997)p. 5-7]:

1. **Prior Studies**. Instruments should never be calibrated against prior empirical studies. As Cooley argued [1997, p. 5], 'Every economic environment has some unique features that are motivated by the questions to be addressed. Those features will usually alter the appropriate calibration'.

2. **Theory**. One should make use of a theoretical framework that facilitates addressing questions about how the economy functions.

3. **Measurement**. By calibrating a particular theoretical framework, one restricts the parameter space. However, further restrictions may be necessary to ensure a model is consistent with observations. More theory may also be necessary to explain new observations.

4. **Match**. Aligning a framework with a measurement goes both ways. First, one must set parameter values so they match the model economy (i.e., a theoretical economy). If one observes that a ratio is more or less constant over a long period of time, the parameter values should be adjusted.

However, it should be stressed that measurement outcomes – or the reported results of a measurement procedure – can never be completely stable and neither do we want them to be. So to demand complete stability from economic models would also be overly onerous. Even the evidence thermometers provide is fallible. These instruments will always carry some degree of sensibility simply because the instruments that are used to manufacture them are fallible. Theories may be revised and simplifying assumptions may be refined. Scholars may ultimately refine some of the simplifying assumptions regarding measurement procedures and outcomes. Furthermore, scholars may ultimately revise entire theories. Chang (2004) showed how such revisions affected how we measure temperature, and Tal (2014) exhibited how such revisions affected how we measure time. In natural sciences, the typical procedure to counter lack of stability would usually 'rely only on uncontroversial and well- tested portions of theory and statistics, and to err on the side of caution when evaluating uncertainties' [(Tal 2016a), p. 5]. As Tal [2016a, p. 4] concluded, however, '[t]he answer is that measurement is not a kind of observation if by that one means that the empirical content of measurement outcomes is epistemically or semantically reducible to relations among qualitative observations.' This is because measurement is complex and relies on multiple factors that goes beyond pure observation.

## 3.4 Concluding Remarks

In this chapter, I set out to identify and describe the conditions that are necessary to take measurements. I rejected the view that measurement can be reduced to relations between observables, mainly due to underdetermination and the possibility of systematic error. Since measurement cannot be reduced to relationships between observables, one can reject the idea that measurement without theory is possible, which provide a model to understand the problems of atheoretical econometrics. Instead, I argued in favour of the model-based approach in which theory plays a crucial role in specifying a set of acceptable values and data play a crucial role too. The model based allows one to discriminate between different readings and solve the problem of underdetermination that RTM faces. Moreover, recognising the role theory plays in measurement will enable scholars to correct for systematic error in solving the second problem. The implications for situations in which a scholar can infer causality is the topic of the next chapter.

# 4

# Inferring Causality by the use of Instruments: The Need for Theory

*In Chapter 3, I argued that as part of the shift from readings of instruments to outcomes, background theory is needed. Thus, the outcomes that measuring instruments provide, or what is often viewed as the evidence measuring instruments produce, is theory-laden. In this chapter, I investigate the implications of the theory- laden nature of evidence, including an investigation into what is needed to infer causality from econometric instruments. I first present a case study on whether money causes income to show the problems that follow without an acknowledgement that theory is needed. I note that the literature is inconclusive and point to the sensitivity of Granger tests. I then move on to argue that this means that acceptable empirical evidence in econometrics should be restricted by background theory. I further note that this means that one can conclude little from Granger tests without a sufficiently robust background theory, due to a problem of stability found in Granger tests. Lastly, I present the case for calibration in econometrics along the lines of Cooley (1997) and Kydland and Prescott (1996a), that makes use of both theory and data. I further note the resemblances to evidential pluralism noting (i) that neither probabilistic dependencies nor mechanisms are sufficient to establish causality in econometrics alone.*

> The fact is, economics is not an experimental science and cannot be.
>
> C.A. Sims 'But Economics Is Not an Experimental Science'

Chapter 2 provided an introduction to the construction of models in econometrics and their philosophical foundation. There, I argued that causal economic models in atheoretical time-series econometrics and their instrumentalist approach to econometrics establish that macroeconomic models should be seen as instruments given their instrumentalist methodology. In Chapter 3, I examined the philosophy of measurement. I argued against the idea that measurement can be reduced to relationships between observations. Further, I argued against an empiricist theory of measurement, arguing that idealisation and systematic error pose a problem as the data underdetermines both the idealisations needed to provide standards and the distribution of systematic errors. Instead, I argued that the model-based account of measurement was a better theoretical choice. I posited that measurement outcomes should be seen as a range of acceptable values consistent with some sort of background theory and not necessarily equal to the reading of an instrument. This background theory should provide a standard, that an instrument could be calabrated according to, providing a set of acceptable values, i.e. outcomes. I further argued that this process is exactly what provides the stability we know from the best instruments we have at our disposal, and exactly what atheoretical econometric instruments lack. For in order for readings of an instrument to approach the same measurement outcomes, i.e. become stable, it is vital that the instrument is calibrated according to a specific background theory. In this chapter, I turn to the evidence that these instruments generate (i.e., measurements). I investigate what is needed to infer causality from measuring instruments, given that measurement is not 'value free', but is instead a theory-laden enterprise. I begin in Section 4.1 with a case study on whether money causes income. This study helps to demonstate some of the problems affiliated with the use of Granger tests. I provide background to the case and discusses the literature. In 4.2 I move on to evidence, noting that 'acceptable evidence' created by measuring instruments is restricted by theory. This is not very different from the evidence created by other instruments. Returning to the example of the thermometer, one would not take a 100 degree Celsius reading of a patient to be an acceptable evidentiary claim about the patient's temperature, since it does not fall within the range of acceptable values (that is, human body temperature cannot reach the boiling point of water, and so the measurement must be incorrect). Instead, one would recalibrate the thermometer to be consistent with background theoretical and statistical assumptions. However, apart from that, I note that neither theory nor evidence restrict causality in the real world. This highlights a previously discussed problem with operationalising Granger tests, in which causality becomes associated with the result of a given procedure. I show that this leads to heavy inflation in the meaning of Granger causality, depending on what is included in any given information set or the standard the instrument is calibrated according to. This leads to the conclusion that a Granger test cannot yield any useful information without a sufficient theoretical background. On the other hand, this supposition reduces Granger tests to a type of test that does not assist in the discovery of causality, but instead tests the strength of given theoretical background assumptions. Lastly, in Section 4.3, I present my proposal, following developments in the philosophy of medicine. First, I introduce evidential pluralism, a concept which provides that both difference-making and mechanisms are crucial to inferring causality in the social sciences. I then try to provide a bridge between evidential pluralism and calibration. Lastly, I note that this view entail that evidence claims produced by measuring instruments are 'model relative', and then note how this can assist in solving disagreements, like we saw in the case study in 4.1.

## 4.1 Case Study: Does Money Cause Income? Evaluating evidence

This section introduces a popular case study in econometrics. The question is whether money causes income. This case study helps us understand some of the problems that Granger tests possess. The main problem being that Granger tests do not provide stable readings. I begin the section by introducing the background before looking at the literature on the case. I provide literature from different economic schools to show precisely how sensible Granger causality is to minor differences in the theoretical background.

### 4.1.1 Does Money Cause Income: The Background

A lot of the Granger-Tests in the literature is partly developed to provide an empirical solution to a quarrel between Monetarists and Keynesians. No example is better than the case of whether money causes GNP or the other way around. The paper with the most impact was Sims (1972), in which Sims was able to show, using Granger causality, that money causes GNP, also known in the literature, as the monetarist hypothesis, which states that:

- **The Monetarist Hypothesis**(MH).
  Variations in the money aggregate produces variations in money income. Often refered to as the new quantity theory of money. As Weintraub noted, the hypothesis rests on the apparent causal relation between $\Delta M$ and $\Delta Y$ with an intervening time lag. The mechanism train runs from the money supply to prices and finally ends at income [(Weintraub 1971), p. 38-41].

On the other hand the Keynesian hypothesis states,

- **The Keynesian Hypothesis**(KH).
  the importance of interest rates. The Keynesian mechanism train runs from money to interest rates, then to investment, before it ends at the final stop, an increase in income [(Weintraub 1971), p. 42-44].

There were mainly two kinds of critical responses to the Sims (1972) article, as argued in Okina (n.d.):

1. The first criticism is the content of the information set [I will return to this problem in 4.2.3] Granger actually do mention this problem himself in Granger (1969).

2. There is a lack of stability in the evidence claims produced by the Sims test in Sims (1972). The lack of stability generally lower our confindence in the trustworthiness of an evidence claim, as noted in Staley (2012) [For more see Section 3.1.4.]

Historically the debate on the monetarist hypothesis dates back to the work of Milton Friedman in the 60s. Milton Friedman and Anna Schwartz argued that it was a variation in the money supply that caused changes in income and not the other way around, as the Keynesians had envisioned [See, Okina (n.d.), p. 131-132]. They did so by showing that the turning point of the money supply show a long lead, the turning point for the money stock, a short lead over the turning point of national income. One of the main critiques was provided in Tobin (1970). The two primary points of criticism put forward by Tobin is:

1. Timing is not evidence for the monetarist hypothesis [(Tobin 1970), p. 302].

2. It is possible that the timing in the turn arounds doesn't appear in a Friedman model [(Tobin 1970), p. 303].

Milton Friedman responded with the following remark in Friedman (1970),

> What does "principal" cause mean? If there were an unambiguous way to count "causes," presumably it would mean, "accounts for more of the variance of money income than any other single cause" - which, if the causes were numerous enough, might be consistent with its accounting for only 1 per cent, say, of the variance of money income [(Friedman 1970), p. 319].

Simply put, it is not possible to count causes. Friedman did agree that the timing argument was of limited use, however Friedman also argue in the paper that these timing arguments doesn't meant much for the actual conclusion. They are to a certain point irrelevant. What the debate on the other hand do suggest is that the model used to interpret timing evidence matter.

The literature since the article published has expanded quite rapidly with inconsistent results. Sims (1980a), had to reject the conclusions of Sims (1972), based on the use of variance decomposition in a tri-variate VAR model. As noted in Okina (n.d., p. 138-139) the selection of information set is crucial for the conclusion reached. For the monetarist hypothesis is only supported, if the information set does not include past values of interest rates. Thus whether Granger causality affirm MH rests on whether we include interest rates in the information set. As Okina (n.d., p. 138-139) state, omitting 'interest rates from its system, does not take proper account of the alternative 'Keynesian hypothesis'.'

### 4.1.2 Does Money Cause Income: What Does the Literature Say?

In Chapter 3, I noted that a good way to evaluate the stability of a given evidence claim is to use a robustness test. A robustness test of the money-GNP case is provided in Feige and Pearce (1979). They raised the following questions:

1. How different is the result, if instead of using af Sims test, a Granger direct test, or a Pierce and Hagh test is used instead?
2. To what effect does the pre-filtering in generating a stationary time series alter the result? Sims selects the value k = 0,75.
3. To what extend does the choice of lags alter the result? Sims select $n_1 = 4$ and $n_2 = 8$.
4. The data used by in Sims (1972) was seasonally-adjusted, does that impact the result?

