The construction of instruments for measuring unemployment

Rodenburg, P.

Publication date
2006

Document Version
Final published version

The Construction of Instruments for Measuring Unemployment

ISBN 90 5170 768 1

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **383** of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

The Construction of Instruments for Measuring Unemployment

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. mr. P.F. van der Heijden
ten overstaan van een door het college voor promoties ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit

op woensdag 5 juli 2006 te 12.00 uur

door

Peter Rodenburg
geboren te Voorschoten

Promotiecommissie:

Promotor: Prof. dr. M.S. Morgan

Co-promotor: Dr. ir. M.J. Boumans

overige leden:

Prof. dr. R.E. Backhouse

Prof. dr. F.A.G. den Butter

Prof. dr. J.B. Davis

Prof. dr. M. Heidelberger

Prof. dr. J. Visser

Faculteit der Economische Wetenschappen en Econometrie

Table of Contents

1 Introduction

1.1 Prologue 1
1.2 The scope of the thesis 2
1.3 The research problem 6
1.4 Outline of the thesis 13
Appendix 1.A: Origin of the Representational Theory of Measurement. 18
Appendix 1.B: Classification of measurement. 19

2 Standardized Quantitative Rules for the Direct Measurement of Dutch Unemployment, 1900-1940

2.1 Introduction 22
2.2 The SQR of Trade Union data 24
2.2.1 The idea of an SQR. 24
2.2.2 Development of the TU SQR 24
2.3 Measurement problems of TU SQR. 34
2.4 The SQR of Labour Exchange data 41
2.5 The measurement problems of LEx SQR. 46
2.6 Conclusions. 49
Appendix 2.A: Accuracy vs. Precision. 54
Appendix 2.B: Employment assistance 1900-1940 56

3 Heidelberger's Correlation Rule and the Measurement of Cyclical Unemployment

3.1 Introduction 59
3.2 Associative measurement and Heidelberger's correlation rule 60
3.3 The development of UV-analysis. 62
3.3.1 Dow and Dicks-Mireaux: the measurement of excess demand. 62
3.3.2 Solow and the measurement of structural unemployment in the USA . 69
3.3.3 NIESR: measurement according to the traditional classification 70
3.4 The correlational structure of UV-analysis. 72
3.5 The abandonment of UV-analysis. 76
3.5.1 The identification problem of UV-analysis. 77
3.6 Conclusions 80
3.7 Epilogue: The place of the UV curve in economic theory 82

4 Two Approaches for Derived Measurement of the NAIRU

4.1 Introduction. 91
4.2 The Phillips curve and the concepts of the NAIRU and the Natural Rate of Unemployment 93
4.2.1 The Phillips curve 93
4.2.2 The Natural Rate of Unemployment and the NAIRU 95
4.3 Measurement of the NAIRU: The structural approach 98
4.4 The Phillips curve as an invariant regularity for measurement?. 101
4.5 The statistical approach: Vector Auto Regression models 109
4.6 The Representational Theory of Measurement and the statistical approach: . . 112
4.7 Conclusions 114
Appendix 4.A: VAR models: A formal treatment. 117

5 Models as Measuring Instruments: Measurement of Duration Dependence of Unemployment

5.1 Introduction 120
5.2 Models as measuring instruments?. 121
5.3 The phenomenon of duration dependence and its causes. 127
5.4 The framework for measurement of duration dependency. 129
5.5 Lancaster's 1979 duration model. 134
5.6 Lancaster’s duration model as a measuring device. 138
5.6.1 The internal principles of the model. 138
5.6.2 The Weibull function as a bridge principle. 140
5.6.3 Calibrating the model 141
5.7 Conclusions 143

6 Conclusions

6.1 The research question reconsidered. 146
6.2 The measuring instruments of unemployment. 152
6.3 Accuracy, concept formation and measurability 156
6.4 The role of measuring instruments. 161

References 164

Summary in Dutch 178

Preface

The cliché goes that writing a dissertation is a matter of hard labour that can easily give rise to mental breakdowns. I have to say, now this thesis is finally finished, that I have to agree. Yes, this thesis is the result of many years of hard labour, sometimes on the verge of a breakdown. However, I feel it is important to stress here that I experienced this period as a very joyful one. Two important factors accounted for this.

First, I experienced the writing of this thesis as an intellectual voyage of discovery. It was great to see the slow progress in the process of writing. The second reason why my PhD-ship became such a pleasant period was because of the wonderful people who I came to meet and who contributed each in their own way to its completion. Without their help and support it may even not have been finished at all.

The foremost persons I am very grateful to are my supervisors. In the first place I would like to thank Mary Morgan for accepting me as her PhD student and for letting me work with her. Though the working relation was sometimes complicated due to the geographical separation (Mary gave up her position at the University of Amsterdam and returned to London, about half a year after I started my PhD), I never felt lost and always felt very privileged to work with her. From Mary I learned to look at the bigger picture and not to worry when a case study didn't seem to fit into my thesis outline. I appreciate the enormous degree of freedom she gave me to pursue my research. In retrospect, I can honestly say that I would not have wanted it in any another way.

I am also greatly indebted to Marcel Boumans, my co-supervisor. He streamlined my thought and forced me to think in terms of measurement. In addition, he went through numerous (and undoubtedly boring) earlier versions. I am grateful for the profound comments he made. Without Marcel this work would have been very different.

To this sequence I definitely have to add Harro Maas. Not only was he formally my co-supervisor for the period of one year (in 2004, when Marcel Boumans was based at the Netherlands Institute for Advanced Study in the Humanities and Social Sciences in Wassenaar), in practice, he served throughout my whole PhD as my third PhD supervisor. He was always sincerely interested in my research and progress, keen on keeping me to my deadlines, and provided me with helpful advice on many occasions.

I felt very lucky to work with these three supervisors and my gratitude to them goes beyond words. Their contributions to this thesis are immeasurable.

Other members of the Research Group in the History and Methodology of Economics are (or were) Geert Reuten, Mark Blaug, John Davis, Hsiang-Ke Chao, Robert Went, Edith Kuiper, Floris Heukelom, Carla van El, Adrienne van den Bogaard, Eric Schliessler, Maarten Biermans, and Dirk Damsma. They all contributed in some way to this thesis and make me look back to my PhD period with great pleasure. I enjoyed the discussions, the open intellectual climate of the research group and the great literature we read and discussed. Geert Reuten, who inspired me in my undergraduate days, and Hsiang-Ke Chao, with whom I have worked for many years on the same topic and whom never stopped encouraging me, deserve special mentioning.

While it is impossible for me, for reasons of space, to mention all staff of the Department of Economics separately here, I would like to thank in particular Robert Helmink and Wilma de Kruijf of the secretariat for their help and support, and Frank Klaassen and Dirk Veestraeten for their company and interesting and entertaining discussions during supper in the university restaurant.

A special thanks goes out to my colleagues and students at the Department of European Studies of the University of Amsterdam. The Department provides a friendly academic climate and enabled me to finish the last bits of my thesis. In particular, I like to thank Michael Wintle for his support and comments on an early version of Chapter 2, and Lia Versteegh and Robbert Büthker for their encouragement and co-operation in teaching.

Others who need to be thanked are Aico van Vuuren, for his helpful comments and suggestions on Chapter 5 on search models of unemployment, and Massimo Giuliodori, for his comments and discussions on VAR models in Chapter 4. Jacques van Maarsseveen of the Central Bureau of Statistics provided great help in my archival work for Chapter 2 and I had the pleasure of writing a paper with him on Dutch unemployment statistics for the *Instituut voor Nederlandse Geschiedenis*.[1]

I would like to thank the Centre for Philosophy of Natural and Social Sciences (CPNSS) of the London School of Economics (LSE) for the facilities they provided during my stay in spring 2002 as a guest researcher. I enjoyed meeting and discussing with Till Gruene, Peter Dietsch, Christoph Schmidt-Petri, Julian Reiss, Roman Frigg, Hasok Chang, Sang Wook Yi, and Sabina Leonelli. Linda Sampson proved to be of great help.

[1] Maarseveen, J.G.S.J. van, and Rodenburg, P. (2006), Statistieken van de werkloosheid in Nederland, 1902-1943, Instituut voor Nederlandse Geschiedenis, deel 6 reeks Bronnencommentaar, ‘Bronnen met betrekking tot armenzorg en sociale verzekering in de negentiende en twintigste eeuw’.

The people of the Tinbergen Institute have a special place in my heart. I had the pleasure of meeting many kind PhD students, and many of them became close friends: Eiko Kenjoh, Ruta Aidis, Sebastiano Manzan, Pedro Cardoso, Yongjian Hu, Bert Hof, Arjen Siegman, Ada Ferrer-i-Carbonell, Jens Grosser, Rutger Hoekstra, Maria Abreu, Neeltje van Horen, Mauro Gastrogiacomo, Joel Noailly, Marije Schouwstra, Naomi Leefmans and Sebastian Buhai. I enjoyed the conversations I had with them and they will not be forgotten easily. Marian Duppen will be remembered for her never-ending support.

The philosophers of the DEMUS-discussion group at the Free University and the EIPE Institute in Rotterdam sharpened my understanding of philosophy, for which I am thankful.

Finally, I like to thank my family: my father and mother, my brother Sander, his wife and daughter Hiromi and Umi, and my close friends from Uilenstede for their love and friendship: Irma Hein, Renate Smithuis and Clarence Sabar, who didn't live out the completion of this book. To him I dedicate this thesis.

Peter Rodenburg
Amsterdam, March 2006

Abbreviations:

CBS	Centraal Bureau voor de Statistiek [*Central Bureau of Statistics*]
CCS	Centrale Commissie voor de Statistiek [*Central Commission for Statistics*]
CPS	Current Population Survey
DDM	Dow and Dicks-Mireaux
DWA	(Rijks) Dienst der Werkloosheidsverzekering en Arbeidsbemiddeling [*National Service for Unemployment Insurance and Employment Exchange*]
DUW	Number of Days of Unemployment per Unemployed person per Week
INU	Index Number of Unemployment
ILO	International Labour Organization
LEx SQR	Labour Exchange SQR
LEx UR	Labour Exchange data Unemployment Rate
LTU	Long Term Unemployment
NAIRU	Non Accelerating Inflation Rate of Unemployment
NIESR	National Institute of Economic and Social Research
NRU	Natural Rate of Unemployment
OECD	Organisation for Economic Co-operation and Development
PU	Percentage Unemployment
RTM	Representational Theory of Measurement
SQR	Standardized Quantitative Rule
TU SQR	Trade Union SQR
TLF	Total Labour Force
UV	Unemployment-Vacancy
VAR	Vector Autoregression Model

Chapter 1

Introduction

"Count what is countable, measure what is measurable and what is not measurable, make measurable"
Attributed to Galileo Galilei

1.1 Prologue

In economics a great deal of work has been done to measure economic phenomena, such as inflation, economic growth, consumption, unemployment, well-being or income distribution. In the history of economics, these instances have been considered as individual, and isolated measurement problems. The establishment of a measurement procedure for each of these properties has been done on a piecemeal basis, and likewise the judgements concerning whether these procedures were satisfactory. The philosophy of science has proved not to be very helpful to scientists in providing practical guidance with respect to measurement.

The philosophy of science has formulated criteria for proper measurement, but they turned out not to be helpful for practitioners of measurement. What seems missing in the practical measurement, for example, in the above-mentioned cases, is a set of criteria of what constitutes satisfactory measurement in practice. These criteria could bridge the gap between the abstract theory of measurement, on the one hand, and more or less ad hoc cases of measurement on the other. The aim of this thesis is to use theories of measurement in order to analyse the philosophical foundations of a number of actual instances of economic measurement and see how far economists' pragmatic methods accord with methodologically sound rules of measurement. This study is therefore a methodological analysis of strategies for the measurement of different forms of unemployment. It is not about the various methods of measuring this, but about the methodological foundations of those methods that economists use to measure things in existing practices. That is, it is a study of how economists have measured the phenomenon of unemployment rather than, for example, an analysis of the comparability of national unemployment figures or how to deal with selective non-response in surveys.

1.2 The scope of the thesis

The word 'unemployment' is now widely used both in economic theory and in everyday language. And though one might expect it to be an old term, it is in fact a fairly new one. Unlike the term 'employment', which was already used in Shakespeare's *Hamlet* in 1602, the English word 'unemployment' came into use not earlier than the mid-1890s (Garraty, 1978: 4; Oxford English Dictionary). It appeared in the *Encyclopaedia Britannica* first in 1911, while the first theoretical work in economics explicitly devoted to the problem of 'joblessness' was Pigou's *Unemployment* in 1913. The term 'unemployment' was introduced towards the end of the 19th century, and once established, we can then look back and identify the phenomenon of unemployment in earlier accounts, or trace notions of 'idleness' of labour even in ancient history.[1] It may come as a surprise that even in the work of important 19th century writers like Marx (1867), who were heavily engaged in the topic of joblessness, – the notion of involuntary unemployment is in fact one of the key concepts in his work – the term itself is absent. Whereas Marx spoke of a "reserve army of labour", "surplus population" or "redundant working population", his contemporaries spoke of "want of employment" or "involuntary idleness" or, what was more often the case, of "laziness" or "pauperism" through "want of work", rather than of unemployment. So, before the first measures of unemployment could be generated at all, not only a definition of unemployment was needed but also the very idea of unemployment as a social phenomenon had to be conceived. Like other more abstract and not empirically easily accessible scientific concepts such as heat or intelligence, unemployment had to be 'discovered' before it could be measured systematically. That conceiving the idea of unemployment was not easy follows from the fact that, for many occupations, temporary idleness was considered as an accepted part of the job. For dockers, day labourers or agricultural workers, for example, being temporarily out of work was inherent to their occupation. Hence separating 'unemployment' or 'underemployment' from 'inherent temporary idleness' was – and still is – a hard nut to crack, but was exactly the enterprise at stake.[2]

---

[1] As for example in the Bible: "And about the eleventh hour he went out, and found others standing idle, and saith unto them, Why stand ye here all the day idle? They say unto him, Because no man hath hired us" (Matt. 20: 6-7).

[2] Part of the problem of conceiving unemployment might be the ambiguity inherent to the concept itself. Work can, apart from being a way of obtaining income, be a powerful means to fulfil human needs like obtaining respect, expressing creativity, establishing social contacts, etc. On the other hand, there is the human desire to be idle or free from work and responsibility from time to time. Unemployment bears the same inherent conflicting elements. Unemployment usually has a negative connotation as it means deprivation of one's income and access to social life. At the same time unemployment bears a sense of leisure depending on the social and cultural environment. The

The first measures of unemployment were – as this study will make clear – collected at the end of the 19th century from establishments like trade unions or labour exchanges or gathered from unsystematic surveys without a clear conception of unemployment, often initiated by local authorities. Measurement of unemployment was in those days basically driven by the relevance of unemployment as a social phenomenon and the desire to alleviate its social consequences such as poverty. Local or national authorities often had the desire to map poverty, and unemployment was often taken as a cause of poverty. Thus, at first, unemployment had an important social meaning but economic scientific theories of unemployment were basically missing at the beginning of the 20th century. So, measurement of unemployment was in general driven by the social relevance of unemployment and the desire to fight unemployment and its consequences, rather than by sincere interest in it as a scientific and theoretical phenomenon. It was, therefore, agents in the social field, like charitable organizations or city councils, put simply 'doers', who were concerned about unemployment or wanted to make policy, who engaged in measurement of unemployment, rather than thinkers like Marx or Saint Simon. The use of the term 'unemployment' in policy and public life therefore predates its scientific use.

Types of unemployment

Throughout the 20th century many efforts have been made, especially by the International Labour Organization (ILO), to make unemployment both accessible for measurement and correspond with our everyday understanding of 'idleness' of labour. In contrast, economic theory has developed over the 20th century highly abstract concepts of unemployment, both on the microeconomic level and the macroeconomic level. Concepts like 'frictional unemployment' or 'cyclical unemployment' are highly abstract and have no clearly observable characteristics, but nevertheless economists have tried to measure them. Intuitively, it may seem that an economic concept like unemployment (and similarly economic concepts such as, investments, output, savings, labour force), is more 'observable' than theoretical concepts such as frictional unemployment, the natural rate of unemployment (or human capital, potential output, equilibrium price, etc.). It is true, for example, that investments are accessible to our senses. We can actually see hydraulic presses, saw machines, factories, construction yards, etc. In the philosophy of science the term 'unobservable' refers to objects or phenomena not directly accessible to human senses, such as black holes, nuclear particles, and the like. Examples in economics are concepts such as human capital, innovation, trade balance, utility, the natural rate

---

French term for unemployment '*chômage*' for example is rooted in the idea of leisure as it is derived from the Latin word '*caumare*' meaning "to take one's ease during the heat of the day" (Garraty, 1978: 4).

of unemployment, etc. As such, they are closely linked to the notion of 'theoretical entity' (Hempel: 1966), as theories suggest the existence of the concept concerned, but accessing them most often goes beyond direct human sensory perception. In this thesis 'unobservable' refers to objects or phenomena that are not *directly* observable but are in principle, (made) measurable by means of other variables that are observable.

On the other hand, a sharp distinction between 'observable' and 'unobservable' concepts of unemployment would also be unsatisfactory. Though we can observe a person doing leisure activities during working hours, we still cannot tell whether the person is unemployed, retired, disabled, or is just having a day off, or is self-employed with irregular hours of work. We still cannot observe an unemployed person through our senses alone. (Un)observability is thus a matter of degree. To avoid these problems a distinction is made in this study between *empirical* concepts of unemployment and *theoretical* concepts.[3] The distinction between these different conceptual levels of notions of unemployment is drawn with respect to the way these concepts come into existence, either in an inductive fashion referred, to as 'classification', or in a deductive fashion, referred to as 'division'.

Characteristic of empirical concepts of unemployment is that they come into existence through a more 'inductive' kind of process. The availability of statistical data gives rise to classification often for administrative purposes. Moreover such empirical concepts can be measured by what is called direct measurement. Direct measurement is "any form of measurement which does not depend upon prior measurement" [4] (Ellis, 1966: 56) (see also Appendix 1.B). Examples of these empirical concepts are "registered unemployment" and "unemployment rates". These kinds of variables are typically collected in two ways: by surveys or by collecting data from organizations. Unemployment in the United States, for example, is measured by conducting surveys, whereas in the Netherlands, until the 1970s, it was traditional to collect monthly data from labour exchanges and social security organizations about unemployed, registered workers.[5] Other labour market indicators are collected from trade unions, business organizations, firms, the tax office, and the like. It can be seen that measurement of this kind of concept is based on counting.

The problem with measurement of this class of empirical concepts lies in defining and agreeing upon the concepts necessary to collect and arrange the data. For example, in the American Current Population Survey (CPS), there is no measure of *underemployment*, since

[3] Hughes and Perlman (1984: 26) refer to these classes as 'statistical' and 'analytical' classes of unemployment.
[4] In contrast with derived measurement, such as measurement of densities or velocities for example, which can only be derived from (direct) measures of volume and mass respectively time and distance.
[5] Currently both methods are combined in the Netherlands.

there is no agreement about what constitutes underemployment, even though there is a good potential device, the population survey, that generates raw data from which it might be measured. And even when consensus can be reached about what an empirical concept should refer to, this needs to be specified in great detail. In the USA, for example, being counted as unemployed requires the person “to look actively for work”, and what that means needs, in turn, to be specified in great detail. Thus, empirically-based concepts of unemployment can be measured directly, that is, counted by surveys or data from establishments like labour exchanges, but their operational definitions are not easy to establish, and they have, in the first place, a social meaning rather than a theoretical one, as they are based not on theoretical considerations but on social conventions. In the Netherlands for example, in order to be counted as unemployed it is required to be a claimant of unemployment benefit. Though this might be fine from a social point of view, it certainly does not follow from a theoretical interpretation of unemployment, as there are no social or economic theories that have this requirement of being a claimant of social welfare. This dilemma is clearly notable in studies of international comparisons of unemployment rates, where differences in social interpretations of unemployment and also in measurement procedure can cause substantial interpretation difficulties. Consequently, the social or institutional status of empirical concepts is clear but their economic, scientific status is not and, moreover, the concepts might be easily subject to change in step with our social institutions or conventions.

Since empirically-based definitions of unemployment are essentially social constructs, they are subject to political pressures that come into play in their construction. Obviously, a variety of groups in society have an interest in the way we agree to define the phenomenon of unemployment. The way unemployment is defined is therefore clearly not a neutral act. However, no matter how important this political process is for understanding the phenomenon of unemployment, it falls outside the scope of this thesis. The political aspects of measurement of unemployment are only treated when they have relevant methodological or epistemological consequences, and (with the exception of Chapter 2) political aspects are absent in the thesis.

Scientific theories, on the other hand, utilize abstract, theoretical conceptions of unemployment. These theoretical concepts play an important role in a specific theoretical framework. Within this framework, these theoretical concepts, such as ‘frictional unemployment’, ‘natural unemployment’ or ‘structural unemployment’, can very often be defined accurately and unambiguously, but are mostly unobservable.

Clearly, these theoretical concepts follow from a particular theory and are therefore theory-laden. In the philosophy of science, in particular Logical Positivists paid a great deal of attention to the epistemological aspects of concept formation.[6] For Logical Positivists science is about formulating general explanatory and predictive principles (laws) expressed in abstract theories that have "great precision, wide scope, and high empirical confirmation" (Hempel, 1952:21). Everyday terms will not suffice for the description of general explanatory and predictive principles since they are not precise enough or have limited scope. Science therefore invokes abstract, technical terms. However, in order to be meaningful for Logical Positivists, a concept's phenomenological perception must correspond with directly observable characteristics (Hempel, 1952: 21-22). That is, the abstract terms – as the theoretical concepts of unemployment – must conceptually be connected with experiential terms that are accessible by immediate observation. This suggests that theoretical terms, like structural or cyclical unemployment, must somehow be connected with observable characteristics (or otherwise they are deemed meaningless) and that might make them measurable.

Attempts to measure theoretical concepts of unemployment in practice however are often not completely satisfactory or even completely fail. They usually cannot be measured directly, but have to be derived from an analytical framework by using specific assumptions and prior measurement of other variables. This study analyses strategies for the measurement of both types of concepts, empirical and theoretical, and uses case studies to explore their methodological problems and issues. However, in order to begin to address the problem, we need to have a theory of measurement.

1.3 The research problem

In the philosophy of science, thinking about measurement is dominated by the *Representational Theory of Measurement* (RTM) (see also Appendix 1.A). This theory of measurement considers measurement as the establishment of a correspondence between a set of manifestations of a property, and the relations between them, and a set of numbers, and the relations between them. The aim of this theory of measurement is to provide science with a fundamental, objective method of investigation. Measurement in this theory is: "The process of empirical, objective assignment of numbers to the properties of objects or events of the real world in such a way as to describe them" (Finkelstein, 1982:6). It follows that some aspects are important for measurement. First, the definition makes clear that measurement is an empirical process. Measurement is not allowed to result from, for example, thought experiments. Secondly, this definition requires that measurement is an objective process. This means that the measurement result is independent of the observer and that the measurement process can be repeated without

[6] Carl Hempel (1952) provides an overview of Logical Positivist thinking on concept formation.

altering the outcome within the limits of error. The ordering of different sets of consumption bundles by individuals is not, therefore, considered as measurement, since it involves subjective ordering. It also rules out classification systems as measurement. Library classification systems for example, though they assign numbers to objects (documents) in an objective way in such a way that the former describe the latter, are not considered as measurement since ordinality is missing in the assignment of the numbers.[7] Finally, measurement deals with the assignment of numbers to objects or the manifestations of properties according to a well-described rule. Measurement is defined in the Representational Theory as showing that "the structure of a set of phenomena under certain empirical operations and relations is the same as the structure of some set of numbers under corresponding arithmetical operations and relations" (Suppes, 1998). It is, in short, about establishing a mapping between an empirical relational structure and a numerical relational structure.

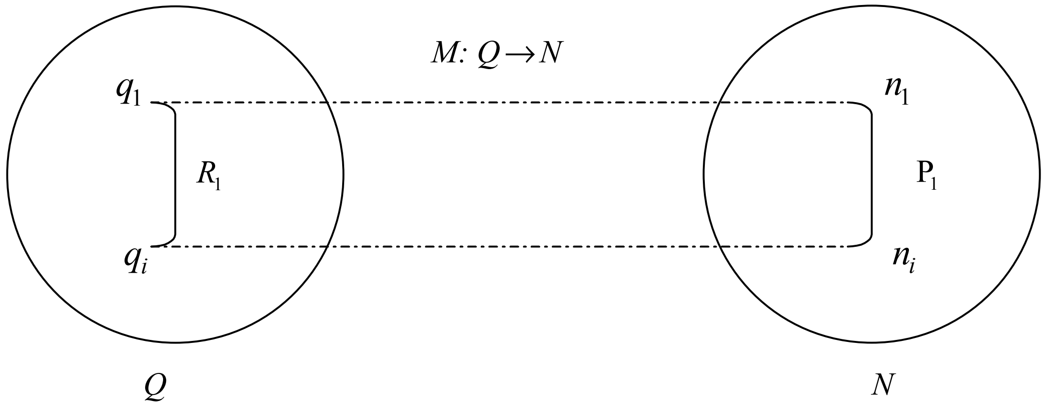


Figure 1.1: Diagrammatic representation of the set-theoretical definition of measurement
Source: Finkelstein, 1975: 105.

Formally, the representational theory of measurement is expressed as follows.

1) Let there exist a class of qualities Q (the measurand class), which consists of elements q_1 to q_i (the measurands). These elements contain the (non-empty) empirical relation R_1 corresponding to a property or a quality.
2) Let there further exist a set of (real) numbers N with the elements n_1 to n_i (the measures), which have the numerical relation P1 .

[7] This point is contested. Ellis (1966) argues that classification can be used as measurement when a rule based on ordinality is involved. Classification of a library can be considered as measurement when, for example, higher library numbers are assigned to thicker books or books with more pages. In other words, when a rule of ordinality is applied.

3) Let there exist a (homomorphical) mapping M of the empirical relational system to the numerical relational system. Thus there exists a mapping with domain Q and a range in N between the empirical relational system of Q with relation R1 and the numerical relational system of N with relation P1 (see Figure 1.1).

Though, in the execution of the actual measurement operation, several sorts of measurement can be distinguished – which in fact have led to different classifications of measurement in the measurement literature[8] – this representational principle of measurement holds for all instances of measurement.

However, in establishing a procedure for the measurement of a quantity – thus in the application of the RTM to the real world – two fundamental problems are encountered. The first one is the *problem of representation*: How can it be ensured that the assignment of numbers to objects or phenomena is justified? That is, how can it be ensured that the numbers are not literally taken in our hands and 'applied' to some physical object, but are done so in accordance with some objective rule? This is done by showing that a phenomenon or object under certain empirical operations contains the same relational structure as a set of numbers under corresponding arithmetical operations and relations. In other words, the relevant structure of the phenomenon must be isomorphic (that is, mapped one-to-one) with an arithmetical structure.[9] [10] Therefore, the RTM has, in order to establish unique measures, an additional requirement known as the *uniqueness condition* (or theorem): "If two or more numerical relational structures, which are isomorphic to a certain empirical relational structure, can be related by a certain permissible transformation, then there exists a unique numerical representation for this empirical relational structure" (Suppes, 1998).

When this representational condition is satisfied, a second fundamental problem arises, known as the *uniqueness problem*. This problem refers in fact to the problem of determining the type of scale for measurement. Temperature for example, can be measured by different measurement operations, which employ different scales: Fahrenheit, Celsius or Kelvin. Now, in order to overcome the problem that these different operations of measurement yield formal

---

[8] See Appendix 1.B

[9] "A simple relation structure (*A, R*) – with *A* being a non-empty set, and *R* a binary relation on this set – is 'isomorphic' to a simple relation structure (*A', R'*) if and only if there is a function *f* such that: (i) the domain of *f* is *A* and the range of *f* is *A*'; (ii) *f* is a one-one function, and (iii) if *x* and *y* are in *A* then *xRy* if and only if *f(x)R'f(y)*" (Suppes, 1998).

[10] For this reason Brian Ellis objects to an earlier definition of measurement by S. Stevens (1959), which lacks this representational condition. Stevens defined measurement as "the assignment of numerals to objects or events according to rule – any rule" (Ellis, 1966: 39). Ellis requires that the rule is determinative and non-degenerating. In the RTM this requirement is formulated as the representational condition.

differences between numbers, the RTM requires a second condition: the *uniqueness condition* (or theorem). This condition states that, for unique measures, two or more isomorphic mappings must be related by certain permissible transformations. An isomorphic mapping is referred to as a 'scale of measurement'. So, several isomorphic mappings, each with their own scale of measurement, are reducible to a unique isomorphic mapping and scale by making certain transformations.[11] The *uniqueness condition* (or theorem) thus assures the uniqueness of the scale used. [12]

However, other problems emerge if we interpret this theory as a practical manual, since the theory of measurement provides no practical guidance to practitioners in the field for establishing appropriate measurement. That is, the RTM bears a strict normative character, as is illustrated by its set theoretical formalization, and though some authors, like Finkelstein, argue that "the gap between the abstract philosopher's approach and that of the pragmatic instrument designer is gradually being bridged" (Finkelstein, 1982: 1), others, like Adams (1966), Heidelberger (1994a, b) and Boumans (2005a), criticize the RTM and argue that it has "turned too much into a pure mathematical discipline, leaving out the question of how the mathematical structures gain their empirical significance in actual practical measurement" (Boumans, 2005a:109). There still seems to be a wide gap between this philosophy, on the one hand, and the practice of measurement by scientists in the field on the other, and, hence the theory seems to be not completely satisfactory for either philosophers or scientists.

Let us now analyse this criticism of the RTM here in more detail. Basically this theory assumes that:

1) The concept to be measured exists.
2) We can execute an operation (based on the isomorphic mapping) such that we can measure its quantities.[13]

However, neither 1 nor 2 are straightforward.

On the first point, the RTM is unhelpful about whether things conceptualized by scientists exist in the first place; their existence is simply taken as given. Of prime importance in the RTM is, as outlined above, the establishment of the correspondence between an empirical relational structure and a numerical relational structure. The theory of measurement is hence

[11] In the RTM, however, the strict requirement of an isomorphism (a one-to-one mapping) is often dropped in favour of the weaker requirement of a homomorphic mapping (many-to-one mapping) since "in many cases of measurement distinct empirical objects have assigned the same number, and thus the one-one relationship required for isomorphism of models is destroyed" (Suppes, 1998).

[12] These internal problems and requirements of representation and uniqueness in the RTM are well known. See, for example, Suppes and Zinnes (1963), Suppes (1998), or Luce and Narens (1994).

[13] Operationalists like Percy Bridgman then take these measures as evidence of the concept's existence.

concerned with the epistemology of an objective correspondence and not the ontological status of the concept to be measured. The theory of measurement takes the Logical Positivist stance here that non-existing entities or concepts are meaningless and attempts to measure them will ultimately fail. In this way the theory of measurement also accords a kind of primary ontological status to quantities. It assumes that the concept or phenomenon exists in the first place, and, moreover, that it is clear and well defined or has an operational definition. However, in many cases of measurement, this is not the case. Often the phenomenon of interest is unclear, and we have at best only ontological conjectures about its existence. In fact, we often have scientific interest in measurement of a concept or phenomenon just because it is unclear to us, and finding a satisfactory measurement procedure might help us to clarify the concept under scrutiny. The case of measurement of unemployment shows this clearly, for neither its ontological status nor its definition was straightforward at the point when it was first measured (see Chapter 2).

On the second point, the RTM provides no guidance for establishing and justifying the homomorphic mapping and how to establish measurement procedures upon that. What philosophers have to say about the development of measurement strategies can be illustrated by Finkelstein:

> "Observation in the real world leads to the identification of empirical relations among these single manifestations [of a quality]. Examples of such relations are similarity, difference and the like. As a result the concept of a quality is formed as an objective rule for the classification of a collection of empirically observable aspects of objects into a single set, together with the family of objective empirical relations on that set" (Finkelstein, 1982:11).

This stance is 'naïve empiricism', and is too simple for several reasons.

First, as outlined above, many concepts we want to measure are not easily observable or accessible to us, or they are accessible in an operational form but the establishment of their isomorphic correspondence with numbers does not simply follow from observations alone. We often need a theory about phenomena. One does not discover quantities of nuclear particles, or relations between them, from merely casual observations. Moreover, since phenomena of interest in science are often unobservable or not easily accessible, we study them by analysing data on the phenomenon. Bogen and Woodward (1988) stress the distinction between data and phenomenon. Theories are about phenomena, that is, stable, repeatable effects or processes, whereas data are observations that serve as evidence for such theories of phenomena. In measurement, we establish data, while we want to measure the phenomenon for which the data is taken as evidence. In science, it can happen that measures are generated, but it is unclear to what phenomenon they correspond, or the data does not correspond to the phenomenon of interest, a problem analysed in full in Chapter 2. In either case, RTM is unable to deal with this phenomenon-data distinction.

Secondly, the RTM takes, as proponents of the theory acknowledge (Suppes, 1998), too little account of the analysis of variability in the quantity measured. Variability in the measures can arise from either variability in the object being measured or in the measurement procedure. The length or the weight of a person, for example, varies over the day and we may want to capture or measure directly that natural variability. Late 19th and early 20th century statisticians, like Pearson, R.A. Fisher, etc., were interested in measuring this natural variability (Hacking, 1983). However, it could also be the case that the variability lies in the measurement procedure. In that case, it is usually attributed to error and, in the 19th century, statisticians, like Lagrange, Laplace and particularly Gauss, developed their impressive statistical framework in order to deal with such measurement errors. Taylor (1982, cited in Suppes, 1998) mentions five kinds of measurement errors in astrological observations, all related to the measurement operation. There are: errors in the measuring instruments; human or personal errors such as errors in the reading of gauges; uncontrollable 'systematic' errors in the measurement conditions (like the earth's magnetic field); uncontrollable 'random' conditions (like meteorological variation); and, finally, 'computational' error, once the numerical observations have been recorded. Statistics thus provides measuring procedures which deal with both kinds of variability.

Thirdly, part of the establishment of a satisfactory measurement procedure involves the finding or construction of suitable measuring instruments, an aspect very underexposed in the RTM. Though the business of constructing measuring instruments is often associated with experimental physics, in the 20th century the social sciences have developed an impressive set of measuring instruments too. Morgan (2001b) for example, provides an inventory of the set of measuring instruments developed by economists, which include, for example, input-out models, balances, social surveys, formulae for the construction of index numbers, accounting rules, etc. Boumans (1999b) suggests using mathematical, economic models as measuring instruments, whereas Porter (1995) stresses the importance of bureaucracies as a measurement device, approaches explored, respectively in Chapters 5 and 2 of this thesis.

Finally, the RTM is also not very helpful in appraising competing procedures for measuring the same concept. Length, for example, can be measured by the use of a measurement rod, beams of light, or by sonic waves. As long as the measurement procedure is well defined and objective, they are equally good, and all three meet the requirements of the RTM. Though the measurement procedure measures the same object,[14] there may be practical or theoretical reasons to prefer one measurement procedure to another.

The aim of this thesis is to investigate the principles upon which the instruments for measuring unemployment are based, and to analyse additional requirements for the construction of measuring instruments. It follows from the above that bringing about the isomorphic mapping required in the RTM, as a basis for a measurement procedure, is not a straightforward affair. It will become clear that a satisfactory measurement procedure at least requires a stable correspondence or representation, so that we can establish an objective rule for measurement, a point stressed by Boumans (1999b). Stable correspondences or representations can be found in different guises, as this thesis will make clear. The remaining chapters can in essence be seen as ways to find, or to construct, stable representations that are partly or completely invariant – or, as Boumans (2005a: 118-119) puts it: representations of relations that are "autonomous as far as possible" – and so each can serve as a principle for the construction of measuring instruments and as a foundation for measurement. However, stability alone is not enough. Somehow the mapping between numbers and phenomenon must be established, and the principle upon which a measuring instrument is based can be interpreted as a way to bring about this mapping. As this study will make clear, the construction of measuring instruments often involves additional requirements, which may not appear to be consistent with the RTM, such as classification, standardization, theoretical or procedural assumptions, conventions, a priori empirical data, and the like, in order for a measuring instrument to work satisfactorily. These additional requirements will be inferred from cases of practical, economic measurement, where we can see different measuring instruments at work. The requirements for constructing measuring instruments will thus not follow from a priori, normative considerations or criteria but from real-world examples of measuring instruments. For this reason, the present thesis bears a positive character.

The reasons for taking the construction of measuring instruments as the main framework for this thesis are threefold. In the first place, there is the justification of the isomorphic mapping of an empirical matter that finds expression particularly in the construction of measuring instruments. Starting from the perspective of the construction of measuring instruments, measurement might regain an empirical dimension. Secondly, it stresses the human involvement in measurement. Nature provides us with no (or very few) measuring devices. Therefore, we have to construct instruments ourselves. In the construction of measuring instruments, however,

[14] P.W. Bridgman argues that, since there are different operations of measurement involved, different concepts are measured.

conventions are unavoidable. Consequently, not only has the construction of measuring instruments a social constructivist dimension, but also all knowledge obtained through them. Thirdly, though in the measurement literature two classification systems of measurement can be distinguished based on different measurement operations; namely Campbell's classification system and Stevens's system (see Appendix 1.B), it was deliberately decided not to use either system as the main framework for this thesis. The reason for doing so was the fear of forcing a case of measurement in such a way to make it fit a particular classification system or particular operations. Instead, it was considered preferable to analyse the cases according to the measuring devices they employ, such as Standardized Quantitative Rules, models, correlation structures, classification systems, and so on. However, there will also be reference to Campbell's or Stevens' classification system occasionally in the thesis when that might be helpful.

1.4 Outline of the thesis

This thesis analyses five case studies where measuring instruments were used for the measurement of concepts of unemployment. The aim is to investigate the principles upon which the measuring instruments are based and to analyse the requirements for constructing the measuring instruments. From the cases, five principles for the construction of measuring instruments were derived. They are:

1) establishing a Standardized Quantitative Rule (Chapter 2);
2) finding a stable correlation where observable variables can 'stand-in' for the unobservable phenomenon (Chapter 3);
3) finding a stable correlation relationship between observables from which the unobservable phenomenon can be derived by regression (Chapter 4);
4) finding a 'invariant regularity' through which the unobservable phenomenon is connected to observable variables (also in chapter 4);
5) finding a representation of a stable mechanism (Chapter 5).

These measuring principles and measuring instruments are derived by carefully analysing a number of historical or methodological case studies of measurement in economics. Each case is used to illustrate a particular measuring instrument and principle and to show how they are used to bring about the required mapping, and thus these cases form the framework of this thesis. Each measuring device is discussed in a separate chapter. The thesis is organized as follows (see Figure 1.2):

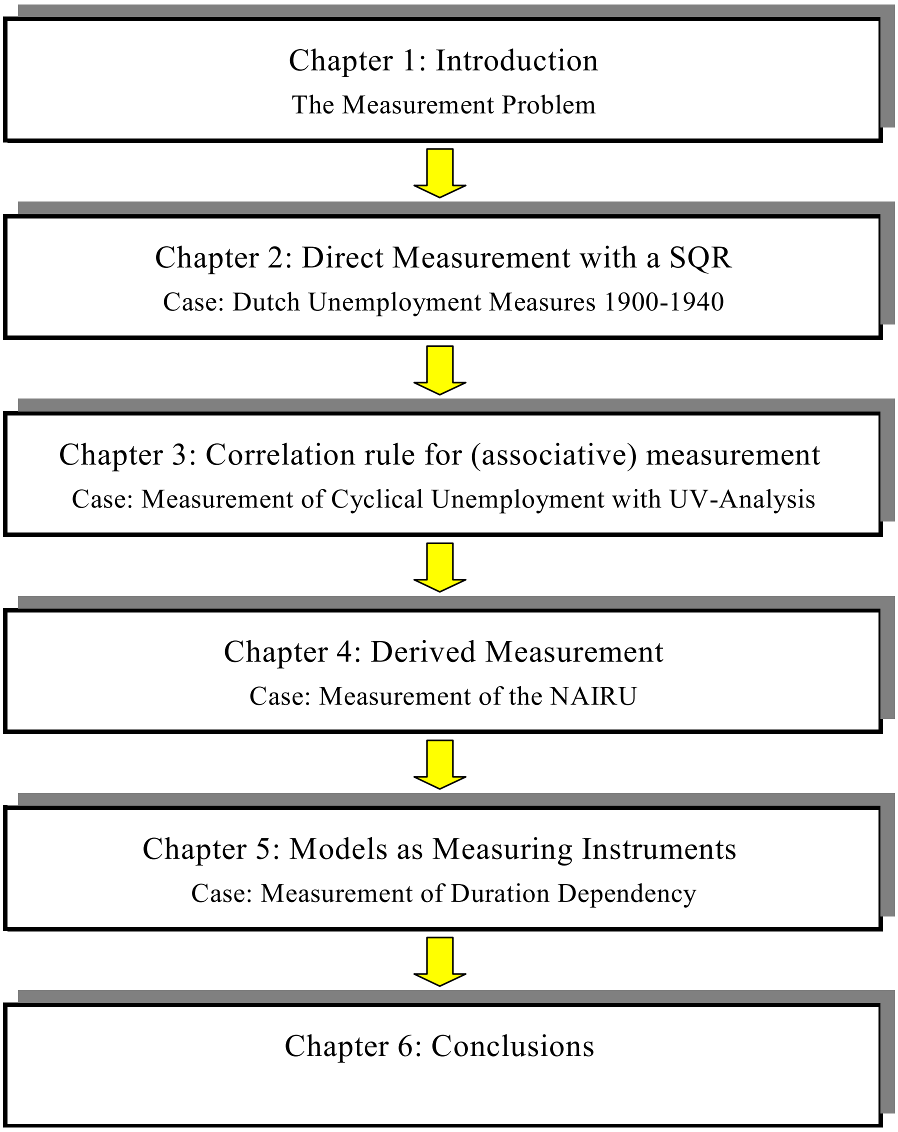


Figure 1.2: Outline of thesis

Chapter 2 discusses the direct measurement of an empirical concept of unemployment. The strategy we can see in this measuring process is that of the establishment of a *Standardized Quantitative Rule* (SQR) for the construction of a measuring instrument. The term is taken from Theodore Porter (1994) who shows how stable, standardized procedures such as, for example, those embodied in bureaucracies, can function as the basis for measurement. This strategy is frequently used in the social sciences, as, for instance, in the measurement of intelligence, which is measured through standardized IQ tests. In this thesis, the SQR strategy is illustrated by the establishment of measures of unemployment in the Netherlands in the period from 1900–1940. At the turn of the 20th century, the concept of unemployment was unclear and the Dutch Statistical Office started to collect piecemeal figures of registered unemployed from individual trade unions and labour exchanges. These statistics were made without a clear conceptualization of unemployment. The case shows how the bureaucratic procedures embodied in the trade union insurance scheme and labour exchanges could provide the stability necessary for measurement, and how bureaucratic procedures function as a measuring instrument to scale measurements. It also shows that measurement by means of an SQR is part of the social construction of the concept of unemployment.

Chapter 3, 4 and 5 discuss the measurement of theoretical concepts of unemployment. Chapter 3 discusses the strategy of finding an invariant correlation in which the unobservable object of measurement correlates with an observable that could serve as a representative or 'stand-in'. This strategy is based on Heidelberger's correlation account of measurement and is typically a case of associative measurement. Heidelberger (1994a) claims a homomorphic mapping can be brought about merely by correlation, as an unobservable might be measured through a stable, correlated relationship, $A \sim B$ (B is correlated with A). A thermometer can be considered as an example of this strategy. Temperature is unobservable and not directly measurable. But it can be measured indirectly, since it correlates with the expansion of mercury or other substance in a thermometer. A change in temperature thus leads to a change in the height of the substance in the column of a thermometer. Therefore temperature is correlated with height of the substance concerned: *temperature ~ height of mercury column* (or other substance). This change in height is, however, idiosyncratic for each and any type of thermometer; it depends on the specific construction and scale of, and substance used in, individual thermometers. Though the thermometer is based upon a causal principle, a correlation between an unobservable and an observable is sufficient for measurement, as long as the correlation itself is more or less invariant for each and any type of thermometer (Chang, 2001, 2004).

The case study that explores this strategy for economics is measurement of structural and cyclical unemployment as was done (mainly in the 1970s) by what is called unemployment and vacancies (UV)-analysis. Data of unemployment and vacancies were plotted in a graph from which structural and cyclical unemployment could be derived from the intersection of a 45º auxiliary line with the UV-plot. The case appears, however, to be more complex than the simple correlation $A \sim B$ outlined above, as both unemployment and vacancies depend upon "aggregate demand" or "the state of the business cycle"[15]. It can therefore be argued that both unemployment and vacancies have a common cause. "*d*" ('deficient demand') causes both U (unemployment) and V (vacancies) to change, and U and V are inversely correlated, but by no means directly causally connected. Thus:

[15] Without going into detail here of what causes or constitutes the "business cycle".

$$d \text{ (“deficient demand”)} \begin{array}{l} \nearrow \\ \searrow \end{array} \begin{array}{c} U \\ \sim \\ V \end{array} \quad \begin{array}{l} \text{(unemployment)} \\ \\ \text{(vacancies)} \end{array}$$

However, it can be argued that it is only deficient demand unemployment (“*Ud*”) that correlates with the “deficient demand”. Thus the full correlational connection upon which UV-analysis is based appears to be as follows:

$$d \text{ (“deficient demand”)} \begin{array}{l} \nearrow \\ \searrow \end{array} \begin{array}{c} U \\ \sim \\ V \end{array} \quad \sim \; Ud \text{ (“deficient demand unemployment”).}$$

Chapter 4 analyses the strategy of finding an invariant regularity, or an invariant, change-relating generalisation between variables, from which we can derive quantities. In Campbell’s classification (see Appendix 1.B), this strategy is considered as derived measurement. Characteristic for this strategy is that – like laws – the relation between variables remains stable or unchanged over a particular domain of changes, both in background conditions, and in changes in the variables figuring in the relation itself. However, the domain over which the regularity remains invariant is, typically, much smaller than that of true laws.

The case study of this chapter is that of measurement of the Non Accelerating Inflation Rate of Unemployment (NAIRU), which is an unobservable, highly theoretical construct. In the economic literature two approaches to its measurement have been established. First, the structural approach seeks to measure the NAIRU indirectly through the use of an invariant regularity, called the Phillips curve, which represents the relation between unemployment and inflation. The structural approach to measurement of the NAIRU, however, turned out not to be completely satisfactory for various reasons. Econometricians have therefore put forward another, way of measuring the NAIRU. This second approach to measurement of the NAIRU is an entirely statistical approach, and the strategy of measurement is based on a stable relationship between observables, from which the unobservable phenomenon is derived by regression. The latter approach involves the use of statistical techniques, such as Vector Auto Regression (VAR) techniques. The aim of this chapter is thus to explore measurement of the NAIRU first as a case of derived measurement from an invariant regularity (the Phillips curve), and, secondly, to contrast that invariant regularity (‘structural’) approach to measurement of the NAIRU with the stable correlation (‘statistical’) approach.

Chapter 5 aims to show how economic models can be used for measurement (see Boumans, 1999a). In economics it is common practice to build models, or, as Cartwright (1999) puts it "blueprints of nomological machines". In many cases, these models generate – when confronted with real-world data – numbers. And, although economists speak rather of 'assessing' or 'estimating' – and in this way make reservations as to the accuracy of the measures – we can think of this generation of numbers as measurement, since it suggests that there is a correspondence between the realm of numbers and phenomena in the real world. In this way the models' internal principles are used as a resource for measurement. This strategy is explained by a case study of models of unemployment that measure duration dependency of unemployment. The statistical framework of the method was first established by Tony Lancaster (1979) and is now used in many studies on long-term unemployment. The particular model we will investigate here is a typical exemplar of this method.

Chapter 6 presents the conclusions.

Finally, it is necessary to make some reservations here. Though economists sometimes explicitly do follow the RTM in constructing their measures, e.g. the axiomatic price index theory by Eichhorn, in many cases economists do not do so. Consequently, we may see the implied representational theory of measurement strategy only partially or only to some degree. For example, contemporary economists take the Phillips curve relation (which lies at the root of measurement of the NAIRU) as a structural relationship without clear causal content, and not necessarily as a numerical law. Thus, though the structural approach to measurement of the NAIRU ultimately fails to provide a good measuring instrument, the intention of this thesis is to make clear the main ideas of the principles upon which we built our instruments to measure unemployment.

Appendix 1.A: Origin of the Representational Theory of Measurement

The foundations of the representational theory of measurement (RTM) as we know it were laid down in the work of Helmholtz (*Zahlen und Messen*, 1887). Before this work, measurement theory was restricted to the expression of measures of a property in a ratio of the magnitude to a standard magnitude taken as unity. So, measurement was taken as the comparing of quantities with an arbitrarily-taken standard, and the measurement result always refers to something that is directly observable. Important contributions to measurement theory were made by Hölder (1901) who axiomatized the measurement of additive qualities (*Die Axiome der Quantität und die Lehre vom Mass*). That is, for the measurement of physical quantities, we can construct an operation that has the formal properties of addition. This approach to measurement came under attack by criticism from social scientists, like S. Stevens, since, in the social sciences, many manifestations of qualities, such as measures of intelligence, alienation or poverty, cannot simply be added up. Stevens carried out much fundamental work on the development of an appropriate analysis of the nature of measurement in his work *On the Theory of the Scales of Measurement* (1946). As this title suggests, his main contribution lies in the classification of scales according to their mathematical properties (see also Ellis, 1966: 58-67). The RTM became accepted by Logical Positivism, as Hempel (1952: 50-78) shows, and under the influence of Logical Positivism the RTM became generally accepted in the philosophy of science. Further recent developments in the RTM mainly involve a further axiomatization of the theory. Pfanzagl (*Theory of Measurement*, 1968) and Krantz, Luce, Suppes and Tversky (*Foundations of Measurement*, 1971) made the main contributions in this domain. The current state of measurement theory is well summarized in Finkelstein (1975 and 1982). A brief overview of the axiomatic approach to measurement can be found in Luce and Narens (1987: 428 - 432).

Appendix 1.B: Classification of measurement

In the relevant literature on measurement, two important classification systems of measurement can be distinguished: Campbell's classification system and the classification system of Stevens.[16] Both classification systems distinguish different types of measurement based on the type of scale involved in the particular measurement operation.[17] In Stevens' system (1946), scales are classified according to their mathematical properties. He distinguishes between, 'nominal', 'ordinal', 'linear interval', 'ratio', and 'logarithmic interval' scales. Campbell's classification system (1928) is based on the operation of measurement. Campbell classifies scales based on the operation of measurement and distinguishes between 'elemental', 'associative', 'derived', and 'fundamental' measurement.

Since Campbell's classification is more relevant for practical strategies of measurement – and therefore for this thesis – there follows a brief exposition of his classification and later modifications. Campbell distinguishes two main categories of measurement: fundamental measurement and derived measurement. Fundamental measurement applies to all measurement that involves no prior measurement. In contrast, for the measurement of velocity or density, for example, (at least) two independent measures are needed. Velocity or density can therefore only be measured in an *indirect* way. In later work, Campbell's interpretation of derived measurement seems to change, and derived measurement seems to be interpreted as measurement by means of constants in numerical laws, that is laws that "state relations between magnitudes" (Campbell, 1928: 57) or "a relationship between two or more quantities under specified conditions" (Ellis, 1966: 111). Calculations can, for example, be applied to the laws of mechanics and the gravity constant and so to derive measures of the acceleration of objects, etc.

Others, like Ellis, have amended Campbell's classification. Starting from that classification, Ellis argues that some forms of derived measurement can be reduced to fundamental measurement. For example, by keeping volume constant, density can be measured directly by measurement of weight. Ellis therefore distinguishes between the main classes: *direct* and *indirect* measurement. Direct measurement applies to any form of measurement, which does not depend upon prior measurement, while indirect measurement involves measurement of one or more other quantities. According to Ellis, Campbell neglects the kind of associative measurement operation exemplified by temperature measurement. Ellis therefore introduces the term 'associative' measurement for the kind of derived measurement operation where the

[16] A third classification system by H. Coombs (1952) is known but used very little in the literature on measurement. Coombs classifies scales by the type of arithmetic that the measurement operations represent.
[17] For a lengthy overview of both classification systems, see, for example, Ellis (1966:52-73).

Campbell's Classification

Class	Definition	Example
* Fundamental measurement	Any form of measurement which does not depend upon prior measurement	Mass, volume, time-interval
* Derived measurement	1) Involves measurement of one or more other quantities 2) Measurement by means of constants in numerical laws	density of substance (mass/volume)

Ellis's modification

Class	Definition	Subclass	Definition subclass	Example
* Direct measurement	Any form of measurement which does not depend upon prior measurement	Fundamental measurement	Equals Campbell's fundamental measurement; involves 'addition'	Mass, volume, time-interval
		Elementary	Applicable to all quantities	Orderning hardness on Moh's scale
* Indirect measurement	Any measurement which involves measurement of one or more other quantities	Associative measurement	Kind of measurement exemplified by temperature measurement	Association of temperature with expansion mercury column
		Derived measurement	Measurement by determination of constants	Measurement based on numerical laws, e.g Ohm's law in multimeter

Figure 1.3: Classification of measurement

measurement of one quantity stands in for the measurement of another quantity. Accordingly, indirect measurement can be divided into two subclasses: ‘associative’ and ‘derived’ measurement, while for direct measurement Ellis introduces the subclasses: ‘elemental’ and ‘fundamental’ measurement.[18] Fundamental measurement corresponds to Campbell’s ‘fundamental’ measurement, which involves the operation of addition. In contrast, some sorts of measures can be established by ordering without addition. For example, by ordering the hardness or ‘scratch resistance’ of minerals (on Moh’s scale), by scratching one mineral on another. In this operation, no addition is involved, and Ellis refers to it as ‘elementary’ measurement. Figure 1.3 summarizes Campbell’s classification and Ellis’s modification.

[18] According to Ellis these kinds of measurement form a hierarchy, as the conditions for their application becomes progressively more stringent, and their range of application becomes progressively less (Ellis, 1966: 57).

Chapter 2

Standardized Quantitative Rules for the direct measurement of Dutch Unemployment 1900-1940

2.1 Introduction

An important ideal of scientific measurement is to escape the bounds of subjective judgements. Not only may impersonal judgements provide a more public form of knowledge but also, more importantly, we are inclined to have more trust in the knowledge generated. In order to rule out subjective judgements in the process of quantification, standardization plays an important role, both in the natural and social sciences. Whereas, in experimental natural sciences, quantification is reached through the use of standardized, mass-produced measurement instruments and standardized procedures for experiments and laboratory use, fitted in standardized protocols, the social sciences rely equally heavily on standardization for achieving measures. In the social sciences, measurement can be achieved by what Theodore Porter (1994: 389) calls Standardized Quantitative Rules (hereafter SQRs), that is, standardized procedures or rules that transform qualities into quantitative measures. Many social science constructs, such as intelligence, quality of life, price inflation, innovations, and unemployment, were successfully made directly measurable by SQRs. Scientists can construct these rules of measurement in science, but, as Porter points out, in the social sciences, SQRs constructed in social or administrative life have been equally important in yielding quantities.

What is required for measurement is that stable entities or concepts are constructed, and standardized procedures or rules may help to achieve this. SQRs require disciplining people, instruments and processes in order to fix conventions in such a way that stable entities, suitable for measurement, are constructed. However, the consequence of this stabilization of entities is that we may lose many of the ways we could understand them. Of the whole rich array of possible interpretations or meanings of the concept, only one, or very few, will survive, upon which people, often specialists or experts, agree by convention. This has happened to many entities in social science, and though they were successfully made measurable by an SQR, they were at the same time understood in a much narrower sense. To quantify qualities therefore involves destruction of meaning, or as Porter puts it: "to quantify qualities is to abstract away much of their conventional meaning" (1994: 396). However, what are abstracted away are often non-measurable or non-operational interpretations of an entity. Nevertheless, the meaning of an entity is often chosen in such a way that it can be made measurable successfully, and hence yield empirical significance. The elusive entity of intelligence in psychology, for example, finally achieved its meaning by simply defining it, according to E.G. Boring's suggestion, as what the IQ test measures. The measurement procedure of an SQR thus reconfigures entities in such a way that they become quantifiable.

This chapter examines how unemployment was reconfigured, standardized and measured in the Netherlands by the use of SQRs in the period before the Second World War. In that time, two different, competing SQRs were constructed, each giving a particular meaning to the abstract idea of unemployment. The two SQRs were based on different administrative procedures. One administrative procedure took place in trade unions, which had set up an unemployment insurance system. This SQR will be referred to in this chapter as TU SQR and resulted in the Trade Union Statistics of Unemployment (TU Statistics). The other SQR was based on the administrative procedure for registering unemployed in labour exchanges and will be referred to as LEx SQR, yielding the Labour Exchange Statistics of Unemployment (LEx Statistics). Of the two SQRs, the TU SQR was considered as the most important one, as it was taken as a representative indicator for the Dutch economy.

This chapter discusses and compares both SQRs. The outline of this chapter is as follows. Section 2.2 is devoted to an analysis of the TU SQR. It clarifies the idea of an SQR and sees how it fits the measurement of unemployment based on trade union data. Section 2.3 focuses on the measurement problems of the TU SQR. Next, the LEx SQR is discussed in Section 2.4, while its measurement problems are analysed in Section 2.5. Section 2.6 finally draws the conclusions.

2.2 The SQR of Trade Union data[1]

2.2.1 The idea of an SQR

Before going into an analysis of how the public measurement system based on trade union data came about and developed through time, first, the account of Standardized Quantitative Rules will be clarified in more detail. What exactly constitutes an SQR? And how does it become standardized? Unfortunately, Theodore Porter is not very specific here. Nowhere does he provide a definition. Instead Porter focuses on political powers. Specific groups in society will have an interest in quantification and hence standardization of quantifying rules, and they will bring about standardization out of self-interest. The name suggests that SQRs are, basically, formal or informal rules, procedures, routines, formulas, or working methods that guide behaviour in institutions or bureaucracies in such a way that they produce numbers. It is the administrative rules laid down in bureaucracies that give rise to reporting and quantification, and these rules could be determined by law, statutory regulations, common practice, rules of thumb, calculating formulas, and so on. As we shall see, the case of the TU SQR involves, in the first place, the regulations for supplying benefits to unemployed trade union members and the procedures for registering and processing these payments, as well as governmental procedures for collecting and processing the unemployment figures. These administrative procedures must be prescribed and enforced in such a way that repetitions of the procedure or routine yield the same quantitative outcome. Without standardization, arbitrary judgments will come into play, which cause measurement errors and a loss of credibility in the numbers generated. The procedure must therefore rule out personal, subjective judgments and act as “a mechanical judgment” (Porter: 1994, 1995).

2.2.2 Development of the TU SQR

The routines or procedures of the TU SQR were subject to changes over time. We can distinguish three different phases in period 1906-1943, in each of which the TU SQR and the interpretation of unemployment remained more less stable.

<u>1906-1911</u>

In the 19th century the Dutch government only had a modest interest in measuring unemployment. Not only was the phenomenon of unemployment not recognized until late in the

[1] Writing this history was complicated by the fact that hardly any internal documents of the Dutch Statistical Office (CBS) are available since its pre-WW2 archives were destroyed during a bomb raid in the Second World War.

19th century (see Chapter 1), but unemployment was also not considered as a serious economic problem.[2] Though censuses had been held in the Netherlands as early as 1849, it was only in 1900 that the Minister of Internal Affairs asked the Central Bureau of Statistics (CBS: the statistical office of the Netherlands, established in 1899) to investigate the feasibility of producing reliable statistics of unemployment (Leunis and Verhage, 1996: 60). Earlier, there was, however, an interest in the quantification of unemployment by social organizations and local authorities who were directly confronted with the consequences of unemployment. In 1893, the Maatschappij tot Nut van 't Algemeen [Society for the General Good] requested the Centrale Commissie voor de Statistiek (CCS [Central Commission for Statistics], established in 1892, and the forerunner of the CBS), to investigate the extent and nature of unemployment in the Netherlands.[3] The Commission replied that:

> For a statistical investigation (..), which would be able to understand unemployment, to its full extent at a given point in time, our Commission regrets to say that, in practice, there is no executable method known[4] (CCS, 1893: 27).

Measurement of unemployment was considered to be extremely difficult and, according to the CCS research, should instead be directed towards its effects, namely, the family-income over a certain time period.[5] The problem of measurement of unemployment was the absence of a method for measuring and the lack of an operational definition of unemployment.[6] A count of unemployed in the city of Utrecht in 1893 made clear that the classification of those out of work was extremely difficult. Yet, the CBS published the first official figures of unemployment a few years later (in 1906), which were obtained from trade unions with an unemployment insurance arrangement.[7]

---

[2] Unemployment was not considered a serious economic problem for two reasons. First, population growth was (in accordance with classical thinking and the doctrine of "laissez faire" which the Dutch government had adopted in the late 19th century) regarded as an endogenous variable and able to adjust to the long term equilibrium state of the economy. Secondly, there was faith in Say's Law, which states that there could be no underutilization of resources by definition (as "supply creates its own demand") and this rules out involuntary unemployment. Apart from some unemployment of those who changed jobs, unemployment could only be created as a result of individual shortcomings, like shirking, drunkenness, laziness, etc. Dutch economists, like C.A. Verrijn Stuart, therefore saw unemployment in essence as a wage problem (Vries, 1976: 6-7).

[3] At the same time of this request for a systematic measurement of unemployment, a few earlier attempts to measure unemployment in the Netherlands were made by local authorities. In the winter of 1893 and the winter of 1894/1895, unemployed workers were counted by order of the City Councils of Utrecht and Amsterdam.

[4] My translation.

[5] An investigation conducted by the CCS in 1894.

[6] The CBS usually established instructions in which statistical concepts were described in detail but they are not known for unemployment for the period 1900-1940. The CBS archive mentions a '*project de definition du chomage*' (project of defining unemployment) in 1921 but the result of this project appears to be missing.

[7] In this way, the CBS followed the practice of the *Bureau van Statistiek* [Bureau of Statistics] of the municipal Amsterdam who collected unemployment data locally in 1899.

In the second half the 19th century in occupational organizations (the forerunners of the trade unions), united workers started to set up private unemployment insurance arrangements for unemployed members in the Netherlands. The first one to establish an unemployment fund in the Netherlands was the professional organization of typographers, which was set up around 1860 (Velthuisen, 1948: 2) At the turn of the century, there were hundreds of these occupational organizations in the Netherlands, operating locally and divided along social and religious lines.[8] Under this insurance arrangement, those trade union members, who became unintendedly jobless[9] and had paid their contribution fees to the unemployment fund, were entitled to a benefit from their organization for a certain period. Benefits were paid on a daily basis. This first unemployment insurance scheme was not, however, insurance to the full extent. For example, the unemployed were only insured for a limited period of time and frequency of unemployment. Further, there was no official agreement between insurer and insured, which included all rights and obligations. Neither was there a separate fund for the payments of benefits. Benefits were paid out of the incoming contributions of the members, and no risk calculations were made.[10] Therefore, there was no guarantee that the unemployment funds could honour their obligations in hard times. Consequently, the unemployment funds could run into severe financial troubles when unemployment increased sharply. This happened at the turn of the 20th century and the trade unions turned to the local authorities for financial support. In 1906, the City Council of Amsterdam established the first municipal unemployment funds in order to relieve those out of work from poverty (followed by the city of Utrecht in the same year). The purpose of these funds was to subsidize local trade unions with an unemployment fund. The municipal unemployment funds thus provided an additional benefit payment to unemployed union members through the trade unions. Unorganized workers did not benefit from this arrangement. By 1912, 32 municipal funds were active in the major urban areas. In this way, the local government gradually became involved in unemployment care in the 20th century. Though the interference of the municipalities was approved by the central government, the government did not consider a national insurance arrangement or the subsidizing of trade unions' private unemployment funds as a governmental duty. Subsidizing unemployment funds, which pursued only the interest of union members, would implicitly support and benefit workers at the expense of employers

[8] Between 1906 and 1911 approximately 2800 of these organizations were active.

[9] It is often argued that the primary aim of unemployment insurance was not to support unemployed members for social reasons, but to prevent unemployed colleagues accepting work under poor conditions, like lower wages, that might affect the interests of other members. See, for example, Velthuisen (1948: 3).

[10] The absence of official statistics of unemployment made the business of calculating the risk of unemployment of course very difficult.

(Velthuisen, 1948: 10). The central government therefore kept aloof from any interference in unemployment assistance.

Each supported trade union had its own terms and conditions for providing benefits to members. Both the trade unions and the municipal unemployment funds kept records of the number and amount of benefit payments to unemployed union members. In 1906, the CBS requested the municipal unemployment funds to provide data once a month, about the number of unemployed trade union members that they subsidized, the number of persons unemployed during each week, and the number of days of unemployment, and, the total benefit payments, all in weekly reviews. These trade union data were published in *Maandschrift*, the monthly publication of the CBS as the statistics of trade union unemployment and were taken as figures of unemployment. Each trade union unemployment fund was mentioned separately, i.e. there was no sense of the aggregate absolute level of unemployment. And, of course, only the subsidized trade unions could be taken into account. Trade unions without an insurance arrangement were not considered.

To summarize, for the period 1906-1911, when the central government initiated attempts to measure unemployment, no operational definition of unemployment was available. The trade union figures that the government took as indicators of unemployment were records of unemployed members and were kept for 'bookkeeping' and administrative ends only. They were not created as measures of unemployment and each trade union had its own, idiosyncratic terms and conditions for considering members as unemployed with respect to the duration and frequency of unemployment spells. Union members did not receive benefits when they were unemployed too long, too short, or too frequently, and were therefore not considered and registered. Consequently, each trade union held its own implicit definition of unemployment. In addition, the idiosyncratic conditions for supplying benefits were subject to changes determined by the trade unions. However, in practice, trade unions applied in a rough sense a more or less similar notion of unemployment, as benefit payment was only considered in case of involuntary idleness due to lack of work, excluding idleness due to labour disputes, illness, strikes and vacations. Voluntary quitters, for example, were not considered eligible for benefit payments. By taking trade union data, the Dutch government implicitly adopted the definitions of the trade unions. However, since measures of unemployment were not collected in a uniform, standardized, way, the TU SQR was not standardized in this period, and it is not quite clear how to interpret these figures of unemployment. It seems best to interpret the unemployment data in this period as local knowledge only defined by idiosyncratic trade union regulations.

1911-1917

The way figures of unemployment based on trade union data were presented and constructed changed radically in 1911. The publication of statistics of individual funds was discontinued, and instead three indicators of unemployment were developed by the CBS, based on trade union data:

- The Index Number of Unemployment (hereafter INU);
- The Percentage Unemployed (PU);
- The Number of Days of Unemployment per Unemployed person per Week (DUW).

This set of indicators was used until 1943, and the figures were published each month, based on the average of four – or five – weekly returns. The most important and most-often used indicator was the INU. We will illustrate how these indicators were calculated by using the following figures taken from the four-weekly period 1 to 27 August 1927.[11]

A.	Number of insured persons:	285,035	(in persons; average per week)
B.	Number of unemployed persons:	18,730	(in persons; average per week)
C.	Number of days of unemployment:	92,360	(in days; average per week)

The INU was defined as the ratio of the number of days of unemployment to the number of days the insured workers could have worked. It was calculated by dividing (C) over the average number insured (A) times the number of working days per week, namely 6. The INU in the example is thus:

$$\text{INU} = \frac{\text{The number of days of unemployment}}{\text{Number of days of potential employment}} = \frac{\text{C}}{6\times\text{A}}\times 100\% = \frac{92,360}{6\times 285,035}\times 100\% = 5.4\%.$$

The second indicator, the PU, was defined as the percentage of the trade union members (in persons) being unemployed (B) of the total number of insured workers (A).[12] In the example the PU is:

[11] Nixon (1928: 643) discusses the Dutch indicators using this example. The figures are based on CBS *Maandschrift* of 31 October 1927.

[12] The PU was calculated slightly differently from the unemployment percentage used in most other countries. Usually the percentage of unemployed is defined as the number of workers unemployed on a given day divided by the number of workers (exposed to the risk of unemployment). In the Netherlands, however, the percentage of unemployed was based on the number of persons who were unemployed at any time during a given week (ending on Saturday). It does not, therefore, represent the number of *cases* of unemployment, but the number of *persons*. For example, a worker who became out of work twice in a week was counted as one case of unemployment.

$$\text{PU} = \frac{\text{Number of unemployed}}{\text{Number of insured persons}} = \frac{B}{A} \times 100\ \% = \frac{18,730}{285,035} \times 100\ \% = 6.6\ \%.$$

The third indicator was the DUW. From the example above, it is calculated as:

$$\text{DUW} = \frac{\text{Number of days of unemployment}}{\text{Number of unemployed}} = \frac{C}{B} = \frac{92,360}{18,730} = 4.93 \text{ days (per person).}$$

The maximum number of days of unemployment/person was obviously 6. The three indicators were calculated for different industries and occupations for which the ratio varied. It can be seen that these three indicators were a well-thought out, systematic set of indicators, since they were related to one another in the following way:

$$\text{PU} \times \text{DUW} = \text{INU} \times 6 \qquad \text{thus: } \frac{B}{A} \times \frac{C}{B} = \frac{C}{6 \times A} \times 6.$$

The CBS considered a greater coverage of workers whose employment status was reported (sample size) of great importance, and was happy to see that the number of counted workers increased sharply after 1914. Soon after the outbreak of the First World War in August 1914, unemployment in the Netherlands increased sharply.[13] As a result, the benefit payments of the trade union unemployment funds rose sharply and the unemployment funds, unequal to their task, ran into financial problems. Already in 1909 – as a reaction to the economic crisis of 1907/1908 – a state committee on unemployment had been established, chaired by Willem Treub. Its task was to investigate the causes of unemployment and propose means to avoid it. In their final report in June 1914, they concluded that unemployment insurance was valuable but should still be left to private initiative. The government was, however, advised to encourage the voluntary insurance of workers; in the first place, by subsidizing trade unions and other organizations that intended to insure workers. In 1914, two months after the presentation of the report of the state committee, the First World War broke out, and the implementation of the committee's recommendations was speeded up. The central government started to subsidize the trade unions on a temporary basis by passing *Noodbesluit 1914* ["Emergency Resolution 1914"]. In this resolution, formally known as the "emergency resolution for the relief of municipal unemployment funds", and also called "Emergency Resolution Treub",[14] the trade union unemployment funds and some unemployment funds of trade unions without insurance were

[13] The INU amounted to 22.2 % in August and 27.2 % in September.

subsidized directly by both the municipal unemployment funds and the central government. In the Circular of 26 August 1914, the government requested municipalities without municipal unemployment funds to establish municipal unemployment funds and support trade union with unemployment funds. As a result of the *Emergency Resolution*, the number of municipal unemployment funds rose sharply from 32 in July 1914 to 87 in December 1914, and to 119 in early 1917. As an overall result, the number of unemployed counted in the statistics rose from 73,000 in Augustus 1914 to 140,000 in December 1914, and to 167,000 early 1917 (Gerwen, 2000: 227).[15]

For the period 1911-1917, the unemployment figures of the TU SQR were no longer taken as totals of individual trade unions, but were now interpreted as a sample of aggregate unemployment. The CBS was careful to present the TU figures as *unemployment of workers at risk of unemployment*. Nevertheless, in political and social discourse they were not interpreted as such but taken as representative for the total labour force, and hence as an indicator of national unemployment. This point will be discussed in full in the next section, where we shall see that the trade unions members were a selective group of workers, a circumstance which had consequences for their representativeness. The introduction of the set of unemployment indicators can be seen as an attempt to make the measurement procedure more standard. But, since the regulations for providing unemployment benefits were still left to the trade unions, and therefore still idiosyncratic and unstandardized, the TU SQR was like that as well. However, the meaning of the unemployment figures in this period shifted from a local to a national indicator (though the CBS also kept publishing the unemployment figures of the individual (major) trade unions on an irregular basis).

<u>1917-1943</u>

After 1917, government interference changed from temporary to permanent. By 1916, unemployment had fallen to more or less ‘normal’ levels, and the necessity for governmental support became superfluous. However, the call for a permanent role for the government in unemployment insurance became louder from, among others, the *Vereeniging van Gemeentelijke Werkloosheidsfondsen* [Association of Municipal Unemployment Funds], *Nederlandsch Verbond van Vakvereenigingen,* [Netherlands League of Trade Unions], and the *Nederlandschen Werkloosheids Raad* [Netherlands Unemployment Council]. The government realized that an

[14] By 1914 Willem Treub was Minister of Agriculture, Industry and Trade and implemented the *Emergency Resolution 1914* himself.

[15] In addition, some occupational organizations without insurance arrangements were supported under Emergency Resolution 1914. Information about unemployment of their members also became public.

abolishment of unemployment subsidy would be a step back in the development of a system of unemployment insurance, hence the transformation of the emergency resolution into a definitive organization of unemployment insurance came up for discussion and, moreover, because the risk of a renewed crisis was not inconceivable. In addition, previous attempts in other countries to set up an insurance system without the involvement of trade unions, as in St. Gallen in Switzerland in 1894, turned out to be unsuccessful (Velthuisen, 1948: 4).

For a successful unemployment insurance, there needs to be: (i) an equal risk of unemployment for the associated workers; and (ii) a system of supervision in order to rule out abuse. The trade union unemployment funds were the only organizations that met these requirements. For the transformation of the temporary unemployment care into a permanent one, the government could therefore either set up their own system with their own supervision system or cooperate with the trade unions. Since the government did not want to set up their own system of unemployment insurance, cooperation with trade unions became inevitable. The Dutch government therefore decided to cooperate with the trade unions and established a permanent system of support to trade unions. Of course, there was criticism of the 1914 Emergency Resolution, since only specific groups (union members) were supported, while others, such as small independent workers and unorganized workers, had to eat into their savings in the event of unemployment. But since the latter were not organized in a professional organization, they were not a political factor of significance.

In 1917, the government passed the "Werkloosheidsbesluit 1917" (*Unemployment Resolution 1917*), which established a permanent role for the government. The municipal unemployment funds were discontinued and the unemployment funds were subsidized directly by the state. With this measure, unemployment insurance became partly a governmental affair; though the insurance arrangements were still left to the trade unions. This resolution was not therefore the start of a compulsory insurance arrangement. With its subsidy, the government wanted to encourage participation in voluntary insurance. The resolution also introduced the 'Danish' system of unemployment insurance. In the old system, known as the 'Gentian' system, subsidy was granted to the payments of the funds to unemployed members, whereas in the new system subsidy was granted on the contributions that the funds members had to pay.[16] This new arrangement, the Danish system, opened the possibility of capital accumulation for the funds, which encouraged the participation in unemployment funds and speeded up the process of centralization of the funds. As a result of the implementation of this new system, the number of insured workers increased.

[16] A comprehensive comparison between both systems can be found in Velthuisen (1948).

The government suspected, however, the trade unions with respect to the abuse of benefit payments, and attached several conditions to the subsidy transfer. The union regulations for benefit payments had to be approved by the Minister of Social Affairs (Article 2 of the Unemployment Resolution 1917). The decision to supply benefits to unemployed members was up to the trade union regulations, though under supervision of both the government and the municipality where the union was based (Article 7). The government and municipality could also give notice of appeal if benefits were supplied in a way not permitted by the rules (Article 5). For the purpose of supervision and execution of the subsidy transfer, the government specially established *Rijks Dienst der Werkloosheidsverzekering en Arbeidsbemiddeling* (hereafter DWA), [National Service for Unemployment Insurance and Employment Exchange]. A first important achievement of the DWA was the standardization of conditions of benefit payments of trade unions. In order to be considered for government subsidy, the trade unions had to adjust their regulations for benefit payment. The DWA designed a standard regulation with 54 Articles, which was adopted by practically all unions, though some deviations were allowed (Velthuisen, 1948: 32). As a result of the government involvement and interference, the trade unions devised similar regulations for supplying benefits. Another obligation imposed on the trade unions funds was the supply of statistical data to the DWA. The data, which the unemployment funds had to supply to the DWA in their weekly reports of the week, concerned the number of unemployed members who received benefit, the number of days over which benefits were paid, the total number of unemployed members, including those not entitled to benefit and the total number of days of unemployment, including the ones over which no benefits were paid. The CBS agreed with the DWA to receive this data from them and used it for calculating and publishing the trade union unemployment statistics, i.e. the indicators of unemployment.

The insurance system remained the same throughout the 1920s and 1930s and so did the method of subsidizing by the government,[17] though in the 1920s, on the initiative of the International Labour Organization, the CBS introduced a subdivision of the indicators of unemployment per industry (20 classes) and per occupation (24 classes). The indicators of unemployment in this period were therefore also understood to represent unemployment at an intermediate, meso-level.

The trade union statistics of unemployment were finally terminated in 1943, as a result of the German occupation in the Second World War. All unemployment funds were united in one

[17] The 1929 Annual report of the CBS mentions proposals to change the method of calculating the INU and the PU. Later annual reports do not, however, show evidence of a change in methods of calculation. It seems, therefore, that the method of calculation was not altered and remained the same throughout the 1930s, though the name of the INU was changed in the 1930s to: ‘days of unemployment in percentages’. Hence, it still presented the average number of days of unemployment relative to the potential number of days of employment.

unemployment fund of the national-socialist *Nederlandse Arbeidersfront* [Netherlands Workers' Front] in 1942. One year later, this unemployment fund was discontinued and the money was transferred to the Dutch government.[18] The trade union unemployment funds had ceased to exist. After the war, the trade union statistics of unemployment were never re-established since the unemployment funds were never re-established. The government developed a system of compulsory unemployment insurance after the war, and the unemployment insurance became entirely a governmental affair.

	1906-1911	**1911-1917**	**1917-1943**
Definition of unemployment	not uniform, determined by trade unions	not uniform, determined by trade unions	implicit, uniform definition determined by government
Standardization SQR	idiosyncratic TU regulations	idiosyncratic TU regulations	"standard" regulations of DWA
Interpretation of unempl. figures	local indicator	national indicator	national indicator, industry and occupation indicator
Rules	subsidy by municipalities	subsidy by municipalities temporary subsidy by state (Danish system)	permanent subsidy by state (Gentian system)
Basis of count	individual trade union records	individual trade union records via DWA	individual trade union records via DWA
Statistics Published	number of unemployed trade union members	INU, PU, DUW	INU, PU, DUW
Presentation of data	totals	calculation of ratios (adjustment for seasons and some professions)	calculation of ratios (adjustment for seasons and some professions)

Table 2.1: Development of the TU SQR, 1906-1943

Thus, after 1917, the TU SQR had changed again. The DWA started to standardize trade union regulations by developing a 'model' regulation. The trade unions adapted their own regulations in favour of the 'model' regulations, and hence a more stable concept of unemployment emerged. The TU SQR for measuring unemployment now became standardized,

[18] A detailed description of the winding up of the funds can be found in Velthuisen (1948).

and unemployment was now defined in a uniform way by the central government, both at macro- and a meso-level. Table 2.1 summarizes the most important characteristics of the development of the TU SQR in the period 1906-1943.

2.3 Measurement problems of TU SQR

The TU SQR had three essential problems concerning the measurement of unemployment.

1. No stable concept of unemployment

As outlined above, when the first attempts were made to measure unemployment, there was no stable conception of what constitutes unemployment. There was no operational definition of unemployment and, by resorting to the bureaucratic data of trade unions, the CBS left defining unemployment to the trade unions, and the unemployment figures acquired a particular interpretation. This was particularly the case for the period 1906-1917. After 1917, when the government became involved in unemployment insurance, the DWA started to stabilize the concept by establishing uniform conditions for benefit payments.

In terms of the diagrammatic representation of measurement from the representational theory of measurement (RTM)(see Chapter 1), this problem can be presented as follows (Figure 2.1).

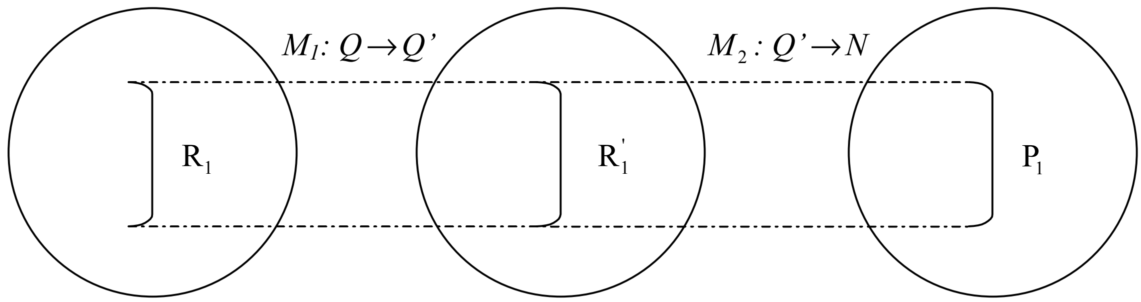


Figure 2.1 Extended diagrammatic presentation of measurement of unemployment

For unemployment, the step from phenomenon to numbers required a process of conceptualization. It was not immediately clear how the abstract idea of unemployment had to be understood, and how it could be transformed into something more directly observable. Therefore, not only is an isomorphic mapping required from the object to be measured to numbers (M_2) but also a stable, operational interpretation of the phenomenon (M_1). As outlined in Chapter 1, the RTM passes over this step of conceptualization, and more or less takes this as given.

The CBS had no idea how to operationalize the concept of unemployment, and took, the trade union data of unemployment. The TU SQR therefore helped to conceptualize unemployment. The process of conceptualization was left to the trade unions as they determined who was considered eligible for benefit payment or not, and this gave unemployment, as Porter argues, a particular meaning. The problem was that the trade union regulations were not standardized, which yielded an unstable conceptualization (M_I) of unemployment that could only be stabilized when then central government intervened in unemployment insurance in 1917 and established uniform conditions for unemployment benefit payment.

2. Selective sample

This problem concerns the accuracy of the TU SQR: Do we measure what we actually want to measure? (see Appendix 2.A). In the TU SQR, unemployed trade union members were considered as a sample of total unemployment. Trade union members are not, however, a random sample, but form a group with specific features, and therefore, the TU SQR defines a particular kind of unemployment. In the first place, only a small percentage of workers were members of a trade union. Most of the workers were not member of a trade union. Robert Kloosterman estimates the share of organized workers to vary between 10-15 % of the total labour force for the period 1920-1939 (Kloosterman, 1985: 25).[19] In this sample group, some specific types of workers were absent or overrepresented. In the Dutch trade unions, skilled workers were overrepresented, but workers who were self-employed, such as small independent business people, or practitioners of free trades (shopkeepers, farmers, handicraftsmen, lawyers, doctors, etc.), were completely absent (Kloosterman, 1985: 19). Other groups of workers that were also excluded were new entrants in the labour market and working family members. Furthermore, in some industries, trade unions were represented much more strongly than in other industries. In 1930, the highest percentages of insured workers could be found in the diamond industry and the printing industry with, respectively 89.4% and 70.0% of the insured (Kloosterman, 1985: 17). This is not surprising, since workers in these industries were the first to organize in the Netherlands. In three major industries, at least a quarter of the workforce was insured: the building industry (45.2%), the metal industry (36,0%) and the food industry (26.6%). In other industries, the insurance rate was far less: agriculture (9.7%), commerce

[19] In its 1925 annual report, the CBS reports: “Most unorganized fall out of scope of investigation, except the few who are members of unemployment funds that are not associated with trade unions. Most of the unemployed are not organized” (my translation)(CBS, 1925: 8-9).

(9.0%), transport (20.5%) and clothing (10.6%). In the clothing industry, the low insurance rate was caused by the structure of the industry: most of the labour was done by home-based workers, or in small firms, and most of the workers were women. As a rule women were underrepresented in the trade union data. It follows that skilled workers, such as diamond manufacturers and printers, were the better organized, whereas unskilled workers, such as agrarian workers and workers in the clothing industry, had a low trade union participation rate and were poorly organized.

Secondly, there seems to be a regional aspect involved in the measurement of unemployment by means of the trade union sample group. Most union members lived in the Western part of the Netherlands, which was (and is) the most urbanized and industrialized part of the country, and there also the industries with high insurance rates were found. Industries with low insurance rates, such as agriculture and clothing, were found in the more rural, Eastern part of the country. In his research on the regional distribution of unemployment for the Dutch inter-war period, Kloosterman finds structural regional differences (Kloosterman, 1985: 280). According to him, the Western part of the country experienced a structurally higher level of unemployment, even after a correction for differences in industries. But, since unemployment in industries in the Western part of the country was more often counted (because of higher insurance rates), as a result, the INU was likely to overestimate unemployment.[20] It will be clear from the above, that the sample on which INU was based was not a random or a representative sample from the total population. In general, the workers whose employment status was reported could be characterized as male, skilled and salaried employees working in the urbanized part of the country. There was an obvious disproportionate representation of certain industries and one sex.

The CBS was aware that this sample of the population was not a representative one, and therefore started an investigation in 1925 into the representativeness of the sample group. The trade unions were questioned about how representative their unemployed members were for total unemployment. Since they, of course, did not have insight into unemployment among non-union members, they could only answer this question with 'considerations of a general nature' (CBS, 1925: 8-9). In 1936 Velthuisen (1936:3) mentions that: "An investigation, undertaken in the year 1925 by the Central Bureau of Statistics, made plausible that, in general, unemployment among unorganized workers is not smaller compared with that of organized, but at least as large".[21]

[20] Van Zanten mentions this uneven distribution of unemployment and its consequences in 1928 (van Zanten, 1928: 315).

[21] My translation

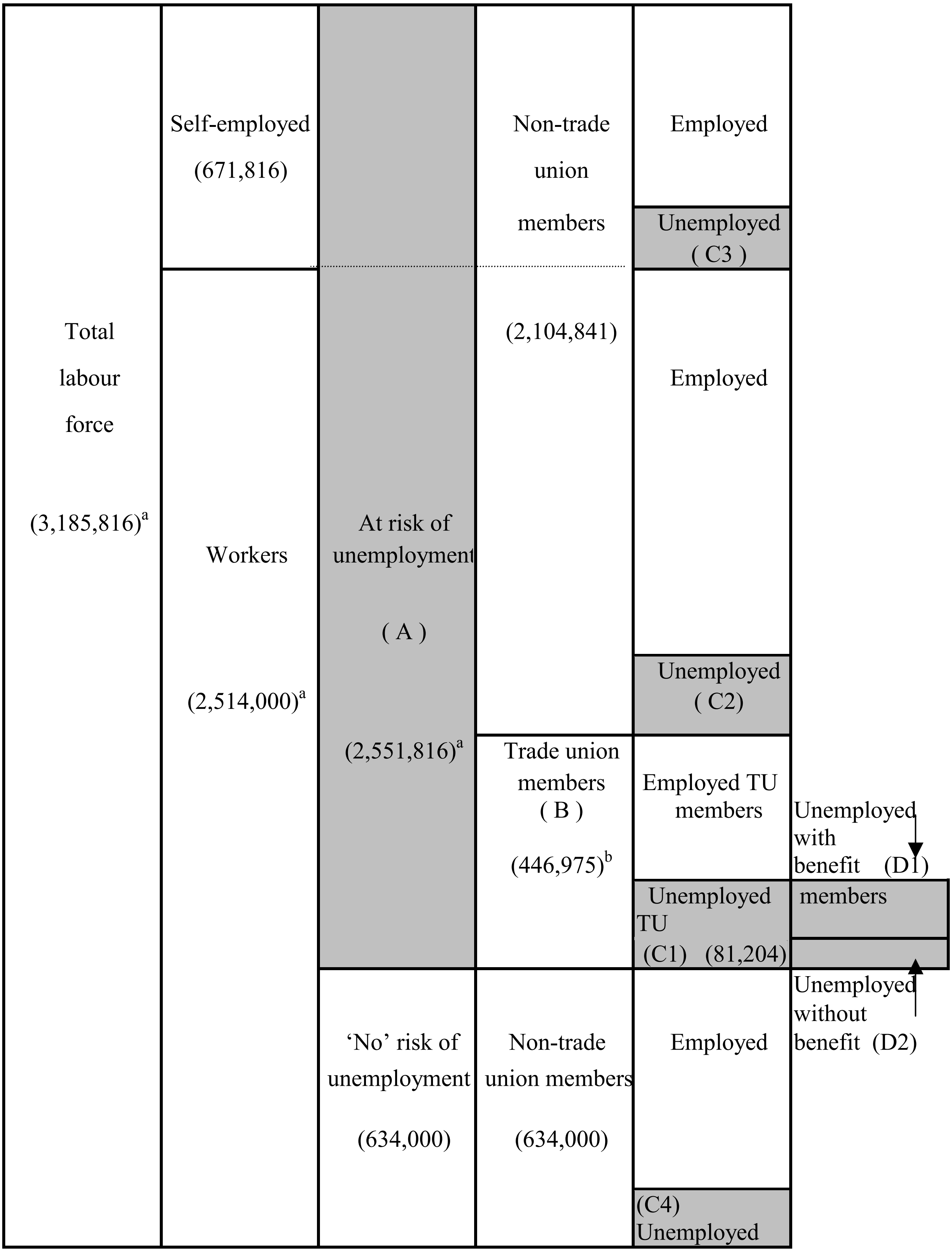


Figure 2.2: Composition of the Dutch labour force, 31 December 1930

[a] Kloosterman, 1985: 24, based on the 1930 census.
[b] CBS Maandschrift, February 1931, p.356; numbers are averages over the month of December.

On this basis, the CBS considered the unemployment among trade union members to be representative: “This investigation would suggest that the unemployment figures are applicable for all workers” (Velthuisen, 1935: 359).[22]

In reality, the CBS took a more sophisticated stance. In order to give the INU an appropriate meaning it was interpreted as *the unemployment rate of workers at risk of unemployment*. Some groups of workers in the labour force, such as civil servants, railway employees, domestic servants, teachers, medical and nursing employees and priests, had substantial job security. They faced ‘no’ (or very little) risk of unemployment as a result of recurrent economic crisis. Therefore, there was no need to insure themselves against the risk of unemployment. As a consequence, these groups were not represented in the INU and this index number of unemployment was interpreted as only being representative of the workers at risk of unemployment.

By using CBS unemployment data, trade union membership data and data of the ten-yearly census of 31 December 1930, we can analyse in closer detail what the TU SQR did define and measure as unemployment (see Figure 2.2). According to the 1930 census, on the 31st of December 1930, the total labour force (TLF) consisted of 3,185,816 persons, of which around 2,514,000 were on the payroll. The remaining 671,816 were self-employed. Of the labour force, approximately 634,000 persons were not exposed to risk of unemployment, as described above. For this reason, this group was left out of the analysis by the CBS. The group at risk of unemployment (group A) had good reasons to insure themselves against unemployment. This group amounted to 2,551,816 persons. Among them were the self-employed. However, these latter had no trade unions and could therefore not insure themselves as workers could, and were consequently not covered by the INU. However, in the analysis by the CBS, this group of self-employed was ignored, most likely because measurement of the size of this group was very difficult too at the time. Only a small proportion of the workers exposed to risk of unemployment were members of a trade union: 446,975 persons (group B). This corresponds to an organization rate of 14% of the total labour force (B/TLF x 100). Unorganized and at risk of unemployment were 2,104,841 persons and their employment status was unknown. The TU SQR counted unemployed trade union members: group C1. On 31 December 1930 this was 81,204 persons. A sizeable fraction of unemployed union members, however, were not entitled to benefit since their entitlement had expired (group D2). The size of this is group was not counted, but had to be estimated by the trade union in order to meet the requirements of the DWA. An example will illustrate that this latter group was likely the largest part of unemployed union members. In 1923,

[22] My translation.

for example, the average number of unemployed persons per week who received benefits was 15,600, while the number of unemployed trade union members without benefits was 22,000. The number of days over which benefits were paid amounted to 72,000. The number of unpaid days of unemployment was 124,800 (Rijksdienst der Werkloosheidsverzekering en Arbeidsbemiddeling, 1925: 5). Of course, estimation of the number of unpaid unemployed members made the statistics less reliable, but there are good reasons to assume that the trade unions were pretty well informed about these members. Usually the trade unions offered certain facilities to unemployed members without benefits, such as employment assistance and partial or complete exemption of contribution payment. The unpaid unemployed therefore had an incentive to register with trade unions as being out of work, even when they were not considered for benefit payment, which made the estimations more reliable.

Figure 2.2 illustrates which part of the total labour force was assumed to be represented by the unemployed union members. It turns out that the number of unemployed union members (small grey area C1), consisting of both registered unemployed with a benefit (D1) and estimated unemployed without a benefit (D2), was taken as a sample of all unemployed at risk (= C1 + C2 + C3). The union members (B) were taken to represent all the workers at risk of unemployment (large grey area A). As already explained, the absence of the self-employed in the sample seems to be ignored in this consideration and also the possibility that workers with ‘no’ risk of unemployment occasionally became unemployed (area C4). They were not present in the sample group. In short, the ratio C1/B, which the trade union statistics of unemployment represent, was assumed to correspond with the ratio (C1 + C2 + C3)/A. And, while D2 was a correct figure, D1 was estimated, so that C1 was itself partly an estimated figure.

<u>3. No standardized sampling</u>

Though the TU SQR became standardized to a high degree (after 1917), the measurement procedure did not yield a standardized sample. The problem was that the characteristics of the sample changed over time, not only as a consequence of variability in the phenomenon (unemployment) but also because of changes in the sample size. It is often argued that, when the sample size of union members increases, there will be an upward trend in union unemployment percentages. Skilled craftsmen usually are the first to organize in a union. So, they are disproportionately represented in the early years. Since they are usually less subject to the risk of unemployment, there is an upward trend in unemployment when the union membership expands, and more (unskilled) workers become organized in unions. Thus, not only were C1 and B (in Figure 2.2) changing (the fraction of unemployed trade union members), but also B relative to A (the fraction of trade union members representing workers at risk of unemployment), so that the ratio C1/B changed as did B/A.

This is different problem from that of a selective sample: it is a problem of precision (see appendix 2.A). The TU SQR suffered from a measurement error due to a disturbing factor it could not control for (sample size). This problem became apparent in the period 1917-1932 when, as can be seen from Figure 2.3, there was a rapid expansion of union membership for the periods 1917–1921 and 1929-1932.

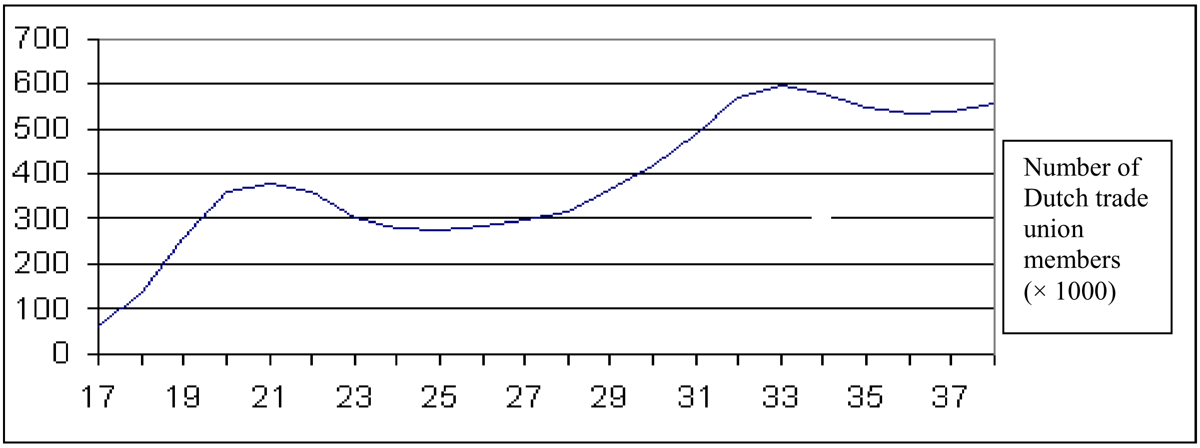


Figure 2.3: Dutch trade union membership, 1917 - 1938
Source: Based on Rijksdienst der werkloosheidsverzekering en arbeidsbemiddeling, *Jaarverslag 1938 (Annual Report 1938),* Algemeene Landsdrukkerij, Den Haag, p.16

The problem of the unstandardized sample/population device was not addressed explicitly in publications by the CBS. It is not quite clear whether it was unknown or ignored, as the CBS was likely to be unable to standardize the bureaucratic procedures from which numbers were drawn such that this problem could be solved.

2.4 The SQR of Labour Exchange data

Measurement of unemployment in the Netherlands involved the use of another bureaucratic recording device for registering unemployment: the SQR of labour exchange data. This SQR (LEx SQR) is based on the administrative procedures at labour exchanges for registering and allocating workers to jobs and on the process of transforming this raw data into a statistic of unemployment called the 'Statistic of Unemployment and Employment Assistance'. At the turn of the 20th century, the vast majority of the workers had to look for work by themselves; this activity was not mediated in any sense at all. Only a small fraction of workers found a job by a form of employment assistance. For the period of analysis, 1900-1940, five forms of employment assistance can be distinguished (see Appendix 2.B), of which public employment assistance through public labour exchanges became the most successful and the foundation of the LEx SQR. The period 1900-1940 can be subdivided into three periods, in each of which the LEx SQR and the corresponding interpretation of unemployment is roughly unchanged. These periods are 1902-1917, 1917-1930 and 1930-1940.

<u>1902-1917</u>

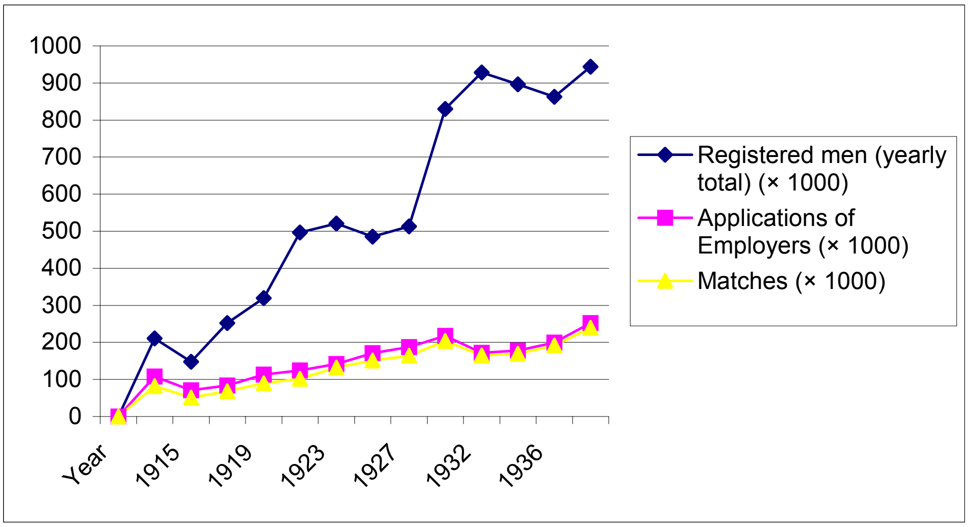
In 1902, a few years after the establishment of the first municipal labour exchange (in 1898 in Amsterdam), the CBS started to collect data of unemployment assistance. The CBS requested labour exchanges to supply data about the number of registered workers and they usually complied. Newly-established municipal labour exchanges were contacted and also requested to provide data. From 1902 till 1917, the provision of data was voluntary and the labour exchanges were free to supply the data in the form they wanted. The labour exchange figures were simply copied from the reporting exchanges and published for each municipal labour exchange separately in the CBS monthly *Tijdschrift* (later named *Maandschrift*). As the number of exchanges increased, it became increasingly inconvenient to publish the data of each exchange separately, and the need was felt to unify the data. The foundation of the *Vereeninging van Nederlandsche Arbeidsbeurzen* [Association of Dutch Labour Exchanges] in 1908 is therefore of importance. The CBS conferred with them in 1910 in order to unify the data. This resulted in the introduction of a new paper form for collecting data in 1916, to be used by the labour exchanges.

Clearly the labour exchanges were in their infancy during the period 1900-1917, and, consequently, the LEx SQR was as well. The labour exchanges were operating only locally, and only a very small fraction of the unemployed working population was registered. The major problem of using the bureaucracy of labour exchanges as a measuring device was that registering was voluntary. This had several consequences both for the concept of unemployment and for the procedure. Though attempts were made to standardize the reporting paper forms of the labour exchanges, the LEx SQR provided an unstandardized quantitative rule in the period 1900-1917.

<u>1917-1930</u>

The 1909 State Commission Treub – which had to advise the central government with regard to unemployment – observed that the public labour exchanges functioned well, but considered only local employment assistance as restrictive. In its final report of 1914, the Commission concluded that involvement of the national government in employment assistance was desirable for achieving an unobstructed functioning of the labour market, as well as for the protection of workers from the abuse of private employment assistance (see Appendix 2.B). In addition, the Commission suggested a greater coverage of support and the establishment of a national network of labour exchanges. In 1916, the Netherlands was divided into 30 districts with exchanges in major places and agencies in minor ones.[23] The *Centrale Arbeidsbeurs* [Central Labour Exchange], established in 1914, became the obvious body to coordinate national employment assistance.

A.



B.

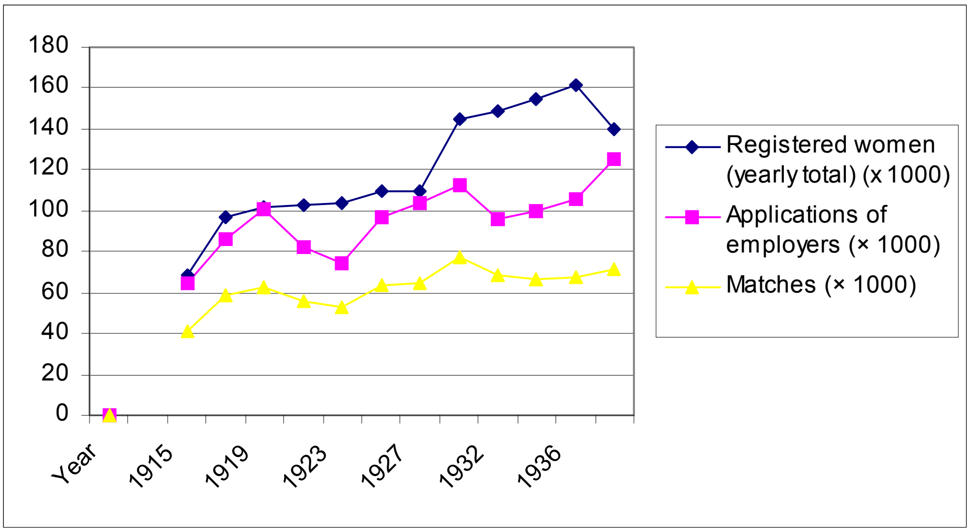


Figure 2.4: Development of registration and applications at Dutch labour exchanges for male workers (panel A), and female workers (panel B), 1914-1938.
Based on: W.L.P.M. de Kort (1940), *De arbeidsbemiddeling in Nederland*, N.Samson N.V., Alphen aan den Rijn, p. 284.

[23] The statutory difference between a labour exchange and an agency is that in a labour exchange a director was in charge.

The Unemployment Resolution 1917, in which subsidies for unemployed union members were arranged and where the *Rijksdienst der Werkloosheidsverzekering en Arbeidsbemiddeling* (DWA) [National Service for Unemployment Insurance and Employment Exchange] was established, had a clause which stated that subsidized unemployed union members had to register at a labour exchange in order to receive benefit. As a result, the number of yearly registrations rose (see Figure 2.4). Also more insight was obtained into the spread of unemployment across the country.

After the approval of the Unemployment Resolution 1917, the LEx SQR changed substantially. The government became involved with employment assistance (basically on economic grounds and because of the success of municipal labour exchanges), and a national network of both public municipal and non-public labour exchanges was set up. In this period, public employment assistance became by far the dominant way of employment assistance. Private employment assistance was disapproved of on ethical grounds by the government,[24] and government policy with respect to private employment assistance was to encourage its gradual disappearance by limiting concessions and by its absorption into other labour exchanges. The most important change in this period – which affected both the TU SQR and the LEx SQR – was that unemployed trade union members were forced to register at labour exchanges in the event of unemployment, in order to claim benefit payments. Most unions had accepted the standard regulation developed by the DWA which involved registration at a labour exchange as evidence of being unemployed, in order to receive benefit. As a consequence, the LEx SQR became more standardized, but, unfortunately, the resolution applied to only a small fraction of the workers: the trade union members. Most workers (around 85%) were simply not members of a union. So, those registered were either unemployed union members or voluntary job seekers. And though the registration of the unemployed was improved and a record was made of registered workers who were still employed, the procedure was still not fully standardized for all workers and some occupations were still more or less absent in the statistics; in particular, those with low union membership, such as agricultural workers, working family members, and the self employed.

Furthermore, the data supplied by the agencies were incomplete, in the sense that only total numbers of registered persons was given, and no classification of the data per occupation or industry. However, it can be concluded that the registration of unemployment became an important element of the function of labour exchanges and the collection of data was set up much more systematic than in the previous period as the country was divided into 30 districts, which enabled a national coverage. The role of the labour exchanges had also changed.

[24] Because of the social abuse involved in private unemployment assistance (see Appendix 2.B)

Mediating between supply and demand of labour was no longer the only important role of the labour exchanges: the compulsory provision of data to the state made registration an end in itself and an important side aspect of the function of labour exchanges.

1930-1940

In order to fight the economic recession of the 1930s, the *Arbeidsbemiddelingswet 1930* [Employment Exchange Act 1930] was passed in 1930. This law, which came into effect on 1 January 1932, involved the establishment of a legal monopoly for public employment assistance and a restricted permit system for private employment assistance. Registration duty for the unemployed was expanded, and it became a legal obligation for every municipality in the Netherlands to establish a labour exchange or an agency. Registration at labour exchanges became compulsory now for four categories of workers:

1. Unemployed trade union members on benefit, (according to Unemployment Resolution 1917).
2. Non-trade union members, who were considered eligible for unemployment benefit payment via private unemployment insurance and, in some municipalities, unemployed workers supported by assistance provided for the poor.
3. Workers involved in unemployment relief work.
4. And, from 21 June 1935 on, family members of the above mentioned categories, provided that they were capable of working, i.e. that they were more than 15 years old.[25]

As a result of this law, the number of labour exchanges or agencies grew to 1064 in 1935 and the number of registered unemployed increased drastically, as can be seen from Figure 2.4. With this measure, employment assistance had developed into a structural anti-unemployment measure. Throughout the year 1932, for example, more than a quarter of all Dutch workers were at some point registered at a labour exchange.

When on 1 January 1933 the CBS, instead of the DWA, again became responsible for producing the registered unemployment statistics,[26] the CBS immediately extended the statistics of registered unemployment by introducing a classification per occupation, which could be presented for the country as whole in 1936, since from that year agencies also supplied data

[25] Though exceptions were permitted, when for example, family members were indispensable in the household.
[26] In 1924, the CBS was confronted with severe budget cutbacks and the responsibility for the compilation of the statistic of unemployment and employment assistance was therefore transferred to the DWA. At the outbreak of the economic crises of the 1930s, the budget cuts were cancelled as the government was in need of reliable unemployment figures. From the 1 January 1933, the CBS again became responsible for the statistic of unemployment and employment assistance.

classified per occupational group. On the proposal of the CBS, it was arranged in the Employment Assistance Law that all public labour exchanges must count the registered unemployed once a month (on the last working day).