The different Granger tests was introduced in 2.3.3. But to expand on the Pierce and Haugh test, it consist of two steps. First, an Arima model is estimated for each series. Secondly, innovations are used to calculate a sample cross-correlation function. This function is then used to draw inferences about the cross-correlation function of the population [(Okina n.d.), p. 145]. Feige and Pearce (1979) concluded their study with the following words 'we are left with the uncomfortable conclusion that an essentially arbitrary choice left to the discretion of individual researchers can significantly affect the nature of the economic conclusions derived from the test procedures [(Feige and Pearce 1979) 532].'

Okina (n.d.) included data from Japan as well and noted that the results found in Oritani (1979) was in agreement with Sims. Data from Japanese tests show that the choice of $k$ in the prefiltering process, and the choice of lag, strongly affect the apparent Granger connection between money and GNP. Komura (1982) actually finds that the influence of GNP on $M2$ is strong, starting from 1972. The floating exchange system became dominant at the same point. Ram (1984) uses a direct Granger test instead of a Sims test, in order to replicate the findings in Komura (1982). The results in Ram are consistent with Komura (1982). The interesting thing is that the relationship between GNP and money become weak, with the introduction of a floating exchange system. This shows how policy change might affect causal direction in the given model over time.

## 4.2 Evidence of What?

Consider two scenarios. One is an experiment in the natural sciences, while the second is an econometric model for mangos. Say there are four different scientists who are performing physical experiments at the same time. One would expect them to arrive at the same measurements within a certain pre-specified margin of error. Hence, one expect stable readings of the instruments used. In the econometric case, however, four econometricians would probably estimate four different elasticities, $\tau_1$, $\tau_2$, $\tau_3$, and $\tau_4$, based on four different models, $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ and $\mathcal{M}_4$, which were conditional on the different ways these econometricians specified their models. The question, in this case, is the same one Mackeprang examined when he ran two separate regressions and arrived at two different elasticities; which model should one choose? This epistemological problem can be formulated in the following way: how do one know which of these models actually measures the real price of mangos? The literature offers many statistical measures of goodness of fit of a model based on sample data. However, a model may also be over-fitted. Many contemporary models used in policy are actually built by groups who advocate for a certain policy; it is thus plausible that these specifications in these models are selected to portray the group's position in the best light possible, a process known as data mining. The problem here, however, is that all of these specifications are equally supported by the data. In other words, it is not possible to identify the correct one. I argue in this section that the only way to solve this issue is to use theory, for as we saw in chapter 3, applying theory is the only way to obtain stable results. Building on the model-based account of measurement, I argue that theory restricts evidence by closing the gap between readings and outcomes, which are underdetermined by empirical data alone. I further note that Granger models may be an improvement over simple regression models. However, they still suffer from the same underlying issue – spurious correlations – and the only way to verify that such relations are non-causal is to be in the possession of range of acceptable values, provided by calibrating the models against our best theories, consistent with background theory and statistical assumptions.

### 4.2.1 Evidence Generated by Measuring Instruments are always Restricted by Theory

One of the key suppositions from the previous chapter is that theory is needed for measurement, because there is a non-trivial gap between an instrument's readings and outcomes which can only be closed by theory. Further, one of the main characteristics of measurement is its stability, and one way to increase the stability of readings is to use calibration to obtain parameter values. As Chapter 2 demonstrated, the main approach dominating econometrics is to take observed data as a given and try to infer the economic structure from it [the DGP]. In this approach, which I referred to as 'atheoretical econometrics', the starting point is the data. From there, one searches for the world that most likely generated that data. On the contrary, calibration, which is a vital component of the model-based view, we treat data and measurements not as given, but as something which must be determined at least in part by theory. This means that one may choose the values of some parameters based on what has been observed, but the way one determines other parameters may be heavily based on theory. Calibration has been criticised by economists in the atheoretical camp (see Sims (1996b)) and other economists (see Hansen and Heckman (1996)) as a substitute to estimation. However, these two should instead be seen as complementary.

When calibrating a Celsius thermometer, one relies on two fixed points: (i) the thermometer reads 0 degrees in ice water and (ii) reads 100 degrees in boiling water. This is based on the theory that the mercury in the thermometer expands linearly within this range of temperatures [for more, see Chang (2004)]. The theory also lays out how to recalibrate the measurement depending on where the thermometer is used; the instrument must be calibrated differently if one is in the mile-high city of Denver or if one is in Quito, both well above sea level. Model economies play that role in economics. They provide fixed points, for some economic questions do have known answers, provided predominantly by economic theories and data. In economics, calibration provides a way to map a framework onto measured data. A framework should provide a set of fixed points that the model could be tested against in order to verify the veracity of the model's answers (when said answers are unknown). In this way, calibration provides a range of idealised and hypothetical values, which one may take to be acceptable given the framework and which one then uses to test a theory against. This is something that even Suppes took to be necessary in order to apply his probabilistic theory of causality. Thus, calibration provides an idealised model of some economic process, fitted to known fixed points, and therefore provide direction as to what to look for and how to do it [For more see Cooley (1997) and Kydland and Prescott (1996b)]. Thus, (i) the values provided by a framework produce a range of acceptable values, as Tal argued, that arise given what one already knows, and (ii) background theory is used to calibrate models such that they mimic the world, conditional on a specific framework [see Tal (2016a), Tal (2017b), and Tal (2019)].

Consider a situation where you wish to take a person's temperature. You apply a thermometer to a person. When you look at the thermometer, the thermometer shows that the temperature of the person is 100 degrees Celsius. Unless the person is actually boiling, you know that the person in question cannot actually be exhibiting a temperature of 100 degrees Celsius, based on the theory that mercury expands linearly and your understanding of human anatomy. Immediately, you would recalibrate the instrument in question, suspecting the reading might be due to a systematic error. Why? Because the reading of the thermometer fell outside of the set of acceptable values for a thermometer in the given situation. In this example, theory and statistical assumptions reveal something about what one would expect the measurement outcome to be. As a consequence, measurement outcome depends on our theoretical commitments. Hence, what counts as acceptable evidence of causality depends on the results one deems to be within the range of acceptable values prior to taking the measurement, making the measurement model-dependent. It is crucial, however, that the set of acceptable values of a measurement is refutable. If not, empirical results would not play a role, and the set of acceptable values would remain static (which is rarely the case). However, the main problem is not that different procedures exhibit different results, but that the set of acceptable values in economic endeavours is not invariable from procedure to procedure. Whether one uses an oral thermometer or an ear thermometer does not change that a patient is considered to be exhibiting a fever if the thermometer reads above 38 degrees Celsius (the outcome being 38+- some random error term). As this thesis discusses later, in the case study, the set of acceptable values differ from one economic school to another, making it almost impossible to use instruments to make economic judgements without theory and without acknowledging that the results obtained will be relative to the background theory in question.

In general, science depends heavily on measurement, and econometrics is no different. It is inappropriate to view econometrics and measurement as independent; they must be understood as mutually dependent enterprises [as argued in Chapter 3]. Hence, the reliability of econometrics depends on our beliefs in the reliability of our instruments to produce correct readings. However, as already noted, when the set of acceptable values does not remain invariant from procedure to procedure, using instruments to obtain reliable readings is essentially futile since these readings are relative to background theory that different schools might disagree about. That is precisely the reason why scholars rarely question the readings thermometers offer. One is not in possession of different background theory before applying a thermometer or different sets of acceptable values when a patient exhibits fever, even when the thermometer is still conditional on plenty of assumptions. Nonetheless, instruments are still the method of choice in contemporary econometrics, as

Chapter 2 demonstrated. It is vital at this point to remember to distinguish between methods and methodologies. Methods are the actual procedures used, while methodology is an overarching view of how science should proceed. In this case, the instruments are the different methods, and 'instrumentalism' is the overarching view of how science should proceed. This methodology represents the deductivist view and methods, being different structural equation models in the Cowles Commission.

Therefore, what is interesting here is not the philosophical implication of the differences in methodology but the difference in the set of acceptable values. The set of acceptable values is determined chiefly at the theoretical level of inquiry, not the empirical level. As a consequence, the problem of instrumentalism in econometrics is based on differences in prior theory. Consequently, one cannot just solve the problem by ignoring it; that would simply be begging the question, which is precisely what occurs in the atheoretical approach. One should not treat the production of evidence in econometrics as a neutral matter. Measurement, and the theory that supports it, restricts what we classify as evidence, and choosing to apply statistical instruments does not solve this problem. The evidence claims produced by statistical instruments are ontologically loaded. They come with commitments. Consequently, statistical tests of causality cannot serve as a decidable algorithm into which one can plug different problems and receive an answer. The answer such an algorithm provides is directly conditional on the underlying framework.

### 4.2.2 Theory and Evidence Do not Restrict Causality

Another thing we should keep in mind is that even though theory restrict measuring instruments, it does not mean that evidence itself restricts causality. In pursuing the aim of causal knowledge the atheoretical view takes data as its starting point and then asks: What can we learn about the world? It is tempting to accept the atheoretical viewpoint and by it accepting that we can move directly from probabilistic dependencies to causality. Following Dyke (2012), I take this to be a fallacy, in so far, as we move directly from readings of our instruments, to conclusions about the fundamental nature of reality. We should in general be careful about mistaking causality with evidence of causality. Correlation might indicate that there is a connection between the variables in question; Granger causality might indicate the direction of such a connection; some underlying mechanism may help explain a possible connection between two variables, but none of them are in itself enough to infer causality. I will return to this point in 4.3.

Inferring the outcome by merely mapping the structure of readings can be seen as a kind of operationalism. Since causality is typically identified with a particular causality test (here, the test for Granger causality), this is the operation one must perform to verify that causality is present [see the case involving time and clocks in Section 3.2.3.1], just like one might identify temperature with the reading on a thermometer. However, as I noted in Chapter 3, this is the wrong approach. Although one can choose from different types of thermometers, and these different types provide different readings, not all of these can be true (i.e., correct) at the same time. The same holds for Granger causality. One may choose different tests, and these tests might provide different readings, not all of which can be correct. Instead, there is an ideal underlying structure that the instrument is said to approximate. Hence, based on the model-based view and given this understanding, the instruments can never provide more than a reasonable approximation of the state of the measurement. Consequently, as I argued in Chapter 3, one should not mistake a reading for the state of a measurand because the reading of an instrument does not assume any notion of reliability and is merely an approximation of some ideal standard. Only measurement outcome does, and one must make a non-trivial inference to bridge the gap between a reading and an outcome. Despite rejecting the operationalist view of causation, Chapter

3 is built on the idea that such an operationalisation of causality is fairly close to similar claims in other fields of science, such as temperature and thermometers or time and clocks.