Though, as a consequence of the Employment Assistance Law 1930 and the economic crisis of the 1930s, the number of workers who registered at labour exchanges increased sharply, the LEx SQR was still largely based on the voluntary registration of the large majority of workers. They had low incentives to register as unemployed (and even less to report as employed again). They were not compelled to register, and some groups of workers, such as agricultural workers, had to travel great distances to labour exchanges to register. Since they also had a low union membership rate (and therefore a low financial incentive to register), they were therefore very incompletely registered. Civil servants of the DWA estimated the mediation by labour exchanges for agricultural workers even in 1939 at most at 2% (Kort, 1940: 329). Nevertheless, by 1940, the statistic of Unemployment and Employment Assistance was considered by the CBS to give a "complete" view of unemployment in 1940.[27]

	1900-1917	**1917–1930**	**1930–1940**
Collection of data	non-systematic	systematic	systematic
Reason for labour exchanges	social	social / economic	economic
Bureaucracy	municipal	government	government
Coverage of labour exchanges	local	regional / national	regional / national
Extent of employment assistance	small	considerable	large
Co-operation in data collection	voluntary	compulsory	compulsory
Responsibility for data collection	CBS	DWA (1924-1933)	CBS (1933-1940)
Basis of count	voluntary	voluntary + compulsory	voluntary + compulsory
Processing of data	'unprocessed'	aggregation	aggregation + classification
Role of labour exchange	mediation	mediation / registration	mediation / registration

Table 2.2: Development of the LEx SQR

During the Second World War, the German occupiers transformed the system of labour exchanges in order to employ Dutch workers for the German war industry. For that purpose on

[27] Anonymous and unpublished internal CBS document *Statistiek Arbeidsbemiddeling en werkloosheid,* CBS archive Map 408, 13.

the 5 October 1940, the Germans established the *Rijksarbeidsbureau* [State Labour Exchange], which consisted of a head office in The Hague and 37 regional labour exchanges. However, since many of the unemployed withdrew from registration because of fear of being deported to Germany, the statistic became more and more unreliable. In 1945, the last year of the war, the CBS suspended all activities. These restrictions had, however, no lasting effect and the statistic was re-established after the war.[28]

Table 2.2 summarizes the major changes in the characteristics of the LEx SQR and the corresponding measures of unemployment.

2.5 The measurement problems of the LEx SQR

Like the TU SQR, the LEx SQR suffered from measurement problems. It had two problems, one with accuracy and one with precision.

1. Standardization of measurement procedure

The first problem was lack of standardization of the reporting procedure, which resulted in a problem with the precision of measurement. Until 1917, registration at labour exchanges was entirely based on voluntariness. But incentives to report at labour exchanges were low for large groups of workers, and the measuring procedure was not fully standardized, thus rendering the LEx SQR an unstandardized quantitative rule. Galenson and Zellner (1957) argue that the usefulness of this sort of data is limited: “When registration is voluntary, employment exchange data are of much more limited value. In such cases, rates of unemployment calculated from them cannot be compared internationally; they can only be used to measure differences in trend from a common base year for which comparative rates of unemployment are available from other sources” (Galenson and Zellner, 1957: 542).

Of importance for the standardization of the LEx SQR is the standardization of the reporting paper form and the Unemployment Resolution 1917. Though, after 1917, the trade union members were forced to register in order to receive benefit, the standardization of the LEx SQR improved only a little, since trade union participation was low (around 10-15%) and the vast majority remained free to choose to register. In addition, it caused a selective population to be measured, as the trade union members became overrepresented in the labour exchange statistics. This point will be discussed in full below.

[28] Basically, the method of measuring registered unemployment remained the same until the 1970s, when surveys were introduced as an additional measurement method.

2. Selective conceptualization of unemployment

The LEx SQR also had a problem with the accuracy of the measurement procedure and in the way it indirectly conceptualized unemployment. In the first place, there were those who registered as job seekers and not as unemployed per se. Also employed workers looking for another job could register themselves at labour exchanges. The procedures in the labour exchanges were, however, aimed at matching supply and demand for labour and were not aimed at distinguishing between those with a job and those without. This distinction was made only later in the 1930s by a separate registration of the 'employed registered'. Secondly, as with the TU SQR, specific groups of workers were more represented than others. Typical workers who registered voluntarily at labour exchanges were manual workers, such as servants, maids, shop assistants, cleaners and tailors. But others, however, were more or less absent. Self-employed, agricultural workers, working family members, and usually women and children did not register and were therefore excluded in the count of unemployment by means of the LEx SQR. Married woman were not allowed to register when their husband was employed. The share of women in registered unemployment was therefore very low.[29] And, finally, persons who were partially unemployed were, in general, excluded from the count of unemployment. A person without a labour contract had to be able and willing to work for a full day in order to be included. A group that was included in the LEx SQR but excluded in the TU SQR were young recent graduates seeking their first employment; they were unemployed and could as such not join a union. It will be clear that, as in the case of the TU SQR, some groups of workers were excluded more or less systematically in the LEx SQR, even though registration was voluntary and free.

In order to get a better impression of the accuracy of the labour exchange unemployment figures, we will analyse them in closer detail for the 31 December 1930 when the ten-yearly census was held, and combine the census data with CBS figures of registered unemployment (see Figure 2.5). It turns out that of the total labour force of 3,185,816 only 156,221 people were registered as unemployed at labour exchanges (group A in Figure 2.5). Of this group, 19,993 persons (12.8% of the registered) had a paid job. Therefore, the number of really jobless was 136,228 persons (group B). Though the LEx SQR only provided absolute numbers, we are now able to calculate a LEx unemployment rate (LEx UR) for the 31 December 1930, if we define it as:

$$\text{LEx UR} = \frac{\text{registered LEx unemployment}}{\text{total labour force}} \times 100\ \% = \frac{B}{TLF} = \frac{136{,}228}{3{,}185{,}816} \times 100\ \% = 4.9\ \%.$$

[29] Until the Second World War, the share of registered women remained very small. On the 31 December 1930, for example, only 9,604 women were registered against 146,617 men: only 6.1 percent. Usual jobs for registered women were housemaid, tailoress, seamstress, maidservant, shop assistant, nanny, or cleaner.

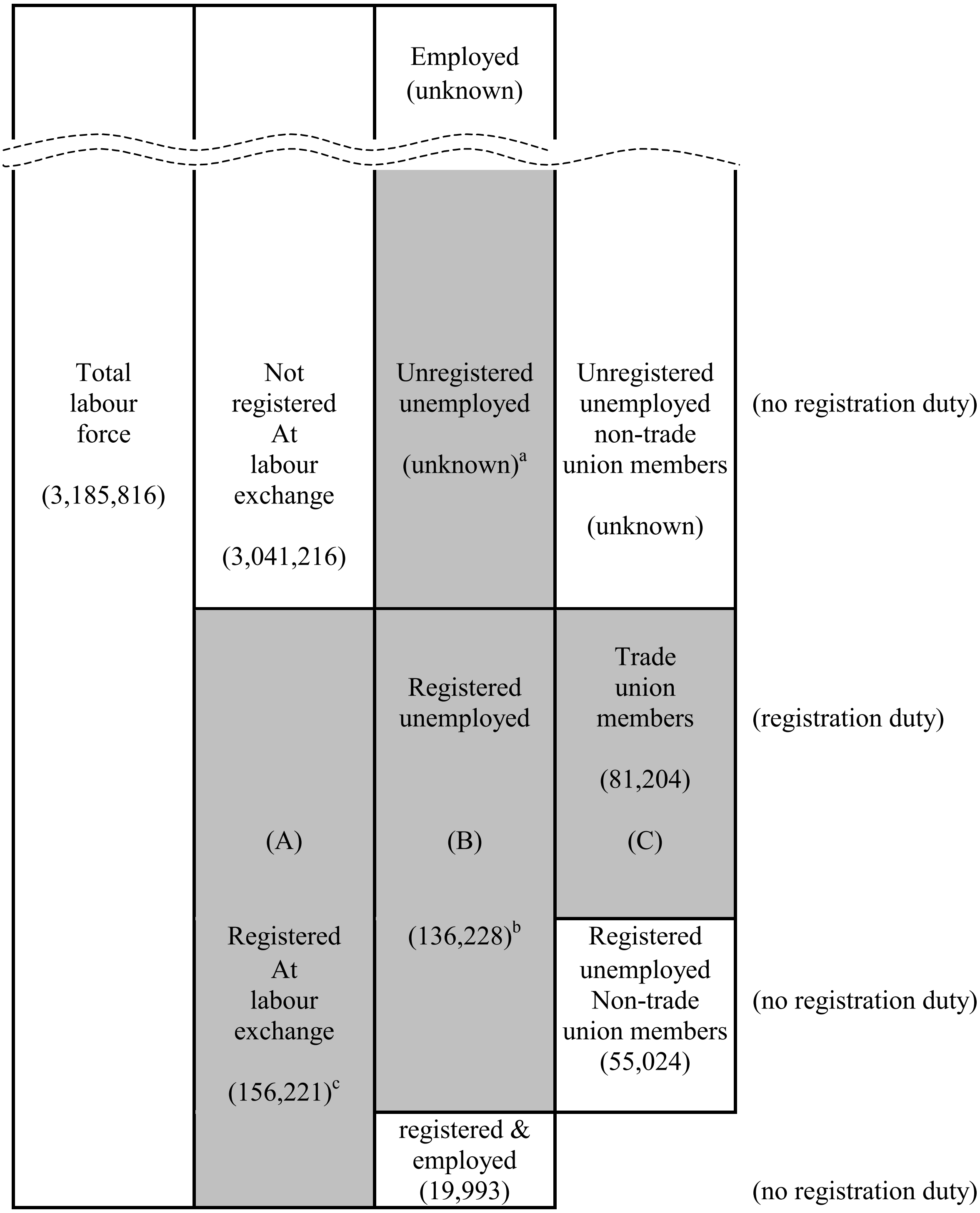


Figure 2.5: Composition of the labour exchange unemployment in the Netherlands on 31 December 1930.[30]

[a] The trade union statistics suggest that this number should be at least 4 times the number of registered unemployed.
[b] CBS Maandschrift, February 1931, p. 362.
[c] CBS Maandschrift, January 1931, p.118.
[30] The areas correspond with the relative size of the different groups.

From CBS data, it is known that, for that date, the TU SQR (see Section 2.2) gave a PU of 18.2 % and an INU of 15.3 %, and that 81,204 trade union members (group C) were unemployed and received a benefit (CBS, February 1931, 356-367). Unemployed trade union members had a registration duty, in order to receive a benefit, and had to report daily at the insurance fund.[31] In return, the unemployed worker got a stamp on his social insurance card (Article 42). This was the well-known 'stamping'. Under the assumption that all unemployed trade union members registered as such at the labour exchanges (since this was a necessary condition for receiving benefit), we find that they form 52.0% of the registered unemployed (the ratio C/A in Figure 2.5). The remaining 55,024 (35.2 % of total registered) thus consists of non-trade union members. It seems, therefore, that the majority of the registered unemployed were members of a trade union since it was compulsory for them to register and voluntary for non-trade union members. Throughout the 1930s, the participation rate of trade unions was between 10-15% (Kloosterman, 1985), and the TU SQR indicates a participation rate of 14 % for 31 December 1930. This means that 14 % of all workers accounted for more than half (52%) of the number of registered workers at labour exchanges. Trade union members were therefore overrepresented in the LEx SQR.

2.6 Conclusions

When the Central Bureau for Statistics (CBS) was asked in the late 19th century to quantify unemployment, no operational definition of unemployment or methods for registering or counting unemployment were known to the CBS. Unemployment was unclear at a conceptual and an epistemic level. In terms of the diagrammatic presentation of the representational theory of measurement (see Figure 2.1), this means that both the conceptualization of unemployment *Q*' and ways to establish the mapping M_2 were unknown. So, the CBS or central government had to develop either an operational definition of unemployment or a method for establishing the mapping M_2. The problem was that, in those days, for many professions, such as day workers, farmhands, dock workers, etc, being temporarily or seasonally out of work was accepted as inherent to the job, and the CBS considered distinguishing the 'unemployed' from these 'temporarily out of work', an impossible task. The CBS did not know how to operationalize the concept of unemployment and resorted instead to bureaucracies that handled administrative data of some kind of unemployment: the trade unions and the labour exchanges.

[31] Article 41 of the Unemployment Resolution 1917 mentions: "The unemployed members are obliged to put their name on a reporting list daily, on a point in time determined by the fund committee" (my translation).

As a result, two mappings M_2 were brought about by two SQRs. In the early period (till 1917), both SQRs were only locally standardized and provided only local knowledge of unemployment. In addition, both SQRs conceptualized unemployment differently. The TU SQR defined the unemployed more or less as 'trade union members on welfare', while the LEx SQR defined unemployed roughly as 'voluntary registered job seekers'. Quantitative rules, however, are not very useful for the measurement of concepts when they are only locally standardized, and hence produce only locally valid numbers. It is necessary to have concepts that have a wide validity and fully standardized rules for measuring them. Porter stresses the importance of the standardization of public measurement systems, when he argues that: "Standardization and proper surveillance are in some way more important to a public measurement systems than is close approximation to true values as defined by elite research laboratories" (Porter 1994: 391). Standardization of both SQRs did not take place until the central government became involved in both unemployment insurance and employment assistance by its acceptance of the Unemployment Resolution 1917. The government attached conditions to the transfer of subsidies to trade unions and established other bureaucracies like the DWA, which helped both SQRs to serve as public measuring systems. Unemployed trade union members were forced to register at labour exchanges in order to receive benefit and, consequently, the characteristics of the registered unemployed changed. For the 31 December 1930, for example, it seems that the majority (about 52%) of the registered unemployed workers at labour exchanges were also included in the sample of the TU SQR. So, as a result of the government interference, both SQRs became tightly connected and more standardized, and the population that the two SQRs measured overlapped to an increasing degree.

When comparing both SQRs (see table 2.3) it can be seen that both SQRs had their own measurement problems. The TU SQR had a problem with both precision and accuracy. First, sampling by the TU SQR was not stable. The characteristics of the sample changed over time not only as a consequence of variability in the phenomenon (unemployment) but also because of changes in the sample size. It was not clear which phenomenon was being captured by this measurement procedure. Therefore, the outcome of the measurement procedure was not precise. Secondly, there were worries about the representiveness of trade union members for the total labour force. This is a problem of accuracy. In its turn the LEx SQR was perhaps not a fullystandardized quantitative rule, since no fully standardized, bureaucratic procedure was set up such that all workers had an incentive or duty to report at labour exchanges. Most workers simply felt outside the scope of the policy measures taken. Nevertheless, it should be noted that public employment assistance was open to all workers and that, particularly during the mass

	TU SQR	LEx SQR
Bureaucracies	TU, DWA, CBS Municipal unempl. funds (1906-1917)	Labour exchanges, CBS, DWA (1924-1933)
Quantitative rule	* Bookkeeping of unemployment benefits by TU's * Administering subsidy by government * Reporting quantities to CBS / DWA * Classification into occupational and industry subgroups * Calculation of index numbers	* Registering supply and demand of labour * Reporting quantities to CBS * Classification into occupational and industry subgroups
Standardization	* High degree of standardization due to standardization TU regulations * Strong financial incentive for getting registered	* Moderate degree of standardization * Registration based on voluntaryness * Low direct incentives for getting registered * Standardization of reporting paper form to CBS
Implicit definition	Trade union members on welfare	Voluntary registered job seekers
Unemployment rate 31 December 1930	18.2 % (PU) 15.3 % (INU)	4.9 %

Table 2.3: Comparison of the main characteristics of the TU and LEx SQR

unemployment of the 1930s, employment assistance became a structural anti-unemployment measure used by large groups of unemployed. In this way, it avoids the problem of biased measurement (accuracy). In addition, the concept that the LEx SQR measured corresponds better to our contemporary understanding of unemployment. In this way, it could be argued that the LEx SQR was more accurate. However, the measurement results – an unemployment rate of 4.9% at the height of the Great Depression – seem very unlikely and indicate a serious problem with the precision of the LEx SQR.

In spite of these measurement problems, and as a consequence of the process of standardization (to a higher degree) of the two SQRs by government involvement, unemployment became not only measurable but also gained validity as a concept. According to Porter, being made measurable with an SQR implies that entities gain empirical significance and as a consequence, become more widely used. Porter gives the example of standardized tests for predicting success in schools, which became increasingly valid beyond the walls of educational institutions (Porter, 1994: 403). In this way, the measures gained empirical meaning, even though these tests are considered less useful outside schools. This also happened to the measurement of unemployment in the Netherlands. The CBS was always careful not to fall into the operationalist trap of defining abstract social science concepts solely as operationally measurable variables, and mentioned explicitly that the statistics concerned not a generic concept of unemployment, but specifically "unemployment of workers at risk of unemployment" or "unemployment among members of trade unions with an insurance scheme" (TU SQR), or "unemployed registered at labour exchanges" (LEx SQR). But, in practice, that is, in social and political discourse, the figures were taken as *the* indicators of unemployment in the Netherlands, and, as such, they became widely used.

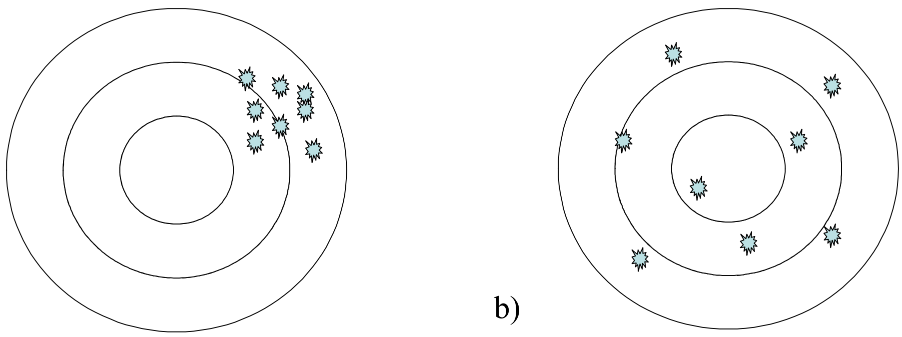
Both SQRs thus indeed created new meaning and took, in spite of measurement problems, the meaning of the measures beyond the limits of their original meaning: trade union and labour exchange administrative registers. However, this was also exactly the problem. In a comparison of both SQRs for the 31 December 1930 – when the economic crisis already was at a high level – we find three completely different unemployment rates from these SQRs: an Index Number of Unemployment (INU) of 15.3%, a Percentage of Unemployment (PU) of 18.2 % for the TU SQR, and a calculated unemployment rate of 4.9% for the LEx SQR. But, what is the empirical significance when we end up with three different measures of the same phenomenon? Both SQRs were standardized to a fairly high degree, seemingly satisfying Porter's requirements for measurement, but yet do not lead to fully satisfactory measurement. Now where does this take us?

For an operationalist, "a concept is nothing more than a set of operations; the concept is synonymous with the corresponding set of operations" (Bridgman 1927: 5). Three different sets of operations for the measurement of an unemployment rate would, for an operationalist, therefore imply, three different concepts of unemployment. Though Porter does not fully embrace this radical stance, his account of measurement bears an operationalists' flavour when he argues that: "there is a strong incentive to prefer readily standardizable measures to highly accurate ones, where these ideals are in conflict" Porter (1994: 391). Clearly, precision is preferred over accuracy. In this case of measurement, we do have more or less standardized measurement, but still feel uncomfortable as to the figures it produces. The question of the accuracy of quantification cannot, therefore, be ignored altogether. Throughout the whole period before the Second World War, the CBS questioned the representativeness of the TU SQR, that is, the accuracy of its measures. We must also make sure that we measure what we intend to measure. And, in the case of the TU SQR, this was not exactly the case. The sample group was clearly a selective sample, which gave rise to doubts regarding the accuracy of the TU SQR. Standardization of the measurement procedure alone is clearly not enough for measurement. However, the CBS was not able to deal with this problem. But, since the CBS also was not able to conceptualize unemployment as a stable entity, measurement would have led nowhere. What is required for measurement is that stable entities or concepts are constructed in the first place, and the SQRs helped to achieve this.

Appendix 2.A: Accuracy vs. Precision

When talking about measurement it is important to make a distinction between the *accuracy* of a measurement and the *precision* of a measurement. Measurement is said to be accurate when the measured value is close to the *true value* (or standard).[32] Precision of measurement, on the other hand, refers to the spread of the measurement errors. It refers to how close the measurement results are to *each other*, and not to a true value. Measurement is said to be precise when the spread is as small as possible.[33]

The distinction between accuracy and precision is often explained by a shooting target analogy. In Figure 2.6a the measures (hits) are precise but not accurate (unbiased). In Figure 2.6b the measures are accurate but not precise.



a) Precise but inaccurate

b) Accurate but imprecise

Figure 2.6: Accuracy and precision of measurement

This distinction between accuracy and precision enables us to analyse measurement errors more systematically (see Figure 2.7). From the above, it follows that a difference between a true value and a measured value can be attributed to two different types of measurement errors, either a lack of precision or a lack of accuracy. When measurement is accurate but imprecise (panel b), the measurement process is unbiased and the errors follow from factors we cannot fully control for, such as uncontrollable 'random' errors, uncontrollable 'systematic' errors, human or personal errors (misreading of gauges, etc), errors in the measuring instrument, and computational errors (see Chapter 1). Consequently, these errors have a stochastic character, and (in experimental sciences) they can be dealt with by appropriate statistical techniques. When measurement is precise but inaccurate, the measurement is biased and the measurement error is due to a systematic error. This is a problem that we – in principle – can control for by calibrating

[32] See, for example, Boumans, Encyclopedia of Social Measurement, (2005b, p. 751-760).

the measuring instrument; that is, bringing about the appropriate, systematic correction.[34] This however, requires knowledge about the true value of the measurand, and unless we are comparing with an (arbitrary) standard, the true value of the measurand is unknown, and calibration is not possible. In economic research, for example, accuracy is assumed in the specification of structural models. The relations these models describe are assumed to hold, ceteris paribus, and are affected by (other than the causal factors involved) random errors with zero means only (and are thus accurate by definition). Consequently, improving measurement concentrates on reducing imprecision (Boumans, 2005b).

Measured value =	*true value* +	(*controllable*) *systematic error* +	*uncontrollable measurement error*
	↓	↓	↓
	unknown	bias	errors in measuring instrument personal errors uncontrollable 'systematic' errors uncontrollable 'random' conditions 'computation' errors
		↓	↓
		accuracy	precision

Figure 2.7: Schematic overview on measurement errors

[33] Ibid.

[34] There are other interpretations of calibration, such as a method of estimation or "tuning", or "testing" a model. See also Boumans (2001); Hansen and Heckman (1996).

Appendix 2.B: Employment assistance 1900-1940

Five forms of employment assistance can be distinguished for the period of analysis, 1900-1940 (Kort: 1940).

1. Private employment assistance

The first form of assistance was private employment assistance with the sole aim of making profit. Private agents were active in small, local parts of the labour market, usually for one or only a few occupations. In 1895, this was the most used method of employment assistance (Kort, 1940: 186).[35] The mediators of labour were often landlords or café owners who kept no registration. After the acceptance of the Emergency Resolution of 19 September 1917 and the establishment of public employment assistance, the share of private employment assistance gradually declined. Private employment assistance, though gradually declining, still continued to exist throughout the 1920s and 1930s.

2. Trade unions and employers organizations

Another form of employment assistance was by trade unions and employers' organizations. In fact, many trade unions at end of the 19th century offered some form of employment assistance, but like private mediators they kept no records, and there was no official bureau for employment assistance. This made their share in employment assistance hard to estimate. Employment assistance was offered in order to make union membership more attractive to workers, but it was also seen by unions as a way to level out wage differences and, more in general, as a means in the class struggle. This form of employment assistance was not, however, open to all; only union members were mediated.

3. Associations of trade unions and employers organizations

Dedicated labour exchange bureaux were established in the late 19th century by associations of trade unions and employers' organizations. They both had a common interest in employment assistance. Both parties acted out self-interest, and no particular social goal was pursued, and, as with private employment assistance and individual trade unions, there was no well-developed registration of workers. Non-public labour exchanges discriminated between union members and non-union members. Some exchanges only mediated for union members and in almost all cases non-union members had to pay a mediation fee. This form of employment assistance therefore

[35] According to an investigation by C.A. Verrijn Stuart.

played a minor role. In 1896, for example, 10 labour exchanges were in existence. However, the share of non-public employment assistance was, and remained, small, as can be seen in Table 2.4.

	Public labour exchanges		Non- public	
Year	Number	Matches	Number	Matches
1903	1	108	5	5,240
1904	1	68	9	5,850
1905	2	473	9	6,567
1906	6	8,609	10	9,063
1907	8	10,664	9	9,146
1908	10	15,602	7	5,388
1909	14	22,419	10	2,038
1910	18	31,925	6	1,666
1911	17	40,374	4	1,775
1912	21	50,782	7	2,947
1913	21	66,082	5	3,724

Table 2.4: Number and matches of municipal and non-municipal labour exchanges in the Netherlands, 1903-1913.
Source: W.L.P.M. de Kort,(1940), *De arbeidsbemiddeling in Nederland*, N.Samson N.V., Alphen aan den Rijn, p. 240.

4. Public employment assistance

The fourth form of employment assistance was by public labour exchanges. The first calls for public assistance go back to newspaper articles in 1885 [36] (Kort, 1940: 231). The underlying idea was to end ignorance about existing vacancies and to stimulate the free working of the market. Others, like C.A. Verrijn Stuart,[37] advocated a governmental role in employment assistance. He emphasized the economic benefits of employment assistance, which in his view could not be attained by employment assistance on moral or philanthropic grounds. In fact, the first public labour exchange bureaux were established, not by the central government, but by the municipalities, mainly those of the major cities, in the first decade of the 20th century, basically as a reaction to the social abuse involved in the private employment assistance. The conditions for the workers involved in private employment assistance were very poor; workers had to pay large mediation fees, or were forced to consume their wage in the café or shop of the mediator;

[36] By civil engineer de Koning, who also published on this topic in *De Economist* in 1885.
[37] Coenraad Alexander Verrijn Stuart (1865-1948) was a leading Dutch economist and was, among other things, chairman of the influential “Nederlandsche Vereeniging voor Staatshuishoudkunde en Statistiek” [Dutch Association for Political Economy and Statistics], editor of the leading Dutch economic journal *De Economist* and, from 1899-1906, the first director of the Central Bureau of Statistics.

mediators lent money at extortionate rates, etc. The first public labour exchange was established in 1898 in Amsterdam and had a tripartite committee of city council, employers and trade unions. Since mediation by municipal labour exchanges was free of social abuse and costless, they soon turned out to be a most successful form of employment assistance. Table 2.4 illustrates the success of the public labour exchanges compared with non-public (dedicated trade union or employers) labour exchanges, both in terms of number of exchanges as well as the number of workers who were successfully mediated: the "matches". Since the municipal labour exchanges were run by civil servants, for whom reporting to their superiors about their activities was part of their work, this meant that much better records of mediated workers were kept by this form of mediation.

5. Charitable institutions

Finally, apart from the four ways of employment assistance described above, there was assistance by charitable institutions (like the Salvation Army). Their aim was to relieve the social burden of unemployment but their share in employment finding was almost negligible.

Chapter 3:

Heidelberger's Correlation Rule and the Measurement of Cyclical Unemployment

3.1 Introduction

The Representational Theory of Measurement (RTM) requires for measurement an isomorphic mapping between a relevant empirical structure and an arithmetical structure. Bringing about this mapping is not easy when the variable of interest is unobservable or in other ways hard to access.[1] For such cases, philosophers of science, like Ellis (1966) and Heidelberger (1994a, b), suggest measuring the unobservable through an observable with which it covaries. In this approach, a more easily accessible variable 'stands in' for the unobservable, and the measurement of one can be reduced to the other since the observable and unobservable are correlated with each other. Brian Ellis (1966) refers to this kind of measurement as *associative* measurement. Michael Heidelberger (1994a, b) provides a more formal treatment of this strategy for measurement, and it is that treatment that we seek to explore in this chapter.

This chapter will discuss Heidelberger's correlation account of measurement, and apply it to the case of measurement of cyclical (or deficient demand) unemployment by unemployment-vacancies (UV) analysis. In the 1960s and 1970s this kind of analysis was used in economics to measure the unobservable and theoretical classes of cyclical, frictional and structural unemployment, which formed up till then the traditional classification of unemployment in the economic literature. In this classification, frictional unemployment is defined as "the amount of

[1] See also Chapter 1 on observables vs. unobservables.

unemployment that corresponds to job vacancies in the same local labour market and occupation" (Macmillan Dictionary of Economics, 1983: 163). Structural unemployment is considered as a form of unemployment that "results from the co-existence but mismatching of unemployed and job vacancies. The unemployed possess neither the right skills nor live in the right places to fill the existing job vacancies" (ibid., 1983: 405). Cyclical unemployment "is short-run demand deficient unemployment" and means that "during certain phases of the business cycle there will be insufficient jobs for the whole of the labour force, no matter how it is trained or deployed" (ibid., 1983: 96). These unobservable classes were made measurable through the use of already available, (and hence 'observable'), unemployment and vacancy data. The method was based on the UV-curve: the inverse relation between the unemployment and vacancies data in an economy.

The chapter proceeds as follows. Section 3.2 will elaborate on the concept of associative measurement and will provide a treatment of Heidelberger's correlation account. Section 3.3 then analyses UV analysis and discusses the main contributions of economists to the development of the UV-curve as a measuring device. Section 3.4 analyses the correlational structure of UV analysis. Finally, in the 1980s, UV analysis was almost completely abandoned. Section 3.5 analyses this sudden fall of UV analysis as a measurement device, and discusses the main methodological drawbacks of UV analysis. Conclusions are drawn in Section 3.6, while Section 3.7 provides an epilogue on the place of the UV-curve in economic theory.

3.2 Associative measurement and Heidelberger's correlation rule

Though Brian Ellis (1966) gave the name *associative* measurement to a particular strategy of measurement, he unfortunately did not make explicit the role of association in the measurement process. Instead, he discusses the example of the thermometer, which is based on association and therefore provides a well-known example of associative measurement. In the measurement of temperature the phenomenon of interest, temperature, is not directly measurable. But it is, however, correlated with the expansion of mercury or other substances in the thermometer, which makes it indirectly measurable. A change in temperature leads to a change in height of the substance in the column of a thermometer. Therefore, temperature is correlated with the height of the substance: *temperature ~ height mercury column* (or other filling). This change in height is, however, idiosyncratic for each and any type of thermometer; it depends on the specific construction, scale of, and substance used in, individual thermometers. Though the thermometer is based upon an underlying causal principle, a correlation between an unobservable and an observable is sufficient for measurement as long as the correlation itself is more or less invariant for each and any type of thermometer (Chang, 2001, 2004) and the measurement device is calibrated. What we thus have to look for is not necessarily laws of nature – as their "claims about what necessarily or reliably happens" are not unproblematic for science (Cartwright, 1983, 1997) – but invariant generalizations, no matter whether they are causal or correlational; a point also stressed by Woodward (2000a).

Michael Heidelberger (1994a) argues the same point as Ellis but explains the role of correlation in measurement. Heidelberger's account draws on the work of Fechner. From its history, Heidelberger distinguishes four different interpretations of the RTM: the classical interpretation; the additive interpretation; the operational interpretation; and the correlative interpretation (Heidelberger 1994a,b). The c*orrelative interpretation*, which Heidelberger favours, was suggested by Fechner but got lost in the course of history. It argues that an unobservable [*uQ*][2] can be measured by "a second, directly observable [and correlated] attribute [*cQ*][3] and a measurement apparatus A that can represent variable values of [*cQ*] in correlation to values of [*u*Q]" (Heidelberger, 1993: 146). The measuring apparatus A thus measures *cQ*, an observable *representative* (Repräsentant), with which the unobservable *uQ* correlates. In this way, *cQ* functions as a replacement or stand-in for the unobservable, and the homomorphic mapping can be brought about merely by correlation. The unobservable might thus be measured through a stable, correlated relationship, $uQ \sim cQ$ (*cQ* is correlated with *uQ*), while *cQ* is measured with apparatus A.

Fechner puts forward the idea of a "measurement formula", that is: "The function that describes the correlation between [*uQ*] and [*cQ*] relative to A (underlying the measurement of [*u*Q] by [*cQ*] in A)" (Heidelberger, 1994: 146). For example, for the measurement of time $\{t\}$, Fechner takes 'distance' $\{d\}$ as a representative, since the time elapsed corresponds to a certain displacement of the hands of a clock or movement of the sun. The relevant measurement formula for the measurement of time (that is, the underlying function of a clock) then becomes $d = kt$ (with k being a constant).

Heidelberger argues that this strategy of measuring an unobservable through an observable variable with which it covariates (associative measurement), as has been succesfully applied to measurement of temperature, does not only hold for the measurement of temperature or time but for *all* measurement:

---

2 "Q" in original.
3 "R" in original.

"..*jede* Messung, nicht nur die Temperaturmessung oder ähnliche andere Fälle, "mit einer gewissen Willkür von besonderen Eigenschaften" abhängig ist. *Jede* Messung beruht auf einer naturgesetzlichen monotonen Kovariation eines Attributs mit einem anderen. Es gibt keine Messung, die nicht von einem Repräsentanten abhängig wäre" (Heidelberger, 1994a: 317).[4]

3.3 The development of UV analysis

In the development of UV analysis, three important contributions will be discussed: Dow and Dicks-Mireaux' seminal 1958 article on the measurement of excess demand in the British economy; Robert Solow's use of the UV framework; and further contributions of the National Institute of Economic and Social Research in London.

3.3.1 Dow and Dicks-Mireaux: the measurement of excess demand

UV analysis originated from the work of two British economists, Dow and Dicks-Mireaux (hereafter DDM) in 1958, at that time associated with the National Institute of Economic and Social Research (NIESR) in London.[5] In their seminal paper, *The Excess Demand for Labour: A Study of Conditions in Great Britain, 1946-1956*,[6] they sought to establish a measure for excess demand, as they were primarily concerned about inflation in the goods market. In the post-war period Keynesianism was the dominant paradigm, and there was a strong belief in aggregate demand control by the government. In the British post-war economy the unemployment rate was very low in the 1950s, around 1.5% on average, and, in such a situation, fluctuations in aggregate demand could easily lead to inflation. DDM therefore sought an indicator that could guide Keynesian fiscal policy in such a way that unemployment could be removed if necessary while avoiding inflation. They suggested using data on vacancies and data on unemployment in order to measure excess demand in the labour market as an indicator for the excess demand in the goods market, since the data on unemployment represents excess supply of labour in the labour market, and that on vacancies represent excess demand for labour. The application of this simple idea was possible for Britain since data for both vacancies and unemployment were available. The collection of (trade union) unemployment data had begun in the 19th century, while the British Government had collected data on vacancies since 1946. By 1958, DDM had available an 11-years' time series of both vacancy and

[4] "… *every* measurement, not only temperature measurement or similar cases 'with some randomness which depends on special properties'. *Every* measurement is based on lawful monotonic covariation of one attribute to another. There is no measurement that does not depend on representatives" (my translation; italics in original)

[5] The National Institute of Economic and Social Research is an independent non-profit organization founded in 1938.

unemployment data for Britain. Data on unfilled vacancies were obtained from notification at labour exchanges. Vacancy data are, however, notorious for their incompleteness for two reasons. First, unlike unemployment, where unemployed workers have a strong financial incentive to register as unemployed, there are no direct financial incentives for firms to notify vacancies or penalties for not reporting vacancies. And, secondly, unlike unemployment, there are no checks on double counting where vacancies are notified at more than one labour exchange, or on undercounting where one vacancy is posted for a number of workers doing the same task. Moreover, firms could be disappointed by the previous mediation of labour exchanges, and look for other ways of hiring workers. William Beveridge (1944: 18) therefore argues that: “the vacancies notified to the exchanges and not filled by them do not necessarily remain unfilled and cannot be taken as a measure of unsatisfied demand; most of them get filled in other ways”. DDM argue, however, that, though there are good reasons to distrust the vacancy numbers, there are also reasons to have a certain confidence in the variation in vacancy numbers. The recording of vacancies might

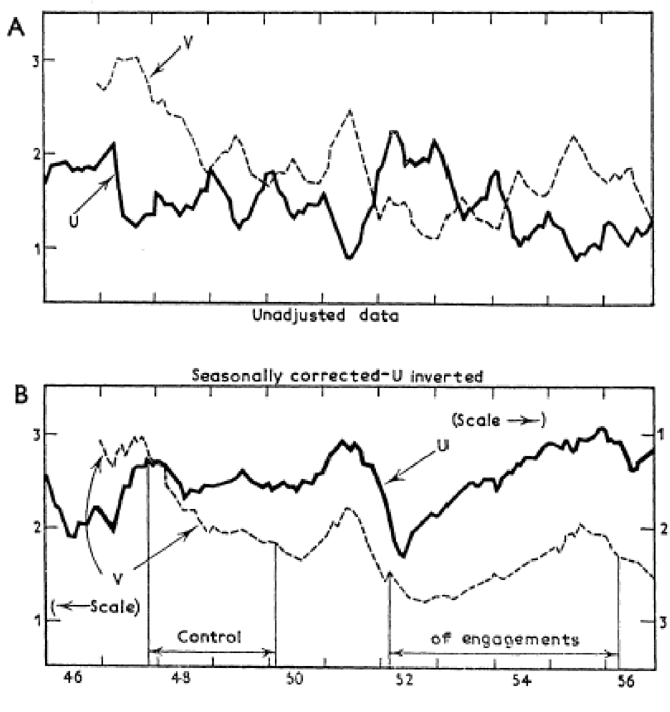


Figure 3.1: Unemployment and vacancy rates, 1946 –1956 (in percentages, Great Britain)
Source: Dow and Dicks-Mireaux, 1958: 3.

[6] A slightly rewritten summary of this article can be found in Dow (1970: 337-343).

be incomplete, but the *behaviour* of vacancies shows that vacancy statistics can be considered as rather reliable indicators. This is especially clear when the behaviour of vacancies is compared with that of unemployment. This is done in Figure 3.1 for British unemployment and vacancy data (in panel B, data are seasonally corrected and the unemployment curve inverted).

The two curves exhibit a remarkable inverse relationship, which leads DDM to conclude: "These observations give one a certain confidence in the vacancy statistics" (DDM, 1958:2). Thus, though vacancy data are probably incomplete, and therefore do not give accurate information about the absolute number of vacancies, they can be used to give an *ordinal* measure of excess demand rather than a cardinal measure. In addition, DDM employed two analyses of unemployment and vacancy data, and found indications that the measurement error of vacancies is fairly stable. First, they compared the effect of seasonal changes on vacancies with the effect of seasonal changes on unemployment. The amplitude of the seasonal variations turned out to be of the same order, suggesting a stable measurement error (DDM, 1958: 26). Second, DDM considered changes in notification of vacancies when the Statutory Regulations for labour exchanges changed. For the period 1946-1956, there were two periods when notification of vacancies to labour exchanges was compulsory in the UK: namely, October 1947-March 1950 and February 1952-April 1956.[7]

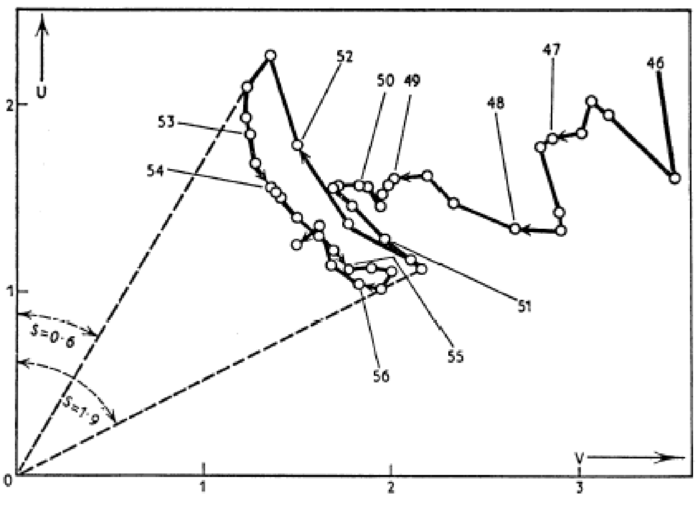


Figure 3.2: Relation between unemployment and vacancy rates (seasonally-corrected quarterly figures)
Source: Dow and Dicks-Mireaux 1958: 4.

[7] These periods are indicated in figure 3.1 as 'Control of engagements'.

However, the change from voluntary to compulsory notification had only a modest effect on the rate of notified vacancies. DDM again interpret this as evidence that the measurement error in measuring vacancies is fairly stable. An important feature that DDM assumed about the behaviour of unemployment is that unemployment above a certain level would be decreasingly sensitive to demand. That is, a further increase in demand should lead to a disproportionately small decline in unemployment rates (and vice versa for vacancies). DDM base that assumption on the rationale that when "demand increases, *u* will decrease continuously (..), but since it cannot shrink below zero it must be supposed to become decreasingly sensitive" (DDM, 1958: 20). DDM point out, however, that it also can be observed empirically in the downswing of 1952 and the boom of 1955. According to DDM, this feature is better observed when the data is presented in Unemployment Vacancy (UV) space and successive observations are connected (Figure 3.2).

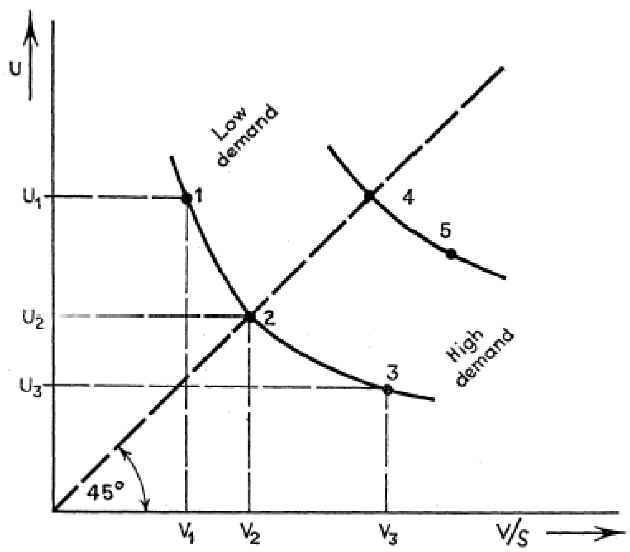


Figure 3.3: Idealized UV relation [where *v/s* = adjusted vacancy rate (see text below)]
Source: Dow and Dicks-Mireaux 1958, 4.

Following this rationale, DDM derive an idealized UV-curve as a rectangular hyperbola (Figure 3.3).[8] The curve shows, first, an inverse relation between vacancy and unemployment rates. When the economy is in recession, it experiences high unemployment rates and low vacancies rates (point

[8] DDM reckon that other curves would be possible, as they argue (about Figure 3.2): "A curve like a rectangular hyperbola seems plausible. Any curve, of whatever shape, which falls from left to right will preserve the correct ranking of observations" (DDM, 1958:22).

1). In an upswing (point 3), the reverse is true: there is a high vacancy rate with a low unemployment rate. Each point on the UV-curve represents a different degree of aggregate demand, and, across the various stages of the business cycle, the economy moves along the UV-curve. Secondly, the hyperbolic and convex shape of the UV-curve represents the feature that a further increase in demand leads to a disproportionately small decline in unemployment rates. This "increasing insensitivity of unemployment", as DDM called it, clearly resembles the neoclassical idea of decreasing returns to input factors found in production and utility functions. Later studies estimate this hyperbolic UV relation as $\log v = \beta_0 + \beta_1 \log(1/u) + \varepsilon$.

The derivation of a hyperbolic-shaped UV-curve is a crucial step in the measurement procedure of excess demand, but this idealized curve does not seem to follow easily from all the observations. The data from 1946 to 1950, for example, do not fit the hyperbolic shaped UV-curve at all (see Figure 3.2). DDM therefore conclude that the observations from 1946 to 1950 do not lie on the UV-curve because of shifts in the level of maladjustment *m* (DDM, 1958:3), i.e. *changes* in the kind of labour supplied and that demanded. DDM attribute the causes of maladjustment to skill mismatch, geographical maldistribution, seasonal variations in demand, and the effect of turnover of labour between firms (DDM, 1958:3). The level of maladjustment *m* is defined as the level of unemployment that equals the adjusted vacancy rate *v/s*.[9] As explained above, the real vacancy rate happens to be subject to measurement error due to under- or overestimation of vacancies. DDM therefore argue that the vacancy rate *v* should be corrected for this measurement error or 'statement error', and that *v/s* instead of *v* should be considered as the true number of vacancies, where *s* is the 'statement error' defined as:[10]

$$s = \frac{\text{unfilled vacancies reported to labour exchanges}}{\text{true number unfilled vacancies}}$$

The level of maladjustment *m* will thus be the level of unemployment, where $u = v/s$ (DDM, 1958: 20), and at this point there no excess demand *d*.

[9] In their seminal article, DDM seem to use different, and conflicting notions of the idea of maladjustment. Whereas they define it as a level of unemployment: namely u_2, they later define maladjustment in accordance with Pythagoras as the square root of *u* times *v/s*, thus as: $m=\sqrt{uv/s}$ (page 22).
[10] Maladjustment is then defined as the amount of unemployment at the point where $u=v/s$ (DDM, 1958: 20).

The convenience of the assumption that the UV-curve is a rectangular hyperbola is that it relates *u, v, s* and *m* in an easy way which makes it possible to derive both excess demand *d* or the level of maladjustment *m*. Given an estimate of *s*, and observations of *u* and *v*, and the corresponding level of maladjustment *m*, then excess demand *d*, which is defined as $d = m - u$, can be calculated. Clearly, (changes in) the value of the statement error *s* could significantly influence the measurement outcomes. For the moment we will assume *s* to be unity.

After deriving an idealized UV-curve, DDM construct a line of zero excess demand. Under the assumption that employers give correct statements about their vacancies (thus conditional on *s*=1), this is an upward sloping line through the origin at a 45º degree angle that separates the areas of excess supply and demand for labour. Each point on that line corresponds to a case where employment equals unfilled vacancies. These two constructs enabled DDM to measure excess demand for labour according to the following general principles (DDM, 1958:5-6):

(i) Vacancies and unemployment rates are plotted in UV space. Vacancies have to be corrected for the degree of over- or underestimating to give the real vacancies rate *v/s*.

(ii) Zero net excess demand is defined as all situations where $u=v/s$, and corresponds to the 45º degree line through the origin.

(iii) Successive points on the 45º degree line correspond to different degrees of maladjustment.[11]

(iv) For any given degree of maladjustment, there will be a series of points corresponding to different degrees of demand and lying on a curve convex to the origin. (Thus, for the degree of maladjustment measured by u_2 (Figure 3.3), there will be points 1, 2, 3…).

(v) Excess demand is measured as vacancies less estimated maladjustment.

Thus, for example, at point 3 on figure 3.3, the excess demand can be measured as $v_3 - v_2$. For situations of net deficient demand, like point 1, deficient demand is measured as estimated maladjustment less unemployment, that is, $u_1 - u_2$. For the year 1956, for example, DDM find that a vacancy rate of 1.7 percent was offset by an excess supply of labour of 1.2 percent, yielding a net excess demand of 0.5 percent. Therefore, vacancies up to 1.2 percent refer to the level of maladjustment, since they can be matched by an equal amount of labour. In the remainder of their

[11] Maladjustment is, however, measured in terms of the corresponding unemployment rate.

article, DDM apply this method to seven sectors of the British economy, and conclude that excess demand for labour in the aggregate was never very considerable in the after-war period 1946-1956.

DDM had some reservations with respect to the precision of their method. In particularly they mention three sources of imprecision, all of which seem to relate to ignorance about the statement error *s*. Lack of knowledge about the 'true' number of vacancies prevents us from knowing this correction factor. First, DDM point out that the statement error affects the net zero demand locus, since it should be redefined as $u = v/s$ when the statement error is taken into account (DDM, 1958:20). When *s* is not unity, the $u = v/s$ locus (or net zero demand locus) will have a slope greater or smaller than 45º. This means that the point of zero excess demand (the intersection of the net zero demand locus and the UV-curve) cannot be identified exactly, without precise knowledge about *s*. Second, DDM note that, apart from changes in demand, changes in the statement error and changes in maladjustment both affect unemployment and unfilled vacancies. However, since the statement error is unknown, it is impossible to distinguish between changes in the statement error and changes in the degree of maladjustment. The observations for 1946 to 1950, for example (in Figure 3.2), were obviously not located on an idealized hyperbolic UV-curve. This indicates, according to DDM, that in that period the British economy was subject to shifts in either the degree of over- or underestimating the real vacancies rate or in the degree of maladjustment. However, it is not possible to identify which one of these two possibilities applied. DDM reckon changes in maladjustment to be the most likely reason, since the UK post-war economy gradually moved from directed labour to a market economy and consequent dislocations had to be overcome. In addition, the data for separate industries do not support steep declines in the statement error (DDM, 1958: 26).

Finally, DDM argue that the statement error itself might be a source of inaccuracy since it might be a function of demand. For example, in times of a slack labour market, employers may not even report vacancies at labour exchanges at all since they are easily filled. However, the two tests on the vacancy data that DDM apply solved most of the disadvantages of their method. Not only do they conclude that the statement error stable, but they also infer that its value is close to unity. A unity value of *s* implies that the zero net demand line can be considered to have a slope of 45º, and that the statement error is unlikely to be a function of demand (DDM, 1958: 28). This enabled them to conclude that the shift in the pre-1950 data was caused by a shift in maladjustment.

3.3.2 *Solow and the measurement of structural unemployment in the USA*

In the post-war period, most governments of European countries and the United States were committed to a policy of maintaining a high level of employment. The social distress of the Great Depression and the belief in Keynesian aggregate demand management led to formulations of target rates of employment, in some countries even explicit formulations of full employment goals. Most European countries experienced very low unemployment rates in the post-war period making any Keynesian fiscal policy unnecessary. The low unemployment rate in Great Britain in the post-war period of around 1.5 percent on average was even lower than the 3 percent target rate of full employment that Beveridge thought was possible in his *Full Employment in a Free Society* and for which he was seriously criticized.[12] In the United States, the situation was different. Throughout the early 1950s, the unemployment rate was at a persistent, higher level, close to 4.5 percent on average. After 1954, the unemployment rate increased even more, and the average rate even rose to 5.5 percent from 1954 through 1963.

This of course raised questions regarding the cause and nature of the high US unemployment, and a debate ensued in the early 1960s that became known as the 'structuralist/deficient demand' debate, or, as it was sometimes referred to, the 'structuralist/antistructuralist' debate. The main issue at stake was the claim of the structuralists that the high unemployment in the US was the result of increased structural unemployment caused by an increase in the skill level of employment.[13] The current unemployed were not able to find jobs because they lacked the required skills. The opponents argued that the unemployment was the result of an insufficiently high level of aggregate demand, and they suggested fighting unemployment by expansionary fiscal policy.

Robert Solow, an 'antistructuralist', points to a strong ideological element involved in the structuralists' argument (Solow, 1964:17). Most of the proponents of the structuralist hypothesis turned out to be right wing-politicians who endorsed this hypothesis in the hope that the government would refrain from any fiscal policy. Another political and even racial element seems to have been involved, in the sense that particular groups in the labour force, most notably blacks and teenagers, were blamed for this increase in unemployment. Solow suggests a test for the structuralist hypothesis to compare vacancy and unemployment data and measure the level of structural unemployment (Solow, 1964: 21). For the framework of this test, Solow refers to an

[12] William Beveridge (1944) considered a 3% unemployment rate as a possible (and even conservative) target rate of unemployment (p. 127-128). It was built up as: 1% seasonal unemployment, 1% change of employment, and 1% frictional unemployment.

article of Hancock (1963). Hancock applies DDM's method to Australia, which suffered from a similar problem to the US, but in testing for an increase in 'maladjusted' unemployment for the period 1947-1962 Hancock only finds a marginal shift of the Australian UV-curve. He refers, instead, to the level of unemployment at the net zero excess demand point as the irreducible minimum of unemployment or 'hard core'.[14]

In 1967 Cohen and Solow applied the first econometric (reduced-form) estimation of the UV-curve. They do not apply DDM's graphical analysis but rather test for shifts of the UV-curve. They find a stable relationship between unemployment and vacancies, and conclude, with respect to the hypothesis that structural unemployment is increasing, that: our answer is "probably not, and almost certainly not much" (Cohen and Solow, 1967: 110).

3.3.3 NIESR: measurement according to the traditional classification

Economists of the National Institute of Economic and Social Research (NIESR) in London, familiar with the UV framework through the work of NIESR-member Dow, made further contributions to the UV-curve framework, and applied it to regional studies.[15] In October 1966 NIESR began a study on regional economic development in Great Britain.[16] The original idea was:

> "to provide a measure of the differences of degree of imperfection between the labour markets of the United Kingdom regions: We wanted to be able to say something about the extent to which the regions' different rates of unemployment could be attributed to differences in labour market imperfections, as opposed to differences in effective demand" (Brown, 1976: 134).

Therefore, they used an analysis of *u* and *v* data for both regions and industries as the framework for these regional studies and called this approach 'UV analysis' (Brown, 1976:134). They identify the level at which unemployment and vacancies are equal for regions or industries as the level of non-demand deficiency unemployment, and consider this a measure of the inefficiency of regional

---

[13] See, for example, Myrdal, *Challenge to Affluence*, 1963.

[14] Hancock defines the 'hard core' doctrine as "a less subtle version of the concept of structural unemployment, which means in broad terms unemployment that exists because the workers available for employment do not possess the qualities that employers with unfilled vacancies require" (Hancock, 1963: 35), and the way he uses this concept seems to bear a resemblance to Friedman's natural rate of unemployment, which was established a few years later.

[15] Members of NIESR in the late 1960s and early 1970s were, among others, Arthur Brown, David Worswick, John Bowers, Paul Cheshire, Edward Webb and Robert Weeden.

[16] The results were presented in Brown (1973), Cheshire (1973), Weeden (1973), Webb (1974), and Weeden (1974), while the framework is explained in close detail in Cheshire (1973).

labour markets. The contribution of NIESR lies in the fact that they further decomposed the non-deficient demand component of unemployment into a structural (u_s) and a frictional (u_f) component of unemployment, so that a classification arises that corresponds to the 'traditional' classification; that is, a division of unemployment into frictional, structural, and deficient demand unemployment.

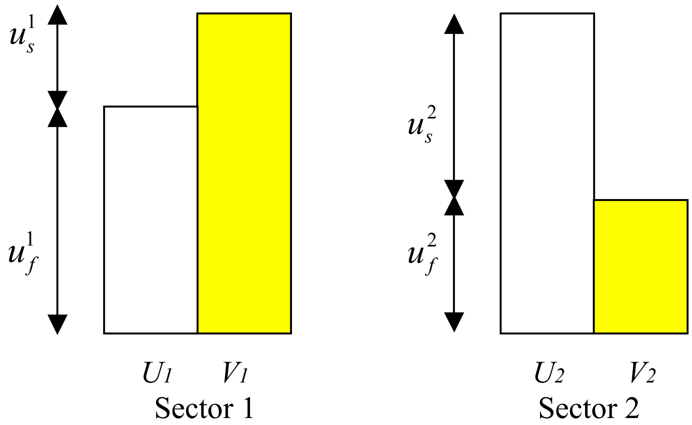


Figure 3.4: NIESR's measurement of frictional and structural unemployment per sector

At the sector level, the level of structural unemployment u_s is determined as the difference between excess supply and demand of labour per sector of the economy (industry or region), while the level of frictional unemployment is the minimum of unemployment and vacancies for each sector (see Figure 3.4). At the aggregate level, for example, for a two industry economy with unemployment U_1 and U_2, and vacancies V_1 and V_2 in the respective industries, and $V_1 > U_1$ and $V_2 < U_2$, aggregate frictional unemployment is defined as the sum of $u_f^1 + u_f^2$ $(= U_1 + V_2)$, and aggregate structural unemployment as the minimum of differences between *Ui* and *Vi*, thus as u_s^1 $(= V_1 - U_1)$ or u_s^2 $(= U_2 - V_2)$, whichever is the smaller. Demand deficiency unemployment is $(U_1 + U_2) - (V_1 + V_2)$ (Cheshire: 1973, 13). Armstrong and Taylor (1980) suggest decomposing structural unemployment even further in a geographical, an occupational, and a simultaneous occupational-geographical component for regional studies. Thirlwall (1969) provides an analytical treatment of this approach. Based on these measures of classes of unemployment, NIESR (Cheshire: 1973) derives a conceptual framework using the simplifying assumption of a fixed level of frictional unemployment. This idea of frictional unemployment as ‘frictions within sectors’ and structural unemployment as ‘frictions between sectors’ is often found in labour market studies in the 1970s.

In the 1970s, UV analysis reached the highest stage of its popularity, and most studies date from this era. UV analysis turned out to be a very simple and easy to use device for analysing the nature of unemployment. In most studies, unemployment was decomposed into two classes: deficient demand and non-deficient demand, so the UV analysis indicated which part of unemployment was caused by deficient demand and could be removed by Keynesian expansionary policy.

Finally, two aspects are noteworthy with respect to the use of UV analysis. Firstly, UV analysis seems to have been most popular in Europe, most notably in Great Britain. For the US there do not seem to be many studies. This might be attributed to the fact that, in the US, the preferred framework for analysis of unemployment was the Phillips-curve and concepts derived from it, such as the natural rate of unemployment. Secondly, UV analysis was particularly popular for regional studies.[17]

3.4 The correlational structure of UV analysis

UV analysis rests on the idea that changes in effective demand will cause “cyclical” or “deficient demand” unemployment. In times of recession, unemployment goes up and changes in line with the business cycle. Thus: changes in “deficient demand” (“*d*”) are responsible for changes in “deficient demand” unemployment (“*Ud*”):

d (“deficient demand”) $\rightarrow$ Ud (“deficient demand unemployment”).

Though this appears a causal relation, from the point of view of measurement the correlation between the two phenomena suffices. However, both these phenomena are abstract, theoretical terms and not directly observable or accessible to our senses. In UV analysis, this problem is circumvented by defining d (“deficient demand”) as excess unemployment over vacancies, thus $d = u - v$, giving a monotonic correlation between d and Ud. Both unemployment and vacancies are abstract terms too, but – as Chapter 2 shows - they might be ‘observed’ directly, for example, by standardized quantitative rules (SQRs) from the registration of the unemployed at labour exchanges. Data on U and V are therefore used to make deficient demand (and subsequently deficient demand

[17] See, for example, Cheshire (1973), Webb (1974), Armstrong and Taylor (1980).

unemployment) operational and in the UV-graph “deficient demand” can be represented as a point on a structural UV-curve.

Changes in the “effective demand” cause a simultaneous movement of *U* and *V* in opposite directions:

$$d \text{ (“deficient demand”)} \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{c} U \\ \sim \\ V \end{array} .$$

U and *V* themselves are inversely correlated but by no means directly causally connected, rather they have a common cause, in the same way as changes in air pressure causally affect both the reading of a barometer and the occurrence of rain:

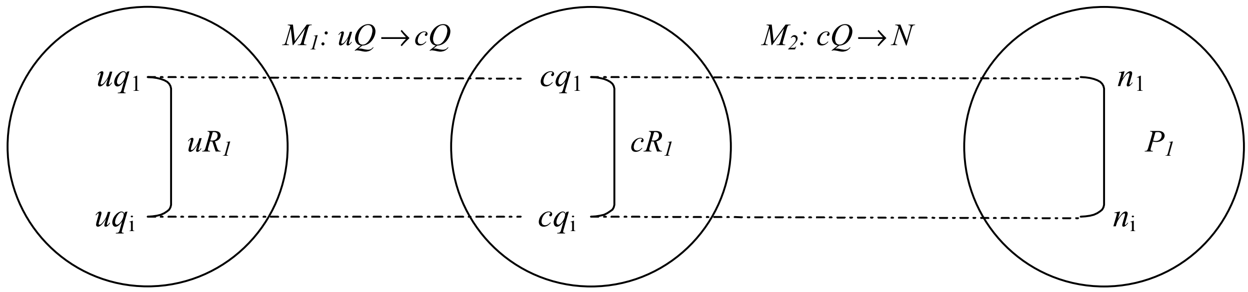
$$\text{Change in air pressure} \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{l} \text{Reading barometer} \\ \sim \\ \text{Rain} \end{array} .$$

U and *V* are thus not the cause of deficient demand unemployment but deficient demand unemployment correlates with the opposite and simultaneous movements in *U* and *V*. The case thus appears to be more complex than the simple $uQ \sim cQ$ correlation outlined above. However, it can be argued that it is only deficient demand unemployment that correlates with the “deficient demand”. Structural unemployment, on the other hand, is thought of being dependent on the institutional organization of the labour market and structural changes in demand, and is hence regarded as independent of the business cycle. And we might use this relationship to identify and quantify deficient demand unemployment. Thus, the full correlational connection upon which the UV analysis is based seems to be as follows:

$$d \text{ (“deficient demand”)} \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{c} U \\ \sim \\ V \end{array} \sim \; Ud \text{ (“deficient demand unemployment”)},$$

where *d* is the cause of movements of both *U* and *V*, and correlates with *Ud*. In addition, *U* and *V* also correlate.

In order to assess to what extent this case yields sound measurement we will analyse the UV analysis in accordance with both the criteria of the RTM and Heidelberger’s criteria for correlation. The requirements of the RTM are satisfaction of the representational theorem and the uniqueness theorem (see Chapter 1). The representational theorem states that: if two or more numerical relational structures, which are isomorphic to a certain empirical relational structure, can be related by a certain permissible transformation, then there exists a unique numerical representation for this empirical relational structure" (Suppes, 1998). The RTM requires that the relevant structure of the phenomenon must be isomorphic in a one-to-one relation between two structures. In measurement theory, one of these structures is an empirical one and is a feature of the real world; the other is an arithmetical structure, and is part of the realm of numbers. In the case of associative measurement (as this case is), the isomorphic mapping runs via the correlated stand-in, deficient demand, which is derived from a structural UV-curve. This case of associative measurement can thus be presented using a diagrammatic representation of measurement derived from the RTM (see Figure 3.5).[18]



Representation of *uQ*: deficient demand (= *u* – *v*)

Correlated phenomenon *cQ*: deficient demand unemployment

Numbers *N*

Figure 3.5: Diagrammatic representation of the UV analysis as associative measurement

The representational theorem requires that both mapping M_1, from the concept of deficient demand to deficient demand unemployment, and mapping M_2, from deficient demand unemployment to numbers, are non-degenerating and strictly isomorphic.

Heidelberger seems to allow for a slightly different interpretation of 'representation'. In the RTM 'representation' is used to stress the strict isomorphic correspondence between the empirical relational structure and the numerical relational structure. For the representation, it is required to have a correspondence of parallel structures, more or less in the same way as a photo is a

[18] See also Boumans (2005a).

representation of some state of the real world. Heidelberger uses the term 'representation' as a less constraining concept than that of isomorphism, as not all representations represent strictly isomorphic structures. The RTM is therefore a correspondence-as-congruence theory. Heidelberger, on the other hand, does not claim a structural isomorphism and his account is therefore rather a correspondence-as-correlation account. The representative ('Repräsentant') is rather an *analogy* that correlates in only one respect (or a few respects) to the unobservable as a whole.[19]

For Heidelberger, this strategy for associative measurement through correlated relationship, *u*Q ~ *cQ* (*cQ* is correlated with *u*Q, while *cQ* is measured with apparatus A) is successful when the following conditions are met (Heidelberger, 1994a: 317):[20]

1) Both *u*Q and *cQ* can be ordered according to the strength of a property in the object of investigation.
2) It is assured by experience that *c*Q correlates monotonically with *uQ*. (Thus when *u*Q increases also *cQ* must increase.)
3) *cQ* is measured by a calibrated measurement instrument A.

Point 1 and 3 of Heidelberger's account seems to work well for UV analysis. When unemployment goes up while structural unemployment is constant, deficient demand unemployment must go up. And, as point 3 demands, considering the Standardized Quantitative Rule for measurement of U and V as a calibrated instrument does not seem to run into deep trouble (see Chapter 2). The problem with Heidelberger's account seems, however, to concern point 2: If the measurand is unobservable how is the correlation to be assured? The RTM experiences a similar problem when it comes to the representational theorem. It requires an isomorphic mapping but leaves out considerations of how to assure one.

---

[19] For this reason Heidelberger prefers the word 'analogy' rather than 'representation' for his account.

[20] My translation. Original in German:
"1) Es muß möglich sein, Gegenstände in Bezug auf die Stärke der Ausprägung von Q [*uQ* in text] und R [*cQ* in text] in eine Ordnung zu bringen.
2) Es muß erfahrungsgemäß gesichert sein, daß R [*cQ*] mit Q [*uQ*] monoton korreliert ist. Wenn Q [*uQ*] stärker wird, wird auch R [*cQ*] stärker, und wenn Q [*uQ*] schwächer wird, wird auch R [*cQ*] schwächer. Das Umgekehrte ist nicht gefordert. Die für eine Menge von Gegenständen durch Q [*uQ*] induzierte Ordnung ist mindestens genauso fein wie die durch R [*cQ*] induzierte. Sie muß sich in der Ordnung von R [*cQ*] wiederfinden lassen.
3) Das messende Attribut R ist an einem Meßinstrument realisiert, so daß feststeht, zwischen welchen Ausprägungen des Attributs am Meßinstrumentgleichen Abstände bestehen sollen. Die verschiedenen Ausprägungen von R müssen an einem Instrument kalibriert sein" (Heidelberger, 1994a: 317).

For both Heidelberger and the RTM this is obviously done by experience, that is, observations in the real world (see Chapter 1), but this may not be obvious. After all, in this case, both correlated phenomena are unobservable. The stand-in for deficient demand unemployment, deficient demand, can only be measured by defining it as *U-V*, while the UV-relation is assumed to be a stationary, structural relation. Or, put differently, the level of maladjustment is assumed to be constant in the short run. It follows that, in this case, measurement is only possible after structures have been "put upon" the world. Some philosophers of science take account of this point and argue that structures are not just 'there outside in nature waiting to be discovered'. Cartwright (1983), for example, argues that phenomena are "prepared descriptions"; that is: "we present the phenomenon in a way that will bring it into theory" (1983: 134). In other words, we have to put structures upon the phenomenon that are consistent with our theories in order to 'see' the phenomenon. Structures therefore, are assigned to phenomena. Since observations are derived from these structures, reason is an essential element in the measurement process.

UV analysis suffered from some methodological shortcomings, all of which had to do with the UV-curve as a structural relation, and so as a sound representation of deficient demand unemployment. As we shall see in the next section, UV analysis was finally abandoned for this reason.

3.5 The abandonment of UV analysis

In the 1980s, studies using UV analysis became rare. While Armstrong and Taylor in 1980 still argue that UV analysis "continues to have considerable potential for further development" (p.100), it was almost completely abandoned only a few years later, as Muysken's article *A Post-Mortem on the UV analysis* (1988) illustrates. This sudden fall is even more remarkable when one considers that the analysis of unemployment was replaced by the new microeconomic search theory, some models of which – most notably matching models of unemployment – rely crucially on the UV-curve. What caused this sudden fall of UV analysis? And why does analysis of unemployment with matching models not suffer from flaws when it uses the same framework? Two reasons in particular seem to account for the fall of UV analysis: a methodological one, and a more paradigmatic one. We will focus here only on the methodological problems of the UV analysis. The paradigmatic reason for its decline is discussed in the epilogue of this chapter (Section 3.6).

3.5.1 *The identification problem of UV analysis*

Advocates of UV analysis considered that it had nevertheless three important problems:

> "The real problem of applying this proposition in practice arises from three sources: the imperfection and possible variability in the recording of both vacancies and unemployment, the fact that some labour markets provide no evidence relating to a time when U and V were equal, and the variability of the true UV relation over time" (Brown, 1976: 140).

Let us analyse these problems in more detail.

The first problem concerns the lack of reliability of vacancy and unemployment data (Brown, 1976; Armstrong and Taylor, 1980). This may seem odd since Dow and Dicks-Mireaux addressed this problem in their seminal article, and argued that it was not a concern. Research in the 1970s, however, showed a substantial underrecording of vacancies, and hence a measurement error of classes of unemployment. A survey by the Manpower Service Commission in 1977, for example, showed that only 36 % of unfilled vacancies were recorded officially (Armstrong and Taylor, 1980: 121). But this problem relates to the direct measurement of *U* and *V*, and concerns the accuracy of measurement apparatus A, not the correlational representation itself. And Cheshire (1973) simply suggests "calibrating" this measurement apparatus by multiplying the vacancies data by a factor of three in order to account for the underrecording of *V* data.[21]

The second problem refers to the fact that, in the application of UV analysis, it occurred that the UV-curve had no intersections with the 45°-degree line, so that the level of deficient demand unemployment was underdetermined. The UV-curve had to be extrapolated in order to determine the level of structural unemployment. However, as will be made clear later, this problem is part of the more fundamental problem of identification of the UV-curve.

The third problem, explored in very great detail, is in fact the problem of the stability of the UV-curve over time. The work of DDM and Solow, as well as theoretical analyses by Holt and David (1966) and Gordon (1966), inspired a series of empirical studies in the late 1960s and throughout the 1970s, starting with Cohen and Solow in 1967, estimating the relation between unemployment and vacancies. Almost immediately after the Cohen and Solow publication in 1967, empirical studies found supposed 'breakpoints' in the UV-curve, suggesting shifts of the curve

[21] The rate of underrecording of vacancies is, however, not likely to be the same for all occupational groups or regions. So, while multiplying the vacancy rate might decrease the measurement error of the structural/deficient demand distinction, it increases the measurement error of structural/frictional unemployment.

further or closer to the origin corresponding, respectively to higher or lower levels of structural unemployment. This obviously raised questions about the stability of the UV-relation and the usefulness of the UV-curve as a structural relation for economic analysis and measurement, and after the 1970s resulted in an enormous amount of empirical studies with an abundance of specifications of UV-curves, all of which incorporate additional variables, dummy variables or lagged variables. The discussion took place roughly speaking following national boundaries. The discussion in the USA focused on the behaviour of the Help-Wanted Index[22] (hereafter: HWI) – as a proxy for vacancies – in relation to unemployment for US data for the period 1951-1966.[23] Cohen and Solow find a systematic pattern connected with business cycle fluctuations. The regression overestimates the HWI during the downswing and underestimates it in an upswing. This phenomenon, confirmed in almost all later empirical studies, reveals 'counter clockwise loops' in the UV-relation. The generally accepted explanation was that vacancies respond much faster to changes in aggregate demand than unemployment does. The finding of these 'counter clockwise loops' was obviously of economic interest, and in a certain way even reassuring, since they could be related to the 'loops' found in the Phillips curve. But making precise observation of shifts was rather difficult.[24]

A more comprehensive discussion focused on the stability of the British UV-curve for the period 1958-1971.[25] These studies all find 'breakpoints' or shifts of the UV-curve for the British economy, such as the outward movement in the period 1967-1969 (see Figure 3.6). Knowledge of the cause of the UV-curve shift does matter. In the way DDM presented the UV relationship, it is regarded as a structural relation and as such it 'stands in' for unobservable deficient demand unemployment, and the representational theorem is only satisfied if this structural relation holds. This is because, if U moves independently of V, deficient demand unemployment correlates not

[22] The HWI of the National Industrial Conference Board is a weighted average of indexes of the number of help-wanted advertisements posted in the leading newspapers of 52 cities in the United States. As such it, can be considered as another example of a Standardized Quantitative Rule for direct measurement.
[23] Cohen and Solow (1967, 1970), Gujarati (1969), Burch and Fabricant (1968, 1971), all in *The Review of Economics and Statistics*.
[24] In order to take these loops into account, Cohen and Solow included a dummy variable, which is equal to 1 in expansions and 0 in contractions. With this specification, they find a stable relationship between HWI and unemployment leading them to conclude that there is no or very little support for the structuralists' hypothesis. Burch and Fabricant (1968), however find a significant outward shift of the UV-curve, while Gujarati (1969) concludes that the HWI-unemployment relation is fundamentally unstable, and that results obtained by disregarding cyclical fluctuations are of questionable value (Gujarati, 1969: 484).
[25] Bowers, et al. (1970), Gujarati (1972a, 1972b, 1973), Taylor (1972), Foster (1973), Knight and Wilson (1974), Evans (1975, 1977), Holden and Peel (1975, 1977), Warren (1977), Parikh (1977), Bewley (1979), almost all in *The Economic Journal* and *Applied Economics*.

only with the business cycle but with other factors as well, and the isomorphic mapping does not hold. The first British studies treated the UV-curve as a structural relationship and attributed the shift to newly introduced social legislation in Great Britain. Gujarati (1972a), for example, argued that the UV-relation is stable over the business cycles[26] and shifts outwards with a constant factor – which he calculated to be 1.443 – to a permanent higher level as a consequence of the new social legalization.

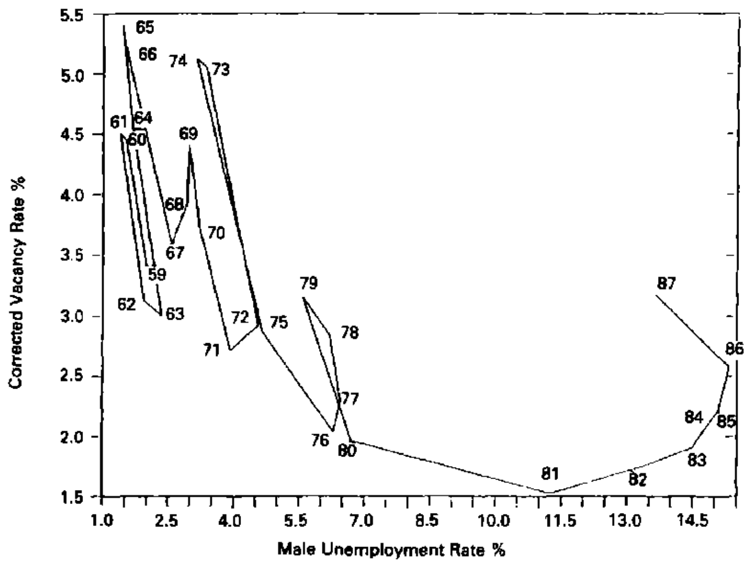


Figure 3.6: Loops and breakpoints in the empirical UV-curve for Great Britain (1959 –1987)
Source: Jackman et al., 1989: 378.

By accounting for the cause of the UV-shift, the UV-curve could be made stable and hence function as a representation of deficient demand unemployment. However, the countless number of econometric studies on the UV-curve provides no clear-cut evidence for the cause (or causes) of the shift and most studies disagree about what causes the shifts.[27] Econometric analysis usually

[26] This, however, seems contrary to his 1969 article in *The Review of Economics and Statistics*.

[27] Among the possible causes of the shift are: the introduction of earning-related unemployment benefit; redundancy payments; selective employment tax; income policy; general ‘structural’ change; devaluation-induced structural change (Bowers, et al., 1970); the Redundancy Payments Act of 1965 and the National Insurance Act of 1966 (Gujarati, 1972a); labour “shake-out” (Bowers, et al., 1970; Taylor, 1972); the efficiency of job search due to demographic changes (Parikh, 1977): increased heterogeneity in labour supply (Kuipers and Buddenberg, 1978); the shift from agricultural to non-agricultural employment; unemployment insurance and demographic changes (Reid and Meltz, 1979); increase in supply of female labour, increase in geographical spread of the labour force (van der Berg, 1982). Layard (1986) tests for an extended set of variables, but finds little or no evidence that can explain the shifts. Pissarides (1989) tests for the replacement ratio, and increased mismatch, but again finds no clear-cut cause, and tends to conclude that the shift is caused by an increase in job choosiness or a fall in search intensity.

provided no, or only very little, evidence for the alternative hypotheses tested, and, after three decades of testing and specifying the UV-curve, Jackman et al. (1989: 392) have to admit that: “we must remain agnostic as to the causes of the change”. Wall and Zoega (2001) finally conclude that the shifts can be attributed to the business cycle and not to only structural variables, thus making the UV-curve a flawed device for analysing structural changes.

The economy turns out to be subject to a set of unknown factors, which affect the UV-curve in three ways simultaneously. First, changes in aggregate demand in the economy cause movements along an otherwise stable UV-curve, while, secondly, structural changes in the economy (whatever they may be) or shocks that affect the labour force cause movements of the curve itself. And, finally, economic policy aimed to make matching in the labour market more efficient shifts the UV-curve inwards deliberately. Since we cannot distinguish between these simultaneous movements, the UV-curve is underdetermined, that is, the exact shape of the UV-curve cannot be estimated, and hence classes of unemployment cannot be determined exactly. This bears resemblance to the identification problem of supply and demand curves, where price-elasticities of supply and demand cannot be estimated when it is unknown whether price or quantity changes are caused by shifts of the supply or of the demand curve.

This identification problem could be overcome if economic theory could provide good reasons for a particular shape of the UV-curve, or could provide us with the conditions where UV-shifts are absent. In that case, it would be possible to estimate the UV-curve and trace its shifts. But unfortunately, economic theory – neither Keynesian or Neoclassical – is able to do so, leaving the UV-curve underdetermined. The absence of an intersection of the UV-curve with the 45˚ degree zero-excess demand line is thus not the real problem. In many studies, the lack of intersections was simply overcome by extrapolating the observations along the UV-curve once the true UV-curve was estimated. It is rather the estimation of the UV-curve itself that causes the trouble.

3.6 Conclusions

Both Michael Heidelberger and the RTM provide an account of measurement. Central to both strategies for measurement is the issue of how to bring about a relation between an empirical structure and a numerical structure that yields sound measurement. In the RTM, this is embodied in the Representational Theorem, which requires this relation to be isomorphic. The RTM can therefore be considered as a correspondence-as-congruence theory. Heidelberger argues for correlation as representation, which is less compelling with respect to requirements for measurement, as the relation is not necessarily required to be isomorphic in nature. Heidelberger's account is therefore a correspondence-as-correlation account. A major problem for both propositions is therefore how to ensure this (isomorphic or monotonic correlation) mapping, especially when the phenomena under scrutiny are unobservable. Both philosophical strands assume that the (isomorphic or monotonic correlation) mapping follows from experience, that is, observation in the real world.

In this case study, the observable of interest – deficient demand unemployment – is unobservable. From experience, we know that deficient demand unemployment is correlated with (deficient) effective demand. After all, in times of recession unemployment goes up, while vacancies go down. This correlational structure is explored in UV analysis. Next, deficient demand is *defined* as $d = u\text{-}v$. Like a thermometer that uses the expansion of liquids or gases as a representative, UV analysis took the joint, inverse behaviour of U and V data, that is, the UV-curve itself, as a representative, and deficient demand is represented as a point on the curve. This assumes, however, that the UV-curve is a structural relationship. Though economists have good reasons to assume this, and the idealized hyperbolic UV-curve still plays an important role in economic analyses (see Epilogue), empirical research has shown that the UV-curve shifted as a consequence of unknown, structural changes in the economy and deliberate policies aimed at making matching in the labour market more efficient. In itself this was not a problem, but the causes of the shift could not be traced and it made the identification of the exact shape of the UV-curve impossible. Consequently, the isomorphic or monotonic correlation was destroyed leaving UV analysis in either account of measurement to inhabit an insufficient 'principle of reliability' and, hence it is an unsound measuring device.

More generally, the philosophical literature seems to be lacking a systematic account of how exactly to establish an empirical mapping, and leaves this field to case studies, probably for the reason that this is too pragmatic an issue. This case study of UV analysis suggests that common sense and good (economic) intuition are just as important as experience.

3.7 Epilogue: The place of the UV curve in economic theory

The UV-curve was put forward by DDM as a practical measurement device to guide economic policy. Its place in economic theory was therefore not immediately clear. The UV-curve obviously had some attractive features, in the sense that it provides a macro-framework that shows that unemployment and vacancies coexist simultaneously in the absence of excess demand, or that some unemployment will exist even at very high levels of demand, but its explanatory power was low since it provided no new insights, other than those that already existed, about what mechanisms caused the simultaneous existence of unemployment and vacancies. More importantly, although the UV-curve was empirically supported, there was no theoretical foundation for it. For example, it was not clear how this UV-curve should be explained in a simple Marshallian supply and demand analysis, or how it related to the Phillips curve. And in addition, both the Phillips-curve and the UV-curve were interpreted to bear (implicit) contradictory notions of the full employment level of unemployment. It is therefore not surprising to see that, in the course of the 1960s and 1970s, studies appeared that tried make clear these issues and tried to incorporate the UV-curve into mainstream economics. Some of these theoretical contributions were obviously driven by the results of empirical UV-studies that emerged from 1967 onwards.

<u>The UV-curve in a Marshallian framework</u>

Hansen (1970) integrates elements of Gordon (1966) and Holt and David (1966) to provide a comprehensive, neoclassical theory of friction in a supply and demand framework, and shows that unemployment and vacancies can coexist. The starting point for Hansen is the division of the labour market into homogenous and frictionless submarkets, where on each submarket there only exists excess supply or excess demand for labour (in the same way that NIESR does). Frictions, however, do exist between submarkets, and unemployment can still exist since excess demand in one submarket cannot be matched by excess supply in another submarket.[28] Submarkets thus have excess supply or demand, and, according to Hansen, this means:

> "that actual unemployment is never *on* the supply curve (if the wage rate is below equilibrium) or the demand curve (when below equilibrium), but let us assume, to the left of both the demand and supply curve" (Hansen: 1970:6).

[28] All unemployment is thus structural unemployment since frictional unemployment is absent.

This is represented in Figure 3.7 where the quantity of labour is plotted on the horizontal axis and the wage level on the vertical axis.[29] If the wage (w_1) is above the market clearing level (w^*), unemployment arises corresponding to the distance KM, while vacancies exist equal to KL. At the market-clearing wage w^*, vacancies and unemployment are equal and correspond to NO, and, for wages below the market clearing level (w_2), an excess demand occurs with PR vacancies and PQ unemployment.

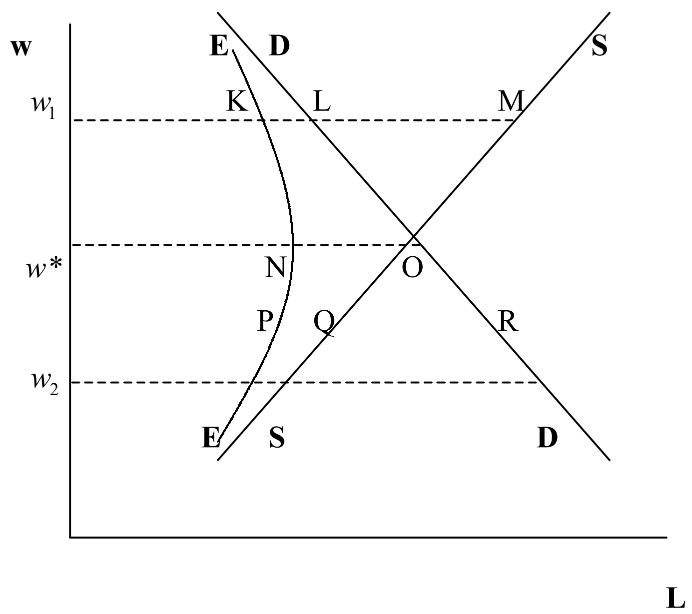


Figure 3.7: Employment function in an S-D diagram
Source: Hansen, 1970: 7.

The locus **EE** is referred to as the 'market clearing path', and it relates the actual quantity of employment to the forces of supply and demand. The shape of the curve is based on the assumption that, as the pressure of demand increases, matching become easier. The distance between **EE** and the supply and demand curve represents DDM's 'degree of maladjustment' in the economy. Plotting the vacancy rates against the unemployment rates yields the UV-curve.

The NIESR presents the same phenomenon by plotting vacancies and unemployment against aggregate demand (Webb, 1974; Brown, 1976), yielding Figure 3.8. When demand for labour is low, such as in O, unemployment occurs corresponding to OA. Unemployment falls, according to line AL, when aggregate demand increases. For every new additional vacancy there are many

[29] Bowden (1980) stresses that the real wage rate w/p is more appropriate.

unemployed workers, and it will be easy to find an unemployed worker who meets the job requirements. However, after a certain point (point M) it will be harder to find appropriate, workers and vacancies will exist as well as unemployment. Hence, in full employment (point L), there will exist an equal amount of unemployment and vacancies. When demand for labour further increases, vacancies will increase according to line *VV* and unemployment falls along line *UU*.

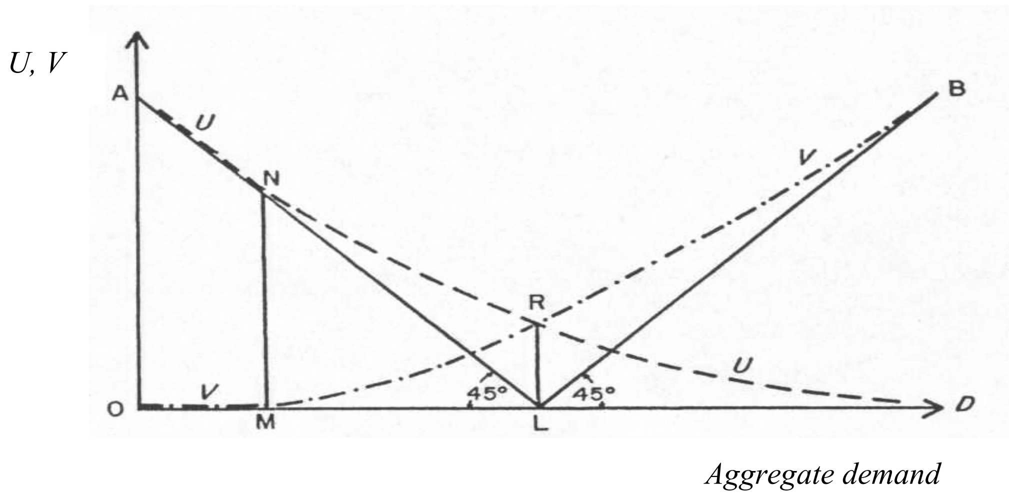


Figure 3.8: Demand for labour, unemployment and vacancies
Source: Brown, 1976: 136.

<u>The relation with the Phillips-curve and notions of full employment</u>

Both the Phillips-curve and the UV-curve were established in 1958, motivated by economists' interest and concern about inflation and its relation with unemployment. The UV-curve was established to provide a suitable measure of excess demand for these inflation-unemployment studies, while the Phillips-curve represents this macroeconomic relationship. The relation between the Phillips-curve and the UV-curve can be formally derived as follows. The UV-curve can be written as:

$$v = \gamma \frac{1}{u} \qquad \gamma > 0 . \tag{1}$$

The increase in wages can be written as a function of excess demand for labour: $\Delta w = f(XD)$. With the difference between vacancies and unemployment taken as excess demand, this relation becomes:

$$\Delta w = \alpha(v - u) + \beta . \tag{2}$$

Substituting (1) in (2) then gives:

$$\Delta w = \frac{\alpha\gamma}{u} - \alpha u + \beta\,, \qquad (3)$$

which is the original Phillips-curve relation, i.e. the relation between wage increases and unemployment.

Both approaches provide competing and incompatible notions of equilibrium and full employment. Though DDM never made the claim explicit, the UV-curve analysis seems to imply a notion of full employment that corresponds to the level of unemployment at the intersection of the UV-curve and the 45 degree line, i.e. the point on the UV-curve where $U=V$. This is not, however, the neoclassical definition of an equilibrium outcome in terms of equilibrium between supply and demand of labour with an implicit reference to the equilibrium wage rate. It is rather more akin to William Beveridge's definition of full employment, which he expressed in terms of unemployment and vacancies (1944: 18).[30]

The Phillips-curve provides other operational definitions of full employment and the way to get there. Two important definitions are Lipsey's and Friedman's. Lipsey (1965) defines full employment as the lowest level of unemployment that could be attained given the government's dual policy targets: both an acceptable rate of inflation and an acceptable rate of unemployment.[31] These dual policy targets can be represented as indifference curves (the dotted lines in Figure 3.9) and full employment is then derived as the minimum combination of both acceptable inflation and unemployment, i.e. where an indifference curve and the Phillips-curve coincide (point *a*). Unemployment rates above the level of point *a* (full employment) are considered as deficient demand unemployment (u_d), which the government can reduce by an expansionary fiscal policy. In Lipsey's account full employment is thus subjectively determined by the government's preferences. In full employment, the remaining unemployment consists of both frictional unemployment (u_f) and structural unemployment (u_s). In Lipsey's account, the government might be able to reduce the level of full employment by the reduction of structural unemployment (u_s). This implies a shift of the Phillips curve to the left until a lower full employment state is realized at a lower indifference curve at point *b*. Shifts of the Phillips-curve to the left could be brought about by, for example, "reducing inequalities in excess demand between various labor markets, and reducing the time taken in

[30] William Beveridge expressed full employment in terms of unemployment and vacancies. According to Beveridge full employment means "having more vacant jobs than unemployed men". (1944: 18).

changing the supply of labor into the form in which it was being demanded" (Lipsey, 1965: 213). The trade-off of costs and benefits of policy measures, such as retraining the unemployed, determines the level of reduction of structural unemployment. Finally, the level of frictional unemployment is exogenous and unavoidable.

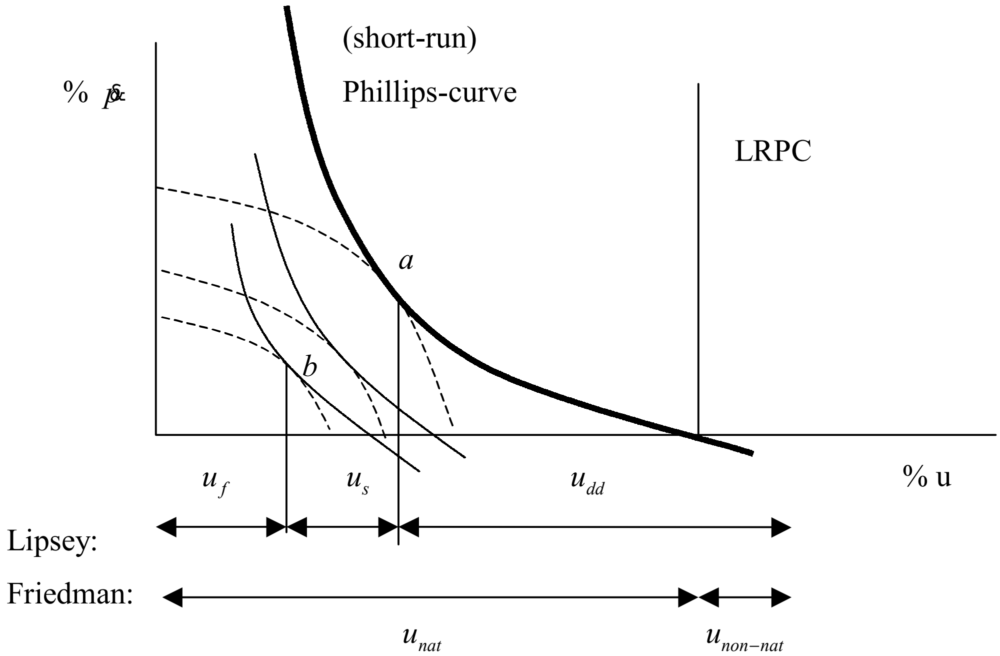


Figure 3.9: Classification of unemployment according to Lipsey and Friedman
Source: Based on Lipsey, 1965.

Friedman (1968) provides another operational definition of full employment. Friedman (and Phelps, 1967) deny the existence of a trade-off between unemployment and inflation. Workers form rational expectations about future inflation, and, when they realize they will be fooled by decreases in real wages as a consequence of inflation shocks, they withdraw labour from the market. The long-run Phillips-curve (LRPC) is therefore vertical and unemployment cannot be pushed below its 'natural rate' (u_{nat}). Long-run equilibrium in the labour market can only exist when there is equilibrium in both the labour and the financial market.

Obviously the different frameworks – UV-curve and Phillips-curve – bear different and conflicting notions about the nature of equilibrium and full employment. Figure 3.10 presents the

[31] In this way, Lipsey tried to bridge the gap between the 'structuralists' and the 'antistructuralists'.

relation between the notions of Lipsey (L), Dow and Dicks-Mireaux (DDM), and Friedman (F), and their corresponding level of full employment.

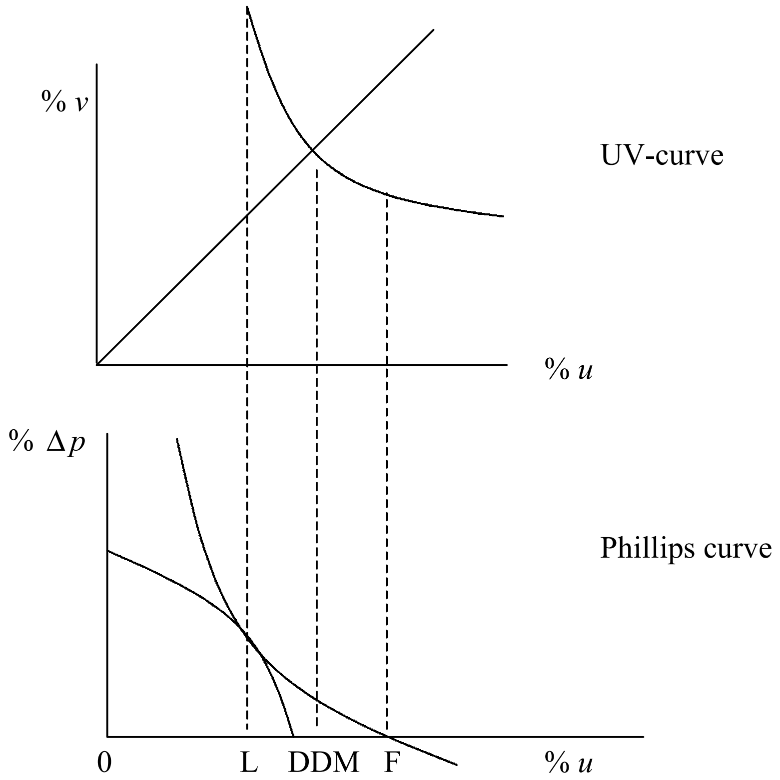


Figure 3.10: Relation between concepts of full employment in UV- and Phillips-curve frameworks
Source: Based on Gordon, 1966.

<u>The paradigmatic element in the fall of the UV-curve</u>

In the 1970s, the field of unemployment theory started to change and a new paradigm, which was labelled as 'search theory', 'flow approach' or 'new microeconomics' took over Keynesian, economic thinking. This new paradigm was motivated by the fact that macroeconomic rational expectations theories of unemployment lacked a microeconomic foundation. Authors like Phelps et al. (1971) sought these foundations in the individual search behaviour of agents. The unemployed make rational decisions to accept a job offer or reject it when they expect higher pay-offs from other future job offers. Unemployment is seen as a productive investment. A characteristic of search theory is that it analyses equilibrium unemployment in terms of flows in and out of unemployment, rather than in terms of a static difference between aggregate stock demand for labour and aggregate stock supply. This framework enables economists to analyse heterogeneous groups of workers with different characteristics in terms of probabilities of in- or outflow. Workers with a higher rate of job finding experience shorter unemployment durations and lower rates of structural unemployment. So, workers and employers with job vacancies are no longer considered as homogeneous, and both therefore have to spend time and resources in order to find a good match. Even in the absence of deficient demand for labour, unemployment and vacancies coexist as a consequence of this time-consuming search process. A variety of models are based on the search approach, e.g. efficiency wage models and matching models.

Central for matching models is the use of a matching function $M = m(U, V)$, from which the (aggregate) number of matches is derived in a similar way as production or utility functions.[32] The key idea of the matching function is that the complicated and stochastic process of job search is captured in one single, well-behaved, aggregate, mathematical function. This matching function is conceptually represented by the idealized, hyperbolic-shaped UV-curve with, most often, constant returns to input factors. Matching models thus also consider the UV-curve as a structural, invariant relationship, but in contrast, do not seem to be hindered by methodological problems (such as underdetermination of the UV-curve) that UV analysis faced, or the observation that the UV relation was unstable or invariant.

This contrast might be explained by a number of reasons. In the first place, UV analysis was set up as a measurement device. However, since the UV-curve could not be exactly identified, its use as a measurement device for classes of unemployment became dubious. Matching models, on the other hand, were established for other purposes: namely as devices for the analytical exploration of mechanisms that generate unemployment. In matching models the identification problem is simply circumvented by constraining the UV-curve to specific forms, such as described by a Cobb-Douglas function. However, even in contemporary economic thinking, shifts of the curve are recognized, and the instability of the UV-curve is taken as an indictor of the performance of the labour market in terms of efficiency in matching unemployed and vacancies. As such, the UV-curve serves an important diagnostic function, more or less like ‘PV’ or ‘indicator diagrams’ do in mechanical engineering (see Figure 3.11). In these diagrams, the pressure and volume in a combustion engine are simultaneously recorded and plotted. The pressure in the cylinder of the engine is a function of

[32] Matching models draw on the pioneering work of Butters (1977), Hall (1979), Pissarides (1979, 1985), Bowden (1980), Diamond (1982b) and Blanchard and Diamond (1989). A survey of the matching function can be found in Petrongolo and Pissarides (2001).

the engine's working cycle, and the indicator diagram is used to diagnose the state of the combustion engine and to detect possible faults in the operation of the machine.

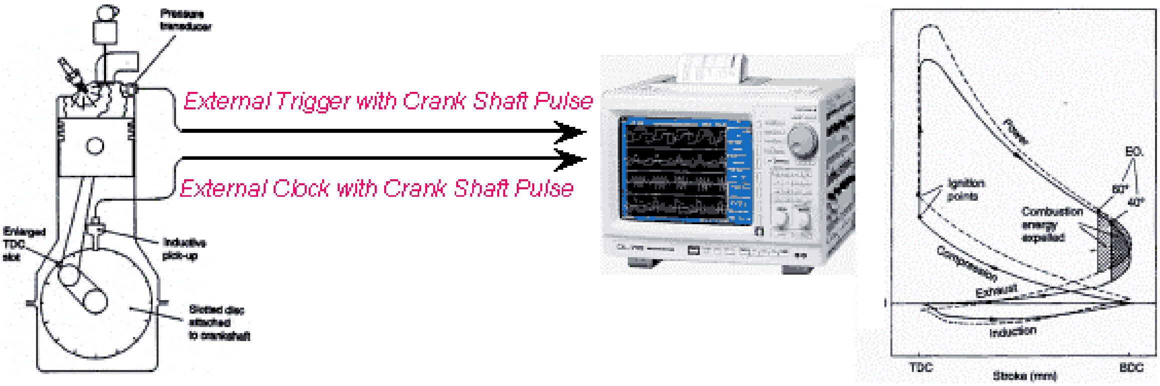


Figure 3.11: The recording of an indicator diagram for diagnostic purposes.
Source: Yokogawa Electric Corporation[33]

Secondly, the distinction of unemployment into classes could be tricky, since "the various categories of unemployment are not always as distinct and separate as has been suggested" (Hughes and Perlman, 1984:32-33). Workers who become unemployed as a consequence of lack of aggregate demand may remain unemployed for the long term and lose the professional and social skills necessary for their occupation. Consequently, they should be considered as structurally unemployed. The distinction between cyclical and structural unemployment is therefore blurred by "various unemployment–persistence mechanisms" (Lindbeck, 1999: 2), and search models try to provide insight into these mechanisms. Secondly, there was the inability to deal with flows in and out of unemployment and changes in unemployment duration. Search theory (of which matching models are a branch) is, however, able to do this, and provides answers to questions that became relevant in the 1980s when long-term unemployment started to occur.

In the decline of UV analysis a paradigmatic element has to be taken into account too. UV analysis was developed at a time when economists had a strong belief in the effectiveness of Keynesian, aggregate demand management. The main concepts that UV analysis aimed to measure, such as excess demand and cyclical unemployment, are important Keynesian concepts, and they were measured in the first place for the guidance of Keynesian economic policy. Since this belief in

[33] http://www.yokogawa.com/tm/pdf/appli/tm-appli_auto12.pdf

Keynesianism was much more profound in Great Britain and continental, Western European countries than in the USA, it is no surprise to see that UV analysis gained popularity almost exclusively in Europe. In the USA, UV analysis was hardly applied. One reason could be the fact that, contrary to European countries, data of unfilled vacancies were not easily accessible for the USA. However, the Help-Wanted Index data could be used instead. It seems, therefore, more likely that the limited use of UV analysis in the USA has to be attributed to a deeper suspicion of Keynesianism in the USA. Indeed, in the 1960s, American economists like Friedman and Phelps attacked the Keynesian disequilibrium theory that formed the foundation of UV analysis. Instead they stressed the compatibility of unemployment with equilibrium by putting forward the idea of a natural rate of unemployment. With the fall of Keynesian thinking in the 1970s, measurement of Keynesian concepts, as offered by UV analysis, became more or less redundant, and contemporary economics became less concerned with cyclical or deficient demand unemployment.

Finally, in retrospect, it can be argued that measurement with the UV-device had a significant impact on economic theory. The place of the UV-curve in economic theory was not immediately clear and competing notions derived from the UV- and Phillips-curve framework dominated macroeconomic thinking, particular in the 1960s. Current matching models of unemployment still apply the idea of structural co-movement of U and V data by using a matching function. The correlation between U and V data is therefore still valid, but economists are no longer seeking causal structure at the macro-level.

Chapter 4

Two Approaches for Derived Measurement of the NAIRU

4.1 Introduction

The previous chapters discussed direct measurement by the use of an SQR (Chapter 2) and correlation as a strategy for measurement (Chapter 3). The first case showed that at the dawn of the 20th century, measurement of unemployment was a process that took place without clear theoretical conceptualizations of unemployment as a social or economic phenomenon. The concept of unemployment developed together with measurement. In contrast, this chapter will analyse a case of measurement where theory went ahead of measurement, and consequently, theoretical considerations come into play. Strictly speaking, theory is not a requirement for measurement in the Representational Theory of Measurement (RTM), as measurement in this theory is about bringing about a correspondence between an empirical relational structure and a numerical one. However, when we try to measure theoretical concepts, theoretical considerations will inevitably come into play in the measuring process.

The strategy we seek to explore in this chapter is that of finding an invariant, change-relating generalization from which we can derive quantities for the purpose of measurement. An early account of this kind of derived measurement can be found in Campbell (1928), where he claims that magnitudes (quantities) of a phenomenon can be derived from 'numerical laws', among which he includes deterministic, causal laws, but also relations without explicit causal content, such as identities. Most of this chapter, however, will draw upon more contemporaneous accounts. In particular, we will explore the account of 'invariant generalization' (or 'change-relating generalization') of Woodward (2000a), or 'invariant regularity' as I prefer to call it, for the purpose of derived measurement.

The case study that will illustrate this strategy is the case of measurement of the NAIRU (the Non-Accelerating Inflation Rate of Unemployment). The NAIRU refers to an unemployment rate where inflation tends to be stable. The concept of the NAIRU, and the concept of the Natural Rate of Unemployment, to which it is conceptually related, are not undisputed, and both have given rise to controversies among theorists and methodologists for several decades. The NAIRU is nowadays considered as an important and useful indicator for policy makers in a variety of ways. For example, deviations from the NAIRU indicate inflationary pressures in the economy or can be used for estimating the output gap. It is sometimes taken as a measure for slackness of the labour market or used for assessing the cyclical component of unemployment. Economists have therefore put a lot of effort into measuring this concept and dozens of studies on this topic can be found. In particular, the OECD gave a strong impetus to measurement of the NAIRU. For these reasons, the NAIRU is accepted as textbook economics, and its measurement became, in its relatively short lifespan, a classic case of measurement in macroeconomics. Though the NAIRU is well defined in theory, we shall see in this chapter that attempts to measure it generate other measurement problems than the kind described in Chapter 2.

Measurement of the NAIRU is based on the Phillips curve relation. In the 1960s, the Phillips curve was widely regarded as a representation of a causal relation between unemployment and inflation, and economists claimed that the Phillips curve could provide a reliable instrument for policy purposes and measurement. Nowadays, many economists believe that the Phillips curve can no longer be interpreted as representing a causal relation, as the empirical Phillips curve relation turned out to break down. Other economists, however, most notably Milton Friedman, reject the idea of a short-run Phillips curve, and hold the belief that the Phillips curve is only invariant and vertical sloping in the long run. Due to the empirical instability that the Phillips curve revealed over the last decades, the deterministic interpretation of the Phillips curve has been relaxed and contemporary economists now take the Phillips curve as a structural relationship without clear causal content. In this chapter, we will interpret the Phillips curve as 'invariant regularity' (or 'invariant generalization' in Woodward's sense), in order to make clear the strategy of deriving measures from an invariant, change-relating generalization.

Since we can no longer interpret the Phillips curve as a causal relation, and the derived measurement of the NAIRU turned out to be not completely satisfactory, other ways of measuring the NAIRU have been pursued; in particular 'non-causal' ways. This latter approach involves the use of statistical techniques, such as Structural Vector Auto Regression (VAR) techniques, which are based on correlation rather than causation. The aim of this chapter is thus twofold: i) to explore measurement of the NAIRU as a case of derived measurement from an invariant regularity (the Phillips curve); and ii) to contrast the structural ('causal') approach used to measure the NAIRU with the 'statistical' approach.

This chapter proceeds as follows. Section 4.2 discusses the Phillips curve and the theory and concept of the NAIRU. Since the concept of the NAIRU is historically and conceptually closely related to that of the Natural Rate of Unemployment, this latter concept is discussed in this section as well. Section 4.3 analyses the structural procedures for the measurement of the NAIRU. Section 4.4 elaborates on the account of invariant, change-relating generalization. It questions whether the Phillips curve can be understood as an invariant, change-relating generalization and discusses the measurement problems associated with the structural approach to measurement of the NAIRU. In addition, the section explores how this case fits the requirements of the RTM. In Section 4.5 VAR models are discussed as a representative of the 'non-causal', statistical approach, while in Section 4.6, VAR models are discussed in relation to the RTM. Finally conclusions are drawn in Section 4.7.

4.2 The Phillips curve and the concepts of the NAIRU and the Natural Rate of Unemployment

4.2.1 The Phillips curve

In the 1950s, a discussion arose in Great Britain about whether inflation there was caused by cost-push or demand-pull factors. The rise of import prices could initiate a wage-price spiral and so cause (cost-push) inflation. On the other hand, (demand pull) inflation could arise as a result of money wage increases in tight labour markets. Phillips (1958) wanted to distinguish between these types of inflation, and he conjectured that there was a causal relation between unemployment and inflation (Wulwick, 1989: 173). Phillips used British time-series data of the rates of change of money wages and unemployment from 1861-1957, which he broke down into three periods (1861-1913, 1913-1948 and 1948-1957). He fitted an empirical curve to a statistical scatter diagram of the data of the first period (1861-1913), and compared the latter periods with the first. The 1861-1913 curve that Phillips fitted was non-linear and negatively inclined, indicating an inverse relationship between unemployment and the rate of change of money wages. The curve featured stationary wage increases at an unemployment rate of 5 ½ percent and, when successive years were connected, counter-clockwise cyclical loops due to trade cycles.

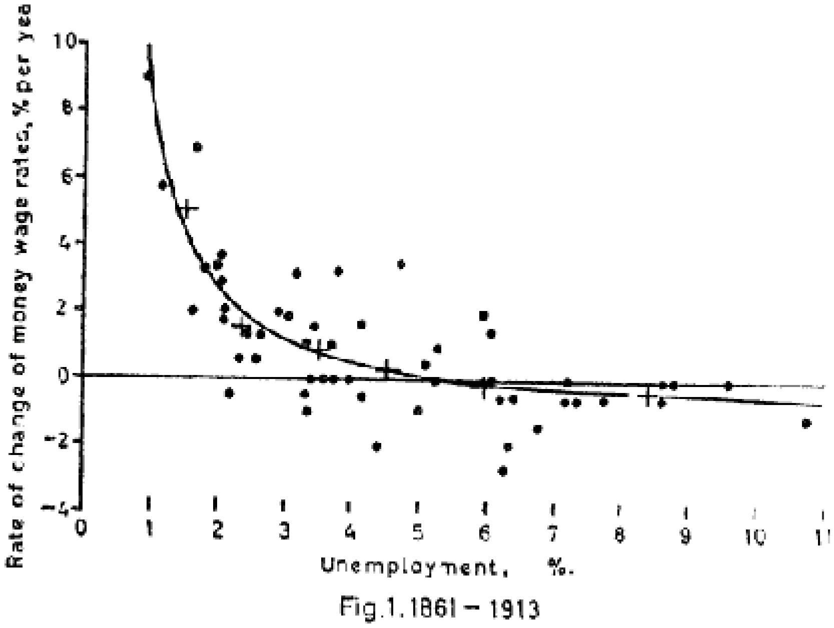


Figure 4.1: Phillips curve
Source: Phillips, 1958: 285.

According to Phillips, in the period 1861-1957, there were "6 ½ fairly regular trade cycles with an average period of about 8 years" (1958: 285). Phillips accounted for the counter-cyclical loops by breaking down the data for each completed cycle. To plot the curve, Phillips used a crude, data grouping technique known as the 'double-averaging procedure'[1] (Wulwick, 1989: 175-176). Phillips formalized the curve as: $\dot{w} + a = bu^{-c}$, where $\dot{w}$ is the growth rate of wages, u the unemployment rate and a, b and c are parameters, and c>0.[2] The Phillips curve thus expressed wage increases simply in a direct relationship with unemployment:

$$\dot{w}_t = f(u_t) \quad \text{with} \quad f' < 0 \text{ and } \quad f'' > 0 . \tag{4.1}$$

Once introduced, the Phillips curve soon became the leading framework of macroeconomic analysis, and rapidly superseded the then dominant IS/LM model. Whereas the IS/LM model was concerned with one policy target, output, the Phillips curve implicitly related two important economic policy targets, unemployment and wage increase (later modified and

[1] In this technique, the horizontal axis of the scatter diagram is divided into 'convenient' intervals, and a curve is drawn through the average of each interval (indicated by the crosses in Figure 4.1).