The view was famously posited by the physicist Percy Williams Bridgman in his book, The Logic of Physics (1927); for an insightful discussion on the depth of Bridgman's work, see Tal (2012). As Chang (2019) noted, Bridgman was influenced by the view that scientists should apply multiple instruments to measure the same phenomenon. Say one wishes to measure distance. If one measures the distance between planets, he or she could measure that in light-years. When measuring a distance in a house, one could instead use a laser measuring tool. According to Bridgman, these two procedures are based on very different concepts of distance. He noted that '[t]o say that a certain star is 105 light-years distant is actually and conceptually an entire different kind of thing from saying that a certain goal post is 100 meters distant' [Bridgman (1927) 1927: 17, 18]. Based on this idea, Bridgman further argued that instead of saying that both the person measuring the distance in space, and the person measuring the length in a house measure the same concept, length, 'the operations by which length is measured should (in principle) be uniquely specified. If we have more than one set of operations, we have more than one concept, and strictly there should be a separate name to correspond to each different set of operations' [Bridgman (1927) 1927: 10 ]. The worry Bridgman had was that if a measurer fails to be specific, he or she would 'get into the sloppy habit of using one word for all sorts of different situations (without even checking for the required convergence in the overlapping domains)' [Bridgman (1927) 1927: 75]. But as argued in Tal (2012)[p. 75], 'If taken literally, conceptual foundationalism [RTM] about measurement scales leads to the same absurd multiplication of quantities already encountered above [Bridgman (1927)].' This problem Tal [p. 75] argues is not only present in Bridgman operationalist view, but provides a problem for RTM too, since systematic errors cannot be transformed away. Thus, leaving us with one conclusion, that different measurement instruments, say thermometers, measure temperature on different scales, but this 'is inconsistent with RTM, according to which both scales are interval scales and hence belong to the same type'. Thus, if one accepts the empiricist view of measurement, in which measurement is reducible to relations between observables, one is left with the idea that different thermometers measure different quantities. Thus, a strict empiricist interpretation of measurement, yields fragmentation of the concepts involved and thus can provide very little insight into the working of measurement in science and econometrics.

By applying the same reasoning to the case-study introduced in the beginning of this section, we would have that contradicting evidence claims provided by two different Granger tests would yield that those two tests was not measuring the same thing, which is hardly useful. As we will see later, the more plausible explanation is that different specifications of a model, different methods, different calibration et cetera leads to difference in results. All things that RTM do not consider. Instead, we should stop seeing as an epistemological theory of measurement. Whether we should even see RTM as a theory of measurement is doubtful. RTM provides the mathematical presuppositions underlying different measurement scales. But, if that is the only goal of RTM, then

> RTM can no longer be considered a theory of measurement proper, for measurement is a knowledge- producing activity, and RTM does not elucidate the structure of inferences involved in making knowledge claims on the basis of measurement operations. In other words, RTM explicates the presuppositions involved in choosing a measurement scale but not the empirical criteria for the adequacy of these presuppositions. RTM's role with respect to measurement theory is therefore akin to that of axiomatic probability theory with respect to quantum mechanics: both accounts supply rigorous analyses of indispen-

> sible concepts (scale, probability) but not the conditions of their empirical application [(Tal 2012), p. 77-78].

In other words, there is no reason to believe that simply mapping the readings of an instrument directly to causality, capture causality, or that different Granger test measure different kinds of Granger causality. A more plausible explanation is that different types of tests produce different evidential claims, simply because the models are calibrated differently. Plural instruments often pick out different things, as we will see in the next section, but it does not entail, that they measure different things. I will return to the problem of disagreement among instruments in the last section of this chapter. However, I will argue that what we find using different tests should be seen as symptomatic, and not in itself constitutive of causality, arguing that we should use multiple sources of evidence.

### 4.2.3 Evidence *Without* Theory: Operationalizing Granger and the Information Set

As I argued in the beginning of this section, I noted that theory restricts the evidence generated by instruments. In this section, we will see how this works in practise. Usually when we operationalise Granger causality, we substitute the information set $\Omega_t$ containing all relevant information with a subset of $\Omega_t$, $B_t$. This is necessary since $\Omega_t$ is larger than what is possible for a human to grasp and even a computer. Therefore, if Granger causality is to need of a set similar to $\Omega_t$, then the definition is not operational. But whatever we choose to include in $B_t$ might affect whether there is a causal connection between two variables. We clearly saw this in the case-study. When adding interest rates the causal connection between money and income disappeared. This raises two problems for the practising economist:

1. Which data should be included in $B_t$?

2. When is the data relevant for the phenomena under investigation?

Both questions are clearly underdetermined by data. Thus, to operationalise Granger causality is exactly to make it relative to a certain model, and we will see in this section how big a role it plays. For different specifications provide different causal claims, that cannot be correct at the same time. Thus, emphasizing the need to distinguish between the reading of the instrument and the outcome of measurement.

One of the most common objections raised in the realm of causality is that the events or variables being dealt with, perhaps $B_t$ and $A_{t+1}$, are actually caused by a third variable, $C_{t-1}$ known as the common cause fallacy. The problem here is that we end up believing that $B_t$ causes $A_{t+1}$ only because we did not take $C_{t-1}$ into account. The problem can be systematised in the following way:

The original definition of Granger causality (See Definition 2.2.5) accounted for such problems by including the large and all-encompassing information set $\Omega_t$, wince all possible common causes are included in $\Omega_t$. This is not the case for $B_t$, meaning that it is possible to choose the wrong $B_t$. Consider the following example:

**Example 2.** Imagine a world in which it has been observed that cows always eats grass just before an earthquake. These are the only two events included in $B_t$. In this case, we may postulate that cows eating grass causes earthquakes. Considerable specialised knowledge is needed to expand the information set $B_t$ in such a way that it invalidates that causal postulate.

In econometrics, the events are included in the information set $B_t$ of the following form [(Milhoj 1985), p. 9]:

$$A_{t+1} = \{X_{t+1} \in A\}. \tag{4.1}$$

$$B_t = \{Y_t = y_t\}. \tag{4.2}$$

in which $X_t$ and $Y_t$ are stochastic variables that denote economic quantities, $y_t$ is an observed value of $Y_t$, and the set $A$ is a collection of possible outcomes of $X_{t+1}$. Thus, in order to operationalise Granger causality, I extend the definition in Chapter 2 (Definition 2.2.5) to the following:

**Definition 4.2.1.** Granger Causality*

1. $X_t$ causes $Y_{t+1}$ iff $\mathrm{P}(Y_{t+1} \in A|B_t) \neq P(Y_{t+1} \in A|B_t - X_t)$ for some A.

This move to weaken the definition of Granger causality by substituting $\Omega_t$ for a weaker set of information,set $B$, is the same move Vercelli (2017a) made in his presentation of $PMC*$ (Definition 2.2.6), which allowed us to utilise Suppe's idea that the determination of causal relationships is relative to a conceptual framework. Another formulations in terms of densities is possible [(Milhoj 1985), p. 10]:

$$\mathfrak{F}(y_{t+1}|B_t) \neq \mathfrak{F}(x_{t+1}|B_t - X_t = x_t) for some A. \tag{4.3}$$

However, a definition of the fact that $Y$ does not cause $X$ does not only demand that $Y_t$ does not cause $X_{t+1}$, but that all past values of $Y$, that $Y_{t-1}, ....$ does not cause $X_{t+1}$. Therefore, we obtain that [(Milhoj 1985), p. 10]:

$$\mathfrak{F}(y_{t+1}|B_t) \neq \mathfrak{F}(x_{t+1}|B_t - X_t = x_t, X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, ...) \text{ for some A.} \tag{4.4}$$

These conditions are about the entire conditional distribution. However, in practice, our interest is typically limited to the prediction provided by $Y_{t+1}$. These predictions are usually the conditional mean value, which give us the following definition [(Milhoj 1985), p. 10]:

**Definition 4.2.2.** Granger Causality**

1. $X_t$=$x_t$ causes $Y_{t+1}$ in the mean iff $E(Y_{t+1}|B_t) \neq E(Y_{t+1}|\{X_t = x_t)$ for some A

Some examples from Milhoj (1985), p. 12,

**Example 3.** Let,

- $X_t = \tau_t$, $Y_t = \tau_t + \varphi_t$, and $Z_t = \varphi_t$

This case is interesting, because the causal connection here relies on which information set we choose. It is immediately visible that there is no connection between $X_t$ and $Z_{t+1}$, since $\tau_t$ and $\varphi_{t-1}$ are independent, if no further information is added to the information set. However, if it is known that $Y_t = y_t$, then we have do have that $X_t$ causes $Z_{t+1}$, since,

$$Z_{t+1} = \varphi_t = y_t - X_t \tag{4.5}$$

Thus,

$$E(Z_{t+1}|Y_t = y_t) = y_t \tag{4.6}$$

And,

$$E(Z_{t+1}|Y_t = y_t, X_t = x_t) = y_t - x_t \tag{4.7}$$

**Example 4.** Let,

- $X_t = \tau_t + \varrho_t$, $Y_t = \tau_t$, and $Z_t = \tau_{t-1} + \varphi$

This case is again interesting, because the causal connection again relies on which information set $B$ we choose. It is immediately visible that $X_t$ causes $Z_{t+1}$, if we do not add any further information to the information set. Since,

$$E(Z_{t+1}) = 0 \tag{4.8}$$

And,

$$E(Z_{t+1}|X_t = x_t) = x_t. \tag{4.9}$$

However, if we learn that $Y_t = y_t$ this no longer holds, since

$$E(Z_{t+1}|Y_t = Y_t) = y_t. \tag{4.10}$$

$$E(Z_{t+1}|Y_t = Y_t \wedge X_t = x_t) = y_t. \tag{4.11}$$

Both examples help to show how important it is that the the economist in question is in possession of the right information. Otherwise, the only possible consequence is a misspecified model. Furthermore, it shows how sensible causal evidence claims are to the content of the information set.

### 4.2.4 Evidence *Without* Theory: What Can be Concluded From Granger Models?

Say there is evidence of a Granger causal relation between two time series, $X_t$ and $Y_t$. The obvious question arises: what can be concluded from this causal relation? Can one conclude that $X$ causes $Y$? Based on what I have discussed thus far in this section, the answer is no. That would be to mistake the evidence for the state of the measurand. Moreover, the Granger connection in question may be accidental. The possibility of accidental Granger connections is supported by an idea discussed in Chapter 2 [Section 2.3.4]: that strong exogeneity is the concept of causality in econometrics, and strong exogeneity holds that Granger causality is necessary, but not sufficient, to establish causality. This is simply because Granger tests are incapable of determining whether a variable is weakly exogenous. Since weak exogeneity is a requirement for strong exogeneity, the conditions are not met. But suppose now that

$$X_t \text{ Granger causes } Y_{t+1}. \tag{4.12}$$

Then we have the following possibilities:

1. $X$ is a cause of $Y$;
2. $X$ may not be a cause of $Y$, but they may share a common cause;
3. $X$ is a cause of $Y$ and $Y$ causes $X$; and
4. $X$ does not cause $Y$.