[2] In his seminal paper, Phillips formalized the curve as: $y + a = bx^c$, with y as the rate of change of the money wage, and x the percentage unemployment. For reasons of consistency with more recent literature, the rewritten more common notation is used here.

replaced by inflation). These dual economic policy targets seemed to be connected by what appeared to be a simple causal relation represented by the Phillips curve itself: that is, low unemployment rates 'cause' inflation. This element was missing in Keynesian theory, and the most important contribution of the Phillips curve was perhaps the recognition that wage inflation could coexist with a considerable amount of unemployment, which undermined the Keynesian ideal of full employment without inflation as the major goal of government policy. In this way, the Phillips curve revealed restrictions on the viability of economic policies that were absent in the IS/LM model. The trade-off aspect of policy targets became apparent after Samuelson and Solow (1960) presented the Phillips curve as a policy menu for governments. Thereby, the Phillips curve was used to clarify the discussion about full employment and how to get there, by providing operational definitions of full employment in terms of unemployment and inflation in the first place. And, given the fact that the Phillips curve was considered as a stable relation that could be measured empirically, theoretical discussions about the exact slope of both the IS and LM curve seemed to have lost their relevance. In addition, the Phillips curve provided an empirically-based way to relax the troubling assumption of fixed prices in the IS/LM model. As a consequence of these novel insights, the Phillips curve itself became the focal point of theoretical and policy discussions, and the attention of economists shifted away from the question of how to derive the appropriate economic policy from the theoretical IS/LM model to discussions about the exact shape of the Phillips curve and a theoretical underpinning of its existence.

4.2.2 The Natural Rate of Unemployment and the NAIRU

The term 'Natural Rate of Unemployment' (hereafter NRU) was coined by Milton Friedman in his 1967 Presidential Address to the American Economic Association (Friedman 1968). Friedman and Phelps (1967) dispute the existence of a trade-off between unemployment and inflation in the long run. In the long run, they argue, the economy tends to an equilibrium unemployment rate – the NRU – that is determined entirely by real factors. Friedman described this unemployment rate as the unemployment rate:

> "that would be ground out by the Walrasian system of general equilibrium equations, provided that there is embedded in them the actual structural characteristics of the labour and commodity markets, including market imperfections, stochastic variability in demands and supplies, the costs of gathering information about job vacancies and labor availabilities, the costs of mobility, and so on" (Friedman, 1968: 8).

Friedman considers the structural and institutional arrangement of the labour market as the cause of natural unemployment. Economies with different structural and institutional arrangements of the labour market will therefore have different NRUs.

Friedman claims to use the term ‘natural’ in the same way as Wicksell used his ‘natural rate of interest’: to separate real shocks from monetary shocks (Friedman, 1968:8).[3] Friedman suggests that actual unemployment can be decomposed into a temporary and a permanent component. The temporary component of unemployment is caused by monetary factors. In the absence of monetary – inflationary – caused unemployment, the remaining unemployment, the natural rate, is permanent. Natural unemployment appears to be a sort of equilibrium search unemployment.

Consistent with this line of reasoning, Friedman argues that the unemployment rate cannot remain below NRU levels. When unemployment is pushed below the NRU level (for example because the government launches a monetary and fiscal expansion in order to reduce unemployment), the economy moves along the Phillips curve and inflation starts to increase. Workers endowed with rational expectations will realize that their real wage has dropped as prices go up and will reduce their labour supply, and unemployment will go up until the NRU is reached again. The economy has now returned to the original unemployment rate, but at a higher rate of inflation. This point lies on a new Phillips curve, above the original one. Friedman’s unemployment dynamics suggests that below-NRU levels are not sustainable and invoked the idea of a long-run Phillips curve that is vertically sloped at the NRU (see Figure 4.2).

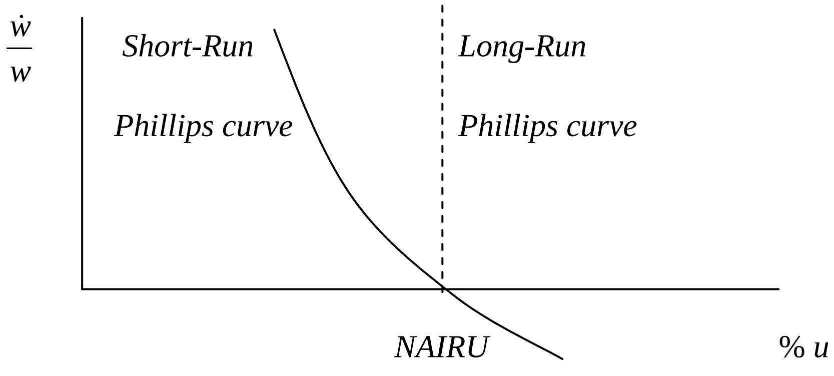


Figure 4.2: Short- and Long-Run Phillips curve and the NAIRU

[3] The term natural rate of unemployment does not therefore seem to be very well chosen, as it does not adequately reflect the idea of distinguishing between real and monetary-caused unemployment. In addition, as Layard et al. (1991) argue, it smacks of inevitability whereas the determinants of the natural rate themselves are partly institutional and therefore not inevitable.

The NAIRU is a concept that resembles the NRU closely. Modigliani and Papademos (1975) were the first to refer to the existence of a NIRU (Non-Inflationary Rate of Unemployment). Later James Tobin (1980) transformed this concept into NAIRU: the Non-Accelerating Inflation Rate of Unemployment.[4] Tobin defines the NAIRU as:

> "the unemployment rate at which the inflation-increasing effects of the excess-demand markets just balances the inflation-decreasing impacts of the excess-supply markets" (Tobin, 1997: 8).

The NAIRU is thus considered as the unemployment rate that is compatible with a stable, non-increasing rate of inflation. It is therefore the unemployment rate at the intersection of the empirical Phillips curve and the horizontal, *u*-axis.

The distinctions between the NRU and the NAIRU are small and subtle, and in recent publications the differences between the NAIRU and NRU have become even more blurred when economists started to distinguish between different kinds of NAIRUs, with as main contenders the short-run and long-run NAIRU.[5] [6] In this chapter, we will refer to the NAIRU as a long-run phenomenon only. But most economists cannot tell the difference between the NRU and the NAIRU, and take them to be more or less the same thing. In particular, monetarists argue that the NAIRU is just another name for the NRU. However, Keynesians and critics of monetarism, like James Tobin, stress that they are definitely distinct concepts.

According to Tobin (1997), three essential features make the NAIRU different from the NRU. In the first place, the NAIRU is a not an equilibrium concept. In contrast with Friedman, who adopts a Walrasian general equilibrium framework in which markets clear by instantaneous prices adjustments, the NAIRU is clearly a disequilibrium concept.[8] The theory

---

[4] The name NAIRU also seems not well chosen. It is often argued that the name gets a derivative wrong and that a more accurate term would be non-*increasing* inflation rate of unemployment.

[5] Richardson et al. (2000: 8), for example, distinguish between:

i) The NAIRU: defined as the equilibrium rate towards which unemployment converges in the absence of temporary supply influences, once the dynamic adjustment of inflation is completed.
ii) The short-term NAIRU: defined as the amount of unemployment consistent with stabilizing the inflation rate at its current level in the next period.
iii) The long-term equilibrium rate (akin to the natural rate): corresponding to a long-term steady state, once the NAIRU has fully adjusted to all supply and policy influences, including those having long-lasting effects.

[6] Some authors, like Cross (1995), argue that, as a consequence of the instability of the Phillips curve, and amendments to the concepts of the NRU and the NAIRU, the NRU and NAIRU became such troublesome concepts that they should better be rejected altogether on empirical grounds. Especially the further distinctions into long- and short-run NRUs and NAIRUs, have made the hypothesis untestable and immune to refutation. According to Cross, the Duhem-Quine problem that hypotheses can be rescued from refutation by suitable amendments of auxiliary hypothesis arises here (Cross, 1995).

[7] The name NAIRU seems also not well chosen. It is often argued that the name gets a derivative wrong and that a more accurate term would be non-*increasing* inflation rate of unemployment.

[8] Tobin stresses that the whole purpose of the NAIRU concept was "to escape the normative equilibrium connotations of the natural rate" (Tobin, 1997: 7-8).

of the NAIRU characterizes markets by simultaneous excess supply and excess demand at prevailing prices. The market is considered as a "never-ceasing inter-sectoral flux of microeconomic demands and supplies" (Tobin, 1997: 9), i.e. a nearly infinite number of submarkets, each being out of equilibrium. For labour markets, excess demand and supply can be observed by the simultaneous occurrence of unemployment and vacancies. Secondly, as a consequence of this approach: "The NAIRU could not be modelled as a single economy-wide market or representative agent" (Tobin, 1997: 8-9). Unlike the NRU, which can be seen as a microeconomic concept, as it is based on rational search behaviour of individual agents, the NAIRU is a macroeconomic concept that has no counterpart in microeconomics. Thus the two concepts are theoretically different. The NAIRU is a Keynesian concept, whereas the NRU fits the Monetarists and New Classical paradigm.

The third distinction, and, for the measurement of the NAIRU, the most important one, follows from the above. Since the determinants of the NRU are the institutional arrangements of labour markets, it follows that each nationally-arranged labour market has its own NRU and that, provided there are no changes in these institutional arrangements, these NRUs must be constant. Changes in the NRU are only possible in the (medium) long run after institutional changes in the labour market. The NAIRU, on the other hand, is clearly a time-varying concept: "The NAIRU varies from time to time as the relationships between unemployment, vacancies and wage changes vary, and as the dispersion of excess demands and supplies across markets changes." (Tobin, 1997, i). Thus, for our purpose of measurement, this is the most relevant distinction to keep in mind: the NAIRU allows for variability, while the NRU is constant in the short term.

4.3 Measurement of the NAIRU: The structural approach

In attempts to measure the NAIRU, two basic approaches can be distinguished: the structural approach, and the statistical approach.[9] Structural methods involve the specification of the wage- and price-determining relationship, and are – in contradiction with Tobin's conception of the NAIRU – most often based on some underlying rational expectations theory of agents' (firms' and trade unions') behaviour. The outcome of these agents' behaviour is the empirical unemployment-inflation relation: the Phillips curve, and the NAIRU is then the unemployment level, which is consistent with stable price and wage inflation, i.e. the level of unemployment where inflation starts to 'take off'. The most favoured way of deriving measures of the NAIRU

[9] A third approach for measurement of the NAIRU that is sometimes used is the reduced-form approach. It combines the two approaches, and uses both structural specification and statistical methods.

is through the use of a structural approach, though recently the statistical approach has gained popularity. We will analyse the structural approach first.

The basic framework for the measurement of the NAIRU is the (simple) Phillips curve relationship. Soon after its introduction, economists made a first modification to the Phillips curve relation by replacing the rate of wage changes by the rate of price changes. The (simple) Phillips curve relationship then expresses the relation between inflation and unemployment as:

$$\pi_t = -\beta(u_t - u^*)\,, \tag{4.2}$$

where π_t is the actual rate of price inflation, u_t the actual unemployment rate, u^* the NAIRU, and β a parameter. The term $(u_t - u^*)$ represents the deviation of the actual unemployment rate from the NAIRU, and is often referred to as the 'unemployment gap'.

However, since Friedman (1968) two other modifications have been made in the representation of the Phillips curve. First, the adaptively (backward-looking) formed expectations of inflation were replaced by rational (forward-looking) expectations. And, secondly, after the experience of the 1973 oil crisis, (exogenous) supply shocks were introduced as a cause of inflation. In the most general form, the *expectations augmented* Phillips-curve relation is formulated as:

$$\pi_t = \pi_t^e - \beta(u_t - u^*) + \gamma X_t + \upsilon_t\,, \tag{4.3}$$

where, π_t^e is expected price inflation, X_t a regressor included to control supply shocks, γ a parameter, and υ_t an error term. When expectations about inflation are realized, that is, when expected inflation coincides with realized inflation, thus when $\pi_t - \pi_t^e = 0$, and in the absence of supply shocks ($X_t = 0$), the NAIRU u^* will coincide with the actual unemployment rate u_t.

In more recent practice, economists follow a slightly different route. Most studies on the measurement of the NAIRU derive the NAIRU from a model of imperfect competitive bargaining that is popularized by Layard, Nickell and Jackman (1991).[10] This model describes the supposed underlying price and wage bargaining process that could lead to wage-price spirals. Critics sometimes refer to this model as the "textbook approach", the "battle of mark-ups" model, or, as the "OECD NAIRU consensus" model,[11] since the OECD frequently uses

[10] Gordon (1982, 1997) popularized a slightly different approach in the USA.
[11] For example by Mitchell and Muysken (2003).

this model. The idea behind imperfect competitive bargaining is that both employers and firms have some market power, since hiring or firing workers is costly for both parties. These costs are, however, a function of the level of unemployment. In these models, firms set prices as a mark-up on expected wages and workers (unions) set wages on expected prices. Firms set prices as a mark up on expected wage ($p-w^e$) as a positive function of employment. The price-setting function of firms then becomes in its most elementary form:

$$p-w^e=\beta_0-\beta_1 u_t, \tag{4.4}$$

where p is log prices, w^e is log expected wages, u_t the actual unemployment rate. The wage-setting function of unions describes wage setting by unions as a negative relation between the unions' wage mark up $w-p^e$ and employment:

$$w-p^e=\gamma_0-\gamma_1 u_t, \tag{4.5}$$

where w is log wages, and p^e is log expected prices. In more extended models, additional variables are added. Usually a regressor X_{1t} is added to both the price and the wage-setting equations, which represents a vector of variables, such as an increase in labour productivity, that exogenously raises prices and/or wages. The wage-setting equation is also often extended with a vector of variables X_{2t}, such as union power. So, the price setting behaviour of firms will then be:

$$p-w^e=\beta_0-\beta_1 u_t+\beta_2 X_{1t}. \tag{4.6}$$

The wage-setting behaviour of unions:

$$w-p^e=\gamma_0-\gamma_1 u_t+\gamma_2 X_{1t}+\gamma_3 X_{2t}. \tag{4.7}$$

For the sake of the argument, it suffices to proceed with only one regressor, X_{1t}, so we will ignore X_{2t} in the remainder of this chapter. A stable inflation will now occur when expectations about future prices and wages are realized, thus when the identities $p=p^e$ and $w=w^e$ hold. When expectations are not fully realized, a wage-price spiral will occur, as wage-setters try to regain the losses imposed on them by price setters, and vice versa, and inflation will start to accelerate. Therefore, equilibrium, and hence non-increasing inflation, exists at the intersection of the wage-setting and price-setting curve, creating a NAIRU-level of unemployment of u* (see Figure 4.3).

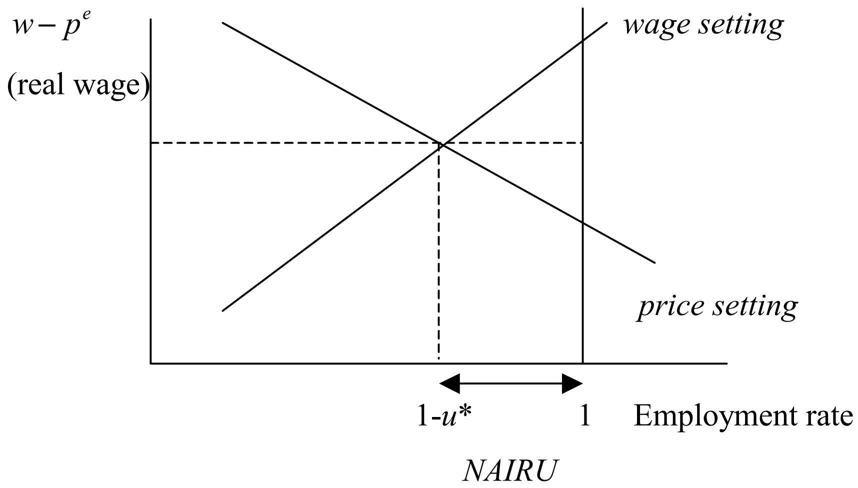


Figure 4.3: Price- and wage- setting equilibrium
Source: Based on Layard, Nickell and Jackman, 1991: 14.

Thus, by using this equilibrium identity, the cumbersome expectations variables cancel out and a more comprehensible equation remains. The equilibrium rate of unemployment in the simplest case is then:

$$u^* = \frac{\beta_0 + \gamma_0}{\beta_1 + \gamma_1}, \qquad (4.8)$$

which we can think of as the NAIRU. In the more elaborated case, the NAIRU becomes:

$$u^* = \frac{\beta_0 + \gamma_0 + (\beta_2 + \gamma_2) X_{1t}}{\beta_1 + \gamma_1}. \qquad (4.9)$$

4.4 The Phillips curve as an invariant regularity for measurement?

According to Campbell (1928), a strategy for the measurement of unobservables is to 'derive' a quantity from a 'numerical law' in which the unobservable of interest participates, and through which it is connected to other variables that are observable (measurable). In Campbell's classification of measurement, he considers this as *derived* measurement (see Appendix 1.B in Chapter 1), which he defines as "measurement by means of constants in numerical laws", where a numerical law "relates the results of measurements of two or more quantities", and “constants” refer to constant parameters in a system, such as the gas constant in the ideal gas law or the constant of gravitation in Newton’s law (Campbell cited in Ellis, 1966: 54). More formally, derived measurement is possible if (Ellis, 1966, 56):

(a) the systems which possess *p* (the quantity to be measured) possess other quantities *q*, *r*, *s*,…, which are independently measurable;
(b) there are laws relating some of *q*, *r*, *s*,…, in which constants occur which vary from system to system;
(c) there is at least one system-dependent constant *c*, which is such that when the systems are arranged in the order of *c*, they are also arranged in the order of *p*.

As to the nature of this physical relation between magnitudes (variables) in the ‘numerical law’, Campbell argues that it “is always some kind of uniform association; it is the relation between things which are such that the presence of one is always an indication of the other” (1928: 58). A necessary conjunction of variables is thus essential. This may seem to suggest a *causal* law, but that stretches the point too far, as a causal law is not necessary for derived measurement. The following example clarifies this point.

An example of a numerical law could be Ohm’s law, which states that the current flowing along a wire (I) is directly proportional to the potential difference along the wire (U), and inversely proportional to its electrical resistance (R), formally written as: $I = \frac{U}{R}$. Multi-meters for the measurement of electric tension, current or resistance are based on this numerical law. While the causal order in Ohm’s law is unclear, it is also irrelevant. That is, for measurement it is irrelevant whether current ‘causes’ tension or the other way around. Identities of the kind $m = \rho \times V$ (mass = density x volume), therefore also fall under ‘numerical laws’ (see Campbell’s example, 1928: 57).

Woodward (2000a) provides a similar, but more sophisticated account. Woodward argues that the notion of ‘invariant generalization’ or ‘change-relating generalization’ can circumvent many of the problems involved in claims about lawfulness. For Woodward, ‘invariance’ means that a relation between two or more variables remains stable or unchanged under changes, in both background conditions and changes (‘interventions’) in variables figuring in the relation itself (Woodward, 2000a: 205). As a consequence, invariant generalizations relate – like causal laws – to changes between variables, and this is the principle we exploit for derived measurement. What distinguishes invariant, change-relating generalizations from laws, however, is not the necessity of changes of the variables participating in the relation but their domain: the range of changes over which the relation remains stable. Whereas lawfulness requires an infinite domain, invariant generalizations are a matter of degree, and typically have a far more limited domain. However, as long as the domain of the invariant generalization is wide enough for the purpose of our measurement, we can use these generalizations to derive our measures.

To summarize, what matters for derived measurement is whether the relation between the variables is *invariant* over a wide enough range of background conditions and variations of the variables involved. That provides the principle of reliability for derived measurement, not the causal ordering.

In Campbell's account, the constants in the mathematical representation determine the scale of measurement, and are therefore referred to as 'scale-dependent constants' of the system. However, we do not always speak of the scale of the quantity measured. Ohm's law for example, does not seem to have a scale-dependent constant (or constants), since we express the law with respect to our Volt, Ampere and Ohm scale. Ohm's law, however, could be postulated as $U = kIR$. The constant k in this mathematical representation is determined by the scale of measurement, and changes in k makes scale transformations according to the uniqueness condition permissible, and thus measurement on other scales possible. When the other variables, I and R, are known or independently measurable, a unique, isomorphic structure follows from Ohm's formula. An invariant, change-relating generalization can thus bring about the isomorphic mapping between an empirical and a numerical relation structure, which the RTM requires.

Let us now analyse how this account applies to the case of measurement of the NAIRU. Of crucial importance for the measurement procedure of the NAIRU – at least for the structural approach – is the Phillips curve. In his seminal article, Phillips is silent as to the nature of the empirical relationship he finds. Probably, Phillips did not consider it a structural relationship. Schwier (2000: 25) argues: "His curve was definitely *not* structural. It was a prediction relation – a crude one but he thought it did the job. We specifically asked about this matter of being structural, and Phillips gave us a very empathic 'no'." Soon after its appearance, however, Samuelson and Solow (1960) presented the Phillips curve as an economic policy menu according to which governments could 'trade off' unemployment for inflation. This paved the way for treating the Phillips curve as a causal, or lawlike relation, and throughout the 1960s the Phillips curve was treated as such: it was "widely interpreted as a causal relation that offered a stable trade-off to policy makers" (Friedman, 1976: 270). In the 1960s, the relation between unemployment and inflation was widely regarded as a simple causal relation represented by the Phillips curve itself: low (that is, below NAIRU) unemployment rates ‘cause’ increasing inflation. However, as is now well understood by economists, the empirical Phillips curve relation broke down, and the causal interpretation of the Phillips curve did not appear to be sustainable. The Phillips curve therefore turns out not to be a deterministic law, in the sense that increasing inflation will necessarily or reliably happen below NAIRU unemployment rates. Therefore, economists currently treat the Phillips curve as a relation without an explicit causal content. In the structural approach, the Phillips curve is modelled as a structural relation, which results from price and wage behaviour-setting of agents.

Recall from equation (4.2) that the simple Phillips curve is commonly expressed as:

$$\pi_t = -\beta(u_t - u^*)\,. \qquad (4.2)$$

In this elementary form, the Phillips curve serves as an invariant regularity. The actual unemployment rate u_t is independently, and directly measurable (by, for example, a standardized quantitative rule (SQR): see Chapter 2), and so is inflation π_t. When we take parameter β as a system-dependant constant, we have established an isomorphic mapping from which variations of the NAIRU follow. The same holds for the more elaborated expectations augmented Phillips curve (equation 4.3).

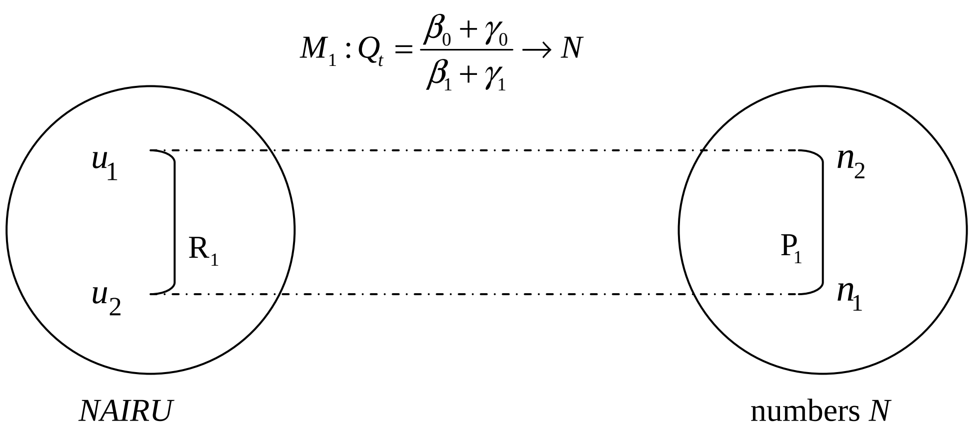


Figure 4.4: Diagrammatic presentation of isomorphic mapping between the NAIRU and numbers, based on the Phillips curve as an invariant regularity

In Chapter 1 of this thesis, we have seen that the RTM has two strict requirements for sound measurement. First, it must been shown that a relevant, empirical structure of a phenomenon is isomorphic with an arithmetical structure. This is known as the representational condition (or theorem). Secondly, the uniqueness condition states that, for unique measures, two or more isomorphic mappings must be related by certain permissible transformations, making transformations of one kind of scale into another. If we interpret the Phillips curve as an invariant regularity, it must provide an isomorphic mapping between the phenomenon (the NAIRU) and numbers, as presented in figure 4.4. Whether such an interpretation is justified must follow from experience. As the previous chapters have shown, this is precisely the problem when the phenomenon of interest is unobservable.

Unfortunately, the structural approach to measuring the NAIRU and the interpretation of the Phillips curve as an invariant regularity encountered a number of problems. Four major problems – some of which are related – will be discussed here.

1 Uncertainty from measurement equation

In the late 1960s, doubts began to arise about the stability of the Phillips curve, as empirical studies revealed large degrees of variance. This became most apparent in the early 1970s, when Western economies experienced stagflation, and their Phillips curve relation shifted outward to the right. Economists considered the relation between unemployment and inflation still invariant, but not stable; that is, subject to shifts over time as a result of influences yet to be determined. Finally economists became convinced that there was not just one invariant Phillips relation, but that, as a result of the shifts over time, a whole series of short-run Phillips curves could exist, each one lying closer to or further from the origin. Econometricians' usual way of dealing with this kind of problem in structural modelling is by 'making' the relation stable by including additional, presumably causal, variables into the specification, in order to account for the shifts. For this reason, the regressors X_{1t} and X_{2t} enter into the wage- and price- setting equations (in conformity with equation 4.6 and 4.7), and through time, variables such as import prices, capacity utilization rate, growth rate of unit labour costs, current profit to sales ratio, etc., have been included in the wage- and price- setting equations. A serious epistemological problem arises here, since we do not know which model for determining the NAIRU is the 'true' one. Alternative specifications are arguably equally plausible but lead to substantially different measures for the NAIRU. Most imprecision in the measurement of the NAIRU arises from this type of uncertainty (Staiger et al., 1997a), which cannot be incorporated in any confidence interval. It does not question the magnitude of parameters, but the definition of operational key variables. Consequently, as Setterfield argues: “The estimates of the NAIRU depend heavily on the specification of the estimated equations, the operational definition of key variables, and the data sample period.” (cited in Horst et. al., 1996: 1). For the Phillips curve to be an invariant regularity implies that we do know that we have the ‘right’ isomorphic structure from which to derive our measures. Needless to say, we do not know this, a problem that causes serious problems for any measuring device.

2 Identification problem of the measurement equation

In the structural approach to measurement of the NAIRU, an identification problem arises. This means, in short, that the system of equations (equations 4.4 and 4.5) contains more variables than equations, which renders the parameters of the set of equations as unknown. This is particularly the case for the wage-setting equation (equation 4.5). In modelling this equation, it became standard practice to include two vectors X_t. One vector X_{1t} for controlling productivity-related variables, and one vector X_{2t} for labour market efficiency related variables.

The implications are that the wage-setting curve is unknown, and hence the NAIRU. For measurement of the NAIRU, the identification problem is solved as follows. It is assumed that the firms are on the labour demand curve (price-setting curve) and, by adding arbitrary exclusion restrictions and/or ad hoc dynamics, the parameters of the model can be identified.

Solving the identification problem and identifying the parameters is thus ultimately only possible by making arbitrary assumptions: “In practice, identification is achieved by arbitrary exclusion restriction (..) and/or ad hoc dynamics. There is a certain act of faith in assuming that one gets sensible results from such practices. But many researchers would argue that unless additional ad hoc assumptions are imposed on the wage-setting process, these problems of identification are unavoidable” (Manning, 1993).

3 Variability of the NAIRU

Identification of parameters became, in addition, very difficult because the NAIRU itself is time-varying. This variability makes econometric estimations of Phillips curves difficult. In the econometric literature, this problem is referred to as the “simultaneity problem” and “lagged endogeneity”. The simultaneity problem exists because of what is called a feedback mechanism: inflation affects wage growth, which feeds back on inflation, ad infinitum.

Simultaneity makes it difficult to evaluate the coefficients linking wage growth and inflation. Lagged endogeneity refers to the fact that current wage inflation is determined in part by its own past values. This presents certain technical complications in the estimation process, because it is difficult to determine how much of the current value of wage growth depends on past wages and how much depends on other factors such as unemployment or inflation.

As a consequence of the difficulty of identifying parameters, the 95% confidence interval for all computations covers a very wide interval for the NAIRU, usually somewhere between 4 percent and 8 percent. Staiger et al. (1997b: 34) ague: "The most striking feature of these estimates is their lack of precision. For example, the 95 percent confidence interval for the current value of the NAIRU based on the GDP deflator is 4.3 percent to 7.3 percent. In fact, our 95 percent confidence intervals for the NAIRU are commonly so wide that the unemployment [of the USA] has only been below them for a brief periods over the last 20 years." Other studies find similar wide confidence intervals. This empirical problem arises from not knowing the parameters of the model concerned. A wide range of values is consistent with the empirical evidence.

4 Role of economic policy

Though the exact cause or causes of the instability of the Phillips curve are not yet fully understood, the fact that the Phillips curve is also used as an important instrument of economic policy certainly adds to the instability of the curve. Phillips curves closer to the origin are taken to represent more efficient labour markets. More efficient and transparent labour markets make it easier for firms to recruit new workers, and the bidding-up process that leads to a wage-price spiral will tend to start at lower unemployment rates. A more flexible and transparent labour market will therefore lead to a low combination of unemployment and inflation. For this reason, governments are eager to 'shift the Phillips curve inwards deliberately' by policies aiming at more efficient matching of supply and demand for labour. However, this contributes to the instability of the Phillips curve, and makes it a flawed device for measurement. This problem is well known in economics, and is sometimes referred to as "Goodhart's law".[12] It states that "any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes" (Goodhart, 1984: 96). Hoskin (1996) restated Goodhart's "law" as: "When a measure becomes a target, it ceases to be a good measure". So, either we can use the

[12] Named after Chief Adviser to the Bank of England, Charles Goodhart.

Phillips curve as an invariant regularity for measurement, or we it can use as an instrument for economic policy, but we cannot have it both ways.

To summarize, in principle the Phillips curve could be explored as an invariant regularity for measurement, and the structural approach tries to do so. The major problem this approach encounters, however seems, to be the epistemological problem of finding the right specification of the Phillips curve. Rather than the simple and unique mapping

$$M_1 : Q_t = \frac{\beta_0 + \gamma_0}{\beta_1 + \gamma_1} \to N \,,$$

we are dealing with the more complex mapping

$$M_2 : Q_t = \frac{\beta_0 + \gamma_0 + (\beta_2 + \gamma_2) X_t}{\beta_1 + \gamma_1} \to N \,,$$

where X_t represents the whole, unknown array of possible variables that could account for the shifts of the Phillips curve: causal variables, policy changes, shocks, etc. Thus:

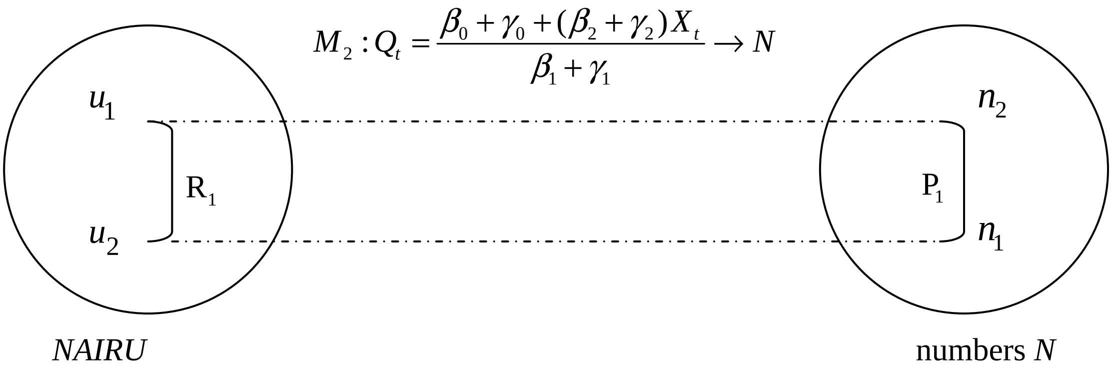


Figure 4.5: Diagrammatic representation of the mapping between NAIRU and numbers in extended structural models

This leaves us with two consequences. First, we have the problem of picking the ‘right’ invariant regularity, as this case leaves us with multiple empirical structures that can bring about mappings between number and the phenomenon. In principle, multiple mappings should not be a problem for measurement. In fact, it is the uniqueness condition that requires that, between unique isomorphic mappings, transformations of structures, and hence transformations of scales, must be permissible. In this case, however, we are not sure which empirical structure captures the phenomenon in the first place. And even worse, we are unsure how the multiple empirical structures relate to one other. We end up with multiple empirical relational forms for the same phenomenon, without the ability to establish unique isomorphic mappings and the relation between these mappings. Needless to say, this violates the uniqueness condition. Secondly, the other side of the coin is that the phenomenon of the NAIRU appears does not appear to be as well defined as we might think, as it leaves space for all sorts of variables to enter into the analysis, a point also indicated by the recent amendments to the concept of the NAIRU.

4.5 The statistical approach: Vector Auto Regression models

The second approach to measurement of the NAIRU started to gain influence in the 1980s. This approach involves the use of purely statistical methods like VAR (Vector Auto Regression) models, filters, such as the Hodrick-Prescott filter and band pass filters, or random walk models. We will focus here on Vector Auto Regression (VAR) models, as they seem most frequently applied for the measurement of the NAIRU. The econometric methodology of this approach was developed by Christopher Sims. Sims (1980) argues that the theoretical restrictions imposed on structural simultaneous models (in the tradition of the Cowles Commission), which are necessary for the identification of such models, are “incredible”. In response to the “Lucas critique”, and the lack of success of large structural equation models in the 1950s and 1960s, Sims suggests the use of a multivariate, autoregressive moving average model. Variables are modelled as vectors and describe the variable dynamics from their own history. In contrast with the Cowles Commission approach, in VAR models there is no a priori division of the variables into endogenous or exogenous and no a priori restrictions for identification are imposed on the variables. This means that all variables, including lagged variables, are considered endogenous and an unrestricted (by theory) VAR model regresses each (non-lagged) variable on a small set of (current and lagged) variables. As a consequence of the refusal to specify exogenous and endogenous variables, and so to specify an a priori causal structure in the model, it has been argued that VAR models are ‘atheoretical macroeconometrics’ (Cooley and Leroy, 1985).[13]

[13] Cooley and LeRoy (1985) argue in their critique of VAR models that such models do not put economic theories to the test, that the mathematics of VAR models is clear but their economic interpretation is not, and that VAR models are not useful for *ex post* policy analysis. They see limited use for VAR models as a tool in *ex ante* policy forecasts and data description.

Let us have a closer look at Sims econometric methodology. A formal treatment of VAR models is presented in Appendix 4.A of this chapter. Less formally, we can summarize Sims methodology as follows (Pagan, 1987:15-19, and similarly Canova (1995:61)):

(i) Decide which variables should enter the model and transform the data to such a form that a VAR can be fitted to it;

(ii) Choose as large a number of lagged variables that is compatible with the size of the data set available, and then fit the resulting VAR;

(iii) Try to simplify the VAR by reducing the number of lags, or by imposing some arbitrary 'smoothness' restrictions upon the coefficients;

(iv) Use the orthogonized innovations [the unexpected part of the variables or disturbances] representations to address the questions of interest.

Step (i) includes setting up the VAR model for the desired purpose, from an unrestricted (by theory) structural model (similar to equation 4.15: see Appendix 4.A), while step (ii) decides the number of lags *p.* The number of lags is usually determined by a statistical selection criterion such as Akaike or Schwarz Information Criterion in order to find a balance between overparametization and oversimplification of the model. Next, it is assumed that the VAR model is hit by shocks (innovations), which serve as the input of the linear dynamic system. In order to do this, we need the Moving Average (MA) representation of the VAR. This can be obtained by application of *Wold's decomposition theorem*.[14] This theorem states that any zero mean, covariance stationary process can be represented as a moving average sum of a deterministic constant (or a constant plus trend), which is a function of time *t*, and white noise processes which are temporal, cyclical deviations from the trend. Formally, we can represent the Wold decomposition theorem as equation 4.10, provided it has zero mean covariance and the weights A_1^i of the error term e_t are square summable:[15]

$$x_t = \mu + \sum_{i=0}^{\infty} A_1^i e_{t-i} \, , \qquad (4.10)$$

[14] Herman Wold (1908-1992) was a Swedish statistician who worked mainly on time-series analysis. He disagreed with Haavelmo's probability approach, and suggested the use of recursive causal chain models. His decomposition theorem followed from his PhD thesis.

[15] See, for example, Greene, 2003: 619.

where μ represents a deterministic constant or a constant plus trend (for the vector x_t), and $\sum_{i=0}^{\infty} A_1^i e_{t-i}$ the white noise process which is considered to represent the temporal, cyclical deviations from the trend. The coefficients of matrix A_1^i are called the *impulse response coefficients* associated with the innovations (shocks) e_t. They describe the propagation of the shocks through the model, and the final effect of the shocks on x_t. For the measurement of the NAIRU, the important assumption is thus made that the actual unemployment rate can be split up, according to Wold's decomposition theorem, into a trend μ (the NAIRU) and the cyclical component $\sum_{i=0}^{\infty} A_1^i e_{t-i}$ ('unemployment gap'), written as:

$$u_t = \mu \text{ (the NAIRU)} \qquad + \qquad \sum_{i=0}^{\infty} A_1^i e_{t-i} \text{ ('unemployment gap').}$$

The unobservable level around which the actual unemployment rate fluctuates is now regarded as the NAIRU, and the NAIRU and 'unemployment gap' are assumed to be uncorrelated. The next steps involve the identification of the shocks (step iii), and the interpretation of the shock impulse response function (step iv).

There are a number of ways to achieve identification. Sims proceeds by analysing the moving average representation of a system with *orthogonalized* innovations (step iv). This means that the error covariance matrix is diagonalized by multiplying the moving average representation of the VAR by a unique triangular matrix with units (1's) on the main diagonal. This approach to identification is known as a *Choleski decomposition*. An alternative way to achieve identification (step iv) is to use the *Blanchard-Quay decomposition* (Blanchard and Quah, 1989).[16] Since this latter approach to identification is often used for the measurement of the NAIRU, and in addition, is more intuitive to the reader, we will leave Sims approach here (step iv) and now follow the Blanchard-Quah approach instead. In the Blanchard-Quay decomposition a Wold decomposition is applied, but, for identification, restrictions are imposed on the coefficients of the VAR by using some specific economic theory of long-run

[16] Blanchard and Quah apply their model to GNP and unemployment, and their model is used for the determination of potential output and the 'output gap'. In recent years, the Blanchard-Quah model was modified for measurement of the NAIRU.

neutrality of one variable on another (here money on employment).[17] The shocks are taken as being transitory. The long-run effects, and hence the long-run dynamics of the VAR model, are then known and used for restricting the coefficients of the model. Conversely, sometimes the short-run dynamics are known and not the long run outcome. In that case, we could use knowledge about the short-run behaviour for imposing restrictions on the coefficients.

Finally, the NAIRU is identified as the part of unemployment, which is inflation-neutral in the long run. That is, the gap disturbance (e_t) has no impact on the NAIRU in the long run. Therefore, the NAIRU is identified by setting the temporary component of u_t (the 'unemployment gap') to zero, $\sum_{i=0}^{\infty} A_1^i e_{t-i} = 0$. The NAIRU is thus operationally defined as the component of the actual unemployment rate that is uncorrelated to inflation in the long run, and the Wold decomposition theorem provides the measuring instrument that, applied to the VAR, turns that set of equations into a measuring instrument for the NAIRU.[18]

4.6 The Representational Theory of Measurement and the statistical approach

Understanding the empirical structure is a crucial step in any measurement procedure and the construction of measuring instruments. Defining the empirical relational structure in the VAR model is not immediately straightforward. VAR modellers explicitly reject the view that we know, a priori, the causal relations needed for measurement, so instead, they aim to use the statistical relations uncovered by the VAR procedure to determine statistically-defined causal relations. Since the VAR model runs regressions on past values of variables, they rather exploit conditional correlations, which are identified with the statistically-defined Granger-causal ordering. A variable *x* Granger-cause *y* (denoted as $x \rightarrow y$) if: "present *y* can be predicted with better accuracy by using past values of *x* rather than by not doing so, other information being identical" (Charemza and Deadman, 1992: 190).[19] So, a weak form of causation – Granger causation – is applied in the VAR model. However, in Sims methodology the test of Granger-

---

[17] Since this step requires specific economic, causal assertions (in a similar way as the Cowles Commission structural approach), Blanchard-Quah type autoregression models are considered as Structural VAR (SVAR) models.

[18] This statistical approach is applied for the measurement of the NRU too. Since the NRU is constant in the short term and independent of inflation, we can decompose unemployment according to the Wold theorem into the NRU and an unemployment gap, and run an autoregression on unemployment data alone.

[19] We speak of Granger causation if the addition of information X (at time t-1) to the set of all past and present information $U_{\text{t-1}}$ (at time t-1) leads to a reduction of the mean square error (MSE) of the unbiased prediction $\tilde{y}_t$. Thus formally there is Granger-causality if: $MSE(\tilde{y}_t|\mathrm{U}_{\text{t-1}}) < MSE(\tilde{y}_t|\mathrm{U}_{\text{t-1}}\backslash \mathrm{X}_{\text{t-1}})$, where MSE stands for the mean square error of predicting $\tilde{y}_t$ conditional on information set $U_{\text{t-1}}$ (Charemza and Deadman, 1992: 191).

causality is only relevant for verifying endogeneity or exogeneity after the model is already constructed, and Sims takes it that Granger non-causality verifies strict exogeneity. All variables in the VAR model are endogenous, but not in the sense of the Cowles Commission who stick to the traditional, Humean interpretation of causality. In Sims methodology, VAR models are essentially correlations, and so is the empirical structure upon which measurement of the NAIRU rests. Though the original VAR specification was deliberately atheoretical, economic theory cannot be eschewed completely. Economic theory enters in the analysis of the shock impulse response function and the variance decomposition. Economic theory tells us, for example, that real (unemployment) shocks will affect inflation in the long run, but not the other way around.[20] In the end, the economist imposes a causal structure upon the otherwise correlational structure of the estimated VAR in order to interpret it.

The Moving Average representation of the VAR, which was obtained by the application of the Wold theorem, forms the direct mapping of the phenomenon – the NAIRU – with numbers. The diagrammatic representation of this case of measurement therefore looks like this: (Figure 4.6).

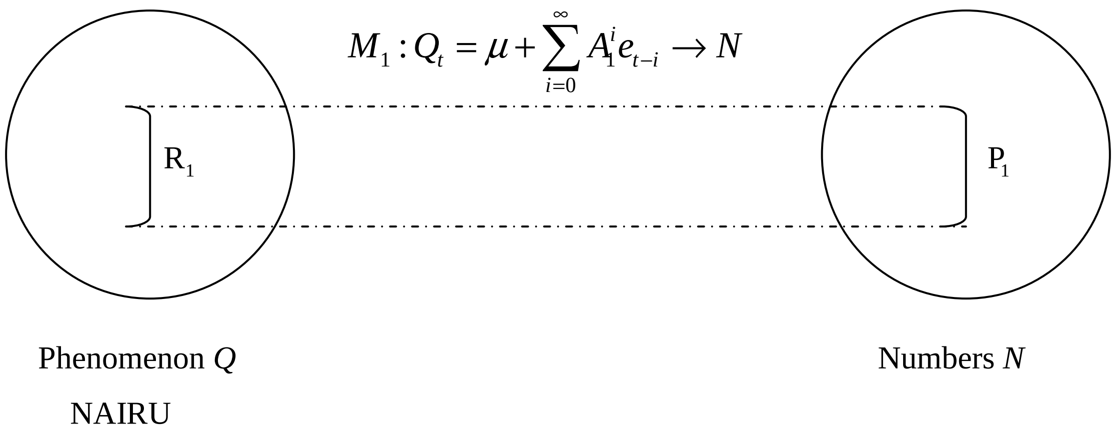


Figure 4.6: Diagrammatic representation of derived measurement by VAR models using the Wold decomposition theorem

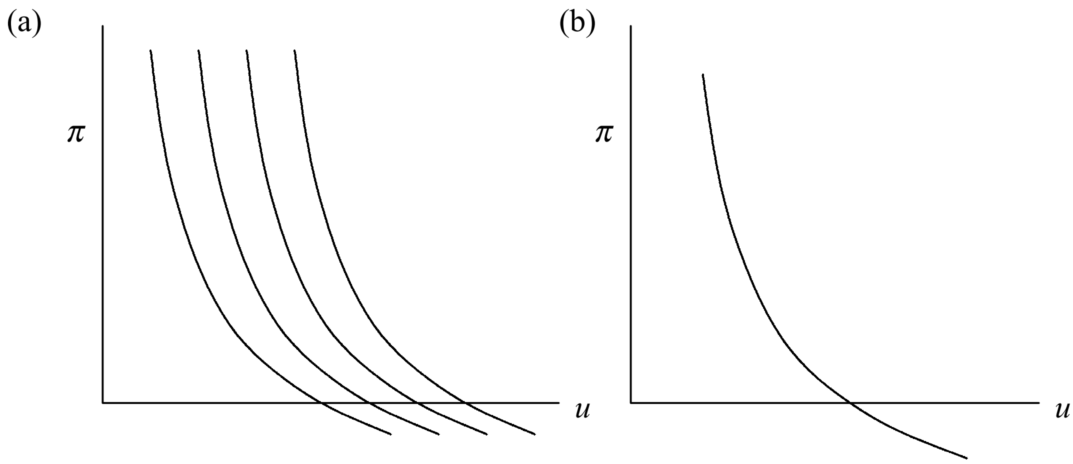
[20] In addition, Cooley and LeRoy (1985) argue that orthogonal innovations need to be treated as exogenous variables. This requires imposition of a prior causal structure upon the system in a similar fashion as the Cowles Commission methodology (see also Qin, 2006).

4.7 Conclusions

Campbell (1928) and Woodward (2000) provide an account of invariant regularities that can bring about a correspondence between phenomena and numbers. Various types of regularities can be used: causal relations, correlations, identies, etc, as long as the physical relation can be expressed in a mathematical relation and is invariant over a wide range of background conditions and changes in the variables figuring in the relation. This may seem to resemble Heidelberger's idea of the use of correlation and a measurement formula that expresses the isomorphic mapping for measurement (Chapter 3). However, in Heidelberger's account the correspondence between phenomenon and magnitudes (numbers) is restricted to association, whereas Campbell, for example, allows a more correlative interpretation that also seems to allow causal relations and identies.

The concept of the NAIRU appeared to be a clearly-defined theoretical concept when it was first introduced. The problem was only finding a satisfactory measurement procedure for it. Two approaches to the measurement of the NAIRU can be distinguished: a structural approach and a statistical approach. Both approaches derive measures of the NAIRU from an apparent invariant regularity, though they both employ a slightly different method for deriving their numbers.

The structural approach explores a causal strategy, and tries to capture causal determinants of NAIRU. It models the wage bargaining process that underlies the Phillips curve. For a long time, this unemployment-inflation relation was considered as a causal relation, and causal relations could be used satisfactory for providing an isomorphic structure. However, this structural approach, though still most often used, is not fully satisfactory from a measurement perspective. The main problem concerns the identification of the Phillips curve, and hence the establishment of a strict isomorphic mapping. The contemporary interpretation is that the Phillips curve represents an unstable relationship and is therefore unsuitable to serve as a measuring instrument of the NAIRU. This is not quite right, as it is not the instability itself that creates the problem. In the case of a perfectly stable Phillips curve, the NAIRU would be a constant variable (see Figure 4.7(b)). However, since the NAIRU is not constant but a time-varying concept, it will be clear that there will a whole set of Phillips curves, that necessarily moves along with the NAIRU. Without movements of the Phillips curve the variability of the phenomenon could not be captured. The problem is that the exact shape of the Phillips curve cannot be identified, and it is therefore not possible to make a clear distinction between the variability of the Phillips curve and the variability of the NAIRU.



Instability of the Phillips curve
Variability of the NAIRU

Stability of the Phillips curve
Invariability of the NAIRU

Figure 4.7: Consequences of interpretations of the Phillips curve for measurement of the NAIRU

This problem gives rise to great variability of NIARU values and wide reliability intervals. This problem was also apparent in Chapter 3 in the case of the UV-analysis. For this purpose, economists have tried to deal with the instability of the Phillips curve by adding explanatory variables into the Phillips curve equation in order to 'make' the Phillips curve stable and so rescue the causal measurement strategy. However, this did not yield the desired result. And since economic theory is not able to provide exclusive information about the causal factors involved, a multitude of alternative, equally plausible specifications exists in the structural approach literature. Just as in Chapter 2, we ended up with 2 SQRs for the same phenomenon, in this case, we end up with multiple mappings for the measurement of a phenomenon. However, in this case, multiple mappings do not follow from uncertainty at a conceptual level of the measurand, but do we find multiple empirical, relational forms.

Given the these problems, it is no surprise to see that the statistical approach is gaining popularity, as it circumvents many of the problems inherent to the structural approach. The statistical approach is thus an attractive alternative to the problems of structural modelling and eschews the search for simple structure. In addition, non-structural methods require fewer assumptions than structural methods, and are therefore to be preferred from a modelling point of view. In the statistical approach, measurement of the NAIRU is based on correlation. The variable of interest is derived from an autoregression on past values of inflation and unemployment. Measurement of the NAIRU in the statistical approach is thus, in essence, based on exploring a correlation, but the way the correlation is used is different from the strategy analysed in the previous chapter. The strategy of Chapter 3 was to look for a correlated observable that could 'stand-in' for the unobservable. In this way, the unobservable was measured indirectly. The case of the VAR models is very different, as it does not rely on a representative or stand-in. The unobservable variable of interest – the NAIRU – is measured more or less 'directly' from an autoregression. Yet, neither is it a case of direct measurement as in Campbell's classification (see Appendix 1.B) since measurement with VAR models requires prior measurement, that is, prior-measured inflation and unemployment.[21] Measures of the NAIRU are derived as in the structural approach, though not from a causal relation but from a correlation, which is exploited as an invariant regularity.

VAR models came under fire in the mid 1980s by supporters of the Cowles Commission approach, and were criticized for being 'atheoretical macroeconometrics' for several reasons, the most important being that the interpretations of the dynamic behaviour are faulty and that for identification ad hoc dynamics are imposed. This criticism of VAR models echoes the sort of criticism that Burns and Mitchell received in the 1930s concerning their work on the measurement of business cycles, which, as Koopmans put it in his classical 1947 article, was "measurement without theory". Cooley and LeRoy foresee only limited use of VAR models, precisely because they are largely unrelated to economic theory. The only benefit from VAR models seems to be their ability to discover statistical correlations in economic data. Nevertheless, it is precisely this feature of VAR models – summarizing correlations in data – that makes them a useful foundation for the construction of measuring instruments. The application of the Wold decomposition theorem to the correlations in data found by VAR models could, provided the relations are stable, well be used as a principle to bring about a correspondence between unobservable and 'more observable' data. The fact that the correlations may not be interesting for economists, or lack an economic underpinning, is not relevant from the perspective of measurement. What matters is that the correspondence they bring about is stable (Chao 2002; Boumans 2005a,) and, as such, they can function well as an invariant regularity for derived measurement.

[21] Recall from Appendix 1.B of Chapter 1 that direct measurement is defined as measurement which does not depend on prior measurement.

Appendix 4.A VAR models: A formal treatment [22]

Main idea:

The NAIRU can be measured by decomposing unemployment in two parts. One is the NAIRU and the other the ‘unemployment gap’ (gap between actual unemployment and the NAIRU). Two disturbances are assumed to affect fluctuations in unemployment: the NAIRU disturbance and the gap disturbance. In this approach the NAIRU is defined as the part of unemployment that is inflation neutral in the long run. It is mathematically derived by setting the long-run effect of the gap disturbance on the NAIRU to zero.

Formal treatment:

VAR models and structural models have a similar structure. Let π be inflation, u unemployment, b and v coefficients, and ε an error term. The unrestricted, structural VAR alternative to the unrestricted structural equations model of the type:

$$u_t = b_{10} + b_{12}\pi_t + v_{11}u_{t-1} + v_{12}\pi_{t-1} + \varepsilon_{ut} \qquad (4.11)$$

$$\pi_t = b_{20} + b_{21}u_t + v_{21}u_{t-1} + v_{22}\pi_{t-1} + \varepsilon_{\pi t} \qquad (4.12)$$

can, by using matrix algebra, be described as:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} u_t \\ \pi_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \begin{bmatrix} u_{t-1} \\ \pi_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ut} \\ \varepsilon_{\pi t} \end{bmatrix} \qquad (4.13)$$

or

$$Bx_t = C_0 + C_1 x_{t-1} + \varepsilon_t, \qquad (4.14)$$

where: $B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}$ $x_t = \begin{bmatrix} u_t \\ \pi_t \end{bmatrix}$ $C_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix}$ $C_1 = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$ $\varepsilon_t = \begin{bmatrix} \varepsilon_{ut} \\ \varepsilon_{\pi t} \end{bmatrix}$.

Multiplying both sides by matrix B^{-1} yields the *standard form* (reduced form) of the VAR as:

$$x_t = A_0 + A_1 x_{t-1} + e_t, \qquad (4.15)$$

[22] This section is loosely based on Enders (2004: 264 -306).

where: $A_0 = B^{-1}C_0$

$A_1 = B^{-1}C_1$

$e_t = B^{-1}\varepsilon_t$.

VAR models concentrate on shocks and therefore we need to identify the relevant shocks so that we can compute the impulse response function. If the first-order autoregressive model (equation 4.14) is stable (that is, the value of A_1 lies inside the unit circle), and the reduced-form residual vector ε_t is normally independently distributed with the variance-covariance matrix Σ, we can apply the Wold decomposition theorem (see also Section 4.5, equation 4.10). That makes it possible to invert the VAR model (equation 4.15) into a Moving Average (MA) form:

$$x_t = A_0 + A_1(A_0 + A_1 x_{t-2} + e_{t-1}) + e_t$$
$$= (I + A_1)A_0 + A_1^2 x_{t-2} + A_1 e_{t-1} + e_t ,$$

where $I = 2 \times 2$ identity matrix. After *n* iterations, this yields:

$$x_t = (I + A_1 + ... + A_1^n)A_0 + \sum_{i=0}^{n} A_1^i e_{t-i} + A_1^{n+1} x_{t-n-1} ,$$

which can be reduced to:

$$x_t = \mu + \sum_{i=0}^{\infty} A_1^i e_{t-i} , \qquad (4.16)$$

where μ represents a deterministic constant or a constant plus trend (for the vector x_t), and $\sum_{i=0}^{\infty} A_1^i e_{t-i}$ the temporal, cyclical deviation from the trend. Equation (4.16) is known as the *Moving Average* representation of the VAR. The coefficients of matrix A_1^i are called the *impulse response coefficients* associated with the innovations e_t. They describe the propagation of the shocks through the model and the final effect of the shocks on x_t. Both π and *u* can thus, according to the Wold theorem (equation 4.15), be decomposed into temporary and permanent components. Applying this Wold decomposition theorem to the vector x_t then yields the following equations for π and *u*:

$$u_t = \mu_u \text{ (the NAIRU)} \quad + \quad \sum_{i=0}^{\infty} A_{12}^{i} e_{t-i} \text{ (the ‘unemployment gap’)}$$

and

$$\pi_t = \mu_\pi \text{ (‘core-inflation’)} \quad + \quad \sum_{i=0}^{\infty} A_{22}^{i} e_{t-i} \text{ (temporary deviation).}$$

The permanent component of unemployment is the NAIRU, while the permanent component of inflation is referred to as ‘core-inflation’.[23] Blanchard and Quay (1989) apply the Wold decomposition and use long-run neutrality assumptions for identification.[24] The NAIRU is identified as the part of unemployment which is inflation neutral in the long run. That is, the gap disturbance (e_t) has no impact on the NAIRU in the long run. Therefore, the NAIRU is identified by setting the temporary component of u_t (the ‘unemployment gap’) to zero, $\sum_{i=0}^{\infty} A_1^{i} e_{t-i} = 0$.

[23] Danny Quah (1995), for example, defines ‘core inflation’ as “that component of measured inflation that has no medium- to long-run impact on real output”.

[24] In contrast with Christopher Sims (1980), who uses the moving average representation of a system with orthogonalized innovations.

Chapter 5

Models as Measuring Instruments: Measurement of Duration Dependence of Unemployment

5.1 Introduction

In economics, models serve many purposes. Models are used to make predictions about possible outcomes and so to evaluate economic policies. Models are used to give detailed descriptions of the economy. Or models are used for their explorative nature, and hence serve as tools in theory construction.[1] An overlooked function of models is that of a measurement device. In many cases, models generate numbers when confronted with real-world data. And, although economists rather speak of 'assessing' or 'estimating' – and in this way make reservations as to the accuracy of the measures – we can, nevertheless, still think of this generation of numbers as measurement, since it suggests that there is a correspondence between the realm of numbers and phenomena in the real world. As explained in the introduction of this thesis, measurement in the Representational Theory of measurement (RTM) is defined as showing that "the structure of a set of phenomena under certain empirical operations and relations is the same as the structure of some set of numbers under corresponding arithmetical operations and relations" (Suppes, 1998). It requires that measuring instruments are capable of bringing about an isomorphic mapping between an empirical relational structure and a numerical relational structure, and the claim of this chapter is that models are capable of doing this. So, what exactly is it in models that make them able to bring about a mapping between phenomena and numbers that makes them suitable as measurement devices? Marcel Boumans (1999a) shows that models contain devices, which enable them to function as measuring instruments since they are representations of stable

[1] A more detailed description of the functions that models serve can be found in Morgan and Morrison (1999).

relationships. This chapter will argue that it is the Logical Positivist idea of internal principles and bridge principles in models that accounts for the stability of these relationships after they have been calibrated.

The aim of this chapter is to provide an analysis of how economic models may function as a measuring device. This is done by a case study of Tony Lancaster's duration model of unemployment published in *Econometrica* as 'Econometric Methods for the Duration of Unemployment' (1979). In this article, Lancaster suggests a statistical method for using duration data of unemployment, which is now widely used in studies on (long-term) unemployment. Many others followed this method specifically for the measurement of duration dependency. In his article, Lancaster also estimates duration dependency of unemployment for a sample of British unskilled, unemployed workers. The topic of measurement of duration dependency is therefore chosen for this chapter.

This chapter proceeds as follows. Section 5.2 will clarify the account of models as measuring instruments. In Section 5.3 the phenomenon of duration dependence and its causes is discussed, whereas Section 5.4 presents the statistical framework for the analysis of duration dependence. The case study concerned, the model of Lancaster, is discussed in Section 5.5. Section 5.6 analyses, respectively, the internal principles and the bridge principles of the model, and the way the model is calibrated. Finally, conclusions are drawn in Section 5.7.

5.2 Models as measuring instruments?

The methodological literature contains several accounts of models.[2] Rather than focus on the differences between these accounts, I will stress the similarities here. Accounts of models seem to have two elements in common. First of all, *models are simplifications.* In order to describe the real world, irrelevant elements are left out, in such a way that the model represents only what we think is the essence of the phenomenon. From large econometric macro-models to game theoretic models, what they all have in common is that they leave out irrelevant aspects of the phenomenon and in this process idealizations are usually unavoidable. The extent to which idealizations take place, however, is a matter of degree. Hence, some models aim to give detailed

[2] There are several, partly overlapping accounts of models, such as the semantic account, the syntactic account, Mary Hesse's analogy account, and Nancy Cartwright's simulacrum account. Daniel Hausman, Ronald Giere and Margaret Morrison and Mary Morgan provide other accounts. For an overview of these accounts see, for example, Morgan (1998).

descriptions of the world, whereas other models are such an idealization that we can think of them as caricatures: they represent some essential feature of the world, only in a distorted way (Gibbard and Varian, 1978). Secondly, *models display structure*. In economic models, this structure consists of assumptions and relations between variables prescribed by, and consistent with, a particular theory, and in most cases this structure is represented mathematically. Structure, however, is not enough. Structure forms an uninterpreted system, and models need interpretation. As Gibbard and Varian (1978:666) put it: “In economists’ use of models, there is always an element of interpretation: the model always tells a story.”[3] Therefore, they claim that a model is “a story with a specified structure” (ibid.).

This account is not contested in the methodological literature. However, here it is useful to add to this account. Models are set up carefully to let them produce a particular behaviour of endogenous variables, and in order to do so, safeguards are needed. Disturbing factors to the model are shielded off by ceteris paribus conditions, and restrictions are placed on parameter values in order not to let variables reach extreme values and so prevent the model from exploding. Thus, if the conditions under which the model operates are regular, the model will operate regularly, and hence will be generating regularities. Models therefore do not just tell a story, they tell the *same* story. And they will do so over and over again since they are designed to do so. Therefore, models are stories with specified structures that produce, when submitted to the same initial conditions, regular behaviour until an end condition is reached. And it is exactly this regularity in behaviour that makes it possible to use models as measuring instruments.

Many of these aspects can be discerned in the work of Nancy Cartwright and Marcel Boumans. According to Cartwright (1999), models are blueprints of nomological machines; machines that produces lawlike behaviour. A nomological machine is “a fixed (enough) arrangement of components, or factors, with stable (enough) capacities that in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behaviour that we represent in our scientific laws” (Cartwright, 1999: 50). Parts of the machine have the virtue – or, in her terminology, the capacity – to make the machine function as it does. The lawlike behaviour of nomological machines is, however, only apparent in situations where confounding influences are properly shielded off, and in real-world economies this is almost never the case, as it is not possible to control all disturbing factors. The assumptions of ceteris

[3] A point also stressed by Mary Morgan (2001a).

paribus are therefore rarely met in the world. As a consequence, models which function with real data in their initial conditions do behave regularly, but are subject to disturbing influences.

Nevertheless, for the purpose of measurement, we need more than just regular behaviour in a model. Models must somehow be connected with actual observations in the world. In Nancy Cartwright's account, this connection works through the Logical Positivist account of internal and bridge principles.[4] Internal principles "present the content of the theory, the laws that tell how the entities and process of the theory behave" (Cartwright, 1983: 132). They bear on the relation between the world and the model and have the character of universal laws. Predictions of what would happen in particular situations follow from the internal principles. My own view would be to think of the internal principle of a model as a *mechanism*, where mechanism can be interpreted in the Machamer et al.-sense: "Entities and activities organized in such a way that they are productive of regular changes from start or set-up to finish or termination conditions" (Machamer et al., 2000: 3).

In many cases, it suffices for economists to investigate the internal principles only. Let us take, for example, Albrecht and Axell's equilibrium search model (1984), as a typical exemplar of a particular approach to model building. In their analysis, they present a model using search theory for the analysis of wage setting and the effect of unemployment benefits on unemployment. After their model is specified mathematically, they run a test of the model with fictitious numbers. For important but unknown variables fictitious values are 'plugged in' and a comparative static analysis is performed. Since there is no real-world input in the model, no actual phenomenon is measured. The test is performed in order to see whether the model generates the right kind of behaviour and Albrecht and Axell's concern is, in the first, place about the internal principles of the model.

Clearly, internal principles alone do not tie models to actual observations in the world. Therefore, they do not bring about a correspondence between an empirical relational structure and a numerical relational structure. What is missing is a way to put in real-world data and so connect the internal principles with real-world observations. In other words, the internal principles lack an operational form that enables them to accommodate real-world data. Therefore, besides internal principles we need bridge principles, which "are supposed to tie the theory to aspects of reality more accessible to us" (Cartwright, 1983: 132). In the Logical

[4] See, for example, Carl Hempel, 1966: 72-75.

Positivist account, bridge principles bear on the relation between the model and the world and have, contrary to internal principles, no universal character. They only hold when the circumstances are ideal. That is, they hold ceteris paribus. The bridge principles are supposed to tie theoretical terms to direct observations, and the mathematical equations used therefore present the model in an operational form with accessible, that is, observable variables. In this way, bridge principles function as the correspondence rules that Logical Positivists require scientific theories to have. In the Logical Positivist tradition, a strict isomorphism between the model and the world is required. That is, essential elements and relations of the phenomenon must be mapped one-to-one to elements and relations of the model, which in their turn must be referable to direct observations. In this way, the mathematical equations of a model serve as correspondence-as-congruence rules, which are established in order to express theoretical terms in direct observational terms. The requirement in the RTM of a non-degenerating isomorphism between an empirical and a numerical relational structure is satisfied in the Logical Positivist account of theory-data connection.[5] It follows that phenomena in the world are tied to empirical data by the twofold set of internal and bridge principles which are both integrated in models. The internal principles provide autonomy by means of the generated regular behaviour, whereas the bridge principles ensure the connection with the real world.

The twofold set of internal principles and bridge principles are thus important building blocks for the construction of measuring instruments, but may not be enough. The internal principles of a model may appear to be robust generalizations, but they may be inexact and inaccurate due to disturbing factors. Boumans (2005a), however, argues that the inexact and inaccurate relations can be transformed into exact and accurate relationships by *calibration* and, in doing so, make the model suitable as a measuring device. In this way, the problem of lack of control over disturbing factors can be circumvented. In the literature, different interpretations of calibration can be found.[6] In this chapter, calibration is taken as 'fine-tuning' of parameter values. Thus, provided that models are adequate under different circumstances: "A measurement model (…) produces exact and precise regularities" (Boumans, 1999a: 398). It is the triplet:

[5] The RTM became accepted by Logical Positivism, as Hempel (1952: 50-78) shows, and, under influence of Logical Positivism, the RTM became generally accepted as the dominant theory of measurement in the philosophy of science.

[6] One interpretation of calibration is that of a method of estimation or "tuning". Another interpretation is that of "testing" a model. (see also Boumans, 2001; and Hansen and Heckman, 1996).

internal principles, bridge principles and calibration, which together make models suitable as *number-generating machines*, and hence measuring instruments.

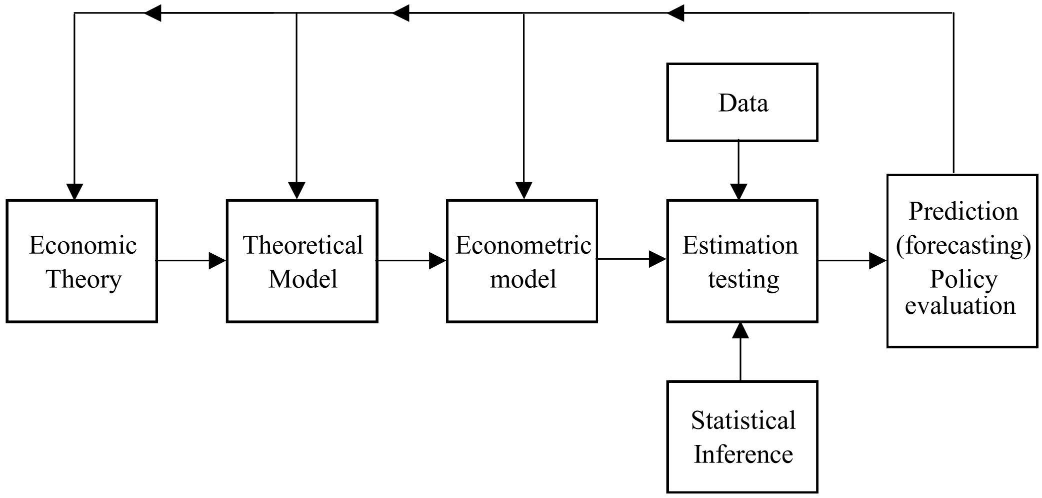


Figure 5.1: Textbook approach to econometric modelling
Source: Spanos, 1986: 16.

This account subsequently begs the question: How do internal and bridge principles come about? In order to answer this question, let us see what the standard textbooks have to say about theory-data confrontation. The standard econometric 'textbook' approach to connecting theory to real-world data looks something like this (see Figure 5.1). A theoretical model – that is, a mathematical formulation of the theory – is derived from economic theory and 'converted' into an econometric model where parameters are estimated and a white noise error term is added in order to take account of measurement errors and/or the effect of excluded variables. In this latter stage, it is assumed that the available data coincides with the theoretical variables that the theory suggests. In this textbook account, economic theory thus provides us with the structure of reality, and when confronted with empirical data we can measure the strength of the relations. This account suggests that economic theory dictates the internal principles, and that the bridge principles are simply an operational form of the internal principles, and necessarily follow from the internal principles.

Many authors (e.g. Morgan 1988) have argued that this textbook account is too simple and therefore to some extent misleading. Cartwright reckons that inevitably unrealistic elements have to be added in the construction of models, that is, in the establishment of internal principles and bridge principles that constitute the model. She argues that laws of nature (i.e. "claims about what necessarily or reliably happens") are scarce, and hard to find since we cannot always control disturbing influences. In her *simulacrum*[7] account of models (1983), she argues that models serve to represent the circumstances where laws apply. Within the special context of a model, regular, lawlike behaviour can be generated that can help us to locate laws of nature, i.e. regularities produced by highly structured arrangements that she characterizes as *nomological machines* (Cartwright, 1999). In this way models serve as blueprints for nomological machines. The unrealistic elements in the construction of models come into play in tying down the models to data. Both internal principles and bridge principles are needed here, and Cartwright stresses that they are both simplifications and inevitably have artificial elements added. Some aspects of the model describe genuine properties, i.e. properties that are essential to the phenomenon, but other aspects of the model are rather 'properties of convenience'.

Strong claims against the econometric textbook account and – derived from that – the way internal principles and bridge principles come about are made by Boumans (1999b). He shows from an analysis of early business-cycle models that, in order to let these models meet certain quality criteria, other ingredients than theory alone have to be incorporated in the model a priori. Ingredients, such as metaphors, analogies, policy views, mathematical techniques and concepts and stylized facts, are taken into account in the design and building stage of the model and subsequently incorporated in the model. In this way, it can be ensured that certain quality criteria are met after confronting the model with empirical data. Thus, the context of discovery and the context of justification are not separate processes in the economic practice of model building but are inextricably intertwined. Boumans therefore argues that justification of the model is built in a priori.

In Boumans' view, model builders already have notions of how to model the phenomenon mathematically and accommodate these notions into the model when designing it (see Figure 5.2), a process he refers to as '*mathematical moulding*' (1999b). In other words, bridge principles do not simply follow from internal principles, which in turn follow from theory, but are established simultaneously, as model builders already have a priori notions of appropriate mathematical representations of internal principles when constructing the model. In Cartwright's

[7] Simulacrum means 'something having merely the form or appearance of a certain thing, without possessing its substance or proper qualities' (Oxford English Dictionary).

view, the bridge principles seem to follow from the internal principles but their status is just convenient, not necessarily true, expressions of internal principles.

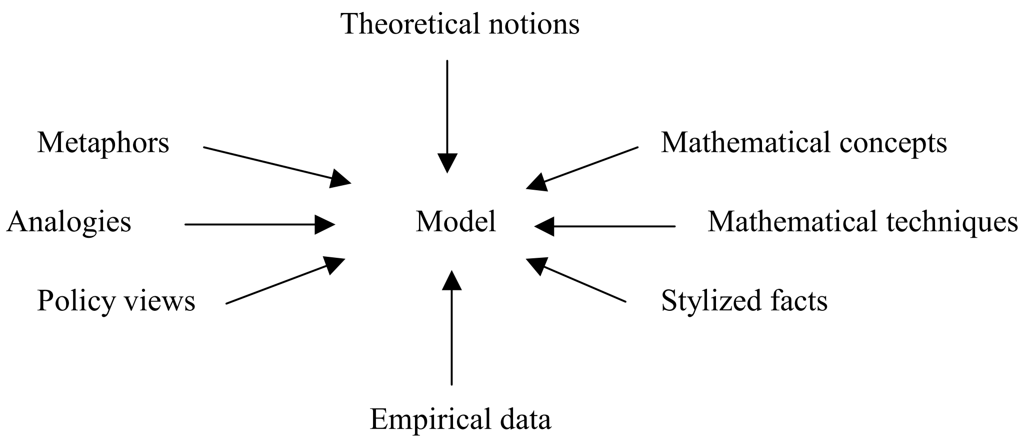


Figure 5.2: Boumans' built-in justification account of models
Source: Boumans (1999b), in: Morgan and Morrison, 1999: 93.

In the next sections, this idea of models as measuring instruments will be explored for the measurement of the duration dependency of unemployment in search models, and a particular model will be investigated by using this framework of internal principles, bridge principles, and calibration.

5.3 The phenomenon of duration dependence and its causes

After decades of prosperous growth, in the 1970s Western economies faced an economic slowdown and a consequent rise in unemployment. This rise in unemployment went hand in hand with a fairly new phenomenon, particularly in Europe, that of long-term unemployment (LTU). Even when the economy recovered, a portion of unemployed were not able to find new employment and remained unemployed for a long period of time.[8] Hence unemployment is disproportionately concentrated on a few individuals who suffer long spells of unemployment. This phenomenon gave a strong impetus to new theories in order to explain the phenomenon of LTU. One line of reasoning is the hypothesis of *duration dependence*. Negative duration dependence is assumed to exist when the probability of leaving unemployment depends on duration. That is, when agents' chances of leaving unemployment decrease because of the very

[8] Mostly taken as longer than 6 or 12 months.

fact that he or she has already been unemployed for a long time. The causes that bring this effect about are multiple. The long-term unemployed might lose professional or social skills, become discouraged, adjust to lower standards of living, or may be stigmatized and become subject to discrimination by employers. As a consequence, of the pool of equally-skilled or experienced unemployed, the worker with the shorter unemployment duration will leave unemployment more quickly. As Machin and Manning put it: “It seems very likely that both high unemployment and a high incidence of LTU have a common cause, an “X” factor or factors, which has resulted in a collapse of exit rates for the unemployed at all durations. The usual suspects for the “X” factor are generous welfare benefits, powerful trade unions, high minimum wage, employment protection, skill-biased technical change, etc.” (1999: 3106-3107). Duration dependence is thus a total outcome effect of a set of multiple, singular causes. In general, however, it is believed that the total outcome effect is negative and that the unemployed who are out of work for a long period of time become even more detached from the labour market. Therefore, it has been argued that high long-term unemployment is itself a cause of high unemployment (Machin and Manning, 1999: 3087).

An important source of duration dependence that is studied very often is a time-varying unemployment benefit structure. In many countries, unemployment benefits change with the duration of unemployment or might end all together after some point. Research indicates that, prior to the depletion of unemployment benefits (at time T in Figure 5.3, when the unemployment benefit level drops from B1 to B2), an increased flow out of unemployment is observed. Researchers infer from this that there is a corresponding drop in the reservation wage (the wage level for which the unemployed are indifferent between accepting a job and staying unemployed) during the two or three weeks preceding time T (Layard, Nickel and Jackman, 1991: 251). The reservation wage is, therefore, assumed to follow the path presented in Figure 5.3. The fall in reservation wage is, however, is a case of positive duration dependence: the approaching depletion of unemployment benefits is assumed to make the unemployed less ‘job choosy’, which increases their chances of leaving unemployment. As a consequence, the reservation wage is thought to decrease with duration and is hence a time varying-variable (Layard, Nickel and Jackman, 1991: 251). Economists have tried to measure this duration dependence – that is the elasticity of a causal factor on unemployment duration – by using empirical data on unemployment duration.

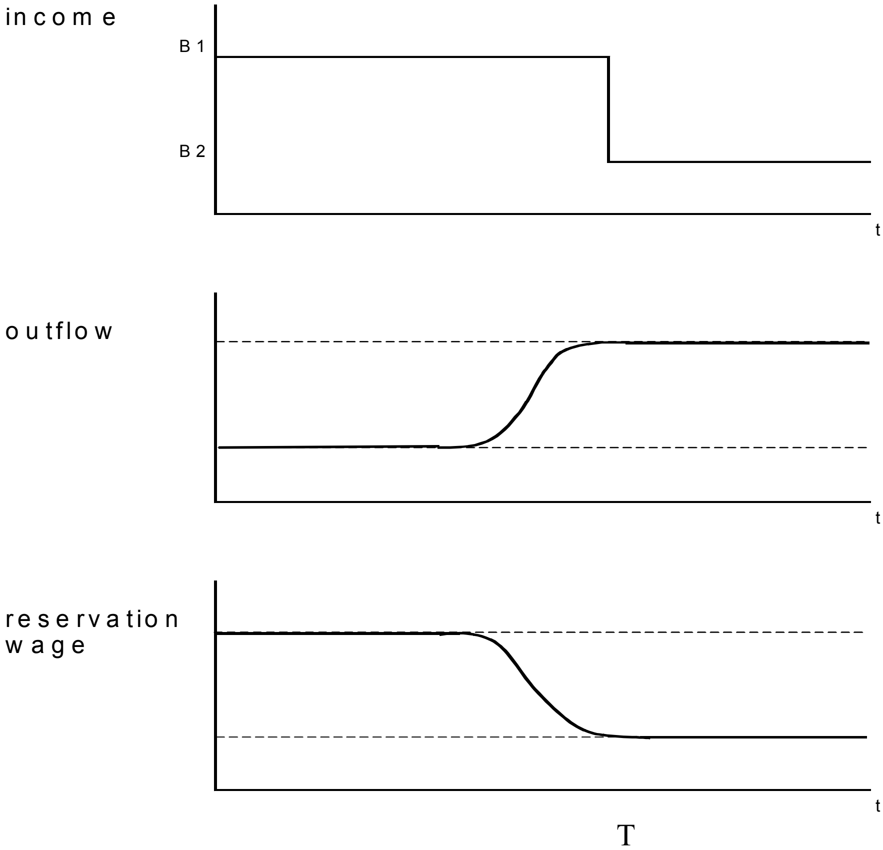


Figure 5.3: Income, outflow from unemployment, and reservation wage in a time varying unemployment structure

5.4 The statistical framework for measurement of duration dependency

Up till the 1960s, the problem of unemployment was primarily analysed in terms of stocks of unemployed. The establishment of the idea of a natural rate of unemployment by Friedman (1968) and Phelps (1967) redirected the attention of economists from macroeconomic approaches to unemployment, as used in Phillips curve analyses, to investigations of the causes of unemployment on a microeconomic level. From there developed a new approach in labour economics, which was based on search theory and analysed unemployment in terms of flows rather than stocks. Peter Diamond (1998: 273) defines search theory as "the analysis of resource allocation with specified, imperfect technologies for informing agents of their trading opportunities and for bringing together potential traders". Search theory analyses rational decision making of (unemployed) agents who optimize their expected flow utility[9] when faced with a known distribution of wages.[10] That means that economic agents are supposed to make a choice between accepting an offered job with a given wage and searching further for a job that pays better and which occurs with a known wage distribution. Unemployment is in this way considered as a productive investment, since a higher wage level will eventually offset the costs of an extended unemployment spell. In these models, the optimal policy of the unemployed is based on the selection of a reservation wage that defines the set of acceptable wage offers given the distribution of wage offers. The reservation wage must be set high enough to make the job acceptable for the unemployed and low enough for the employer to offer the job. This general idea of search theory has resulted in a variety of models, such as search models, matching models and efficiency wage models.

A variety of ideas are used in search models of unemployment. Early search models assume the existence of wage setting by firms with a certain known distribution. When a job arrives, identical workers decide to take it or to reject the wage offer. Unemployment in these models is thus caused by the rejection of wage offers by workers. Later models include all kinds of heterogeneities in the labour market, such as heterogeneity in information, skills of workers, reservation wage, location etc. These make it necessary for both firms and workers to spend resources to find productive job matches and are also the cause of mismatches. What search models have in common is that they assume the steady state condition for the economy, and that the flow into employment is matched by an equal, loss of existing jobs, and hence inflow equals outflow.

However, when we confront theoretical search models using this framework with empirical data, we find that this theoretical framework is not well suited for actual measurement of the effects of duration dependence. Important variables like the reservation wage are not directly observable, or are even impossible to derive from duration data.[11] Moreover, in the case of duration dependence, the reservation wage is not constant but a time-varying function of the unemployment duration, making observation even more complicated. Data we can 'observe' from the unemployment statistics involve accessible characteristics, such as age, work

[9] The summed and discounted value of expected future income.
[10] Or prices. Search theory has its counterpart in the analysis of consumer behaviour in the goods market.
[11] Just as Flinn and Heckman (1983), labour economists try to infer the reservation wage from wage data where the lowest accepted wage is taken as the reservation wage, so do contemporary researchers make use of surveys in order to study the reservation wage. Both methods are, however, very sensitive to the exact specification.

experience, education and unemployment duration. Historians and philosophers of science like Brian Ellis (1966) and Michael Heidelberger (1994a, b) show that one way to deal with such a problem is to try and find another variable with which the phenomenon is functionally correlated, in the same way as changes in temperature are correlated with expansion of a mercury column (see Chapter 3). So, if we want to have search models to generate real world numbers, we could perhaps resort to an operational model in which the unobservable is replaced by a covarying variable, which we can confront with observable data. Lancaster (1979) provides such an econometric method that is consistent with standard search theory. It is based on an econometric specification of the outflow rate out of unemployment in terms of conditional probabilities of an unemployed individual leaving unemployment, and it makes use of three related probability distributions (see Figure 5.4).

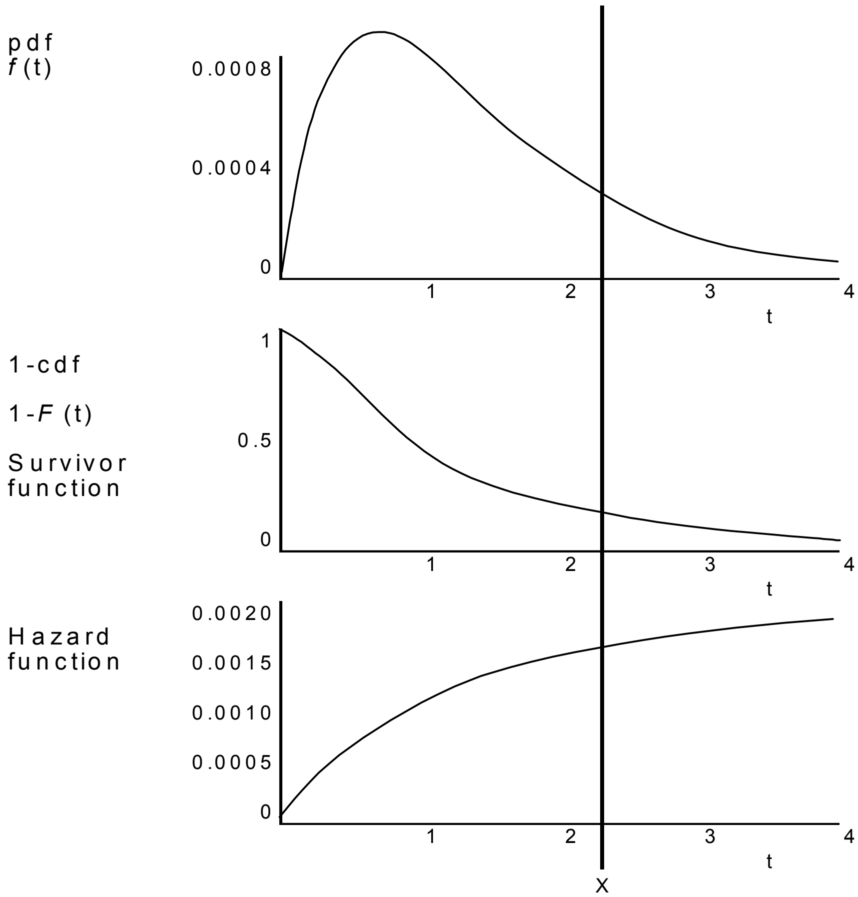


Figure 5.4: Examples of probability distributions from duration data

The first one, the probability distribution function (in short: *pdf*) *f*(t)) can be established from data on unemployment spells. The function determines the percentage who failed to find a job in time X to X + an (infinitely small) interval Δt. Therefore, it gives the proportion of all unemployed who find a job within Δt of a duration of say, 18 months. This probability is referred to as the 'unconditional probability of leaving unemployment' as it is not conditional on the state of the unemployed. It is uninformative about which of the unemployed has left the pool of unemployed. This probability is therefore not very illuminating in the analysis of duration data.

More informative is the second distribution function, the cumulative distribution function (*cdf*), or more accurately, the function 1-*cdf*. The cumulative distribution function *F*(t) determines the total percentage who'failed' by time X. Consequently, the function 1-*cdf* (or 1-*F*(t)) determines the total percentage of a cohort remaining after time X. For that reason, this function is referred to as the *survivor function*, and it can be obtained from the *pdf* by integration. The relation between *pdf* and *cdf* is:

$$f(t) = dF(t)/dt. \qquad (5.1)$$

The third relevant distribution function is the *hazard rate* or *hazard function* (or *instantaneous failure rate*).[12] This function became the main framework for duration analysis, and in economics it is common to refer to the hazard function θ as the 'exit' or 'outflow' rate. The hazard function is used intensively in life-cycle problems, where objects or agents move from one state to another after some time. Light bulbs, for example, move from the state 'working well' to the state 'broken down' after some time. The process of entering or leaving unemployment can similarly be considered as the transition of one (employment) state to another.[13] The hazard function θ is the function that represents the *conditional* probability of moving from one state to another. It gives the probability that a worker unemployed for, say, 18 months will find a job. The hazard function thus gives the probabilities of being unemployed

---

[12] In full: "reemployment probability function hazard of failure rate" (Lancaster 1979: 940). Other names under which the function appears in scientific literature are, among others, 'age specific death rate', 'force of mortality', 'failure rate', 'hazard', etc.

[13] Unfortunately, since this kind of duration analysis originates from mechanical engineering, which studied the breakdown or failure of physical objects in time, it follows that the (joyful) event of finding a job is interpreted as a 'failure'.

during the different stages of their unemployment spell, and is therefore a more useful concept for the analysis of duration dependence. The hazard function is determined as:

$$\theta(t) = \frac{f(t)}{1 - F(t)}. \qquad (5.2)$$

The hazard function is thus the probability distribution function divided by 1 minus the cumulative distribution function:

$$\theta(t) = \frac{pdf}{1 - cdf}. \qquad (5.3)$$

In search models, the hazard rate or the probability of leaving unemployment θ is often considered as the product of the probability of finding a job (or 'arrival rate' of jobs) λ and the probability of accepting the job $p(z)$. The job will be accepted when it pays an amount greater than or equal to the worker's reservation wage *z*. The probability $p(z)$ is therefore a function of the distribution of wages. The arrival rate is considered to be a function of the labour market tightness, that is, the unemployment-vacancy ratio (U/V). The outflow probability is then written as:

$$\theta_i = \lambda_i \left(U / V \right) p(z_i). \qquad (5.4)$$

The hazard rate can be made observable from duration data of unemployment, which are relatively easy to obtain from, for example, surveys at labour exchanges, and may concern completed or uncompleted spells of unemployment. The hazard function can, however, have very different shapes, and be either a positive, negative or constant function. Figure 5.5 shows the different shapes the hazard function can exhibit. Time dependency is absent when the hazard function is constant over time (as in Figure 5.5c). In the case of a decreasing hazard function (Figure 5.5b), the unemployed are 'getting better' at staying in their unemployment state,[14] i.e. unemployment is negatively duration dependent, as the probabilities of leaving will collapse. Increasing hazard functions indicate positive duration dependency, meaning that the probability of leaving increases with duration.

[14] Again, the terminology is unfortunately counterintuitive.

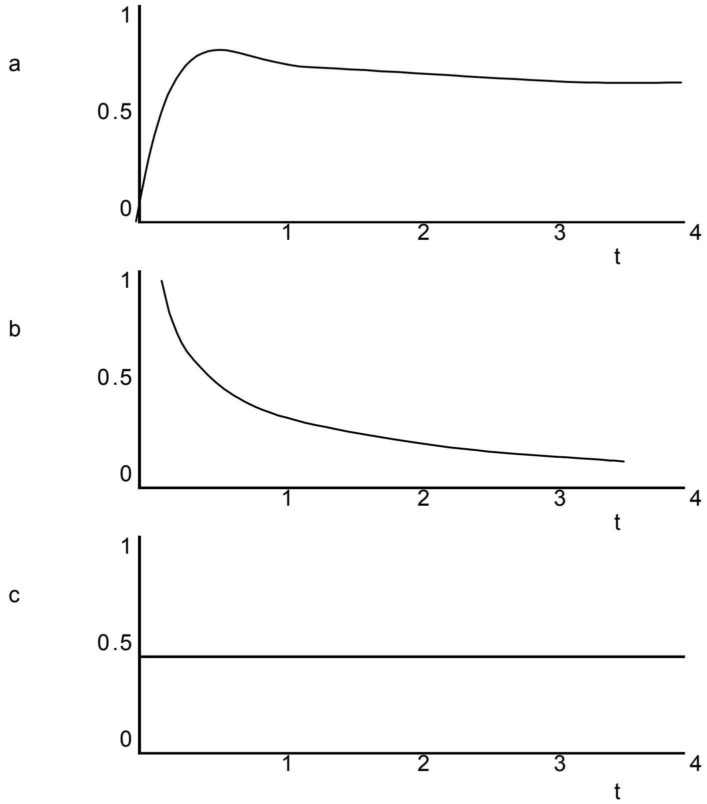


Figure 5.5: Three hazard functions
Source: Lancaster, 1990:10.

So, in the approach suggested by Lancaster, the analysis of the reservation wage variable, which is unobservable from the duration data, is replaced by the analysis of the hazard rate. This hazard rate can be derived from observable duration data and is taken as a sound proxy of the reservation wage. In the econometric model building-process the central issue then becomes the finding of an appropriate specification of the outflow process.

In order to get a better understanding of this approach we will study the econometric specification Lancaster that applied in his model in more detail.[15]

5.5 Lancaster's 1979 duration model

Lancaster demonstrates his method by calculations of the temporal variation of the chances of an unemployed individual returning to work, which obviously is duration dependent. His method

[15] A overview of this method can be found in Layard, Nickell and Jackman (1991) and Machin and Manning (1999).

became very influential, and many other economists followed his approach. However, whereas some researchers, like Narendranathan, et al. (1985), follow a structural approach in order to establish the determinants of the duration of the spell of unemployment, Lancaster follows a reduced-form approach, which he subsequently tried to give a 'structural interpretation'. Many structural duration models (like that of Narendranathan, et al., 1985) suffer from measurement difficulties at the structural-form level due to identification problems and the use of the reduced form is a classic econometric response to such difficulties.

In his model, Lancaster distinguished two opposite time varying-effects on the probability of outflow: first, the effect of a declining reservation wage due to time-varying unemployment benefit structure; and secondly, a decline in the arrival rate of jobs, which causes the worker's outflow probability to deteriorate. For Lancaster's model to work, he now needs to specify the hazard. According to Lancaster, the mathematically most attractive way to proceed is to reduce this set of variables to a function of time and a function of other regressor variables, x (Lancaster, 1979: 945). He, therefore, specifies the outflow probability as a function of observable characteristics (x) and time (t): [16]

$$\theta(t) = \psi_1(x)\psi_2(t)\,. \qquad (5.5)$$

The most obvious function to use for the specification of *pdf* would be the exponential function. But an exponential distribution yields constant hazards rates, and is therefore not very helpful in cases where we deal with duration dependence. The exponential distribution function is, however, a special case of a function known as the *Weibull distribution function*. The Weibull function is a function widely applied to reliability and life data problems. It is often used in biometrics, medicine and mechanical engineering for the analysis of life behaviour, such as the administration of a treatment; weight gain after birth; the disappearance of symptoms; the ability to return to work; and duration of light bulbs or ball bearings.[17] It is widely used in duration analysis of unemployment, since the function can easily take duration dependency into account. The Weibull pdf function can be expressed in different forms. In its most general form, it is written as:

[16] This implies that the re-employment probabilities of the unemployed follow the same time path, referred to in the literature as '*proportional hazards*'.

[17] W. Weibull originally established the function in 1937 in Sweden for the analysis of metal fatigue in ball bearings.

$$f(t) = \alpha\beta t^{\alpha\text{-}1}\exp(\text{-}\beta t^{\alpha}) \qquad (5.6)$$

$$F(t) = 1 \text{ - } \exp(\text{-}\beta t^{\alpha}) \qquad \theta(t) = \alpha\beta t^{\alpha\text{-}1}. \qquad (5.6a)$$

In Lancaster's model α represents the coefficient of duration dependence, and β is the vector of coefficients associated with the observable characteristics x. We can think of β as the elasticity of the hazard with respect to the reservation wage z. In the simpler case of β =1, the Weibull function is reduced to:

$$f(t) = \alpha t^{\alpha\text{-}1}\exp(\text{-}t^{\alpha}). \qquad (5.7)$$

As a result, Lancaster specifies the conditional probability function $\theta(t)$ with a Weibull function for the time-varying component of (5.5) as: [18]

$$\psi_2(t) = \alpha t^{\alpha-1}. \qquad (5.8)$$

In these equations, α is known as the 'slope parameter'. It can easily be seen that the factor $t^{\alpha-1}$ is responsible for the degree of duration dependence. The consequences of different values of α for the Weibull *pdf* and hazard rate are presented in Figures 5.6 and 5.7. In the special case of α = 1, there is no duration dependence. The Weibull function then will yield a common exponential function, and the Weibull hazard rate (Figure 5.7) will be constant (and identical to Figure 5.5c). Values of α >1 determine positive duration dependence and hence increasing hazard rates; while values of α <1 determine negative duration dependence and decreasing hazard rates.

---

[18] The observable characteristics of ψ_1 in Lancaster's model are written as:
$\log\psi_1 = \beta_0 + \beta_1 \log AGE + \beta_2 \log UNEMPLOYMENT + \beta_3 \log REPLACEMENT$.

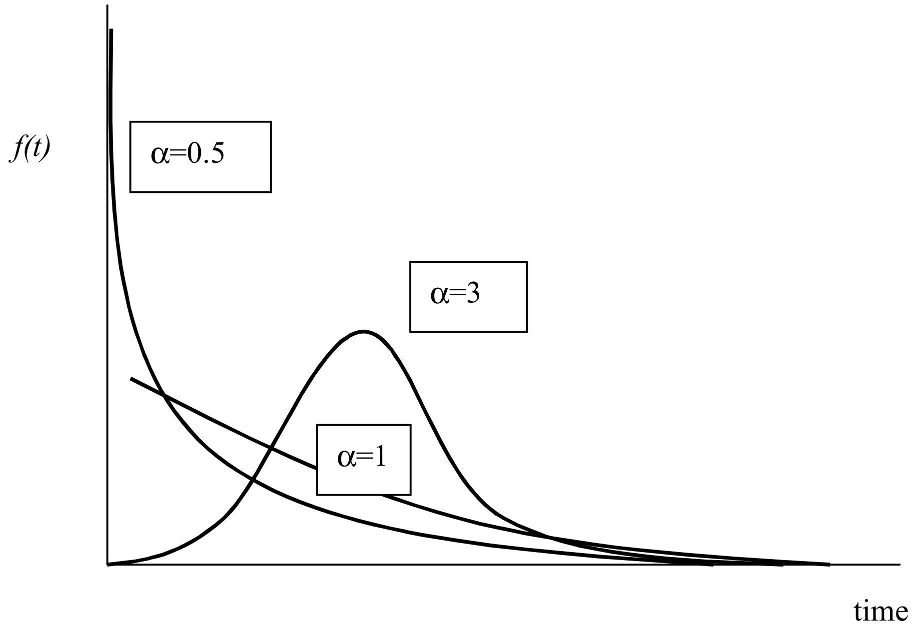


Figure 5.6: Weibull *pdf* with different values of α

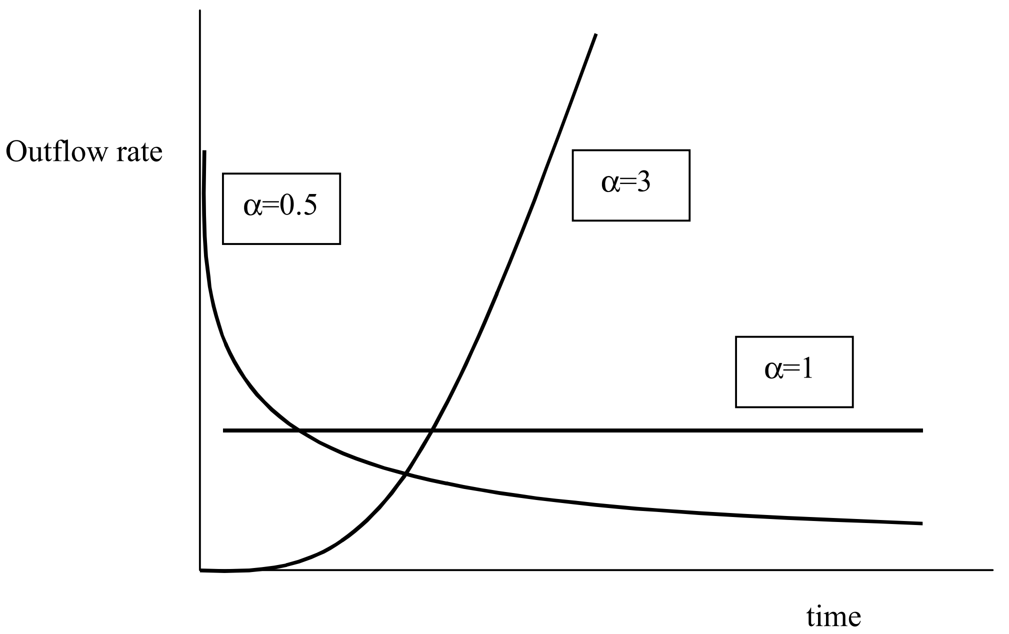


Figure 5.7: Weibull hazard rate with different values of α

In economics, it is assumed that the total outcome effect of duration on unemployment is negative, and that α is therefore smaller than 1. In that case, both the Weibull *pdf* and the hazard function are decreasing functions, which represent the decreasing probabilities for the long-term unemployed to find new employment. With his model, Lancaster finds a value for α of 0.77, which indicates a negative duration effect of unemployment. Of the two opposite effects, the drop in the stream of job offers is the strongest and fails to be offset by a drop in the reservation wage (caused by a time-varying unemployment benefit structure).

In the conclusions of his seminal article, Lancaster casts some doubts on the method he suggests. In particular, he worries about the arrival rate of jobs, as this rate is a function of the labour market tightness that is defined in terms of unemployment and vacancies, i.e. variables we cannot control for. The arrival rate therefore might not be constant over time and thus might influence the probability of leaving unemployment. Lancaster warns that: "It should be noted that studying θ" [*outflow*] "is very much a second best to studying $\overline{w}$ [*the reservation wage*] itself since even though $1-F(\overline{w})$ is monotonic in $\overline{w}$, θ and $\overline{w}$ need not vary in the same direction over time due to the possibility of variation in λ, the rate at which opportunities of employment present themselves" (Lancaster, 1979: 941). Lancaster therefore concludes: "My own view is that the study of duration of unemployment data is probably not going to be a very helpful way of testing those predictions of search theory which concern themselves with the way individuals vary their reservation wage as the time passes" (Lancaster, 1979: 956).

5.6 Lancaster's duration model as a measuring device

Let us now see how Lancaster's model constitutes a measuring device, and analyse the model in terms of the triple: internal principles, bridge principles, and calibration.

5.6.1 *The internal principles of the model*

Economic theory tells us that the driving principle (internal principle) of search models is the process of reservation wage determination (or cut-off price in the case of the goods market) and the tendency of the reservation wage to go down (cut-off price to go up) for various reasons. The internal principles of the search models, however, relate to variables, such as the reservation wage, which are unobservable from duration data. Duration data is relatively easy to obtain, and, for this reason, economists consider another internal principle to which they can apply this sort of data: namely, the outflow process, as a replacement of the process of reservation wage adjustment. The behaviour of the main variables in the two mechanisms seems to be correlated, in the sense that they exhibit opposite behaviour. It can be observed from Figure 5.3, but follows also from equation (5.4) $\theta_i = \lambda_i p(z_i)$ that: with a constant arrival rate λ, θ (the hazard rate) and z (the reservation wage) move in opposite directions.

With respect to measurement, the relevant point here is that contemporary economists take the outflow process of unemployment as a stable process or stable regularity, and that quality can serve to bring about a non-degenerating correspondence between empirical phenomena and numbers. Though unemployment obviously changes over aggregate demand cycles, the vast majority of the labour force is employed. At a 5% unemployment rate, 95% of the labour force is employed. In addition, the stock of unemployment is very small compared with total turnover (flow) in the labour market. Thus, although there are fluctuations in the unemployment rate, they are small compared with the stock of unemployment, and show the labour market to be a regular working device. Viewed in this light, the labour market itself can be considered as a nomological machine that generates more or less regular behaviour (outflow) and the models that describe this process can be considered as its blueprint. That qualifies models to embody a principle of reliability that we can explore for the purpose of the construction of measuring instruments.

The case concerned also illustrates how the textbook approach to econometric modelling is an idealization, and shows that arbitrary, un-theoretical assumptions or decisions are needed, in both in the internal and the bridge principles. The theory used here, search theory, analyses the search behaviour of agents but provides only a broad framework, covering a wide variety of search models that represent different allocation mechanisms in the labour market. But, the theory is not helpful in all aspects of model building. For example, in search models the gains of a productive match have to be distributed between the worker and the firm. In the literature, the function that describes this process of allocation of gains is referred to as the 'sharing rule'. Search theory is silent, however, about what this function should look like, and the fact that often an equal sharing is used is more likely explained by its ease of use, and the fact that it results in Nash equilibria rather than that a solid theoretical foundation can be found.[19] Theory provides no compelling reason as to why the gains should be split equally. Other arbitrary decisions in model

[19] See for example Diamond (1982).

building include for example, on-the-job search, job duration, etc. In Narendranathan, et al.'s model (1985), the valuation of leisure is left out, since the marginal utility of leisure is likely to vary with personal characteristics *x*. Narendranathan, Nickell and Stern argue that "arguments of this kind are one of the reasons why it is difficult to identify structural models of search behaviour without making arbitrary prior assumptions." (Narendranathan, et al., 1985: 311). For this reason, economists often resort to reduced-form models. At the same time it shows that important elements in the model – which constitute the internal principles - may not necessarily be 'real' properties of the modelled object, but rather pure works of fiction.

5.6.2 *The Weibull function as a bridge principle*

The internal principle (outflow process) is connected to the world by imposing a Weibull specification upon the hazard function, where observable characteristics of the unemployed can easily be accommodated. Recall (from equation 5.5) that the outflow probability is defined as $\theta(t) = \psi_1(x)\psi_2(t)$. Two causal factors are considered: time (*t*), and the set of personal characteristics (*x*). In this way, the Weibull function intermediates between the relation between the model and the mathematical representation, and connects its observables to theoretical terms. The Weibull function thus brings about the mapping between the phenomenon and numbers, and the diagrammatic representation of measurement then looks like this (see Figure 5.8).

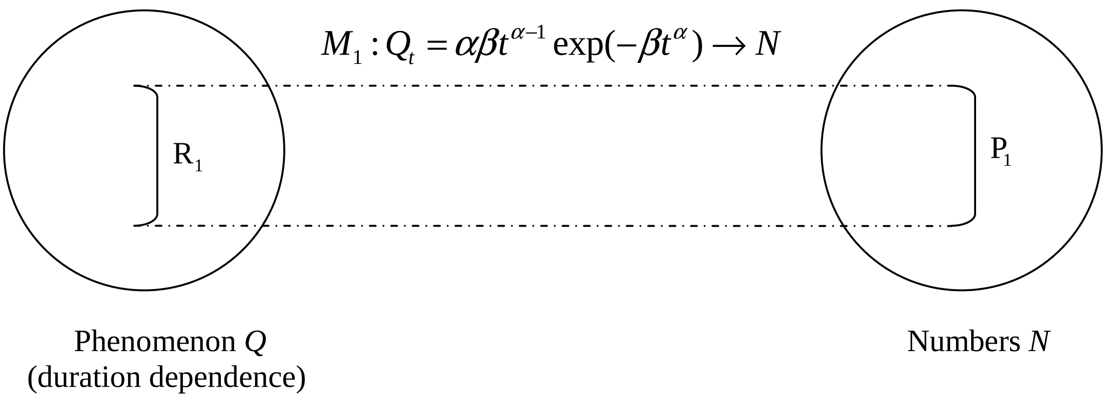


Figure 5.8: Diagrammatic presentation of the Weibull function as an isomorphic mapping

In Logical Positivism, the bridge principles are interpreted as strict isomorphism between the model and the world. This view requires that the Weibull function is a true and exact description of the outflow process of the labour market. But this claim seems much too strong. The outflow process could be described by a whole set of other distribution functions, such as the log-normal, Gompertz, Inverse Gaussian, generalized F, and Gamma distributions, etc. There are no compelling reasons for using the Weibull function, other than that it functions well over a wide variety of cases. The other distributions mentioned might be useful only in specific circumstances (Kiefer, 1988: 655).

Cartwright's interpretation of bridge principles as abstractions therefore seems more appropriate.[20] The representation is not necessarily a true one, and the bridge principles are more like conventions. Bridge principles are not necessarily true, but must work well in a great variety of cases. They are chosen for their convenience since we do not want to develop a bridge principle for each unique situation. As Cartwright puts it: "the 'right kind of description' for assessing an equation is seldom, if ever, a 'true description' of the phenomenon studied; and there are few formal principles for getting from 'true descriptions' to the kind of description that entails an equation. There are just rules of thumb, good sense, and, ultimately, the requirement that the equation we end up with must do the job." (Cartwright, 1983: 133). The widespread use of Weibull functions in order to operationalize search theory seems only to be inspired by its convenience of use and the belief that the function does the job quite well, i.e. that the Weibull function exhibits decreasing hazard rates. And it is exactly this that Cartwright calls *real* bridge principles. "To do quantum mechanics, one has to know how to pick the Hamiltonian. The principles that tell us how to do so are the real bridge principles of quantum mechanics" (Cartwright, 1983: 136). Or analogously, the principles that tell us how to pick the right equation for the hazard are the true bridge principles of search theory.

5.6.3 *Calibrating the model*

Finally, as a last step in the measuring procedure, the Weibull function has to be calibrated. That is, the model parameters need to be adjusted in such a way that the model fits the empirical data well and thus reveal the negative duration dependence. The value of α is therefore necessarily limited to values smaller than 1. After calibration, two things have been accomplished. In the first place, the outflow process is presented as an invariant generalization. The Weibull function, which was chosen for its convenience, turned out to be very sensitive to its parameter values (as

[20] Her account seems more akin to the *semantic* account of models, which considers models as representations.

can been seen in Figures 5.6 and 5.7), but calibration fixed the parameter value, and, then the function becomes robust over a wide variety of circumstances and independent of disturbing factors. Secondly, the parameter value provides the rate of duration dependency we were after.

Unfortunately, a serious problem arises in the calibration of α, which is inherent in this measurement procedure of duration dependence. In his seminal paper, Lancaster already points to the problem of *unobserved heterogeneity*. The problem is as follows: if we observe decreasing probability of outflow associated with duration, is the cause then *'true'* duration dependence, or is the phenomenon caused by other variables that are not observable to the researcher? Unobservable variables such as the workers' mental state or work attitude are not captured in data sets but might account for long-term unemployment (LTU) as well. So, the problem is in short: Is duration dependence the cause of LTU, or is it something else? Economists try to deal with this problem in the following way.[21] The Weibull function for the hazard (equation 5.6a) is re-specified as:

$$\theta(t) = \upsilon \exp(x, \beta)\psi_2(t), \qquad (5.9)$$

where υ is the variable that captures all unobserved variables, such as work attitude, motivation, mental state, etc. It basically means that (for the simpler case of β = constant) we write θ as a function of a set of unobservables (υ), observables (x), and time (t). Next, a distribution for υ is specified: $H(\upsilon)$. This is considered as an 'error' distribution and is usually taken as a Gamma function. It is then possible to determine the duration dependency by imposing restrictions on the functional form. The fundamental problem with this method is thus how to discriminate between duration dependency α and unobserved heterogeneity υ. This manifests itself as the problem of how to model the baseline hazard $\psi_2(t)$ and the 'error' function $H(\upsilon)$. Identification of equation (5.9) is possible only by making specific and untheoretical assumptions about the baseline hazard and the 'error' function and the assumption of mixed proportional hazards: the assumption that υ is strictly independent from x and t. As Lancaster puts it: "Identification can be achieved when certain functional form restrictions can be assumed. Unfortunately these functional form restrictions generally have little or no economic-theoretical justification. There is no known economic principle that implies that hazard functions should be proportional …still less does

[21] An overview of the existing literature that deals with this problem can be found in Machin and Manning (1999: 3107-3111).

economic theory imply Weibull models" (Lancaster, 1990: 157). Machin and Manning therefore conclude that: "It does not really seem possible in practice to identify separately the effect of heterogeneity from that of duration dependence without making some very strong assumptions about the functional form which have no foundation in any economic theory" (Machin and Manning, 1999: 3111).