The improvement here over a pure correlation test is that the possibility of a causal connection between Y and X can be excluded, due to Axiom T [Section 2.3.2]. Thus, the Granger approach to causality in fact captures the causal direction missing in a simple correlation. However, based on the literature on Granger causality, all four are possibilities when the instrument reading suggests that the null should be rejected. Thus, this yields the problem that an instrument reading may yield four different interpretations, possibly creating inconsistent results without any ability to discriminate between them since the operation verifies Granger causality. The connection can, in theory, be spurious (Scenario 4) due to non-linearity, as argued in D. L. Roberts and Nord (1985), time-series cointegration, as noted in H.-Y. Lee, K. Lin, et al. (2002), or common cause fallacy, see Sargent and Sims (1977). Further, it could mean that there is instant Granger causality between $X_t$ and $Y_t$ and $Y_t$ and $X_t$. Yet another problem occurs if the model is misspecified; in that case, error will render the test uninterpretable since it 'will imply that a point null hypothesis will certainly be rejected if the sample size is large enough' point null hypothesis will certainly be rejected if the sample size is large enough. [(Jacobs et al. 1979), p. 401]. The problem here is that whether a model is misspecified relies heavily on background beliefs, such as the choice of economic theory. There is no 'point from nowhere' where one can decide whether a given model is misspecified. Even if one obtains a negative answer to a Sims test or a Direct Granger test, this may indicate that there is a non-linear Granger causality between $X_t$ and $Y_t$ as Dufour and Taamouti (2010) noted. Even instant causality is possible if time series are non-stationary, as H.-Y. Lee, K. S. Lin, et al. (2002) argued. Yet another and more pressing problem is that Granger causality inherited the problem of sporiusity from simple correlation tests. Chowdhury (1987) managed to show that GNP causes sunspots with a direct Granger test and that prices caused sunspots with a Sims test. The study did not validate any other hypotheses. Thus, one might come to the conclusion that:

> Granger causality does not meet the requirements of an investigator who uses this method due to epistemic reasons (i.e. in order to discover what variable is the cause and what is the effect) and does not possess prior knowledge on the phenomenon considered [(Maziarz 2015), p. 98].

The main idea here is that these statistical properties still leave work left to be done, which supports the idea that there is no direct mapping from readings to outcomes. In order to apply causality tests like the Granger tests, it is vital to have some theoretical background, and with it some knowledge of the underlying mechanisms.

## 4.3 A Modest Proposal

At the heart of this proposal is the idea that realism and reductionism are directions, not positions. To infer causality in econometrics, both mechanisms and evidence of difference-making are crucial. Both extremes contain problems; consider the idea that causes can be inferred from mechanisms alone. This hardly comports with how causality in economics is often viewed. For the causes, econometricians are primarily interested in are the one that exhibit some element of probabilistic dependence. Difference-making can provide evidence for such, by showing that the historical values of one value can increase the predictability of another. There is clearly some epistemology in this as well since it is precisely the element of difference-making that instruments seek in econometrics. Additionally, causes that exhibit an aspect of probabilistic dependence are causes that can be exploited. It should be noted here, however, that the search for such probabilistic dependencies by evidence of difference-making is not inconsistent with non-reductionist positions. The idea is simply that one cannot wholly separate causes from probabilistic dependencies. However, the other extreme is just as indefensible. Causal relations cannot be reduced to probabilistic dependencies entirely as some practitioners in econometrics suggest. When econometricians speak of causality in economics, they are referring to something more than simple 'constant conjunction'. Furthermore, structures or mechanisms (as opposed to empirical regularities) appear to be the foundation of explanatory success in the sciences, something that can scarcely be explained purely by probabilistic dependencies. Thus, the old Kantian dictum, that 'concepts without percepts are empty; percepts without concepts are blind' can be rewritten here: 'mechanisms without probabilistic dependencies are empty; probabilistic dependencies without mechanisms are blind'. This means that mechanisms without any empirical content are too strong a claim, but difference-making without any theoretical knowledge appear to be too weak. Thus, I present a pluralist case in this section, arguing that both are required to infer causality. In the model-based approach, a theory provides a range of acceptable values for an instrument, and the instrument provides readings consistent with the relationship between outcomes and readings. I begin the section by introducing evidential pluralism. After which, I discuss the weight that should be assigned to each component (i.e., evidence of difference-making and evidence of mechanisms). In the second part of the section, I note one outcome of such an approach: evidential claims become model-dependent. Another outcome is that this approach allows one to better explain how disagreement arises. Furthermore, I revisit the case study introduced in Section 4.1.

### 4.3.1 Rejecting Both Approaches: The Case for Pluralism

I provide the case for evidential pluralism here, mainly based on the works of Federica Russo and Jon Williamson in the philosophy of medicine, and I present the Russo-Williamson thesis that simply put argues that neither evidence of difference-making nor evidence of mechanisms are sufficient to infer causality. What is not always entirely clear is just how much weight should be ascribed to either. I argue here that it depends on whether an agent does have knowledge of an underlying mechanism. If so, following the model-based approach to measurement, such knowledge should be assigned more weight than evidence of difference-making, and that such knowledge is always essential to infer causality.

#### 4.3.1.1 Embracing Pluralism in Evidence

Both 'difference-making', as we saw in the probabilistic account of causality in chapter 2 [see Section 2.1.1.2.0.2], in which the cause increases the probability of the effect, as well as, 'mechanistic' evidence, as we saw in the Cowles approach, is important to establish causality in econometrics. A new line of thought in social sciences argue exactly this. Evidential pluralism, see Claveau (2012) and Moneta and Russo (2014b). They support a particular kind of pluralism, which take causal claims in the social sciences to be the correlations that can be mechanically explained. In the philosophy of medicine, this view is often referred to as the 'Russo-Williamson Thesis (RWT)', which states that,

> To establish causal claims, scientists need the mutual support of mechanisms and dependencies. The idea is that probabilistic evidence needs to be accounted for by an underlying mechanism before the causal claim can be established [(Russo and Williamson 2007), p. 159].

Understanding the two components in the following way [(Russo and Williamson 2007), p. 159],

1. **Probabilistic Evidence**. Probabilistic evidence is important in econometrics simply because causal claims are used for prediction. Econometricians want to know how a certain policy, say raising the minimum wage, will work in the real world before choosing to introduce the policy. Further, difference-making is crucial since

what econometricians are interested in are exactly those things that exhibit some aspect of regularity.

2. **Mechanism**. The mechanisms, on the other hand, are crucial in two respects: (i) they explain the probabilistic dependencies that show up in the data, and (ii) they yield evidence of stability (which I argued was a crucial aspect of causality in Chapter 2).

Claveau (2012) is sceptical to the idea that evidential pluralism is correct. Based on unemployment case studies from the OECD, see OECD (1994), Claveau argues that different sources of evidence support different causal conclusions, arguing that both types of evidence are not needed to support every causal conclusion out there. On the other hand, Moneta and Russo (2014b) claim that both are needed to support causal conclusions, in other words, RWT holds in economics as well. I go in a slightly different direction here. I want to bridge the essence of the evidential pluralist view, that different kinds of evidence is important with the calibration view in the philosophy of measurement. For as I mentioned earlier, the calibration view in economics is also based on the idea that both data and theory is important.

I agree with the evidential pluralists that the main problem in both the theoretical and the atheoretical approach is exactly that they do not see the value in both *difference-making* and *mechanisms* to establish causal claims. As we saw in Chapter 2 [see Section 2.1.1.1,2.1.1.0.1 and 2.2], the theoretical approach causal relations are postulates based on pure theorizing, thus the probabilities do not enter the stage before at the epistemological level, but only to measure the strength of the relation in therefore play no role in establishing it. On the other extreme, you find the atheoretical instrumentalist approach, that apply causality tests [see Section 2.3.3], and do not make use of mechanisms to establish causality. Mainly due to the subjective nature of theorizing making uninterpreted statistical models the only way to obtain credible results in econometrics, [see especially Sims (1987), p. 53.] Thus, the atheoretical approach mistakes readings for outcomes, and with it comes problems of sensitivity and spuriousity as we saw in the case study earlier. Apart from just agreeing on the problem at hand, the calibration view and the evidential pluralist share two things in its solution too,

1. Both evidence of difference-making, as well as theoretical knowledge, in the form of mechanisms should play a role in inferring causality.

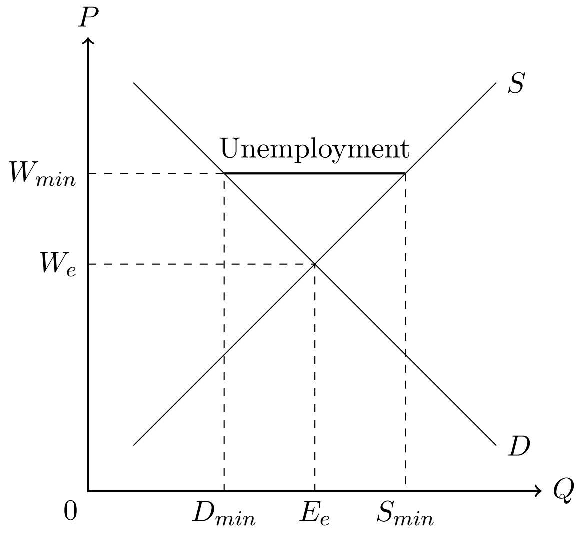
2. Evidence of difference-making, is restricted by some framework, in the form of mechanisms, and the framework should in part be informed by evidence of difference-making.

For as we have seen plausible mechanisms may indicate that a particular Granger relation between two time series only arises in a very specific population, or that it might be due to spuriousity [See Section 4.2.4 for an example]. Known mechanisms can exactly play the role of fixed points in a model, which allow us to calibrate our instruments accordingly. For calibration is exactly a comparison of a reading under some test, and a standard of known accuracy, in this case the mechanism. Thus, mechanisms should play the same role in causal models, as the freezing point and the boiling point play for a thermometer. For known mechanisms can help in two ways, (i) a mechanism can help us correct for systematic errors, for if we know which variables should be included, we can avoid the omission of variables that are relevant in a given case and (ii) mechanisms often comes with theoretical commitments that may help us in choosing assumptions, as well as, providing us with a set of acceptable outcomes. Thus, the goal of calibration here would be to mimic the most stable mechanisms we know to take on questions for which we do not have any answers. By mimicking stable mechanisms and further theoretical constraints, we can improve the stability of evidence produced by certain instruments. However, it should be noted that such calibrated models are of course conditional on a specific framework, that we by no means can know whether is true or not. But as noted in Chang (2004), who coined the term 'epistemic iteration', we adopt a framework, without any assurance, that it is correct. We then calibrate our instrument to this framework to create stable readings, after which, we start doing multiple inquiries and even correcting the system, that could be by changing assumptions or ontological commitments. What Chang (2004) shows is that this provide a self-correcting process, which secure self-correcting progress in science. But a proccess that needs both theory and data.

#### 4.3.1.2 Types of Evidence and the Weight of Each: How Calibration Works

I noted in the previous section that there are two kinds of evidence that are important to establishing causal connections. Evidence of difference making can help provide mechanisms with empirical content by measuring the strength of a relation (as laid out in the theoretical view). Moreover, they can help support policy by offering qualitative connections in quantitative terms. Econometricians not only want to know that $X$ causes $Y$, but if $X$ happens, they also want to know what impact it has on $Y$. Say a group of politicians wishes to enact a policy, P. Assume that P hikes the minimum wage from £10–15. Classical economics provides that there is a mechanism, $M$, that indicates that if such policy is enacted, it may harm employment. However, not only is it important to know that it might harm employment, but also how much it might harm employment. In this section, I intend to show how we all calibrate beliefs and how it impacts our understanding of evidence. This provide a way to understand some of the topics that we have discussed so far in this chapter including, the impact of model dependence, disagreement and how we weight the evidence produced. But it also show us how we may tackle these problems, as we will see in the upcoming section.

The question here is, how much weight should be assigned to each kind of evidence? Say we share the monetarist view, as in the case-study, and decide to calibrate our instruments accordingly. When should we in light of data revise our beliefs? I argue that it is clear that it is the theoretical background – in this case, the mechanism – that decides the weight assigned to the evidence generated by some atheoretical instrument, since it is the theoretical framework, through calibration, that generates the range of acceptable values. However, the range of acceptable values will vary among different agents. A set of acceptable values will be larger in cases where the credence assigned to a given mechanism is lower. Similarly, if the credence assigned to a mechanism is higher, the set of acceptable values will be smaller. Consider the minimum wage case once more. Classical economics establishes that if the minimum wage exceeds the equilibrium in a given market, unemployment will rise. Consider the following figure:




Based on this figure, and the mechanism derived from this figure, a hike in the minimum wage will lead to increased unemployment. If a particular economic agent's degree of belief in this economic mechanism is high, that economic agent will probably not believe studies that show that a minimum wage hike did not lead to an increase in unemployment. This is because such studies would be inconsistent with that agent's theoretical background, and as a consequence, outside of the range of his or her acceptable values. However, the lower the degree of an agent's belief in a particular mechanism, the more receptive that agent is towards findings that are inconsistent with the mechanism in question. The problem here is that there are multiple empirical studies on both sides. Some indicate that minimum wage hikes do not cause increases in unemployment, while some indicate that they do. The degree of belief is what leads one economic agent to believe the conclusions of one or the other body of literature on the topic. Empirical evidence does not; whether one accepts something as evidence is directly affected by one's theoretical background. Thus, a single agent do calibrate his beliefs to suit background theory.

### 4.3.2 What This Means for Econometrics: Bridging Atheoretical and Theoretical Econometrics

In the first part of this chapter, I argued that the main problem of the case study was disagreement and how to choose between competing evidence claims. I have argued the calibration view may provide an explanation to why disagreement arises and showed this view has similiarties with another view evidential pluralism. In this section, I discuss two consequences of this.

#### 4.3.2.1 Model Dependence: Revisiting the Model-Based View

It follows from the discussion above that the calibration view in measurement entails that causal claims become valid relative to a given framework - not very different from the view defended by Patrick Suppes [see Section 2.1.1.2.0.2.]. This does not entail that the phenomena themselves are relative. It shares these two properties with the evidential pluralist view [see Moneta and Russo (2014b)]. Instead, calibrated models reinforce the view that every part of the measurement enterprise is important and that measurements are not a given, as noted in Tal (2016a) and Mari (2005). Measurement relies on assumptions, both statistical and theoretical, as well as causal knowledge. Due to the importance of calibration in establishing measurement outcomes, evidence produced by instruments is valid relative to the framework it is calibrated according to. It is precisely the 'fine tuning' of parameters that establishes the stability of evidential claims produced by a given instrument.

At the beginning of this section, I introduced a case study. I asked the question, whether money causes income. According to some economists, this provides an economic regularity, well-substantiated by empirical analysis, especially argued in Friedman and Schwartz (1965) and Friedman and Schwartz (2008). Further, as we saw in Section 4.1.2, some Granger tests seem to support MH as well. Thus, MH builds on numerous investigations and is supported by an underlying mechanism, as well as applied statistical modelling. However, in Section 4.1.2. it is seen that whether the evidence supports MH depends on what is included in the information set. Thus, the way we specify our model does matter. Furthermore, how we calibrate our instrument has an impact.

Therefore, we should be careful with the way we compare results produced by instruments calibrated according to different standards, i.e. different frameworks. How can we compare such results then? The answer here is, we should not do a 1-1 comparison. Cooley touched the subject of 'restricting the mapping,' in this case, between instrument readings and measurement outcomes in Cooley (1997). According to the background theory, restricting the mapping between those allows us to generate the outcomes that display the desired properties. However, the instrument cannot investigate the features of the economy that we use to calibrate it. The standards of such a framework is given; for 'the output of a calibration exercise is the answer to the question the model economy was designed to answer'. In this case, whether money causes income.

Boumans (2015) further argued what macroeconomists do is to calibrate their instruments, or 'fine tuning' parameters, to isolate variables from external influences. By doing this, the economist does not have to control the entire system. This maneuver aims to produce invariant relations between readings and outcomes that mimic those we know from stable instruments in the natural sciences. The main goal here is the same as in the evidential pluralist view; to produce stable predictions by multiple sources of evidence. Here, the focus is more narrow to produce reliable measurements.

#### 4.3.2.2 Disagreement: Returning to Whether Money Causes Income

The model-dependence that follows from the calibration view assists us by providing a useful way to understand the disagreement between different instruments observed in the case study. It provides a better way to solve the problem of disagreement between different instruments. Disagreement in the case provided in 4.1.1 can be reduced to three factors,

1. **Model Specification**. The importance of model specification is argued in Feige and Pearce (1979) and Geweke et al. (1983). Apart from that, I went on to show the importance of the information set in 4.2.3.

2. **Calibration**. Choosing parameters and frameworks (that work as a standard).

3. **Method Choice**. Choosing a Granger model.

As noted in Moneta and Russo (2014b),

> Friedman and Schwartz (1982) estimated to establish evidence concerning the money demand and the influence of money on income and prices in UK, they used phase-averaged data, considered the parameters in the money demand equation as constant, treated money as exogenous and, more in general, ignored

> the mutual interdependence of money, income, prices, and interest rates [p. 70-71].

The last point was stressed in Okina (n.d.). Sims used seasonally-adjusted data in Sims (1972) to show that money granger causes income. It should be noted here that to be skeptical toward the result of a given empirical test is to discuss the different phases, and especially the calibration of the instrument. This necessarily leads to a discussion of the background theory. Once again, imagine that a thermometer exhibits a 'too high' reading. The apparent move in such a case would be to question and discuss the different phases of the instrument. In this case, one could argue that the model was misspecified, claiming that interest rates should be included, or that the choice in Friedman and Schwartz (2008) to hold the parameters in the money demand equation constant was a mistake. One might further question the functional form of Granger instruments. D. L. Roberts and Nord (1985) argued that the functional form affects the discoveries of Granger instruments. The casual relation observed in non-linearised data disappears in the transformation. Thus, the findings are not insensitive to logarithm transformation. Glasure and A.-R. Lee (1998) argued that Granger models are sensitive to the specific procedure used to discover it, and H.-Y. Lee, K. Lin, et al. (2002) showed that Granger tests are sensitive to cointegrated time series. In such cases, Granger tests are biased to reject the null more often. Dunne and R. P. Smith (2010) argued that Granger causality tests could be susceptible to the number of variables included in the Granger test. However, not only to the number of variables but also sensitive to the number of lags and the choice of parameters. Dunne and R. P. Smith (2010) noted that a consequence of this might be 'data-mining, searching for results in accord with one's beliefs; there is an issue about how results should be reported' [(Dunne and R. P. Smith 2010), p. 439]. The result in Geweke et al. (1983) supports the sentiment in Dunne and R. P. Smith (2010) by showing that different Granger reject the null by different frequencies and Conway et al. (1984) noted that

> Subsequent Monte Carlo tests offered suggestive results indicating differences in the power of various causality tests and showing that one could easily produce conflicting conclusions by employing a battery of causality tests on the same data sets [p. 2].

Conway et al. (1984) refer to Geweke et al. (1983) and Nelson (1981). All of this help to support the need to discuss all phases of the modelling process, for since

> there are competitive specifications of the statistical model, one should always question which alternative is better justified [Moneta and Russo (2014b), p. 71]

These questions are not, as pointed out in Moneta and Russo (2014b), strictly solved by a priori reasoning. Still, a priori reasoning does play a role in discussing the different phases in the modelling process since the choice of further specification and framework is underdetermined by data. Discussing these phases gives us a point of departure in solving the apparent disagreement and denounces the idea that these instruments provide 'neutral' results. It might not solve any dispute, but it becomes more clear where the disagreement lies, and 'surely open and honest confrontation at the level of the model should lead the debate forward, and thereby enhance our understanding of phenomena' [Moneta and Russo (2014b), p. 71].

## 4.4 Concluding Remarks

In this chapter, I identified and described the conditions necessary to infer causality by instruments in econometrics, with the primary goal being stable results. I started by introducing a case study that did help elucidate some of the problems that hunt Granger models. I argued that since measurement cannot be reduced to relationships between observables, theory is indispensable. In chapter 3, I argued in favor of the model-based approach in which theory plays a crucial role in specifying a set of acceptable values. In chapter 4, I argued in favor of calibration. I did show, following Boumans (2004), that to obtain stable readings in econometrics, it is necessary to calibrate instruments against stable facts or theories that provide a standard that allows us to estimate a range of acceptable evidence. I did also show that there are some resemblances between the calibration view and the evidential pluralists. Both views see the importance of using both data and theory in order to infer causality in econometrics.

# 5

# Conclusion

The starting point for econometrics is not great. The data is often passive, and Models were introduced as a way of trying to circumvent this fundamental condition. The econometric models, this thesis argues, should be seen as measuring instruments. However, using instruments does not change the underlying data. I have argued that data in econometrics is simply too bad to stand on its own legs. It needs theory to make substantial claims, and it needs theory to get useful information out of the data. This is clearly shown in the fact that Granger models seem incapable of producing stable and reliable evidence to either one side or the other in the big economic questions, as exemplified in the money and income case.

There is a discussion in econometrics on 'taking the con out of it'; however, that misunderstands the very foundation of econometrics. It is a movement that wants to eradicate the subjective part of econometrics, but that is hardly possible. As seen in this thesis, it is background theory that makes it possible to extract causal information from data. This naturally entails that modelling in econometrics contains a subjective element.