It seems that a high degree of 'mathematical moulding' is involved in this case of construction of bridge principles and determination of duration dependency, but, as we have already seen, the procedure for measuring duration dependence runs into trouble as a consequence of other unobservables or not easily accessible variables.

5.7 Conclusions

In Logical Positivist philosophy of science, the path from theory to data is taken as a top-down process that runs as follows:

Theory
⇓
Internal principle
⇓
Bridge principle
⇓
(Direct) observations

The stance seems to be that, once we have established the internal principle(s) from theory, the bridge principle(s) will easily follow, as we can always find a convenient mathematical expression that fits the internal principle and data well. In economic science, models often function as mediating devices between theory and data (Morgan and Morrison, 1999), and they constitute both internal principles and bridge principles. The representational theory of measurement requires for sound measurement a strict isomorphic mapping between an empirical relational structure and a numerical relational structure. In economic models, the internal principles can bring about this isomorphic mapping, as they are simplifications of essential structures in the world and have proven to be more or less stable (or in any case stable enough) over a variety of circumstances. Accordingly models can be used as measuring instruments. The internal principles are expressed in convenient, but not necessarily true, mathematical descriptions, which are known as the bridge principles. Since the internal principles of a model may be inexact and inaccurate due to disturbing factors, they can be transformed into exact and accurate relationships by calibration. Thus, for measurement by models, the triple: internal principle, bridge principle, and calibration is required.

The idea of a top-down path from theory to data also underlies the textbook approach to econometric model building. The case at hand shows that this account is not completely appropriate here. To begin with, the problem is that there is an insuperable theory-data gap (see, for example, Morgan 1990). The driving force in search theory, which forms the internal principle of search models, is adjustment of the reservation wage (or price) of heterogeneous workers, i.e. workers who have differences in productivity, valuation of leisure, information, etc. However, there appeared to be no real-world data that could be applied to search models in order to make the crucial variables in the model, such as the reservation wage or flow utility, observable to the researcher. In addition, search theory does not provide us with strategies for making the theory operational. On the other hand, data of duration of unemployment is relatively easy to obtain from, for example, surveys, and the statistical methods for analysing it are well known from biometrics and mechanical engineering. The innovative aspect of Lancaster's approach thus appears to be the replacement of the internal principle by an entirely different framework, that of the outflow process of unemployment. Contemporary economists consider the outflow out of unemployment as a relatively stable process; i.e. only a small fraction of the turnover in the labour market is unemployed, and this fraction is more or less stable over different circumstances. In addition, this process it associated with reservation wage setting, the internal principle of search theory. The step from a theoretical economic model to an econometric model is not simply a matter of 'filling in' the parameters. Key variables in the economic model are completely absent in the econometric model, and, instead, a completely different mechanism is operationalized in order to analyse the implications of search theory. In this case, the path from theory to data appears to be a bottom-up process: available data drives the search for an internal principle that is more or less consistent with search theory.

As a consequence of this 'internal principle shift', duration models of unemployment cannot test the behavioural implications of search theory, a point well taken by Lancaster. Theoretical explanations take place in the domain of search theory. In curve-fitting the outflow process, there is nothing that gives rise to an explanation of the behaviour of economic agents or even the phenomenon of a regular outflow. So, where search theory tries to provide an explanation for why there is still unemployment even when supply and demand for labour are equal (the natural rate of unemployment), the duration models do not do so. Nevertheless, they can still be used for measuring a certain aspect of unemployment, namely its duration dependency.

Finally, in the case of measurement of duration dependence, the measurement procedure depends critically on the establishment of an appropriate mathematical representation of the outflow process. Out of a set of possible functions, the Weibull function is chosen for reasons of ease and because it provides decreasing hazard rates. However, as shown, there are no compelling economic reasons for using this particular function, nor are there any reasons to believe that it corresponds to real-world outflow processes of unemployment. The Weibull function just does the job of measuring duration dependence, and therefore it serves as a *real* bridge principle, as Cartwright puts it. Nevertheless, as a consequence of unobserved heterogeneity (LTU might be attributed to a characteristic unobservable to the researcher), duration analysis of unemployment runs into trouble. By imposing strong restrictions and assumptions upon the specification used, it was possible to obtain measures of duration dependence, but they turned out to be very sensitive to the precise specification of the Weibull function and the underlying assumptions. So, if one is to rescue duration analysis, untheoretical assumptions have to be made, for which even the principle of convenience for choosing bridge principles cannot be helpful. Thus, while the analysis of duration data was set up in the first place to circumvent the problem of making an unobservable explanatory variable operational, new unobservable variables emerge in this approach that prevent this method from being fully satisfactory.

Chapter 6

Conclusions

6.1 The research question reconsidered

The dominant theory of measurement in the philosophy of science is the Representational Theory of Measurement (RTM). Since this theory is an important framework for this thesis, its main assertions will now be briefly recapitulated. In this theory of measurement, measurement is about establishing an isomorphic mapping between an empirical relational structure and a numerical relational structure. The finding of isomorphic structures is therefore of great importance for any measurement procedure. The first fundamental problem of measurement is to justify the chosen isomorphic mapping. The way this is done is by showing that the phenomenon under scrutiny has the 'same structure' in empirical situations as the arithmetic of numbers. After all, isomorphic refers to that fact that the empirical structure and the numerical structure are mapped one-to-one, and thus necessarily have the 'same structure'. This requirement for measurement is formalized as the Representational Theorem.

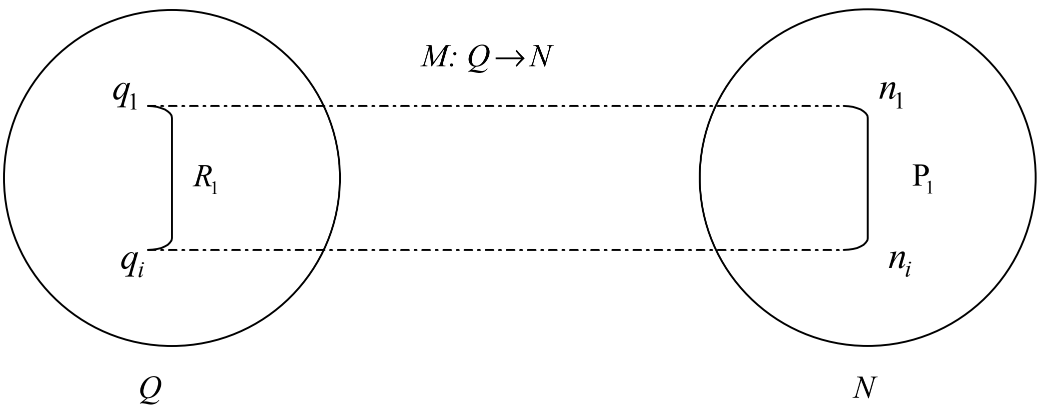


Figure 6.1: Isomorphic mapping in the Representational Theory of Measurement
Source: Finkelstein, 1975: 105.

Once we have established this isomorphic mapping, and, in addition, have assured that we can allow for transformation of measurement scales – the second requirement of the RTM, known as the Uniqueness Theorem – then we have according to the RTM, established a sound foundation for measurement.

The problem now is that, in the RTM, the justification of the choice of isomorphic mapping for measurement, is only done formally and set-theoretically. However, the justification of isomorphic mappings is an empirical, experimental affair. Judging whether a thermometer really measures the phenomenon it is supposed to measure, heat, is a matter of experimentation, empirical judgements and good common sense. For this reason, the formalism of 'isomorphism' cannot be taken as a sufficient, testable criterion for good measurement; that would be circular. Sound measurement of a phenomenon requires an empirical proof of isomorphism and not a formal one. Substantial parts of this thesis are dedicated to how economists bring about, and justify, a correspondence between a phenomenon and numbers in practice. Most philosophical work on the RTM of the last decades has, however, focused on the axiomatization of the two theorems for various kinds of empirical situations, and not on the development of empirical or experimental guidelines for how to establish these isomorphic mappings. As a consequence, the RTM not only has no practical guidance to offer to researchers in the field, but has also become detached from contemporary philosophical debates on empirical research and experimentation, like those initiated by Hacking (1983), Cartwright (1983), Latour (1987), and Bogen and Woodward (1988).

The reason for this neglect of the empirical aspects of measurement and instrument design in the RTM seem to me a heritage from its philosophical roots, Logical Positivism. In Logical Positivist thinking, theoretical constructs, which are abstract, unobservable entities suggested by theory, must, in order to be meaningful, be expressed in observational terms.[1] These connections, which must guarantee a strict mapping between the unobservable theoretical concept and observables accessible to our senses, are called 'correspondence rules'. The observational terms can finally be ordered quantitatively, and hence measured. In Logical Positivism the path from theory to numbers thus runs as:

[1] See, for example, Hempel (1952).

Theory

↓

Theoretical concepts

↓

Observational terms

↓

Data (Measures)

This philosophical stance is nowadays widely rejected as, in the latter half of the 20th century, under the influence of Thomas Kuhn, a debate arose about the theory-ladenness of observations. However, some of the underlying ideas turn out to be very persistent and continue to have a deep influence on the philosophy of science.[2] The requirement of isomorphism in the RTM is undoubtedly indebted to the Logical Positivist requirement of the 'correspondence rule' for meaningfulness. It may, therefore, not come as a surprise that the RTM was fully endorsed by Logical Positivists as the sole adequate theory of measurement. Unfortunately, such a Logical Positivist account employs a rather naïve understanding of scientific instrumental design and use. Scientific instruments are simply taken as 'amplifiers', or 'magnifiers', of our physiologically-limited sensory capacities, and have therefore no independent epistemic role of their own in the exploration of the world. The epistemic importance of instruments is simply reduced to the epistemology of commonsense-experience, and for this reason empiricist philosophers have given scant attention to the epistemic role of instruments and scientific instrument design.

For this final analysis of the cases of measurement in this thesis I prefer to reformulate the problem of establishing satisfactory measurement in a slightly different way without being inconsistent with the RTM: namely, as the problem of constructing measuring instruments in order to establish quantitative facts about (known or unknown) phenomena. Measuring instruments serve to obtain facts about a phenomenon, and the data, the measures we take, by means of our measuring instrument, are taken as these facts (see also Boumans, 2005a). Therefore:

Data → Measuring instrument → Facts about phenomenon

[2] Such as, for example, the idea of unity of science, reductionism, and the theory-data distinction.

Framing measurement in this way, by applying a phenomenon-data distinction rather than a theory-data distinction, seems to have three important benefits.

In the first place, it employs the phenomenon-data distinction, as suggested by Bogen and Woodward (1988), and that means that a theory of measurement can be applied to phenomena that are not (yet) fully understood. Measurement is only possible when:

i) it is known what is the essential, qualitative property that defines a phenomenon; and,

ii) this property can be ordered quantitatively according to a rule of ordinality, that is, we can talk in a meaningful way about 'more' or 'less' with respect to this property.

And, in the RTM, this process is confined to bringing about a strict isomorphic mapping of the essential, qualitative property to a numerical relational structure. Unfortunately the essential, qualitative property that defines a phenomenon is not always clear. This was obvious in Chapter 2, in the case of unemployment in the 19th century in the Netherlands (and many other countries). But let us take, as another example, the phenomenon of 'managerial efficiency'. The phenomenon makes sense, since some people are obviously much better managers than others, and are paid so accordingly. In addition, people talk in a meaningful way about 'better' or 'worse' managers, which suggest that some sort of ordering can be made. Yet, there is no objective rule for classifying some aspect of observable objects as manifestations of managerial efficiency. There are no objective empirical relations, such as precedence or succeeding, with respect to managerial efficiency. Ordering of managerial efficiency is therefore impossible. Any classification of managerial efficiency is based on value judgements. Finkelstein therefore concludes that: "The search for measuring such a conceptual entity as 'managerial efficiency' must fail until the concept is clarified" (Finkelstein, 1982: 10). And, since the clarification of such a concept falls beyond the scope of measurement theory, that seems to be the end of the story: 'managerial efficiency' cannot be measured within the framework of the RTM.

Bogen and Woodward, on the other hand, provide a more encompassing account of phenomena. In their account phenomena are the subject of scientific theories and scientific inquiry. Phenomena are considered as stable, repeatable features of the world (see Figure 6.2). However, in general, phenomena are not directly observable to the researcher. What we do observe are readings of gauges, interpretations of photos, scatter diagrams, and so on. In other words, what we observe are data, which serve as evidence for the phenomenon of interest. However, during the process of establishing the data, i.e. the measurement process, disturbing, confounding influences take place (see also Appendix 2.A). The establishment of data is thus subject to unsystematic measurement errors, idiosyncratic to the measurement or experiment setup. Therefore, establishing data provides us with a scatter of measurement results, and in scientific investigations these standard errors, calculated from the variance of the distribution of measurement errors, are usually reported. From these observations, phenomena can be detected by discovering patterns in the data, a point emphasized by McAllister (1997). Bogen and Woodward do not rule out that theory suggests or indicates phenomena. The notion of a melting point for lead, for example, might well follow from theory. The exact determination of this melting point can, however, only be done by inference from the data, and not from theory. Scientific instruments thus function as epistemic intermediars between the phenomenon and the data, and not the theory and the data as Rothbarth and Slyden (1994) suggest.

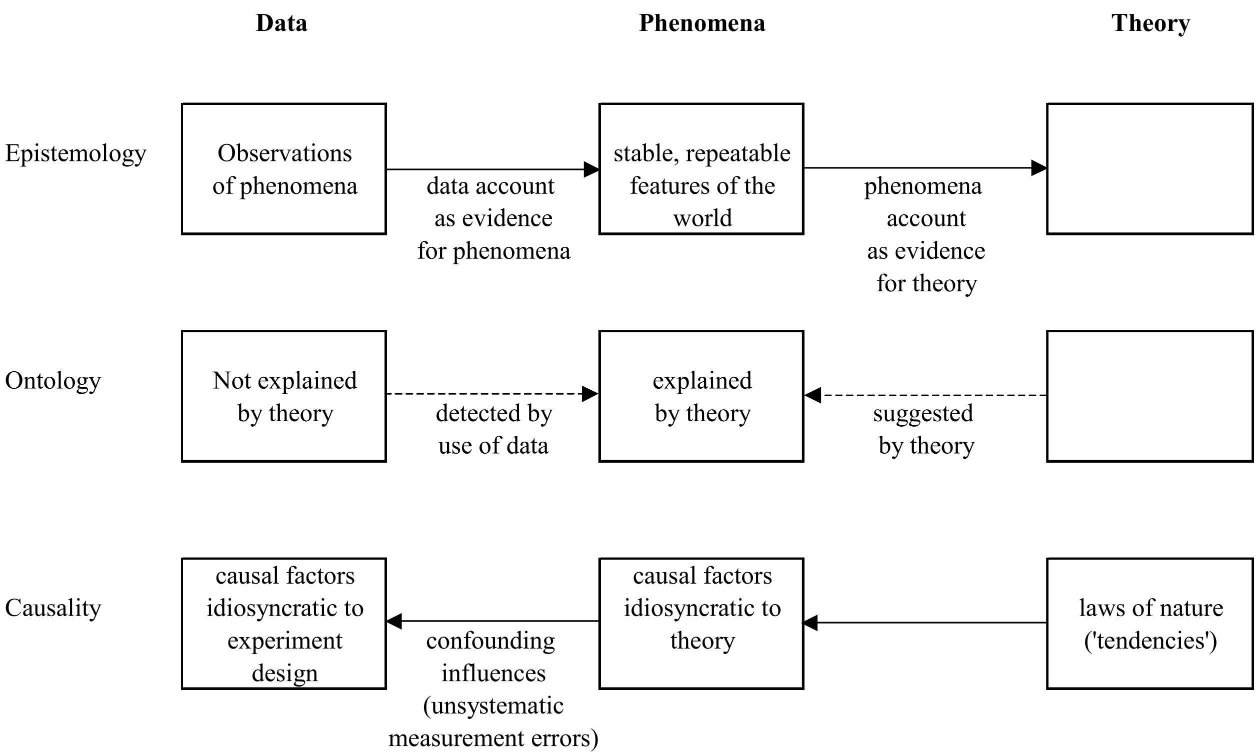


Figure 6.2: Phenomena-data distinction
Source: Based on Bogen and Woodward, 1988

Thus, by measurement we establish the data, while at the same time we want to measure the phenomenon for which the data is taken as evidence. Consequently, we must not only obtain the facts about a phenomenon but also ensure that what we measure is indeed the phenomenon under scrutiny and not something else. The first benefit of this phenomenon-data approach is that it calls for a close examination of the accuracy of measurement. This point will be discussed in more detail in Section 6.3.

The second advantage of using this phenomenon-data distinction is that it allows for a systematic account of measurement errors. The RTM has a problem with measurement errors as it does not account for them. The problem is that, in a strict isomorphic mapping, repetitions of the measurement procedure necessarily must yield exactly the same outcomes, as confounding influences do not feature in the mapping itself. In every measurement setup, however, disturbing causal factors which cause unsystematic measurement errors are inevitable, but, since they are not accounted for in the axiomatized isomorphic mapping of the RTM, all measurements must necessarily have exactly the same values. However, as is well known, as a result of unsystematic measurement errors, the measurement of the melting point of lead will yield a scatter diagram of measurement results from which we must infer the 'true' melting point. Luce and Narens (1994:227) mention this problem of the RTM with measurement errors as a second class of criticisms in their list of 15 problems concerning the RTM, and they suggest solving it by a probabilistic approach to measurement errors compatible with the RTM. This seems not completely convincing, and possibly a phenomenon-data distinction, as suggested by Bogen and Woodward, provides a richer and more promising way to proceed and to incorporate a theory of measurement errors into a theory of measurement.

Finally, the third benefit of interpreting measurement by applying a phenomenon-data distinction, and more general, frame measurement in a construction-of-measuring-instrument framework, is that it makes more explicit the empirical and experimental requirements of measurement, as well as the (epistemic) role of measuring instruments. Measuring instruments must bring about the isomorphic mapping between the empirical relational structure and the numerical relational structure (the Representational Theorem). That is, they must constitute some empirical satisfactory rule of ordinality, or as Brian Ellis calls it, a "set of ordering relationships" (Ellis, 1966).

In constructing measuring instruments, we also have to determine a scale of measurement (= Uniqueness Theorem). Since this is an important requirement in the RTM, here this element will be incorporated in the construction-of-measuring-instrument framework for the assessment of the cases of measurement. So, since measuring instruments establish data, which serve as evidence for a phenomenon, the final framework, runs as:

$$\text{Data} \quad \rightarrow \quad \begin{pmatrix} \text{Measuring instrument} \\ + \text{ determination scale} \end{pmatrix} \quad \rightarrow \quad \text{Facts about phenomenon}$$

Both requirements of the RTM have thus to be incorporated in the construction of measuring instruments. By framing the establishment of an isomorphic mapping in the framework of a construction-of-measuring-instrument, measurement regains an empirical and experimental dimension.

Finally, it is necessary to remark that not all authors agree with Bogen and Woodward's phenomenon-data distinction. Bruce Glymour for example, disagrees and argues that "no such distinction between data and phenomenon is needed, and that the distinction which is needed, [theory-data distinction] is already well established in the relevant literature" (Glymour, 2000: 32). However, I hope to have made clear in the above discussion that the phenomenon-data distinction provides a richer framework for the assessment of measurement and scientific measuring instrument design than the RTM does.

By applying the phenomenon-data distinction, measurement becomes more closely connected to current debates in the philosophy of science, while it can easily be seen that framing measurement in this way is not inconsistent with the RTM. What is called a class of qualities Q in Figure 6.1 is simply denoted as a phenomenon, while the set of numbers N refers to the data. It is, however, necessary to make the implicit assumption that a phenomenon (such as managerial efficiency or unemployment) can be ordered with respect to at least one particular qualitative property. Otherwise, the phenomenon is indeed rendered immeasurable.

Next, this framework will be used for the assessment of the cases of measurement and to draw conclusion.

6.2 The measuring instruments of unemployment

In the previous chapters, four cases of measurement have been analysed. The phenomena measured in these cases were, respectively, Dutch unemployment from 1900-1940; cyclical unemployment; the NAIRU; and duration dependence. All the cases finally yielded measurements. But what constituted the measuring instruments in these cases of measurement? Let us analyse them in closer detail.

In the literature on measurement, a distinction is sometimes made between *observed variables* (variables that can be observed) and *latent variables* (unobservable variables). None of the phenomena in this thesis turned out to be directly observable or accessible to our senses, and could therefore be considered as latent variables. In general, latent variables are considered to be accessible and hence measurable through observable variables. However, this distinction between 'observable' and 'unobservable' seems rather uncomfortable to me. Economic variables such as savings, investments, price inflation, human capital, and unemployment etc, all appear to be not directly observable. The concept of unemployment in Chapter 2 turned out to be unobservable, but was used as 'observable' input in the measurement of the NAIRU. Observability therefore seems to be a matter of degree and rather forms a continuum, rendering the distinction between observable and unobservable not very fruitful.[3] For the sake of clarifying the argument, however, the distinction between observable and unobservable will be maintained here.

For the measurement of unobservable phenomena, we thus have to establish a principle of measurement that relates the unobservable phenomenon of interest to related observable characteristics. In the case of Dutch unemployment 1900-1940 (Chapter 2), for example, the unobservable phenomenon of unemployment was accessed through data from the registrations of jobless workers. These registrations (data), such as the Index Number of Unemployment (INU), the Percentage of Unemployment (PU), the number of Days of Unemployment per Unemployed per Week (DUW), and registered unemployed at labour exchanges, were taken as facts about the phenomenon of unemployment and served as evidence for it. In this case, bureaucracies, labour exchanges and trade union unemployment funds, served as the measuring instrument.

However, we must somehow be confident that our observable measures are produced reliably and that they stand in for the unobservable phenomenon of interest. Brian Ellis requires for sound measurement "assignment of numerals to things according to any determinative, non-degenerate rule" (Ellis, 1966: 41), where *determinative* refers to the reliability of the measurement: that under the same conditions the same numbers are assigned; and *non-degenerate* means we must avoid degenerate rules, such as: 'assign the number 2 to everything'. In other words, we need some principle that reliably relates unobservable

[3] In the philosophy of science the argument that a distinction between observable and unobservable is fruitless and theoretically vague has been made by Grover Maxwell (1962). See also Bogen and Woodward (1988). Ian Hacking refers to the stance, that there is no significant distinction between observable and unobservable entities, as a 'conservative (realistic) response to the theory-observation distinction' (Hacking, 1983: 170-171).

phenomena to observable characteristics, upon which we can build measuring instruments. Here, this will be referred to that as the 'principle of reliability' of a measuring instrument. Perhaps the most obvious candidate for such a principle, which connects observables with unobservables, are causal principles. After all, causal connections imply that effects necessarily and reliably happen when their cause arises. Thus, when unobservable phenomena have known, observable effects, we can utilize that for the construction of measuring devices. The thermometer for the measurement of heat, for example, is built upon an underlying causal principle: namely, that mercury expands when heat rises. As is well understood, causality is the result of interaction between the triple: cause, effect, and background conditions. For the construction of measuring instruments, this implies that disturbing background influences need to be shielded off as much as possible. Thermometers with gas fillings are therefore preferred over liquid thermometers, since the disturbing influence of container expansion is minimalized. Deterministic causal relations are one way to reliably bring about the strict isomorphic

	Phenomenon	**Measuring instrument**	**Principle of reliability**	**Facts about phenomenon**
Empirical Concept	Dutch Unemployment 1900-1940	Bureaucracy	TU- SQR LEx-SQR	PU, INU, DUW, Registered unemployed
Theoretical Concepts	Cyclical Unemployment	UV-curve	Correlation	U-V as % of total unemployment
	NAIRU	a) Structural model of Phillips curve b) VAR model	Causal relation Regression	% total unemployment
	Duration dependence	duration model of unemployment	stable mechanism	α $(0 < \alpha < n)$

Figure 6.3: Overview of measuring instruments and principles of reliability

mapping that the RTM requires. The cases of measurement in this thesis showed that measuring instruments also rely on other principles of reliability than causal principles.

In the case of Dutch unemployment 1900-1940, the principle of reliability, the principle that brought about the connection between the unobservable phenomenon of 'unemployment' and something observable (stamping, registering etc.) were the Standardized Quantitative Rules (SQRs) embedded in these bureaucracies. They brought about the correspondence between the phenomenon and the data and constitute the principle upon which the measuring instrument of this case of measurement rests. In the 1930s, Burns and Mitchell used similar SQR's for the construction of business cycle barometers in their research of business cycles (van den Bogaard, 1998). It was the standardization that gave us trust in numbers, and not the 'causality' that the unemployed go to stamp for benefit.

In Chapter 3, the unobservable of 'cyclical unemployment' was accessed through the difference between unemployment and vacancy data, which were taken to correlate with the business cycle and cyclical unemployment. The principle of reliability here was 'correlation'. As in the case of thermometer, there are causal factors at work in this instance. But, as Heidelberger (1994a, b) argues, and the chapter shows, we do not need to know all of them, and can build measuring instruments upon mere correlation, as long as the correlation is invariant over the domain of operation.

In the case of measurement of the NAIRU, two approaches were considered. In the structural approach, measurement was based on capturing the underlying casual factors of wage and price setting in structural models of the Phillips curve relation. The statistical approach explored regression as a way to establish a mapping between phenomenon and numbers. Regression seems some halfway house between correlation and causality. It is neither correlation nor causality, but features a certain lawlike quality that was exploited by means of Vector Auto Regression (VAR) models. The often heard critique of VAR models that: they are 'atheoretical macroeconometrics'; they use ad hoc dynamics for identification; and their correlations have no proper economic interpretation, does not matter from the point of view of measurement, as long as the correspondence they bring about is stable.

Duration dependence was measured by a duration model (instrument) that represents a stable mechanism: namely, the outflow process out of unemployment. The unobservable phenomenon of duration dependency was connected to observable transition data by the models' internal and bridge principles. Figure 6.3 provides an overview of the measuring instruments used in the four case studies and the measuring devices and principles of reliability upon which they rest.

6.3 Accuracy, concept formation and measurability

The Representational Theorem requires that measuring instruments constitute isomorphic mappings, and this thesis showed how such mappings can be brought about and embodied in scientific instruments in different ways. A practical requirement of the isomorphic mapping is that it consists of a stable correspondence, or 'as stable as possible', and so addresses the invariance problem (see, for example, Boumans, 2005b: 752). But, how do we know the mapping is justified? How do we know the mapping yields unbiased measurements, as there seems no way to assess this isomorphic mapping directly in order to evaluate the measuring instruments? The problem is that both the true value of the measurand and (possible) systematic errors are unknown, and we therefore cannot ensure from the data alone that we are making a correct inference about the phenomenon, and judge whether a systematic error is present or not. This is the problem of accuracy of measurement, also referred to as validation of measurement. How was this problem of accuracy addressed?

In the RTM, the occurrence of systematic errors are simply ruled out on formal grounds. A determinative, non-degenerating rule or correspondence is postulated between the measurand and numbers. The case of the measurement of Dutch unemployment with an SQR (Chapter 2) showed that a determinative, non-degenerating rule, as Ellis puts it, is not enough. In that case, two standardized and non-degenerating, quantitative rules were established for the measurement of unemployment. Throughout the period 1900-1940, statisticians raised questions about whether what was measured was, in fact, what was intended to measure. The concept of unemployment was not clear at a theoretical level, and statisticians questioned trade unions about the representativeness of unemployed trade union members for the total labour force.

Another way to address the problem of measuring unemployment could be setting up independent measurement operations for the same phenomenon. Again, in Chapter 2 such an approach was attempted for the measurement of unemployment by two independent SQRs, but they appeared mutually inconclusive. Addressing this problem of accuracy turned out to be very difficult, since both the true value of the measurand and the systematic error are unknown, and the justification of the accuracy of the chosen principle of reliability seems grounded on the practical experience of experts rather than on some objectifiable criteria. For example, the UV relation (Chapter 3) was grounded on good economic judgement by knowledgeable economists. And, clearly, theoretical considerations do play a role here, as the UV relation was assigned a particular shape on theoretical grounds. But convention is also an important element. The case of the measurement of Dutch unemployment with an SQR (Chapter 2) showed that a stable correspondence could also be obtained from a convention. The construction of an SQR for unemployment in the Netherlands was the outcome of the interplay between trade unions, local and central government. Given the problem of assessing accuracy, the practices of measurement focus on reducing the random error instead. Accuracy is thus obtained by aiming at precision (Boumans, 2005b: 754).

This seems a rather unsatisfactory state of affairs. Thus, while in the practice of measurement, scientists focus on precision rather than accuracy, the theory of measurement avoids making qualitative judgments about phenomena, and consequently argues that a phenomenon, such as, for example, 'managerial efficiency' cannot be measured. The problem thus seems to be that the phenomenon we want to quantify is not clear or lacks an operational definition. This was precisely the case for the phenomenon of unemployment when it was first measured in the 19th century.

The point I want to make here is that – contrary to what the RTM suggests – clarification of phenomena is an important aspect of measurement, and is not totally independent of the construction of measuring instruments. Let us examine this point in closer detail for unemployment in economics.

In the economic literature, many different concepts of unemployment occur. Hughes and Perlman (1984: 26), for example, make note of no fewer than 70 different types of unemployment in the economic literature. Most of these different classes are completely or partially incompatible or incommensurable with each other, and economists reckon that no single classification can be considered as the definite or true one. This plain fact about classes or types of unemployment raises many questions. Why do so many different types of unemployment exist? Let us start with analysing how classes of unemployment come into existence as phenomena in the first place. Concepts of unemployment come into existence, basically as the consequence of two different approaches of inquiry: an inductive, empirical approach, and an analytical, deductive approach. This clearly reflects the two modes of scientific inquiry: induction and deduction.

When classes of unemployment follow from theories, they are typically called into existence in order to denote the nature of unemployment and the causes of this particular type of unemployment. They are referred to as theoretical concepts in this thesis (see also Figure 6.3). Since many different analytical approaches (theories) have been applied to the problem of unemployment, many different theoretical classes of unemployment have been invoked as well, such as 'Keynesian', 'Marxian', or 'Classical unemployment'. This process of going from universals to particulars is called '*division*' (Benjamin: 1925). "Division means the process of specifying or enumerating the individual members of a class" (Benjamin: 1925, 459), and therefore makes clear which subgroups with the same properties are implicit in any class. An aggregate object or quantity is "analysed" or "broken up" into smaller subsets of objects, which have the same characteristics or properties in common. The outcomes of this process are identifiable as particulars.[4] In the economic literature, the term 'decomposition' is more often used instead of division (see, for example, Muysken, 1989).[5]

Concepts of unemployment, however, may also come into existence from an inductive, empirical approach. Data of particular unemployed are collected, by means of measuring instruments and grouped, classified or processed, in such a way that new classes of unemployment come into existence, where the members of each group share the same empirical characteristics or properties. Examples are classes like 'long-term unemployment', 'unemployment in the manufacturing industry', 'unemployment among highly educated workers', etc. In this thesis they are referred to as 'empirical concepts' of unemployment (see also Figure 6.3). The process by which data are ordered or "put together" from particulars to universals, is called '*classification*' (Benjamin: 1925).[6] This is the converse of the strict meaning of the term 'division'. The main purpose of a classification is, according to Sokal (1974, 1116): "to describe the structure and relationship of the constituent objects to each other and to similar objects, and to simplify these relationships in such a way that general statements can be made about classes of object". One reason for classification is that we think that the classes represent the 'natural' structure of the world from which we hope to learn the laws governing the behaviour of the objects (Sokal: 1974, 1116).[7] Another reason for invoking separate classes is the imposition of external constraints on the data, for example by institutions or for administrative reasons by SQRs. Thus, classification is directed by the possibility or the desire to classify data of unemployment into separate categories. In any

[4] Therefore, Benjamin also refers to this process of division as 'particularization'.

[5] In the remainder of the chapter, the outcome of the process of division is referred to as a 'classification', which may seem confusing but is more in line with the economic literature.

[6] It is also frequently used to denote the result of classification itself. Thus, the result of classification is a classification (Sokal: 1974, 1116). In this chapter, the term 'classification' is used for both the process of classification and the outcome of this process.

[7] Three stands on classification can be distinguished: naturalism; modified conventionalism; and radical conventionalism (Hollinger: 1976, 319). The first view, naturalism, seeks a 'natural' classification of objects belonging to natural kinds, i.e. objects with essential properties. The second view, modified conventionalism, states that a classification of natural kinds alone is impossible, both because the "real essence" of objects is unknown and because of human interests. As a consequence, a classification is at least partly a social or linguistic convention, and hence subject to subjective judgements. Finally, radical conventionalism rejects the idea that a meaningful distinction between 'natural' and 'conventional' can be made, and claims therefore that all classifications are conventions of some sort.

case, through bringing about patterns in data (classification), we try to capture particular phenomena, and unless it is possible to find a 'natural' classification, classifying remains a subjective matter in which we rely on conventions and definitions of the objects, as we saw in Chapter 2 of this thesis.

Though classification and division are presented here as strictly distinct approaches, this distinction is not always clear in practice. First of all, as is often the case, some exceptions or in-between cases can be found. 'Seasonal' unemployment, for example, seems to fit best in the 'causal' classification, since there are implicit notions about its (set of) cause(s). It is, however, sometimes determined inductively by classification as a pattern in unemployment data, and it is not derived from any specific economic theory. Secondly, in some instances division is only possible when classification precedes division. The classes, in which we can divide tables in a room, for example, depend upon the classes we have established by classification. For unemployment, this is the case, for example, when one wants to break up unemployment by duration. In these cases "one never divides a class into sub-classes without at the same time looking over the different sorts of objects which might be put into these sub-classes, *i.e.* classification" (Benjamin, 1925: 461). Finally, the distinction between classification and division is hardly made in the economic literature: The term 'classification' is used for both division and classification.

In the measurement literature, it is sometimes argued that classification is a rudimentary form of measurement. Stevens (1959) set the stage for this view with his influential definition of measurement which states that "measurement is the assignment of numerical to objects or events according to rule – any rule" (Stevens, cited in Ellis, 1966: 59). If one is to adopt this definition of measurement, it follows that the assignment of numbers to football players in a team is a way to measure the football players, as each player is assigned a number. In this way, classification, the forming of a class of similar objects is a way of rudimentary measurement. Even though the numbering of football players often happens according to a more or less standardized rule (that is, increasing from the goalkeeper to the forwards), the important point here is that, for the numbering of football players, the order of numbering is irrelevant and one might as well use other rules (remember: *any* rule according to Stevens). Thus, the numbers may be permuted in any way we like; it still renders satisfactory numbers for football players. By numbering the players, we measure them on a certain scale and, even if we permute the rule of assignment, we can always execute some permutation transformation so that we get the original numbering. Thus, “the scale remains invariant under any permutation transformation” (Ellis, 1966:59). Watanabe (1996: 60) proves Stevens’ point that classification is a kind of nominal measurement, set-theoretically.

Stevens was, however, seriously criticized for his definition and its implications, for instance, that we can consider classification as a way of measurement.[8] The main argument is that, in his definition no ordering principle (or rule of ordinality) is required, so that we cannot speak meaningfully about ‘more’ or ‘less’. In a class of similar objects, classification provides no ordering principle, and hence measurement of the objects by classification alone is not possible. Stevens’ definition is, therefore, considered as too broad and classification is generally not seen as measurement.

Nevertheless, classification is an important step in any measurement procedure that defines the set of properties Q to be measured (Figure 6.1). The team of football players, for example, forms a subclass of all football players, and before we can assign numbers to the players of the team at all, it is necessary to identify, out of the class of all football players, the players who belong to the same team. It is, therefore, true that “classification procedure precedes all measurement” (Watanabe, 1996: 59), and in the theory of measurement this classification process seems rather ignored “on the grounds that it concerns the qualitative aspects of objects rather than their quantitative aspects” (ibid.). Figure 6.4 summarizes the main characteristics of these kinds of concepts.

Clarifying a concept is about classification and division, and necessarily involves qualitative judgments. Classification and division are therefore important aspects of measurement. They are necessary for and precede *all* measurement The research in this thesis suggests that clarifying a concept is a process of classification and division, and that classification (and division) are a process of clarifying a concept. They are mutually constitutive.

---

[8] See, for example, Ellis (1966).

	Empirical concepts	Theoretical concepts
Examples	Registered unemployment unemploym. percentage registered vacancies	Frictional, cyclical, search natural, NAIRU etc.
Defined by	Convention	Theory
Origin	Classification	Division
Type of Measurement	Direct measurement	Indirect measurement Associative measurement
Measuring Strategy	Bureaucracies SQRs	Model devices "Real" bridge principles

Figure 6.4: Characteristics of empirical and theoretical concepts of unemployment

6.4 The role of measuring instruments

The reason why concepts of phenomena are unclear is because the essential, qualitative property that defines them is unclear. One reason for this could be that the phenomenon can only be defined by a whole set of essential, qualitative attributes. There is not just one attribute, but a whole set of such characteristics that define the phenomenon. Managerial efficiency for example, is likely not to be defined by one single statement. Good managers are good communicators, have knowledge about what is happening on the shop floor, motivate employees, have long-term management vision, and so on. It is the whole set of such statements that defines managerial efficiency. Ellis calls such phenomena 'cluster concepts' (Ellis, 1966: 35) and argues that quantity concepts are usually cluster concepts. Many social science concepts like, price inflation, human capital, and unemployment are indeed defined by a set of essential features, and that makes them difficult to measure. In economics, we see attempts to measure these concepts by specifying the constituent features. For example, price inflation is measured by the Consumer Price Index (CPI) that is taken as a weighted average of those essential attributes that we take to define consumer consumption. Or, in econometrics, we use measurement formulae like $y_t = c + \alpha_1 x_1 + \alpha_2 x_2 + ... a_n x_n + \varepsilon_t$, where we take the set of endogenous variables x_1 to x_n to constitute the phenomenon.

For the measurement of unemployment, there seems an additional complication, as there is a moral dimension involved. Social, political, and scientific discourse about unemployment centres on the question to what extent the unemployed themselves are responsible for their employment status. In other words, to what extent is unemployment involuntary? And this makes measurement of unemployment so difficult: it is ultimately a moral issue for which we do not even have a satisfactory, operational, economic theory (see, for example, de Vroey: 2005).

As explained above, clarification of phenomena falls beyond the scope of the RTM. However, this does not seem to do justice to the role of measuring instruments in practice. Often measuring instruments play an important role in the process of clarification of phenomena. Theoretical disagreement about essential features of unemployment or about the causes of unemployment has not stopped statisticians, politicians and bureaucrats from constructing measuring instruments and simply measuring unemployment. Porter (1995) teaches us that though standard rules are essentially conventions, we can use them for quantification and let phenomena gain significance. Bureaucracies bring into being classifications that do not exist in a pure state, like classes such as "juvenile unemployment", or "unemployment among low-educated workers", and so on. Morgan (2001b: 248-249) shows how, through measurement, phenomena become detached from the specific theory from which they originate and become accepted in everyday life as theory-neutral (but not value-free) measures. Though this perhaps seems to have the flavour of operationalism, I am inclined to think it is not so. Sometimes there are ways to assess the accuracy of measurement, such as measuring the same phenomenon by independently different measurement operations. But even when that is not possible, it is important to realise that measurement is inevitably an iterative process, from which clarification of phenomena is inseparable. Even supporters of the RTM seem to realize this epistemic role of measuring instruments. As Finkelstein argues: "the accumulation of data by measurement leads to the clarification and re-evaluation of the quality concept" (Finkelstein, 1982: 11), a remarkable point of view, considering that in the RTM, measurement of a phenomenon requires a strict isomorphic mapping in the first place, and a re-evaluated phenomenon another strict isomorphism.

The relation between the measuring instrument and the phenomenon is not only a one-way relation; it works the other way as well. In the 1960s, the Phillips curve became such a central issue in macroeconomics that, in theorizing about this empirical unemployment-inflation relationship, new concepts such as the natural rate of unemployment and the NAIRU came into being. These concepts are so inextricably connected to the Phillips curve that attempts to make them operational almost automatically follow from the description of this empirical relationship. Without the Phillips curve, it would not probably make sense for economists to talk about such a thing as the NAIRU. Also, bureaucracies create new entities. The relation between phenomena and measuring instruments is therefore a reciprocal relationship:

Data ↔ Measuring instrument ↔ Facts about phenomenon

Clarification of phenomena is thus an important aspect of measurement, and is not totally independent of the construction of measuring instruments. Measuring instruments do play an important epistemic role in science. They are not simply neutral 'amplifiers', or 'magnifiers', of our physiologically-limited sensory capacities. This is not because they embed systematic measurement errors of some kind and, in this way, are responsible for biased access to phenomena. And nor is it because scientists are struggling with the sort of facts they produce and have difficulty in forcing them into conformity with a particular theory or paradigm as Thomas Kuhn (1961) argues. Rather it is because measuring instruments act as mediators between data and phenomena, and as such not only provide us with quantitative facts about phenomena, but also help to clarify phenomena in the first place.

References

Abraham, K.G. (1983a), Structural/Frictional vs. Deficient Demand Unemployment: Reply, *The American Economic Review*, vol. 76, no. 1, 273-276

Abraham, K.G. (1983b), Structural/Frictional vs. Deficient Demand Unemployment: Some New Evidence, *The American Economic Review*, vol. 73, no. 4, 708-724

Abraham, K.G. and Katz, L.F (1986), Cyclical Unemployment: Sectoral Shifts or Aggregate Demand Disturbance, *The Journal of Political Economy*, vol. 94, no. 3, 507-522

Achinstein, P. (1964), Models, Analogies, and Theories, *Philosophy of Science*, vol. 31, 328-350

Adams, E., (1966), On the Nature and Purpose of Measurement, *Synthese*, vol. 16, 125-169

Albrecht, J. W. and Axell, B. (1984), An Equilibrium Model of Search Unemployment, *The Journal of Political Economy*, vol.92, no. 5, 824-840

Angrist J.D. and Kreuger, A.B. (1999), Empirical Strategies in Labor Economics, in: Ashenfelter, O., and Card, D. (eds.), *The Handbook of labor Economics*, Volume 3a, ch. 23

Armstrong H. and Taylor, J. (1980), The Measurement of Different Types of Unemployment, in: Creedy, J. (ed.), *The Economics of Unemployment in Britain*, Buttersworths, London

Backhouse, R.E., (ed.), (1997), *New Directions in Economic Methodology*, Routledge, London

Bailer-Jones, D. M. (2002), *It is still Mechanisms that make us Understand*, working paper

Belderbos, R.A. and Teulings, C.N. (1989), *A New Interpretation of the Relation between Unemployment and Vacancies,* SEO research memorandum no. 8904, University of Amsterdam

Benjamin, A.C., (1925), Classification and Division, *The Journal of Philosophy*, vol. 22, no. 17 458-463

Berg, D.J. van der (1982), Specification and Estimation of the Unemployment-Vacancy Curve in the Period 1956-1979, *De Economist*, vol. 130, no. 3, 397-419

Beveridge, W. (1909) *Unemployment: A Problem of Industry*, Longmans, Green and Co, London

Beveridge, W. (1930), *Unemployment: A Problem of Industry*, Longmans, green and Co, London

Beveridge, W. (1944), *Full Employment in a Free Society*, George Allen & Unwin Ltd, London

Bewley, R.A. (1979), The Dynamic Behaviour of Unemployment and Unfilled Vacancies in Great Britain: 1958-1971, *Applied Economics*, vol. 11, 303-308

Bie, R. van der, (1995), *Een doorlopende groote roes; De economische ontwikkeling van Nederland, 1913-1921*, Tinbergen Institute PhD thesis, no. 98, Amsterdam

Blanchard, O. J., and Quah, D. (1989), The Dynamic Effects of Aggregate Demand and Supply Disturbances, *The American Economic Review*, vol. 79, no. 4, 655-673

Blanchard, O.J. and Diamond, P.A. (1989), The Beveridge Curve, *Brookings Papers on Economic Activity*, vol. 1989, no. 1, 1-60

Blaug, M. (1980), *The Methodology of Economics, or How Economists Explain*, Cambridge University Press, New York

Blaug, M. (1985), *Economic Theory in Retrospect*, Cambridge University Press, Cambridge

Bleakley, H. and Fuhrer, J.C. (1997), Shifts in the Beveridge Curve, Job Matching and Labor Market Dynamics, *New England Economic Review*, sept/oct., 5-19

Bogaard, van den A. (1998), *Configuring the Economy; The Emergence of a Modelling Practice in The Netherlands, 1920-1955,* University of Amsterdam, PhD thesis

Bogen J. and Woodward, J. (1988), Saving the Phenomena, *The Philosophical Review*, vol. 67, no. 3, 303-352
Boianovsky and Trautwein (2003), Wicksell, Cassel and the Idea of Involuntary Unemployment, *History of Political Economy,* vol. 35, no. 3, 385-436
Borsch-Supan A.H. (1991), Panel Data Analysis of the Beveridge curve: Is there a Macroeconomic Relation between the Rate of Unemployment and the Vacancy Rate?, *Economica,* vol. 58, no. 231, 279-298
Boskin, M.J., e.a. (1998), The Consumer Price Index, and the Cost of Living*, Journal of Economic Perspectives*, vol. 12, no. 1, 3 –26
Boumans, M.J. (1992), *A Case of Limited Physics Transfer; Jan Tinbergen's Resources for Re-shaping Economics*, Tinbergen Institute PhD thesis, no. 38, Amsterdam
Boumans, M. (1999a), Representation and Stability in Testing and Measuring Rational Expectations, *Journal of Economic Methodology*, vol. 6, no. 3, 381-401
Boumans, M. (1999b), Built-in Justification, in: Morgan, M. S. and Morrison M. (eds.), (1999), *Models as Mediators*, Cambridge University Press, Cambridge
Boumans, M. (2001), Measure for Measure: How Economists Model the World into Numbers, *Social Research*, vol. 68, no. 2, 427- 453
Boumans, M. (2005a), *How Economists model the World into Numbers*, Routledge, London
Boumans, M. (2005b), Economics, Strategies in Social Sciences, in: Kempf-Leonard, K. (ed.) *Encyclopedia of Social Measurement*, vol. 1, Elsevier
Bowden, R.J. (1980), On the Existence and Secular Stability of u-v Loci, *Economica,* vol 47, no. 185, 35-50
Bowers, J.K., Cheshire, P.C. and Webb, A.E. (1970), The Change in the Relation between Unemployment and Earning Increases: A Review of some Possible Explanations, *National Institute Economic Review*, no. 54, 44-63
Bowers, J.K. (1976), Some Notes on Current Unemployment, in: Worswick, G.D.N. (ed.), *The Concept and Measurement of Involuntary Unemployment*, George Allen & Unwin Ltd, London
Brain, R. M., and Norton Wise (1999), M., Muscles and Engines; Indicator Diagrams and Helmholtz's Graphical Methods, in: Biagioli, M. (ed.), *The Science Studies Reader*, Routledge, New York and London
Bridgman, P.W. (1927), *The Logic of Modern Physics*, MacMillan, New York
Bridgman, P.W. (1938), Operational Analysis, *Philosophy of Science*, vol. 5, 114-131
Brink, van der J.R.M. (1942), *Maatschappijstructuur en werkgelegenheid,* N.V. Dekker&Van der Vegt, Utrecht
Brown, A.J. (1973), *The Framework of Regional Economics in the United Kingdom*, National Institute of Economic and Social research, Regional Papers I, Cambridge University Press
Brown, A.J.(1976), UV Analysis, in: Worswick, G.D.N. (ed.), *The Concept and Measurement of Involuntary Unemployment*, George Allen & Unwin Ltd, London
Budd, A., Levine, P. and Smith, P. (1987), Long-term Unemployment and the Shifting U-V Curve, A Multi-Country Study, *European Economic Review*, vol. 31, 296-305
Burch, S.W. and Fabricant, R.A. (1968), A further Comment on the Behavior of Help-Wanted Advertisement, *The Review of Economics and Statistics,* vol. 50, no. 2, 278-281
Burch, S.W. and Fabricant, R.A. (1971), Cyclical Behaviour of Help-Wanted Index and the Unemployment Rate: A Reply, *The Review of Economics and Statistics,* vol. 53, no. 1, 105-106
Butler, R. (1997), Stories and Experiments in Social Inquiry, *Organizational Studies*, vol. 18, 927-948
Butter, F.A.G. den, and Morgan, M.S., (eds.), (2000), *Empirical Models and Policy-Making: Interactions and Institutions*, Routledge, London and New York

Butters, G.R., (1977), Equilibrium Distributions of Sales and Advertising Prices, *The Review of Economic Studies*, Vol. 44, 465-491
Campbell, N.R. (1928), *Measurement and Calculation, An Account of the Principles of,* Longmans, Green and Co. Ltd., London
Canova, F. (1995), The Economics of VAR models, in: Hoover, K. (ed.), *Macroeconometrics; Developments, Tensions and Prospects*, Kluwer Academic Publishers
Carnap, R. (1991), Logical Foundations of the Unity of Science, in: Boyd, R. Gaspers, Ph. and Trout, J.D. (eds.), *The Philosophy of Science*, MIT Press, Cambridge
Cartwright, N. (1983), *How the Laws of Physics Lie*, Clarendon Press, Oxford
Cartwright, N. (1989), *Nature's Capacities and their Measurement,* Oxford University Press, Oxford
Cartwright, N. (1997), Models: the Blueprints for Laws, *Philosophy of Science*, vol. 64, proceedings of the 1996 Biennial Meetings of the Philosophy of Science Association, part II, 292-303
Cartwright, N. (1998a), Causation, in: Edward Craig (Ed), *Routledge Encyclopedia of Philosophy*, Routledge
Cartwright, N. (1998b), Capacities, in Davis, J. B., Hands, D.W. and Maki, U., *The Handbook of Economic Methodology,* Edward Elgar, Cheltenham, UK
Cartwright, N. (1999), *The Dappled World*, Cambridge University Press, Cambridge
CBS, (1941), De ontwikkeling der openbare arbeidsbemiddeling 1900-1941, *Maandschrift van het Centraal Bureau voor de Statistiek*, april 1941, 241 - 244
CBS, (1929,1930,1931), *Maandschrift van het Centraal Bureau voor de Statistiek*, Centraal Bureau voor de Statistiek, Voorburg
CBS, (1901, 1902, 1915–1932, 1935–1966), *Jaarverslag*, Centraal Bureau voor de Statistiek, Voorburg
CBS, five anonymous, undated internal papers in CBS archive (written 1936-1945), CBS-Archive, Voorburg, Map 408.
CCS, Centrale Commissie voor de Statistiek (1892/1893), *Jaarverslag 1892/1893,* Centrale Commissie voor de Statistiek, Voorburg
Chang H., (2001), Spirit, Air, and Quicksilver: The Search for the "Real" Scale of Temperature, *Historical Studies in the Physical and Biological Sciences*, vol. 32, no. 2, 249-284
Chang H. (2004), *Inventing Temperature*, *Measurement and Scientific Progress*, Oxford University Press, Oxford
Chao, H.K. (2002), *Representation and Structure; The Methodology of Econometric Models of Consumption*, Tinbergen Institute PhD thesis, no. 289, Amsterdam
Charemza, W.W. and Deadman, D.F. (1992), *New Directions in Econometric Practice*, Edward Elgar, Cheltenham
Cheshire, P.C. (1973), *Regional Unemployment Differences in Great Britain*, National Institute of Economic and Social research, Regional Papers II, Cambridge University Press, Cambridge
Chiarini, B. and Piselli, P. (2001), Identification and dimension of the NAIRU, *Economic Modelling*, vol. 18, 585-611
Churchland, P.M. and Hooker, C. A., (eds.), (1985), *Images of Science*, University of Chicago Press, Chicago
Cohen, M.S. and Solow, R.M. (1967), The Behaviour of the Help-Wanted Advertising, *The Review of Economics and Statistics,* vol. 49, no.1, 108-110
Cohen, M.S. and Solow, R.M. (1970), The Behaviour of the Help-Wanted Advertising: A Reply, *The Review of Economics and Statistics,* vol. 52, no. 4, 442-443
Commissie ingesteld door het hoofdbestuur der Nederlandse Maatschappij voor Nijverheid en handel te Haarlem (1940), *Registratie van arbeidskrachten*, N. Samson N.V., Alphen aan den Rijn