I suggested that a way to map background theory onto data was to make use of calibration. It is important to concede that we will never be in possession of an exact function that decides the information set, why a choice has to be made. I choose an articulated choice based on theory above an atheoretical choice - whatever that means. There is not much difference between the situation of an economist and a painter. The painter cannot measure everything down to the smallest detail either. He has to rely on skill, experience, and informed judgment in making hard choices in the process of creating and interpreting a model - and here, every step of the modelling process matter. If we want econometrics, statistics, and theory to reach a higher level together, it needs the same skillset. As my old economics professor said, 'a Marxist may not get much out of a model build on neoclassical assumptions, but an atheist does not get much out of church art either'. That is how it needs to be, as long as those assumptions are stated clearly, to begin with. This is closely related to the aim of this thesis; to acknowledge the theory-laden nature of results in the social sciences obtained through measurement.

This thesis developed criticism against the atheoretical perspective in econometrics as I aimed to do. It was done by providing an in-depth philosophical analysis of its foundation and by characterising models in atheoretical econometrics as measuring instruments. This provided a link to the philosophy of measurement. Here I explored what characterised measurement and concluded that it was not observation; That measurement could not be reduced to relations among observables since there was a non-trivial inference from reading to an outcome. I went on to investigate what was needed to infer causality by measuring instruments. Here I argued that if measuring instruments should provide a reliable way to infer causality theory was needed. Calibration provided a way in which we could map a theoretical framework into the data by providing a standard from which we could obtain an acceptable range of values. It also allowed me to defend the need for both empirical data and background theory. Something that is often lost, unfortunately. Further, this allowed me to introduce the works of Federica Russo and Jon Williamson. Their evidential pluralism provided a great amount of inspiration in writing the last chapter. Apart from these two great philosophers, I discussed the works of other great philosophers and economists, among them Eran Tal, Patrick Suppes, Kevin Hoover, Christopher Sims, Clive Granger, Tjalling Koopmans, and Trygve Haavelmo, which were all critically analysed. This allowed us to see important differences in the theoretical and the atheoretical tradition in econometrics, and it further allowed me to identify the weaknesses of both approaches.

A great deal of the argument being made in this thesis needs to be explored further. Both the role of measurement in econometrics, applying the calibration view in econometrics, and evidential pluralism in the social sciences, need further study. There is a project on the latter at the University of Kent led by Jon Williamson, which hopefully will produce some valuable results in this area. Another part that needs further study is the role of judgment in modelling, which is something I hardly touched upon. But this is something Marcel Boumans have been working on. This leaves a lot more to be done in the future.

# Bibliography


Aldrich, J. (1989). "Autonomy". In: *Oxford Economic Papers* 41.1, pp. 15–34.

Backhouse, R. E. (2007). "Representation in economics". In: *Measurement in economics: A handbook*. Academic Press/Elsevier London, pp. 135–52.

Barlow, R. (2002). "Systematic errors: facts and fictions". In: *arXiv preprint hep-ex/0207026*.

Beebee, H. (2016). "Hume and the Problem of Causation". In: *The Oxford Handbook of Hume*. Oxford University Press, pp. 228–249.

Boland, L. A. (2014). *Model Building in Economics: its purposes and limitations*. Cambridge University Press.

Boumans, M. (2004). *How Economists Model the World into Numbers*. Routledge.

– (2010a). "The problem of passive observation". In: *History of Political Economy* 42.1.

– (2010b). "The problem of passive observation". In: *History of Political Economy* 42.1.

– (2012a). "Mathematics as quasi-matter to build models as instruments". In: *Probabilities, laws, and structures*. Springer, pp. 307–318.

– (2012b). "Observations in a Hostile Environment: Morgenstern on the Accuracy of Economic Observations". In: *History of Political Economy* 44.Supplement 1, pp. 114–136.

– (2015). *Science Outside the Laboratory: Measurement in Field Science and Economics*. Oxford University Press.

– (2016). "Suppes's outlines of an empirical measurement theory". In: *Journal of Economic Methodology* 23.3, pp. 305–315.

Boumans, M. (2005). "Measurement in economic systems". In: *Measurement* 38.4, pp. 275–284.

Bridgman, P. (1927). *The logic of modern physics*. Vol. 3. Macmillan New York.

Burns, A. F. and W. C. Mitchell (1946). *Measuring business cycles*. burn46-1. National bureau of economic research.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Clarendon Press.

– (2008). "In praise of the representation theorem". In: *Representation, Evidence, and Justification: Themes from Suppes*, pp. 83–90.

Chang, H. (2001). "Spirit, air, and quicksilver: The search for the" real" scale of temperature". In: *Historical studies in the physical and biological sciences* 31.2, pp. 249–284.

– (2004). *Inventing temperature: Measurement and scientific progress*. Oxford University Press.

Chang, H. (2019). "Operationalism". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Winter 2019. Metaphysics Research Lab, Stanford University.

Chao, H. (2002). *Representation and structure: The methodology of econometric models of consumption*. Thela Thesis.

– (2005). "A misconception of the semantic conception of econometrics?" In: *Journal of Economic Methodology* 12.1, pp. 125–135. ISSN: 1350178X. DOI: 10.1080/1350178042000330931.

– (2007). "Structure". In: *Measurement in economics: A handbook*. Academic Press/Elsevier London.

– (2020). *Representation and structure in economics: The methodology of econometric models of the consumption function*. Routledge.

Chao, H., S. Chen, and R. L. M. (2013). *Mechanism and Causality in Biology and Economics*. Vol. 3, pp. 19–34. ISBN: 978-94-007-2453-2. URL: http://link.springer.com/10.1007/978-94-007-2454-9.

Chowdhury, A. R. (1987). "Are causal relationships sensitive to causality tests?" In: *Applied Economics* 19.4, pp. 459–465.

Christ, C. F. (1966). *Econometric models and methods*. Wiley.

– (1994a). "The Cowles Commission's contributions to econometrics at Chicago, 1939-1955". In: *Journal of Economic Literature* 32.1, pp. 30–59.

– (1994b). "The Cowles Commission's contributions to econometrics at Chicago, 1939-1955". In: *Journal of Economic Literature* 32.1, pp. 30–59.

Claveau, F. (2012). "The Russo–Williamson Theses in the social sciences: Causal inference drawing on two types of evidence". In: *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43.4, pp. 806–813.

Conway, R., P. Swamy, J. Yanagida, and P. von zur Muehlen (1984). "The impossibility of causality testing". In: *Agricultural Economics Research* 36.1489-2016-125328, pp. 1–19.

Cooley, T. F. and S. F. Leroy (1985). "Atheoretical macroeconometrics: A critique". In: *Journal of Monetary Economics* 16.3, pp. 283–308. ISSN: 03043932. DOI: 10.1016/0304-3932(85)90038-8.

Cooley, T. F. (1997). "Calibrated models". In: *Oxford Review of Economic Policy* 13.3, pp. 55–69.

Cuddington, J. T. (1980). "Simultaneous-equations tests of the natural rate and other classical hypotheses". In: *Journal of Political Economy* 88.3, pp. 539–549.

Culp, S. (1994). "Defending robustness: The bacterial mesosome as a test case". In: 1994.1, pp. 46–57.

Domotor, Z. and V. Batitsky (2008). "The Analytic Versus Representational Theory of Measurement: A Philosophy of Science Perspective". In: *Measurement Science Review* 8.6, pp. 129–146.

Dufour, J.-M. and A. Taamouti (2010). "Short and long run causality measures: Theory and inference". In: *Journal of Econometrics* 154.1, pp. 42–58.

Dunne, J. P. and R. P. Smith (2010). "Military expenditure and granger causality: A critical review". In: *Defence and Peace Economics* 21.5-6, pp. 427–441.

Dyke, H. (2012). *Metaphysics and the representational fallacy*. Routledge.

Engle, R. F. and C. W. J. Granger (1987). "Co-integration and error correction: representation, estimation, and testing". In: *Econometrica* 55.2, pp. 251–276.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983a). "Exogeneity". In: *Econometrica: Journal of the Econometric Society*, pp. 277–304.

– (1983b). "Exogeneity". In: *Econometrica: Journal of the Econometric Society*, pp. 277–304.

Epstein, R. J. (2014). *A history of econometrics*. Elsevier.

Farrance, I. and R. Frenkel (2012). "Uncertainty of measurement: a review of the rules for calculating uncertainty components through functional relationships". In: *The Clinical Biochemist Reviews* 33.2, p. 49.

Feige, E. and D. Pearce (1979). "The casual causal relationship between money and income: Some caveats for time series analysis". In: *The Review of Economics and Statistics*, pp. 521–533.

Fennell, D. J. (2005). "A philosophical analysis of causality in econometrics." PhD thesis. London School of Economics and Political Science (United Kingdom).

Freeman, J. R. (1983). "Granger causality and the times series analysis of political relationships". In: *American Journal of Political Science*, pp. 327–358.

Friedman, M. and A. J. Schwartz (1965). "Money and business cycles". In: *The state of monetary economics*. NBER, pp. 32–78.

Friedman, M. (1970). "Comment on Tobin". In: *The Quarterly Journal of Economics* 84.2, pp. 318–327.

Friedman, M. and A. J. Schwartz (2008). *A monetary history of the United States, 1867-1960*. Princeton University Press.

Frigerio, A. E. a. (2010). "Outline of a general model of measurement". In: *Synthese* 175.2, pp. 123–149.

Frisch, R. (1933). "Editor's Note". In: *Econometrica* 1.1, pp. 1–4. ISSN: 00129682, 14680262. URL: `http://www.jstor.org/stable/1912224`.

– (1970). "From utopian theory to practical applications: The case of econometrics". In: *Economic Planning Studies*. Springer, pp. 1–39.

Fullbrook, E. (2008). *Ontology and economics: Tony Lawson and his critics*. Routledge.

Geweke, J., R. Meese, and W. Dent (1983). "Comparing alternative tests of causality in temporal systems: Analytic results and experimental evidence". In: *Journal of Econometrics* 21.2, pp. 161–194.

Geweke, J. (2017). "Endogeneity and Exogeneity". In: *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan UK, pp. 1–4. ISBN: 978-1-349-95121-5. DOI: `10.1057/978-1-349-95121-5_118-2`. URL: `https://doi.org/10.1057/978-1-349-95121-5_118-2`.

Giedymin, J. (1976). "Instrumentalism and its critique: a reappraisal". In: *Essays in Memory of Imre Lakatos*. Springer, pp. 179–207.

Gilbert, C. and D. Qin (2007). *Representation in econometrics: A historical perspective*. Tech. rep. Working Paper.

Glasure, Y. U. and A.-R. Lee (1998). "Cointegration, error-correction, and the relationship between GDP and energy:: The case of South Korea and Singapore". In: *Resource and Energy Economics* 20.1, pp. 17–25.

Good, I. J. (1959). "A theory of causality." In: *British Journal for the Philosophy of Science* 9, pp. 307–310.

– (1961). "A causal calculus I." In: *British Journal for the Philosophy of Science* 11, pp. 197–217.

– (1962). "A causal calculus II." In: *British Journal for the Philosophy of Science* 12, pp. 43–51.

Grabner, C. (2016). "From realism to instrumentalism-and back? Methodological implications of changes in the epistemology of economics". In:

Granger, C. W. J. (1969). "Investigating causal relations by econometric models and cross-spectral methods". In: *Econometrica: journal of the Econometric Society*, pp. 424–438.

– (1980). "Testing for causality: a personal viewpoint". In: *Journal of Economic Dynamics and control* 2, pp. 329–352.

– (1988). "Some recent development in a concept of causality". In: *Journal of Econometrics* 39.1-2, pp. 199–211. ISSN: 03044076. DOI: `10.1016/0304-4076(88)90045-0`.

– (1999). *Empirical Modeling in Economics: Specification and Evaluation*. Cambridge University Press.

Haavelmo, T. (1944). "The probability approach in econometrics". In: *Econometrica: Journal of the Econometric Society*, pp. iii–115.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge university press.

Halvorson, H. (2012). "What scientific theories could not be". In: *Philosophy of Science* 79.2, pp. 183–206. ISSN: 00318248. DOI: `10.1086/664745`.

Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.

Hansen, L. P. and J. J. Heckman (1996). "The empirical foundations of calibration". In: *Journal of economic perspectives* 10.1, pp. 87–104.

Hanson, N. (1965). *Patterns of discovery: An inquiry into the conceptual foundations of science*. CUP Archive.

Hausman, D. M. (1981). "John Stuart Mill's philosophy of economics". In: *Philosophy of Science* 48.3, pp. 363–385.

Haynes, S. and W. O'Brien (2000). "Basic Concepts of Causation". In: *Principles and Practice of Behavioral Assessment*. Springer, pp. 159–170.

Heckman, J. J. (2000). "Causal parameters and policy analysis in economics: A twentieth century retrospective". In: *The Quarterly Journal of Economics* 115.1, pp. 45–97.

Heilmann, C. (2015). "A new interpretation of the representational theory of measurement". In: *Philosophy of Science* 82.5, pp. 787–797.

Hoover, K. D. (1988a). *The New Classical Macroeconomics: A Skeptical Inquiry*. Basil Blackwell New York, NY.

– (1988b). *The New Classical Macroeconomics. A Skeptical Inquiry*. New York: Basil Blackwell.

– (2001). *Causality in macroeconomics*. Cambridge University Press.

– (2004). "Lost causes". In: *Journal of the History of Economic Thought* 26.2, pp. 149–164.

– (2007). "Econometrics as observation: The lucas critique and the nature of econometric inference". In: *The Philosophy of Economics: An Anthology*, pp. 297–314.

– (2008). "The New Palgrave Dictionary of Economics Online causality in economics and econometrics". In: *Palgrave*, pp. 1–8.

Hoover, K. D. and M. E. Dowell (2001). "Measuring causes: Episodes in the quantitative assessment of the value of money". In: *History of political economy* 33.5, pp. 137–161.

Huemer, M. and B. Kovitz (2003). "Causation as simultaneous and continuous". In: *The Philosophical Quarterly* 53.213, pp. 556–565.

Israel-Jost, V. (2011). "The epistemological foundations of scientific observation". In: *South African Journal of Philosophy* 30.1, pp. 29–40.

Jacobs, R. L., E. E. Leamer, and M. P. Ward (1979). "Difficulties with testing for causation". In: *Economic Inquiry* 17.3, pp. 401–413.

JCGM (2008). "Evaluation of measurement data—Guide to the expression of uncertainty in measurement". In: *Int. Organ. Stand. Geneva ISBN* 50, p. 134.

Johansen, S. (1992). "A representation of vector autoregressive processes integrated of order 2". In: *Econometric theory* 8.2, pp. 188–202.

Kaergaard, N. (1984). *Den Oekonometriske Metode og Kritikken Heraf*. Koebenhavns Universitets Oekonomiske Instituts.

Keane, M. (2010). "Structural vs. atheoretic approaches to econometrics". In: *Journal of Econometrics* 156.1, pp. 3–20. ISSN: 03044076. DOI: 10.1016/j.jeconom.2009.09.003. URL: http://dx.doi.org/10.1016/j.jeconom.2009.09.003.

– (2013). "Experimentalist Michael R Keane". In: 24.2, pp. 47–58.

Keynes, J. (1939). "The League of Nations Professor Tinbergen's Method". In: *The Economic Journal* 49.195, pp. 558–577.

Keynes, J. M. (1921). *A treatise on probability*. Macmillan and Company, limited.

Kirchgassner, G., J. Wolters, and U. Hassler (2012). *Introduction to modern time series analysis*. Springer Science & Business Media.

Klein, L. (1977). "Comments on Sargent and Sims' "Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory"". In: *New methods in business cycle research* 1, pp. 203–208.

Komura, C. (1982). “Money, income, and causality: The Japanese case”. In: *Southern Economic Journal*, pp. 19–34.

Koopmans, T. (1937). *Linear regression analysis of economic time series*. Vol. 20. De erven F. Bohn nv.

– (1941). “The logic of econometric business-cycle research”. In: *Journal of Political Economy* 49.2, pp. 157–181.

– (1947). “Measurement without theory”. In: *The Review of Economics and Statistics* 29.3, pp. 161–172.

– (1949). “Identification problems in economic model construction”. In: *Econometrica, Journal of the Econometric Society*, pp. 125–144.

Koopmans, T. C., H. Rubin, and R. B. Leipnik (1950). “Measuring the equation systems of dynamic economics”. In: *Statistical inference in dynamic economic models* 10, pp. 53–237.

Koopmans, T. (1953). “Identification problems in economic model construction”. In: *Studies in Econometric Method, Cowles Commission Monograph*. John Wiley and Sons, pp. 27–49.

Koopmans, T. and W. Hood (1953). “The estimation of simultaneous linear economic relationships”. In: vol. 14. John Wiley and Sons, pp. 112–200.

Krantz D., E. a. (1971). “Foundations of measurement, Vol. I: Additive and polynomial representations”. In: *Dover Publications*.

Kursteiner (2016). “Granger-Sims Causality”. In: *Palgrave Dictionary*. Palgrave, pp. 10291–11304.

Kydland, F. E. and E. C. Prescott (1996a). “The computational experiment: An econometric tool”. In: *Journal of economic perspectives* 10.1, pp. 69–85.

– (1996b). “The computational experiment: An econometric tool”. In: *Journal of economic perspectives* 10.1, pp. 69–85.

Lagueux, M. (1994). “Friedman’s ‘instrumentalism’ and constructive empiricism in economics”. In: *Theory and Decision* 37.2, pp. 147–174. DOI: 10.1007/bf01079264.

Lawson, T. (1989). “Realism and instrumentalism in the development of econometrics”. In: *Oxford Economic Papers* 41.1, pp. 236–258.

Leamer, E. (1983). “Let’s take the con out of econometrics”. In: *The American Economic Review* 73.1, pp. 31–43.

Lee, H.-Y., K. Lin, and J.-L. Wu (2002). “Pitfalls in using Granger causality tests to find an engine of growth”. In: *Applied Economics Letters* 9.6, pp. 411–414.

Lee, H.-Y., K. S. Lin, and J.-L. Wu (2002). “Pitfalls in using Granger causality tests to find an engine of growth”. In: *Applied Economics Letters* 9.6, pp. 411–414.

Leijonhufvud, A. (1997). “Models and theories”. In: *Journal of Economic Methodology* 4.2, pp. 193–198.

Lucas, R. E. (1976). “Econometric policy evaluation: A critique”. In: *Carnegie-Rochester conference series on public policy*. Vol. 1. North-Holland, pp. 19–46.

Lucas, R. E. and T. J. Sargent (1981). *Rational expectations and econometric practice*. Vol. 2. U of Minnesota Press.

Mackeprang, E. P. (1906). *Pristeorier: en statistisk undersøgelse over forholdet mellem pris og efterspørgsel*. Bagge.

Maki, U. (2001). *The economic world view: Studies in the ontology of economics*. Cambridge University Press.

Malinvaud, E. (1988). “Econometric Methodology at the Cowles Commission: Rise and Maturity”. In: *Econometric Theory* 4.2, pp. 187–209. ISSN: 14694360. DOI: 10.1017/S0266466600012020.

Mari, L. (2003). “Epistemology of measurement”. In: *Measurement: Journal of the International Measurement Confederation* 34.1, pp. 17–30. ISSN: 02632241. DOI: 10.1016/S0263-2241(03)00016-2.
– (2005). “The problem of foundations of measurement”. In: *Measurement: Journal of the International Measurement Confederation* 38.4, pp. 259–266. ISSN: 02632241. DOI: 10.1016/j.measurement.2005.09.006.
Maziarz, M. (2015). “A review of the Granger-causality fallacy”. In: *The journal of philosophical economics: Reflections on economic and social issues* 8.2, pp. 86–105.
Milhoj, A. (1985). *Kausalitet eksogenitet*. Universitetets Statistiske Institut.
Mill, J. S. (1836). *On the definition of political economy; and on the method of investigation proper to it*. Vol. 4. October. University of Toronto Press, pp. 120–164.
– (1906). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific investigation*. Longmans, Green.
Moneta, A. (2005a). “Causality in macroeconometrics: some considerations about reductionism and realism”. In: *Journal of Economic Methodology* 12.3, pp. 433–453.
– (2005b). “Some Pecularities of the Concept of Causality in Macroeconometrics”. In: *History of Economic Ideas*, pp. 57–82.
Moneta, A. and F. Russo (2014a). “Causal models and evidential pluralism in econometrics”. In: *Journal of Economic Methodology* 21.1, pp. 54–76. ISSN: 14699427. DOI: 10.1080/1350178X.2014.886473.
– (2014b). “Causal models and evidential pluralism in econometrics”. In: *Journal of Economic Methodology* 21.1, pp. 54–76.
Morgan, M. S. (1990). *The History of Econometric Ideas*. Cambridge University Press.
– (2012). *THe World in the Model*. Cambridge University Press.
– (1988). “Finding a satisfactory empirical model”. In: *The Popperian Legacy in Economics*. Cambridge University Press, Cambridge, pp. 199–211.
Morgenstern, O. (1963). *On the accuracy of economic observations*. Princeton University Press.
Mouchart, M., F. Russo, and G. Wunsch (2010). “Inferring causal relations by modelling structures”. In: *Statistica* 70.4, pp. 411–432.
Nell, E. and K. Errouaki (2013). *Rational Econometric Man: transforming structural econometrics*. Edward Elgar Publishing.
Nelson, C. R. (1981). “Adjustment lags versus information lags: A test of alternative explanations of the Phillips curve phenomenon”. In: *Journal of Money, Credit and Banking* 13.1, pp. 1–11.
OECD (1994). *Jobs Study: Evidence and Explanations*. Paris: OECD Publishing.
Okina, K. (n.d.). “Reexamination of Empirical Analysis Using the Granger Causality—‘Causality’between Money Supply and Nominal Income”. In: *Monetary and Economic Studies* 3.3 (), pp. 129–162.
Oritani, Y. (1979). “Relationship Between Money Supply, Government Expenditure and Nominal GNP: A Test of Monetarist Hypothesis in Japanese Economy”. In: *Kinyu Kenkyu Shiryo, No. l, Special Economic Studies Department, The Bank of Japan*.
Pesaran, M. H. and R. Smith (1992). “The interaction between theory and observation in economics”. In: *Economic and Social Review* 24.1, pp. 1–23.
– (1995). “The role of theory in econometrics”. In: *Journal of Econometrics* 67.1, pp. 61–79.
Pheby, J. (1991). *Methodology and economics: a critical introduction*. ME Sharpe.
Press, C. (1981). “John Stuart Mill ’ s Philosophy of Economics Author ( s ): Daniel M . Hausman Source : Philosophy of Science , Sep ., 1981 , Vol . 48 , No . 3 ( Sep ., 1981 ),

pp . 363-385 Published by : The University of Chicago Press on behalf of the Philosophy of Scie". In: 48.3, pp. 363–385.

Qin, D. (1993). *The formation of econometrics: A historical perspective.* Clarendon Press.

Ram, R. (1984). "Money, Income and Causality in Japan. Supplementary Evidence: Comment". In: *Southern Economic Journal*, pp. 1214–1218.

Reiss, J. (2001). "Natural economic quantities and their measurement". In: *Journal of Economic Methodology* 8.2, pp. 287–311.

– (2012). "IDEALIZATION AND THE AIMS OF ECONOMICS: THREE CHEERS FOR INSTRUMENTALISM". In: *Economics and Philosophy* 28.3, pp. 363–383. DOI: 10.1017/s0266267112000284.

– (2016). "Suppes' probabilistic theory of causality and causal inference in economics". In: *Journal of Economic Methodology* 23.3, pp. 289–304.

Reziti, I. and A. Ozanne (1997). "An error correction model of output supply and input demand in Greek agriculture". In: *Discussion Paper Series-School of Economic Studies, University of Manchester (United Kingdom).*

Roberts, D. L. and S. Nord (1985). "Causality tests and functional form sensitivity". In: *Applied Economics* 17.1, pp. 135–141.

Roberts, F. (1979). *Measurement Theory with Applications to Decisionmaking, Utility, and the Social Sciences.* Addision- Wesley, Reading MA.

Rodenburg, P. (2004). "Models as Measuring Instruments". In:

Russo, F. (2009). "Methodology of causal modelling". In: *Causality and Causal Modelling in the Social Sciences: Measuring Variations*, pp. 53–90.

Russo, F. and J. Williamson (2007). "Interpreting causality in the health sciences". In: *International studies in the philosophy of science* 21.2, pp. 157–170.

Salmon, W. C. (2003). *Causality and Explanation.* Oxford University Press.

Sargent, T. J. and C. A. Sims (1977). "Business cycle modeling without pretending to have too much a priori economic theory". In: *New methods in business cycle research* 1, pp. 145–168.

Sargent, T. J. and N. Wallace (1976). "Rational expectations and the theory of economic policy". In: *Journal of Monetary economics* 2.2, pp. 169–183.

Sawyer, K. R., C. Beed, and H. Sankey (1997). "Underdetermination in economics. the duhem-quine thesis". In: *Economics and Philosophy* 13.1, pp. 1–23. ISSN: 14740028. DOI: 10.1017/S0266267100004272.

Scott, D. and P. Suppes (1958). "Foundational aspects of theories of measurement". In: *The journal of symbolic logic* 23.2, pp. 113–128.

Silberstein, M. (2012). "Emergence and reduction in context: Philosophy of science and/or analytic metaphysics". In: *Metascience* 21.3, pp. 627–642. DOI: 10.1007/s11016-012-9671-4.

Simon, H. A. (1952). "On the definition of the causal relation". In: *The Journal of Philosophy* 49.16, pp. 517–528.

– (1953). "Causal ordering and identifiability". In: vol. 14. John Wiley and Sons, pp. 49–75.

Sims, C. A. (1972). "Money, income, and causality". In: *The American economic review* 62.4, pp. 540–552.

– (1980a). "Comparison of interwar and postwar business cycles: Monetarism reconsidered". In: *The American Economic Review* 70.2, pp. 250–257.

– (1980b). "Macroeconomics and reality". In: *Econometrica*, pp. 1–48.

– (1981). "What Kind of a Science Is Economics? A Review Article on Causality in Economics by John R. Hicks". In: *Journal of Political Economy* 89.3, pp. 578–583.

Sims, C. A. (1987). "Making economics credible". In: *Advances in Econometrics: Fifth World Congress*. Vol. 2, pp. 49–60.

– (1996a). "Macroeconomics and methodology". In: *Journal of Economic Perspectives* 10.1, pp. 105–120.

– (1996b). "Macroeconomics and methodology". In: *Journal of Economic Perspectives* 10.1, pp. 105–120.

Smith, A. (1982). *The Wealth of Nations: Books I-III*. Penguin Classics.

Spohn, W. (1983). "From Hume via Suppes to Granger". In: *Causalita e modelli probabilistici*, pp. 69–87.

Staley, K. (2004). "Robust evidence and secure evidence claims". In: *Philosophy of Science* 71.4, pp. 467–488.

– (2012). "Strategies for securing evidence through model criticism". In: *European Journal for Philosophy of Science* 2.1, pp. 21–43.

Strawson, G. (2014). *The Secret Connexion: Causation, Realism, And David Hume*. Revised. Oxford University Press.

Summers, L. H. (1991). "The Scientific Illusion in Empirical Macroeconomics". In: *The Scandinavian Journal of Economics* 93.2, p. 129.

Suppe, F. (1991). *The Semantic Conception of Theories and Scientific Realism*. University of Illinois Press.

– (2000). "Understanding scientific theories: An assessment of developments, 1969-1998". In: *Philosophy of Science* 67.3 SUPPL. ISSN: 00318248. DOI: 10.1086/392812.

Suppes, P. (1966). "Models of data". In: *Studies in logic and the foundations of mathematics*. Vol. 44. Elsevier, pp. 252–261.

– (1973). "A probabilistic theory of causality". In: *Amsterdam: North-Holland.*

– (2002). *Representation and invariance*. College Publications London.

Suppes, P. and D. H. Krantz (2007). *Foundations of measurement: Geometrical, threshold, and probabilistic representations*. Vol. 2. Courier Corporation.

Suppes, P. and J. L. Zinnes (1962). *Basic measurement theory*. Citeseer.

Swistak, P. (1990). "Paradigms of measurement". In: *Theory and decision* 29.1, pp. 1–17.

Tal, E. (2011). "How accurate is the standard second?" In: *Philosophy of Science* 78.5, pp. 1082–1096.

– (2012). *The epistemology of measurement: A model-based account*. University of Toronto (Canada).

– (2014). "Making Time: A Study in the Epistemology of Measurement". In: *British Journal for the Philosophy of Science* 39.1, pp. 297–335.

– (2016a). "How Does Measuring Generate Evidence? The Problem of Observational Grounding". In: *Journal of Physics: Conference Series* 772.1. ISSN: 17426596. DOI: 10.1088/1742-6596/772/1/012001.

– (2016b). "How does measuring generate evidence? The problem of observational grounding". In: *Journal of Physics: Conference Series* 772.1.

– (2017a). "A Model-Based Epistemology of Measurement". In: *Reasoning in measurement*. Ed. by N. Mößner and A. Nordmann. Routledge, pp. 233–253.

– (2017b). "Calibration: Modelling the measurement process". In: *Studies in History and Philosophy of Science Part A* 65-66, pp. 33–45.

– (2019). "Individuating quantities". In: *Philosophical Studies* 176.4, pp. 853–878.

Tinbergen, J. (1939). *Statistical testing of business cycle theories: Part i: A method and its application to investment activity*. League of Nations.

Tobin, J. (1970). "Money and income: post hoc ergo propter hoc?" In: *The Quarterly Journal of Economics*, pp. 301–317.

Tooley, M. (1990). "Causation: Reductionism versus realism". In: *Philosophy and Phenomenological Research* 50, pp. 215–236.

Turner, D. (2005). "Local underdetermination in historical science". In: *Philosophy of Science* 72.1, pp. 209–230. ISSN: 00318248. DOI: 10.1086/426851.

Turner, S. (1987). "Underdetermination and the Promise of Statistical Sociology". In: *Sociological Theory* 5.2, p. 172. ISSN: 07352751. DOI: 10.2307/201938.

Vercelli, A. (1991). *Methodological foundations of macroeconomics: Keynes and Lucas.* Cambridge University Press.

– (2017a). "Causality and Economic Analysis: A Survey". In: *Macroeconomics: A Survey of Research Strategies.* Ed. by A. Vercelli and N. Dimitri. Routledge, pp. 393–422.

– (2017b). "Probabilistic Causality and Economic Models: Suppes, Keynes and Granger". In: *Nonlinear and Multisectoral Macrodynamics.* Ed. by A. Vercelli and N. Dimitri. Palgrave Macmillan, pp. 224–246.

Vining, R. (1949). "Koopmans on the Choice of Variables to be Studies and the Methods of Measurement". In: *The Review of Economics and Statistics*, pp. 77–86.

Von Mises, L. (1996). *Human action.* Auburn, L. von Mises Institute.

Weintraub, S. (1971). "Keynes and the Monetarists". In: *The Canadian Journal of Economics/Revue canadienne d'Economique* 4.1, pp. 37–49.

Williamson, J. (2009). "Probabilistic theories of causality". In: *The Oxford handbook of causation*, pp. 185–212.

Willink, R. (2013). *Measurement uncertainty and probability.* Cambridge University Press.

Wold, H. O. (1969a). "EP Mackeprang's Question Concerning the Choice of Regression a Key Problem in the Evolution of Econometrics". In: *Economic Models, Estimation and Risk Programming: Essays in Honor of Gerhard Tintner.* Springer, pp. 325–341.

– (1969b). "Mergers of economics and philosophy of science". In: *Synthese* 20.4, pp. 427–482.

Yalçin, Ü. D. (2001). "Solutions and dissolutions of the underdetermination problem". In: *Nous* 35.3, pp. 394–418. ISSN: 00294624. DOI: 10.1111/0029-4624.00303.

Yule, G. U. (1926). "Why do we sometimes get nonsense-correlations between Time-Series?–a study in sampling and the nature of time-series". In: *Journal of the royal statistical society* 89.1, pp. 1–63.

Zellner, A. (1979). "Causality and econometrics". In: *Carnegie-Rochester conference series on public policy.* Vol. 10. Elsevier, pp. 9–54.

– (1988). "Causality and causal laws in economics". In: *Journal of econometrics* 39.1-2, pp. 7–21.