Commissie van onderzoek naar de omvang en feitelijke beteekenis der werkloosheid te Amsterdam (1927), *Rapport der Commissie van onderzoek naar de omvang en feitelijke beteekenis der werkloosheid te Amsterdam*, Stadsdrukkerij, Amsterdam
Collier, A. (1998), Critical Realism, in: Edward Craig (Ed), *Routledge Encyclopaedia of Philosophy*, Routledge
Cooley, T.F. and LeRoy, S.F. (1985), Atheoretical Macroeconomics; A Critique, *Journal of Monetary Economics*, vol. 16, no. 3, 283-308
Coombs, C.H., (1952), A Theory of Psychological Scaling, *Engineering Research Institute Bulletin*, no. 34, University of Michigan Press, Ann Arbor
Cross, R. (1991), Alternative Accounts of Equilibrium Unemployment, in: De Marchi, N. and Blaug, M. (ed.) *Appraising Economic Theories; Studies in the Methodology of Research Programs*, Edward Elgar Publishing Ltd, Hants
Cross, R. (ed). (1995), *The Natural Rate of Unemployment; Reflections on 25 Years of the Hypothesis*, Cambridge University Press, Cambridge
Davis, J.B., Wade Hands, D., and Mäki, U. (eds.), (1998), *Handbook of Economic Methodology*, Edward Elgar, Cheltenham
Davis, J.B., Marciano, A., and Runde, J. (eds.), (2004), *The Elgar Companion to Economics and Philosophy,* Edward Elgar, Cheltenham
De Marchi, N. and Gilbert, C. (1989), *History and Methodology of Econometrics*, Clarendon Press, Oxford
Diamond, P.A. (1982), Aggregate Demand Management in Search Equilibrium, *Journal of Political Economy*, vol. 90, no. 5, 881-894.
Diamond, P.A. (1982), Wage Determination and Efficiency in Search Equilibrium, *Review of Economic Studies*, vol. 61, no. 3, 457-475
Diamond, P.A. (1984), *A Search-Equilibrium Approach to the Micro Foundations of Macroeconomics*, MIT Press, Cambridge
Diamond, P. A. (1998), Search Theory, in: Eatwell, J., Milgate M. and Newman, P. (eds), *The New Palgrave Dictionary of Economics*, Macmillan Press Ltd., London
Dow, J.C.R. and Dicks-Mireaux, L. (1958), The Excess Demand for Labour: A Study of Conditions in Great Britain, 1946-1956, *Oxford Economic Papers*, vol. 10, no. 1, 1-33
Dow, J.C.R. (1970), *The Management of the British Economy, 1945-1960*, Cambridge University Press, Cambridge
Ellis, B. (1966), *Basic Concepts of Measurement*, Cambridge University Press, Cambridge
Enders, W. (2004), *Applied Econometric Time Series*, John Wiley & sons, Inc., Hobroken
Erwich B., and Maarseveen, J.G.S.J. van (eds.), (1999), *Een eeuw statistieken. Historisch-methodologische schetsen van de Nederlandse officiële statistieken in de twintigste eeuw*, Amsterdam: Stichting beheer IISG
Evans, G.J. (1975), A Note on Trends in the Relations between Unemployment and Unfilled Vacancies, *The Economic Journal*, vol. 85, no. 337, 135-139
Evans, G.J. (1977), Notes on the Changing Relationship between Registered Unemployment and Notified Vacancies, 1961-1966 and 1966-1971, *Economica*, vol. 44, no. 174, 179-196
Fisher, I. (1922), *The Making of Index Numbers; A Study of Their Varieties, Tests, and Reliability*, August M. Kelley Publishers, New York
Finkelstein, L. (1975), Fundamental Concepts of Measurement: Definitions and Scales, *Measurement and Control*, vol.8, March 1975
Finkelstein, L. (1982), Theory and Philosophy of Measurement, in: Sydenham P.H. (ed.): *Handbook of Measurement Science*, John Wiley & Sons Ltd.
Flinn C.J. and Heckman J.J., (1983), Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States, *Journal of Labor Economics*, vol. 1, no. 1, 28-42

Folmer, Anth. (1916), De Intercommunale Arbeidsbemiddeling, *Tijdschrift der nationale vereeniging tegen de werkloosheid*, vol. 5, 237-277, and 424-429.
Foster, J.I., (1973), The Behaviour of Unemployment and Unfilled Vacancies: Great Britain, 1958–1971; A Comment, *The Economic Journal*, vol. 83, no. 329, 192-201
Foster, J.I., (1974), The Relationship between Unemployment and Vacancies in Great Britain, 1958-72: Some further Evidence, in: Laidler, D. and Purdy, D.L. (eds.), *Inflation and Labour Markets*, Manchester University Press, 164-196
Franklin, A., et. al., (1989), Can a Theory-Laden Observation Test the Theory?, *The British Journal for the Philosophy of Science*, vol. 40 no. 2, 229-231
Friedman, M. (1953), The Methodology of Positive Economics, in: *Essays in Positive Economics*, University of Chicago, 3-43
Friedman, M. (1968), The Role of Monetary Policy, *American Economic Review*, March, no.1
Friedman, M. (1975), *Unemployment versus Inflation; An evaluation of the Phillips curve*, The Institute of Economic Affairs
Friedman, M. (1976), *Inflation and Unemployment*, Nobel Memorial Lecture, University of Chicago, Illinois, December 13
Frigg, R., (2003), *Re-presenting Scientific Representation*, Ph D. thesis, London School of Economics
Galenson, W. and Zellner, A. (1957), International Comparison of Unemployment Rates, in: National Bureau Committee for Economic Research, *The Measurement and Behaviour of Unemployment; A Conference of the Universities,* Princeton University Press, Princeton
Garraty, J.A. (1978), *Unemployment in History*, Harper Colophone Books, New York
Garside, W.R. (1980), *The Measurement of Unemployment in Great-Britain 1850-1979; Methods and Sources*, Basil Blackwell, Oxford
Gerritsz, J. (1916), Het ontstaan van den Nationale werkloosheids-Raad, *Tijdschrift der nationale vereeniging tegen de werkloosheid*, vol. 5, 777-781.
Gerritsz, J. (1916), De taak der Arbeidsbeursen, *Tijdschrift der nationale vereeniging tegen de werkloosheid*, vol. 5, 1139-1141.
Gerwen, J. van, (2000) *De ontluikende verzorgingsstaat; overheid, vakbonden, werkgevers, ziekenfondsen en verzekeringsmaatschappijen, 1890-1945*, deel 3, Verbond van verzekeraars, Den Haag/NEHA, Amsterdam
Gewin, B. (1898), *Arbeidsbeurzen*, A.J. v. Huffel, Utrecht
Gibbard, A. and Varian, H. (1978), Economic Models, Journal of Philosophy, vol. 75, 664-77
Gilpatrick, E. (1966), On the Classification of Unemployment: A View on the Structural-Inadequate Demand Debate, *Industrial and Labor Relations Review*, vol. 19, no. 2, 201-212
Glennan, S. S., (1996), Mechanisms and the Nature of Causation, *Erkenntnis*, vol. 44, 49-71
Glymour, B. (2000), Data and Phenomena: A Distinction Reconsidered, *Erkenntnis*, vol. 52, 29-37
Goedhart, G.J. en Roeterink, A. (1929), *De ontwikkeling der arbeidsbemiddeling in Nederland*, H.J.W. Becht, Amsterdam
Goodheart, C.A.E. (1984), *Monetary Theory and Practice; The UK Experience*, Macmillan Press, London
Gordon, A.R. (1966), *The Goal of Full Employment*, John Wiley & Sons Inc., New York
Gordon, R.J. (1982), Inflation, Flexible Exchange Rates and the Natural Rate of Unemployment, in: Bailey, M. (ed.), *Workers, Jobs, and Inflation*, Brookings Institute, Washington
Gordon, R.J. (1997), The Time-Varying NAIRU and its Implications for Economic Policy, *Journal of Economic Perspectives*, vol. 11, no. 1, 11-32

Greene, W. H. (2003), *Econometric Analysis*, Prentice Hall, New Jersey
Gregg, P. and Petrongolo, B. (1997), *Random or Non-random Matching? Implications for the use of the UV Curve as a Measure of Matching Effectiveness*, Centre for Economic Performance, discussion paper no. 348
Gujarati, D. (1969), Cyclical Behaviour of Help-Wanted Index and the Unemployment Rate, *The Review of Economics and Statistics,* vol. 51, no. 4, 482-484
Gujarati, D. (1972a), The Behaviour of Unemployment and Unfilled Vacancies: Great Britain, 1958-1971, *The Economic Journal*, vol. 82, no. 325, 195-204
Gujarati, D. (1972b), A Reply to Mr. Taylor, *The Economic Journal*, vol. 82, no. 328, 1365-1368
Gujarati, D. (1973), A Reply to Mr. Foster, *The Economic Journal*, vol. 83, no. 329, 201-203
Haavelmo, T. (1944), The probability Approach in Econometrics, *Econometrica,* vol. 12, issue supplement July, iii-vi and 1-115
Hacking, I. (1983), *Representing and Intervening*, Cambridge University Press, Cambridge
Hacking, I. (1990), *Taming of Change*, Cambridge University Press, Cambridge
Hacking, I. (1991), Experimentation and Scientific Realism, in: Boyd, R. Gaspers, Ph. and Trout, J.D., (eds.), *The Philosophy of Science*, MIT Press, Cambridge
Hall, R. E. (1979), A Theory of the Natural Unemployment Rate and the Duration of Unemployment, *Journal of Monetary Economics,* vol. 5, 153-169
Hall, R.A. and Yellen, J. (1989), The Beveridge Curve. Comment and Discussion, *Brookings Papers on Economic Activity*, vol.1989, no. 1, 61-76
Haltiwanger, J. (1987), Natural Rate of Unemployment, In: Eatwell, J., Milgate, M., and Newman, P., (eds.), *The New Palgrave; a Dictionary of Economics,* Macmillam Reference, London
Hancock, K.J. (1963), Unemployment in Australia, *Summer School of Business Administration*, vol. 8, 35-54
Hansen, B. (1970), Excess Demand, Unemployment, Vacancies, and Wages, *Quarterly Journal of Economics,* vol. 84, no.1, 1-23
Hausman, D.M. (1984), *The Philosophy of Economics*, Cambridge University Press, Cambridge
Hausman, D.M. (1992), *The Inexact and Separate Science of Economics*, Cambridge University Press, Cambridge
Heckman, J. and Hansen L., (1996), The Empirical Foundations of Calibration, *Journal of Economic Perspectives*, vol. 10, no. 1, 87-104.
Heidelberger, M. (1993), Fechner's Impact for Measurement Theory, *Behavioral and Brain Science*, vol. 16, no. 1, 146-148
Heidelberger, M. (1994a), Alternative Interpretationen der Reprasentionstheorie de Messung, in: Meggle, G. and Wessels, U.(eds.), *Analyomen 1, Proceedings of the 1st Conference "Perspectives in Analytical Philosophy*", Walter de Gruyter, Berlin
Heidelberger, M. (1994b), *Three Strands in the History of the Representational Theory of Measurement*, Working Paper, Humbolt University Berlin
Helmholtz, H. von (1887), *Zählen und Messen erkenntnistheoretisch betrachtet*, Nachdruck Darmstadt: Wiss. Buchgesellschaft
Hempel, C.G. (1952), *Fundamentals of Concept Formation in Empirical Science*, University of Chicago Press, Chicago
Hempel, G.C. (1962), Explanation in Science and History, In: R.G. Colodny (ed), *Frontiers of Science and Philosophy,* University of Pittsburgh Series in the Philosophy of Science, vol. I., Pittsburgh,
Hempel, G.C. (1966), *Philosophy of Natural Science*, Prentice-Hall, Englewood Cliffs, N.J.
Hendry, D.F., and Morgan, M. S., eds., (1995), *The Foundations of Econometric Analysis,* Cambridge University Press, Cambridge
Holden, K. and Peel, D.A. (1975), The Determinants of Unemployment and UV Relationship, *Applied Economics,* vol. 7, 251-255

Holden, K. and Peel, D.A. (1977), The Relationship between Unemployment and Vacancies: Reply, *Applied Economics,* vol. 9, 81-82
Hölder, O. (1901), *Die Axiome der Quantität und die Lehre vom Mass*, Sächsische Berichte 53
Hollinger, R. (1976), Aspects of the Theory of Classification, *Philosophy and Phenomenological Research*, vol. 36, no. 3, 319-338
Holt, C.C and David, M.H. (1966), The Concept of Job Vacancies in a Dynamic Theory of the Labor Market, in: *The Measurement and Interpretation of Job Vacancies: A Conference Report of the National Bureau of Economic Research*, NBER, New York, 1966
Hoover K. D. (1998a), *Econometrics and Reality,* University of California, Davis, working paper
Hoover, K. D. (1998b), Causality, in: Davis, J.B., Hands, D.W. and Mäki, U. (eds.), *The Handbook of Economic Methodology*, Edward Elgar Publising Ltd., Cheltenham
Hoover, K. D. (2001), *The Methodology of Empirical Macroeconomics*, Cambridge University Press, Cambridge
Hoover, K. D. (2001), *Causality in Macroeconomics*, Cambridge University Press, Cambridge
Horst, A. van der, Jacobs, J. and Schoonbeek, L. (1996), *Is there a NAIRU for the Netherlands?* working paper, Universiteit van Groningen
Hoskin, K. (1996), The 'Awful Idea of Accountability': Inscribing People into the Measurement of Objects, in Munro, R. and Mouritsen, J. (eds.), *Accountability: Power, Ethos and the Technologies of Managing*, International Thomson Business Press, London, 265-282.
Hughes, J. J. and Perlman, R. (1984), *The Economics of Unemployment*, *A Comparative Analysis of Britain and the United States*, Wheatsheaf Books Ltd, Brighton
Hughes, R.I.G. (1997), Models and Representation, *Philosophy of Science*, 325-336
International Labour Office (1922a), *Statistics of Unemployment in Various Countries, 1910 to 1922*, Studies and Reports Series C, no. 6
International Labour Office (1922b), *Methods of Compiling Statistics of Unemployment: Replies to Governments,* Studies and Reports Series C, no. 7, Geneva
International Labour Office (1925a), The Second International Conference of Labour Statisticians, *International Labour Review*, vol. 12, no. 1, 491-512
International Labour Office (1925b), *Methods of Statistics of Unemployment: Report for the International Conference of Labour Statisticians,* Studies and Reports Series N, no. 7
International Labour Office (1934), National and International Index Numbers of the General Level of Unemployment, *International Labour Review*, no. 4, 557-571
International Labour Office (1939), World Index Numbers of Unemployment, *International Labour Review*, no. 1, 118-129
International Labour Office (1943), *The International Standardisation of Labour Statistics; A Review of the Statistical Work of the International Labour Office and the Various International Statistical Conferences*, Studies and Reports Series N, no. 25, Montreal
Jackman, R., Layard, R. and Pissarides, C. (1989), On Vacancies, *Oxford Bulletin of Economics and Statistics,* vol. 51, no. 4, 377 - 394
Jossa, B. and Musella, M. (1998), *Inflation, Unemployment and Money: Interpretations of the Phillips Curve*, Edward Elgar, Cheltenham
Kalachek, E. and Westebe, R. (1961), Rates of Unemployment in Great Britain and the United States, 1950-1960, *The Review of Economics and Statistics*, vol. 43, no. 4, 340-350
Katz, L. (1988), Some Recent Developments in Labor Economics and Their Implication for Macroeconomics, *Journal of Money, Credit and Banking,* vol. 20, no. 3, 507-522
Kiefer, N.M. (1988), Economic Duration Data and Hazard Functions, *Journal of Economic Literature*, vol. 26, no. 2, 646-679
Klep, P.M.M., and Stamhuis, I., eds., (2002), *The Statistical Mind in a Pre-Statistical Era: The*

Netherlands, 1750-1850, Askant, Amsterdam
Kloosterman R. (1985), *Werkloosheid in Nederland, 1920 –1939*, H&S, Utrecht
Knight, K.G. and Wilson, R.A. (1974), Labour Hoarding, Employment and Unemployment in British Manufacturing Industry, *Applied Economics*, vol. 6, 303-310
Koopmans, Tj. (1947), Measurement without Theory, *Review of Economics and Statistics*, vol. 29, 161-179
Kort, de W.L.P.M. (1940), *De arbeidsbemiddeling in Nederland*, N.Samson N.V., Alphen aan den Rijn
Krantz, D.H, Luce, R.D., Suppes, P., and Tversky, A. (1971), *Foundations of Measurement*, Vol. I Academic Press, New York
Kuhn, Th. (1961), The Function of Measurement in Modern Physical Science, *Isis*, vol. 52, no. 2, 161-193
Kuipers, S.K. and Buddenberg F.H. (1978), Unemployment on Account of Market Imperfections in the Netherlands since the Second World War, De Economist, vol. 126, no. 3, 390-412
Lancaster, T. (1979), Econometric Methods for the Duration of Unemployment, *Econometrica*, vol. 47, no. 4, 939-956
Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge
Latour, B. (1987), *Science in Action: How to Follow Scientists and Engineers Through Society*, Open University Press; Milton Keynes
Layard, R. and Nickel, S. (1986), Unemployment in Britain, *Economica*, vol. 53, 121-169
Layard, R., Nickell, S. and Jackman, R. (1991), *Unemployment: Macroeconomic Performance and the Labour Market*, Oxford University Press, Oxford
Leeson, R. (2000), *A.W.H. Phillips: Collected Works in Contemporary Perspective*, Cambridge University Press, Cambridge
Leunis, W. P. and Verhage, K.G. (1996), *Arbeidsrekeningen; Kern van het statistisch bestel*, Centraal Bureau voor de Statistiek, Voorburg
Lier, Th. Van, (1932), *De arbeidsbemiddelingswet 1930 en hare uitvoering*, N. Samsom NV, Alphen aan den Rijn
Lindbeck, A. (1999), *Unemployment – Structural*, Institute for International Economic Studies, Stockholm University, seminar paper no. 676
Lindberg, John (1932), An Attempt to Construct International Measures of Unemployment, *International Labour Review*, no. 4, 491-512
Lindberg, John (1934), Some Problems in the Construction of Index Numbers of Unemployment, *International Labour Review*, no. 4, 472-499
Lipsey, R.G. (1965), Structural and Deficient-demand Unemployment Reconsidered, in: McCormick, B.J. and Owen Smith E. (eds.), *The Labour Market*, Hammondsworth, 245-269
Lubbers, G. (1926), *De statistiek van het arbeidsloon en van de werkloosheid,* H.J. Paris, Amsterdam
Luce, R.D. and Narens, L. (1987) Theory of Measurement, in: Eatwell, J., Milgate, M. and Newman, P. (eds.), *The New Palgrave: A Dictionary of Economics*, MacMillan Press Ltd, London
Luce, R.D. and Narens, L. (1994), Fifteen Problems concerning the Representational Theory of Measurement. In: Humphreys, P. (Ed.), *Patrick Suppes: Scientific Philosopher,* vol. 2, 219-249, Kluwer Academic, Dordrecht
Luce, R.D., Krantz, D.H., Suppes, P.,and Tversky, A. (1990), *Foundations of Measurement*, vol. III, Academic Press, New York
Maarseveen, J.G.S.J. van, and Gircour, M.B.G., eds., (1999), *A Century of Statistics. Counting, Accounting and Recounting in the Netherlands*, Amsterdam, Stichting beheer IISG

Maarseveen, J.G.S.J. van, Gircour, M.B.G., and Schreijnders, R., eds., (1999), *A Century Rounded up. Reflections on the History of the Central Bureau of Statistics in the Netherlands*, Amsterdam, Stichting beheer IISG
Maarseveen, J.G.S.J. van, and Rodenburg, P. (2006), *Statistieken van de werkloosheid in Nederland, 1902-1943*, Instituut voor Nederlandse Geschiedenis, deel 6 reeks Bronnencommentaar, 'Bronnen met betrekking tot armenzorg en sociale verzekering in de negentiende en twintigste eeuw'
Maas, H. (2005). *William Stanley Jevons and the Making of Modern Economics*, Cambridge University Press, Cambridge and New York
Machamer, P., Darden, L. and Craver, C. (2000), Thinking about Mechanisms, *Philosophy of Science*, no. 67, 1-25
Machin, S. and Manning, A. (1999), The Causes and Consequences of Long-Term Unemployment in Europe, *Handbook of Labor Economics*, vol. 3, 3085-3139
Macmillan Dictionary of Modern Economics, (1986), David W. Pearce (ed.), London
Malinvaud, E. (1977), *The Theory of Unemployment Reconsidered*, Basil Blackwell, Oxford
Malinvaud, E. (1994), *Diagnosing Unemployment*, Cambridge University Press, Cambridge
Manning, A (1993), Wage Bargaining and the Phillips Curve: The Identification and Specification of Aggregate Wage Equations, *The Economic Journal*, vol. 103, no. 416, 98-118
Mario, V, di. *The Estimation of the NAIRU and the Effect of Permanent Sectoral Employment Reallocation;* The Italian Evidence, Istituto di Studi Economici, Universita Delgi di Napoli, working paper no. 7, 2002
Marx, K. (1867), *Das Kapital; Kritik der Politischen Oekonomie*, Verlag von Otto Meissner, Hamburg,
Maxwell, G. (1962), The Ontological status of Theoretical Entities, in: Feigl, H. and Maxwell, G. (eds.), *Scientific Explanation, Space, and Time*, volume III, Minnesota Studies in the Philosophy of Science, University of Minnesota Press
McAllister, J. W. (1997), Phenomena and Patterns in Data Sets, *Erkenntnis*, vol. 47, 217-228
Michell, J. (2003), Epistemology of Measurement: the Relevance of its History for Quantification in the Social Sciences, *Social Science Information*, vol. 42, no. 4, 515 - 534
Mill, J.S. (1925), *A System of Logic*, Longman, Green, and Co., London
Mitchell, W. and Muysken, J. (2003), *Misrepresentation and Fudge – The OECD NAIRU Consensus*, Centre for Full Employment and Equity, Working Paper no. 03-11
Modiglianin, F. and Papademos, L. (1975), Monetary Policy for the Coming Quarters: The Conflicting Views, *The New England Review*, March- April, 2-35. Reprinted in: *The Collected Papers of Franco Modigliani*, (1980), vol. 3, MIT Press, Cambridge
Morgan, M.S. (1988), Finding a Satisfactory Empirical Model, in: De Marchi, N. and Blaug, M. (eds), *The Popperian Legacy in Economics.* Cambridge University Press, Cambridge
Morgan, M.S. (1990), *The History of Econometric Ideas*, Cambridge University Press
Morgan, M. S. (1998), Models, in: (Davis, J.B., Wade Hands, D. and Maki, U, eds.), *The Handbook of Economic Methodology*, Edward Elgar, Cheltenham
Morgan, M. S. and Morrison M. (1999), Models as Mediating Instruments, in: Morgan, M. S. and Morrison M. (eds.) (1999), *Models as Mediators*, Cambridge University Press, Cambridge
Morgan, M.S. (2001a), Models, Stories and the Economic World, *Journal of Economic Methodology*, vol. 8, no. 3, p. 361-384
Morgan, M.S. (2001b), Making Measurement Instruments, In: Klein, J.L. and Morgan, M.S. (eds.), The Age of Economic Measurement, Annual Supplement to Volume 33, *History of Political Economy*, Duke University Press, Durham and London
Morgan, M.S. (2003), *Business Cycles: Representation and Measurement,* in ed: Dominique

Ladiray, *Monographs of Official Statistics: Papers and Proceedings of the Colloquium on the History of Business Cycle Analysis* (Luxembourg: Office for Official Publications of the European Union)
Morgenstern, O. (1950), *On the Accuracy of Economic Observations*, Princeton University Press, Princeton
Morren, H.J. (1936), Wat er te leeren valt uit werkloosheidsstatistieken, *Tijdschrift van den Nederlandschen Werkloosheids-Raad*, Amsterdam
Morren, H.J. (1940), Het Rijksarbeidsbureau, *Economisch Statistische Berichten*, Rotterdam, 6 November 1940
Morren, H.J. (1941), *De werkloosheidsverzekering, haar grondslag en haar ontwikkeling in Nederland*, N. Samsom NV, Alphen aan den Rijn
Muysken J. and Meijers, H.(1988), *A Post-Mortem on UV Analysis*, Research Memorandum 88-031, University of Maastricht
Muysken, J. (1989), Classification of Unemployment: Analytical and Policy Relevance, *De Economist,* vol.137, no. 4, 397–424.
Muysken, J., Bierings, H. and De Regt, E. (1990), *Structurele werkloosheid in Nederland, 1960-1988*, Organisatie voor Strategisch Arbeidsmarktonderzoek (OSA), research paper W 82
Myrdal, G. (1963), *Challenge to Affluence,* Pantheon, New York
Narendranathan, N., Nickell, S., Stern, J. (1985), Unemployment Benefits Revisited, *The Economic Journal*, Vol. 95, no. 378, p. 307-329
Neubourg, C. de (1987), *Unemployment, Labour Slack and Labour Market Accounting; Theory, Measurement and Policy*, PhD thesis, Groningen
Nishizaki, F. (1997), *The NAIRU in Japan, Measurement and its Implications*, OECD Working paper, no.173, Paris
Nixon, J.W. (1928), The Measurement of "Risk" in connection with Labour Statistics, *International Labour Review*, vol. 17, no.5, 633-650
OECD (1997), *Concept, Measurement and Policy Implications of the NAIRU – Perspective from Belgium,* OECD working paper no. 185. Paris
Oxford English Dictionary, Oxford University Press, Oxford
Pagan, A. (1987), Three Econometric Methodologies: A Critical Appraisal, *Journal of Economic Surveys,* vol. 1, 3-24
Parikh, A. (1977), The Relationship between Unemployment and Vacancies: Comment, *Applied Economics,* vol. 9, 77-80
Pearl, J. (2000), *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge
Petrongolo, B. and Pissarides C.A., (2001), Looking into the Black Box: A Survey of the Matching Function, *Journal of Economic Literature*, vol. 39, 390-431
Phanzagl, J. (1968), *Theory of Measurement*, Physica-Verlag, Würzburg, Wien
Phelps, E. S.(1967), Phillips Curves, Expectations of Inflation, and Optimal Unemployment over Time, *Economica*, vol. 34, no. 135, 254-81
Phelps E.S. (1968), Money-Wage Dynamics and Labor-Market Equilibrium, *Journal of Political Economy,* vol. 76, 678-711
Phelps E.S., and Holt, C., (1970), *Microeconomic Foundations of Employment and Inflation Theory*, W.W. Norton, New York
Phelps, E.S. (1987), Philips Curves, In: Eatwell, J.. Milgate, M., and Newman, P. (eds.), *The New Palgrave; a Dictionary of Economics,* Macmillam Reference, London
Phelps, E. (1992), A Review of Unemployment, *Journal of Economic Literature*, 1476-1490
Phelps, E.S. (1995), The Origins and further Development of the Natural Rate of Unemployment, In: Cross, R (ed.), *The Natural Rate of Unemployment, Reflections on 25 Years of the Hypothesis*, Cambridge University Press, Cambridge

Phillips, A.W. (1958), The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957, *Economica*, vol. 25, no. 100, 283-299
Pigou, A.C. (1913), *Unemployment*, The home university library, Williams and Norgate, London
Pigou, A.C. (1933), *The Theory of Unemployment*, Macmillan, London
Pissarides, C.A. (1979), Job Matching with State Employment Agencies and Random Search, *The Economic Journal*, vol. 89, no. 356, 818-833
Pissarides, C.A. (1985), Short-Run Dynamics of Unemployment, Vacancies, and Real wages, *The American Economic Review*, vol. 75, no. 4, 676-690.
Pissarides, C.A. (1986), Unemployment and Vacancies in Britain, *Economic Policy: A European Forum*, vol. 1 no. 3, Cambridge University Press, Cambridge
Pissarides, C.A. (1994), Search Unemployment with On-the-job Search, *Review of Economic Studies*, vol. 61, no. 3, 457-475
Pissarides, C.A. (1999), *Equilibrium Unemployment Theory*, MIT Press, Cambridge
Porter, T. M. (1986), *The Rise of Statistical Thinking, 1820-1900,* Princeton University Press, Princeton
Porter, T. M. (1994), Making Things Quantitative, *Science in Context*, vol. 7, 389-407
Porter, T. M. (1995), *Trust in Numbers, The Pursuit of Objectivity in Science and Public Life*, Princeton University Press, Princeton
Porter, T. M. (2001), Economics and the History of Measurement, In: Klein, J.L. and Morgan, M.S. (eds.), The Age of Economic Measurement, Annual Supplement to Volume 33, *History of Political Economy*, Duke University Press, Durham
Pribram, K. (1926), The Scope of Labour Statistics, *International Labour Review*, vol. 14, no. 4, 476-488
Qin, D. and Gilbert, C.L. (2001), The Error Term in the History of Time Series Econometrics, Econometric Theory, vol.17, 424-450
Qin, D. (2006), *VAR Modelling Approach and Cowles Commission Heritage*, Queen Mary University of London working paper, no. 557
Quah, D.T. and Vahey, S.P. (1995), Measuring Core Inflation*, The Economic Journal*, vol. 105, no. 432, 1130-1144
Quah, D.T. (1992), The Relative Importance of Permanent and Transitory Components: Identification and Some Theoretical Bounds, *Econometrica*, vol. 60, no. 1, 107-118
Reese, T.W. (1943), Norman Campbell's Theory of Fundamental Measurement, *Psychological Monographs*, vol. 55, no. 3, 6-20
Reichenbach, H. (1956): *The Direction of Time*, University of Los Angeles Press, Berkeley
Reid, F. and Meltz, N.M. (1979), Causes of Shifts in the Unemployment-Vacancies Relationship: An Empirical Analysis for Canada, *The Review of Economics and Statistics*, vol. 61, no. 3, 470-475
Reiss, J., (2001), Natural Economic Quantities and their Measurement, *Journal of Economic Methodology*, vol. 8, no. 2, 287-311
Reiss, J., (2002), *Epistemic Virtues and Concept Formation in Economics,* Ph.D thesis, London School of Economics
Resnik, D.B. (1994), Hacking's Experimental Realism, *Canadian Journal of Philosophy,* no. 24, p. 1169-1185
Robinson A. (1945), Sir William Beveridge on Full Employment, *The Economic Journal*, vol. 55, no. 217, 70-76
Rodenburg, P. (2001), *Tracing the Changing Measures of Unemployment in Dutch Unemployment Statistics 1900-1940*, Tinbergen Institute Discussion paper, TI 2001-061/1, Tinbergen Instituut, Amsterdam
Rodenburg, P. (2002), *Diagnosing Unemployment: The 'Classification' Approach to Multiple*

Causation, Centre for Philosophy and Economics Discussion Paper Series, DP MEAS 24/02, London School of Economics
Rodenburg, P. (2005), Models as Measuring Instrument: Measurement of Duration Dependence of Unemployment, *Journal of Economic Methodology*, vol. 12, no. 3, 407-431
Rothbart, D., and Slayden, S.W., (1994), The Epistemology of a Spectrometer, *Philosophy of Science*, vol. 61, no. 1, 25-38
Rijksdienst der werkloosheidsverzekering en arbeidsbemiddeling (1925), *Communication sur la méthode de la statistique de chômage aux Pays-Bas,* internal paper, july 1925, CBS-Archive, Voorburg, Map 407.
Rijksdienst der werkloosheidsverzekering en arbeidsbemiddeling (1938), *Jaarverslag van den Rijksdienst der werkloosheidsverzekering en arbeidsbemiddeling over 1938 betreffende werkloosheidsverzekering en wachtgeldregelingen,* Algemeene Landsdrukkerij, Den Haag
Richardson, P., Boone, L., Giorno, C. et. al. (2000), *The Concept, Policy use and Measurement of Structural Unemployment: Estimating a Time Varying NAIRU across 21 OECD Countries*, OECD economic department working paper no. 250, OECD, Paris, France
Salop, S. C., (1979), A Model of the Natural Rate of Unemployment, *American Economic Review*, vol. 69, no. 1, 117-125
Samuelson, P.A. and Solow, R.M. (1960) Analytical Aspects of Anti-Inflation Policy, *American Economic Review*, vol. 50, no. 2, Papers and Proceedings of the Seventy-second Annual Meeting of the American Economic Association. (May, 1960), 177-194.
Schlick, M. (1985), *General Theory of Knowledge,* Open Court, La Salle, Illinois
Schlick, M., (1991), Positivism and Realism, in: Boyd, R. Gaspers, Ph. and Trout, J.D. (eds.), *The Philosophy of Science*, MIT Press, Cambridge
Schwier, A. (2000), Playing around with some Data, in: Leeson, R. (ed.) *A.W.H. Phillips: Collected Works in Contemporary Perspective*, Cambridge University Press, Cambridge
Sent, E.-M. (1998), The Evolving Rationality of Rational Expectations: An Assessment of Thomas Sargent's Achievements, Cambridge University Press, Cambridge
Sims, C.A. (1980), Macroeconomics and Reality, *Econometrica*, vol. 48, no. 1, 1-48
Sokal, R.R. (1974), Classification: Purposes, Principles, Progress, Prospects, *Science*, vol. 185, no. 4157, 1115-1123
Solow R.M. (1964), *The Nature and Source of Unemployment in the United States,* (Wicksell Lecture 1964) Almqvist & Wiksell, Stockholm
Spanos A., (1986), *Statistical Foundations for Econometric Modelling*, Cambridge University Press, Cambridge
Staiger, D., Stock, J.H. and Watson, M.W. (1997a), How precise are Estimates of the Natural Rate of Unemployment? In: Romer, C.D. and Romer, D.H., *Reducing Inflation: Motivation and Strategy*, National Bureau of Economic Research, University of Chicago Press, Chicago
Staiger, D., Stock, J.H. and Watson, M.W. (1997b), The NAIRU, Unemployment and Monetary Policy, *Journal of Economic Perspectives*, vol. 11, no. 1, 33-49
Staatscommissie over de werkloosheid (1914), Eindrapport van Staatscommissie over de werkloosheid, ’s-Gravenhage.
Stevens, S.S. (1946), On the Theory of the Scales of Measurement, *Science*, vol. 103, 677-680
Stevens, S.S. (1959), Measurement, Psychophysics and Utility, in: Churchman, C.W. and Ratoosh, P. (eds.), *Measurement: Definitions and Theories*, Wiley, New York, 18-63.
Stilting, A.J. (1907), *Werkloosheid,* J.H. de Bussy, Amsterdam
Stigum, B. P. (2003), *Econometrics and Philosophy of Economics, Theory-Data Confrontations in Economics*, Princeton University Press, Princeton
Storer, P. (1994), Unemployment Dynamics and Labour Market Tightness: An Empirical Evaluation of Matching Function Models, *Journal of Applied Econometrics*, vol. 9 no. 4, 389-4519

Stiglitz, J. (1997), Reflections on the Natural Rate Hypothesis, *Journal of Economic Perspectives,* vol. 11, no. 1, 3-10
Suppes, P and Zinnes, J.L. (1963), Basic Measurement Theory, in: Luce, R.D., Bush, R.R. and Galanter, E. (eds.), *Handbook of Mathematical Psychology*, vol. 1, 1-76, Wiles, New York
Suppes, P., Krantz, D.H., Luce, R.D., and Tversky, A. (1989), *Foundations of Measurement*, vol. II, Academic Press, New York.
Suppes. P. (1998), Theory of Measurement, in: Edward Craig (Ed), *Routledge Encyclopedia of Philosophy*, Routledge
Swijtink, Z. G. (1987), The Objectification of Observation: Measurement and Statistical Methods in the Ninetheenth Century, in: Krüger, L., Daston, L.J., and Heidelberger, M. (eds.), *The Probabalistic Revolution*, vol. I, MIT Press, Cambridge
Sydenham P.H. (1982), Measurements, Models, and Systems, in Sydenham P.H. (ed.): *Handbook of Measurement Science*, John Wiley & Sons Ltd., Hobroken
Taylor, J. (1972), The behaviour of Unemployment and Unfilled Vacancies: Great Britain, 1958 – 1971, An Alternative view, *The Economic Journal*, vol. 82, no. 328, 195-204
Thirlwall, A.P. (1969), Types of Unemployment, with Special Reference to "Non-Demand Deficient" Unemployment in the UK, *Scottish Journal of Political Economy*, vol. 16, no. 1, 20-49
Tinbergen, J. (1925), De roulering in het werklozenleger, *De Economist*, vol. 77, 772–782
Tobin, J. (1980), Stabilization Policy Ten Years after, *Brookings Papers on Economic Activity*, No. 1, 19-71
Tobin, J. (1997), Supply Constraints on Employment and Output: NAIRU versus Natural Rate, *Cowles Foundation Paper* 1150, Yale University, New Haven
Vanderkamp, J. (1966), An application of Lipsey's Concept of Structural Unemployment, *The Review of Economic Studies,* vol. 33, no. 3, 221-225
Velthuisen, A.B. (1935), Bevolking en werkloosheid, *Tijdschrift van den Nederlandschen Werkloosheids-Raad*, Amsterdam
Velthuisen, A.B. (1936), *De statistiek der werkloosheid*, internal paper CBS, 8 December 1936 CBS-archive, Voorburg, Map 408
Velthuisen, A.B. (1948), *Werkloosheidsverzekering in Nederland,* N.V.Uitgevers-Mij A.E. Kluwer, Deventer
Verlinden, J. (1997), *Concept, Measurement, and Policy Implications of the NAIRU, Perspectives from Belgium*, OECD working paper no. 85, Paris
Verbeek, M. (2001), *A Guide to Modern Econometric*, John Wiley & Sons, Hobroken
Verrijn Stuart, C.A. (1917), *Inleiding tot de beoefening der statistiek*, De erven F. Bohm, Haarlem
Verweij, R.A., (1940), Werkloosheid, *Economisch Statistische Berichten*, Rotterdam, 6 November 1940
Visser, J. (1989), *European Trade Unions in Figures; 1913-1985*, Kluwer, Deventer and Boston:
Vries, H. de, (1976), *Economenwerk en werkloosheid; Nederlandse economen over het werkloosheidsbeleid voor 1940*, Universitaire Pers leiden, Leiden
Van Fraassen, B.C. (1980), *The Scientific Image*, Clarendon Press, Oxford
Vos, R. (1995), The Logic and Epistemology of the Concept of Drug and Disease Profiles, *Poznan Studies in the Philosophy of the Sciences and Humanities*, 69-86
Vroey, de M. (2004), *Involuntary Unemployment: The Elusive Quest for a Theory*, Routledge, London
Wall, H.J. and Zoega G. (2001), *The British Beveridge Curve: A Tale of Ten Regions*, CEPR discussion paper, no. 1771
Wanningen, G.H. (1998), *De NAIRU en inflatie: een empirische analyse voor Nederland*, Dutch Central Bank, onderzoeksrapport WO&E nr. 548
Warren, R.S. (1977), The Behaviour of Unemployment and Unfilled Vacancies in Great Britain: a

Search-Turnover View, *Applied Economics*, vol. 9, 237-242
Warren, R.S. (1980), Unemployment and Unfilled Vacancies in U.S. Manufacturing: A Search-Turnover View, *The Review of Economics and Statistics*, vol. 62, no 2
Watanabe, H. (1996), Theory of Classification of Objects by Predicates, *Measurement*, no. 1, 59-69
Webb, A.E. (1974), *Unemployment, Vacancies and the Rate of Earnings: A Regional Analysis,* National Institute of Economic and Social research, Regional Papers III, Cambridge University Press, Cambridge
Webb, J. N. (1939), Concepts used in Unemployment Surveys, *Journal of the American Statistical Association*, vol. 34, no. 205, 49-61
Weeden, R. (1973), *Interregional Migration Models and their Application to Great Britain*, National Institute of Economic and Social research, Regional Papers II, Cambridge University Press, Cambridge
Weeden, R. (1974), *Regional Rates of Growth of Employment: An Analysis of Variance Treatment*, National Institute of Economic and Social research, Regional Papers III, Cambridge University Press, Cambridge
Woodward, J. (1989), Data and Phenomena, *Synthese*, vol. 79, 393-472
Woodward, J. (2000a), Explanation and Invariance in the Special Sciences, *British Journal for the Philosophy of Science*, vol. 51, no.2, 197-254
Woodward, J. (2000b), Data, Phenomena, and Reliability, *Philosophy of Science*, vol. 67, issue supplement, 163-179
Worswick, G.D.N. (ed.) (1975), *The Concept and Measurement of Involuntary Unemployment,* George Allen & Unwin Ltd., London
Wilde, H. de (1909), *De Gemeente en werkloosheid*, G.J.A. Ruys, Utrecht
Wilks, S.S. (1961), Some Aspects of Quantification in Science, *Isis*, vol. 52, no.2, 135-142
Wulwick, N.J. (1989), Phillips' Approximate Regression, Oxford Economics Papers, new series, vol. 41, no. 1, 170-188
Yellen, J. (1989), The Beveridge Curve, *Brookings Papers on Economic Activity*, vol. 1989, no. 1, 1-76
Yokogawa Electric Corporation, *Simple Creation of Indicator Diagram (P-V Diagram)*, (application note), http://www.yokogawa.com/tm/pdf/appli/tm-appli_auto12.pdf
Zanten, J.H. van (1927), *Leerboek der statistische methode*, N. Samsom, Alphen aan de Rijn
Zanten, H. van (1928), Hoe groot is de omvang der werkloosheid in Nederland?, *Economisch Statistische Berichten*, Rotterdam, 11 April 1928
Zhao, H., *Measuring the NAIRU – An Structural VAR Approach*, working paper

Summary in Dutch

Het gebruik en de constructie van meetinstrumenten heeft in de wetenschapsfilosofie tot nu toe weinig systematische aandacht gekregen. De reden hiervoor is dat meetinstrumenten door wetenschapsfilosofen doorgaans slechts gezien worden als neutrale ‘versterkers’ of ‘vergroters’ voor onze beperkte zintuiglijke waarnemingen. Zo bouwen we dus bijvoorbeeld thermometers omdat onze zintuigen niet goed genoeg zijn om gedetailleerde, incrementeel kwantitatieve waarnemingen te doen van het fenomeen hitte.

Deze gangbare opvatting vindt zijn weerslag in de huidige dominante theorie over meten, de *Representational Theory of Measurement* (RTM). Voortgekomen uit het *Logisch Positivisme* hanteert deze theorie een verzamelingstheoretische benadering waarin meten wordt gedefinieerd als een non-degeneratieve isomorfische (strikte één-op-één) afbeelding van een verzameling van empirische relaties op een verzameling van numerieke relaties. Echter, door het strikt formele en normatieve karakter biedt deze theorie weinig praktische leidraad voor wetenschappers in het veld over hoe nu een dergelijke isomorfische afbeelding tot stand moet worden gebracht. Bovendien biedt de RTM geen verklaring voor meetfouten (immers, een strikt isomorfische afbeelding impliceert dat herhaling van een meting tot exact dezelfde uitkomsten zal moeten leiden, anders betreft het geen isomorfisme), terwijl gebruikers van meetinstrumenten in de praktijk altijd te maken hebben met meetfouten, en juist groot belang hechten aan accuratesse (‘unbiasedness’) en precisie van meetinstrumenten. Mede als gevolg daarvan is de RTM geïsoleerd geraakt van de huidige wetenschapsfilosofische debatten over empirisch onderzoek en experimenteren, zoals die geïnitieerd zijn door onder andere Hacking (1983), Cartwright (1983), Latour (1987) en Bogen and Woodward (1988). Tevens gaat deze theorie ervan uit dat meetinstrumenten behalve het produceren van numerieke feiten geen verdere epistemische rol spelen in de wetenschap.

Om betekenisvol te kunnen meten in de wetenschap zullen we dus, zo leert de RTM ons, meetinstrumenten moeten ontwikkelen die op de een of andere manier een isomorfische afbeelding tot stand brengen tussen een empirisch relationeel systeem (variaties in het fenomeen) en een numeriek relationeel systeem (variaties in nummers). Meetinstrumenten zouden dus zulke isomorfische afbeeldingen moeten constitueren omdat ze anders niet als goede meetinstrumenten beschouwd kunnen worden. Door nu te kijken hoe in de praktijk meetinstrumenten worden geconstrueerd kunnen we leren langs welke wegen onderzoekers deze isomorfische afbeeldingen tot stand brengen, welke additionele benodigdheden hierbij noodzakelijk zijn in de praktijk en langs welke weg eventueel een praktische, positieve theorie van meten ontwikkeld zou kunnen worden die empirisch relevant is voor wetenschappers in de praktijk.

In dit proefschrift wordt daartoe een aantal cases bestudeerd waarin sociale wetenschappers – economen of statistici – trachten bepaalde verschijningsvormen van werkloosheid te meten. Deze verschijningsvormen die men tracht te meten (behandeld in hoofdstuk 2 t/m 5) zijn respectievelijk: werkloosheid in Nederland 1900-1940, cyclische werkloosheid, de NAIRU (Non-Accelerating Inflation Rate of Unemployment), en duurafhankelijkheid van werkloosheid (zie figuur 1).

Hoofd stuk	**Fenomeen**	**Meetinstrument**	**Betrouwbaarheids principe**	**Feiten over het fenomeen**
2	Werkloosheid in Nederland, 1900-1940	Bureaucratie	TU- SQR LEx-SQR	INU, PU, DUW Geregistreerde werkloosheid
3	Cyclische werkloosheid	UV-curve	Correlatie	U-V als % van totale werkloosheid
4	NAIRU	a) Structureel model van Phillips curve b) VAR model	Causale relatie Regressie	% van totale werkloosheid
5	Duur afhankelijkheid	duur model van werkloosheid	stabiel mechanism	α ($0 < \alpha < 1$)

Figuur 1: Overzicht van de meetinstrumenten en betrouwbaarheidprincipes voor het meten van werkloosheid

Door het bestaan van meetfouten die optreden in alle gevallen van meten fungeert het vereiste van een isomorfisme in feite als een idealisatie, zoals ook door voorstanders van de RTM wel wordt erkend. In de praktijk blijkt dat het vereiste van een isomorfisme losgelaten en vervangen wordt door een representatie van een zo invariant mogelijke relatie tussen variatie in het fenomeen en variatie in nummers. Als zodanig fungeren deze representaties van invariante relaties als principes van betrouwbaarheid: we hebben vertrouwen in de meetuitkomsten doordat we vertrouwen hebben in het principe waarop het instrument is gebaseerd. De constructie van meetinstrumenten is dus feitelijk een zoektocht naar (locale) invariantie. De cases behandeld in de respectievelijke hoofdstukken laten verschillen manieren zien hoe sociale wetenschappers deze invariantie pogen te vinden, en een variëteit aan representaties van invariante relaties zullen worden besproken: causale relaties (hoofdstuk 4), correlaties (hoofdstuk 3) en regressie vergelijkingen (hoofdstuk 4 en 5) , representaties van stabiele mechanismen (hoofdstuk 5), diagrammen en modellen (hoofdstuk 3, 4 en 5) en gestandaardiseerde kwantitatieve regels (hoofdstuk 2). Hoewel men in de wetenschap voor het vinden van invariante relaties doorgaans op zoek gaat naar wetmatige verbanden, laten de cases (alle hoofdstukken) zien dat economen een heel scala aan andere principes toepassen om hun metingen, die plaatsvinden onder omstandigheden waarin verstorende invloeden moeilijk kunnen worden afgeschermd, toch nog betrouwbaar te maken. Sterker nog, aangezien wetmatige verbanden schaars zijn in de economische wetenschap zijn ze daar doorgaans toe gedwongen.

Op de vraag langs welke wegen de theorie van het meten aangevuld of verbeterd zou kunnen worden, biedt dit proefschrift enkele suggesties. Ten eerste concludeert het proefschrift dat een fenomeen-data distinctie, zoals de filosofen Bogen en Woodward (1988) voor ogen hebben, goede aanknopingspunten biedt voor het ontwikkelen van een theorie over meten, onder andere omdat bijvoorbeeld meetfouten eenvoudig in een dergelijk raamwerk geaccommodeerd kunnen worden (hoofdstuk 6).

Ten tweede demonstreert dit proefschrift enkele belangrijke inzichten met betrekking tot conceptualisatie en classificatie van verschijningsvormen van werkloosheid. Zowel conceptualisatie als classificatie zijn in de RTM geen onderwerp van onderzoek. De RTM neemt de measurand als gegeven en kent zo een ongeproblematiseerde ontologische status toe aan deze concepten. Concepten zullen vanzelf als niet-bestaand worden beschouwd als metingen ervan zullen blijven falen. De casestudies in dit proefschrift laten zien dat de werkelijkheid veel gecompliceerder is. Werkloosheid werd in de 19e eeuw als onmeetbaar beschouwd door statistici, maar wordt nu, ruim 100 jaar later, als een onproblematisch en betrekkelijk eenvoudig meetbaar concept gezien. De mate van observeerbaarheid van de verschillende concepten en klassen van werkloosheid lijken dus een continuüm van onobserveerbaar tot volledig observeerbaar te vormen.

In aanvulling hierop beargumenteert dit proefschrift dat classificatie (en feitelijk ook divisie) – dat een belangrijke rol bleek te spelen in de totstandkoming van de theoretische concepten van werkloosheid – als een (rudimentaire) vorm van meten beschouwd dient te worden, een punt dat niet in de RTM wordt behandeld, omdat kwalitatieve uitspraken over de measurand vermeden dienen te worden. Echter, zoals ook Watanabe (1996) al verzamelingstheoretisch aantoont, laat dit proefschrift zien (onder andere in hoofdstuk 2, 3 en 4) dat definiëring van een (onduidelijk) concept, bijvoorbeeld werkloosheid zoals dat in de 19^{e} eeuw werd gezien, een proces van classificatie is. Zij zijn wederzijds constitutief en een voorwaarde voor meten. Classificatie en concept formatie kunnen niet los worden gezien. Dus hoewel classificatie strikt genomen geen kwantiteiten toekent aan een measurand laat dit proefschrift zien dat kwalitatieve uitspraken (classificatie) over de measurand onvermijdelijk zijn en een rudimentaire vorm van meten betreft. In dat licht bezien zou Alfred Binet's adagium over het meten van IQ, "we do not measure, we classify", vervangen dienen te worden door "we classify, therefore we are measuring".

Tot slot wordt uit bovenstaande de conclusie getrokken dat meetinstrumenten niet alleen slechts neutrale, passieve 'registeerders' of 'opnemers' zijn. Dit is zoals gezegd de gangbare opvatting onder wetenschapsfilosofen, en zelfs een sociaal-constructivistische filosoof als Kuhn (1961) benadrukt slechts deze passieve, paradigmatische aspecten van meten (namelijk dat de uitkomsten van metingen in overeenstemming moet worden gebracht met het dominante paradigma) en niet andere, meer actieve rollen, zoals bijvoorbeeld een epistemische rol, die meetinstrumenten zouden kunnen spelen in het krijgen van toegang tot theoretische of statistische concepten. Door de manier waarop wetenschappers meetinstrumenten construeren, laten wetenschappers ze soms tevens een actieve epistemische rol spelen. Niet omdat meetinstrumenten systematische fouten kunnen bevatten (bijvoorbeeld doordat de lens van de telescoop niet perfect bol is zien we een planeet op een andere plaats dan waar die zich in werkelijkheid bevind). Ook niet omdat wetenschappers proberen meetresultaten zo te interpreteren dat ze commensurabel zijn met een bepaald dominant paradigma, zoals Kuhn beweert. Maar omdat instrumenten fungeren als mediators tussen data en fenomenen, en in die rol actief mee kunnen helpen om toegang te krijgen tot (nog onduidelijke) fenomenen en ze zo te bestuderen. Dit wordt het duidelijkst geïllustreerd door de geschiedenis van het meten van werkloosheid (hoofdstuk 2), waar mensen uit de praktijk die bezorgd waren om de sociale consequenties van werkloosheid, zoals gemeentebesturen en charitatieve instellingen, zonder duidelijk theoretische conceptualisatie de meting van werkloosheid ter hand namen en zo vat kregen op het fenomeen van werkloosheid.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

332. A. MULATU, *Relative stringency of environmental regulation and international competitiveness.*
333. C.M. VAN VEELEN, *Survival of the fair: Modelling the evolution of altruism, fairness and morality.*
334. R. PHISALAPHONG, *The impact of economic integration programs on inward foreign direct investment.*
335. A.H. NÖTEBERG, *The medium matters: The impact of electronic communication media and evidence strength on belief revision during auditor-client inquiry.*
336. M. MASTROGIACOMO, *Retirement, expectations and realizations. Essays on the Netherlands and Italy.*
337. E. KENJOH, *Balancing work and family life in Japan and four European countries: Econometric analyses on mothers' employment and timing of maternity.*
338. A.H. BRUMMANS, *Adoption and diffusion of EDI in multilateral networks of organizations.*
339. K. STAAL, *Voting, public goods and violence.*
340. R.H.J. MOSCH, *The economic effects of trust. Theory and empirical evidence.*
341. F. ESCHENBACH, *The impact of banks and asset markets on economic growth and fiscal stability.*
342. D. LI, *On extreme value approximation to tails of distribution functions.*
343. S. VAN DER HOOG, *Micro-economic disequilibrium dynamics.*
344. B. BRYS, *Tax-arbitrage in the Netherlands evaluation of the capital income tax reform of January 1, 2001.*
345. V. PRUZHANSKY, *Topics in game theory.*
346. P.D.M.L. CARDOSO, *The future of old-age pensions: Its implosion and explosion.*
347. C.J.H. BOSSINK, *To go or not to go…? International relocation willingness of dual-career couples.*
348. R.D. VAN OEST, *Essays on quantitative marketing models and Monte Carlo integration methods.*
349. H.A. ROJAS-ROMAGOSA, *Essays on trade and equity.*
350. A.J. VAN STEL, *Entrepreneurship and economic growth: Some empirical studies.*
351. R. ANGLINGKUSUMO, *Preparatory studies for inflation targeting in post crisis Indonesia.*
352. A. GALEOTTI, *On social and economic networks.*
353. Y.C. CHEUNG, *Essays on European bond markets.*
354. A. ULE, *Exclusion and cooperation in networks.*
355. I.S. SCHINDELE, *Three essays on venture capital contracting.*
356. C.M. VAN DER HEIDE, *An economic analysis of nature policy.*

357. Y. HU, *Essays on labour economics: Empirical studies on wage differentials across categories of working hours, employment contracts, gender and cohorts.*
358. S. LONGHI, *Open regional labour markets and socio-economic developments: Studies on adjustment and spatial interaction.*
359. K.J. BENIERS, *The quality of political decision making: Information and motivation.*
360. R.J.A. LAEVEN, *Essays on risk measures and stochastic dependence: With applications to insurance and finance.*
361. N. VAN HOREN, *Economic effects of financial integration for developing countries.*
362. J.J.A. KAMPHORST, *Networks and learning*.
363. E. PORRAS MUSALEM, *Inventory theory in practice: Joint replenishments and spare parts control.*
364. M. ABREU, *Spatial determinants of economic growth and technology diffusion.*
365. S.M. BAJDECHI-RAITA, *The risk of investment in human capital*.
366. A.P.C. VAN DER PLOEG, *Stochastic volatility and the pricing of financial derivatives.*
367. R. VAN DER KRUK, *Hedonic valuation of Dutch Wetlands*.
368. P. WRASAI, *Agency problems in political decision making.*
369. B.K. BIERUT, *Essays on the making and implementation of monetary policy decisions*.
370. E. REUBEN, *Fairness in the lab: The effects of norm enforcement in economic decisions.*
371. G.J.M. LINDERS, *Intangible barriers to trade: The impact of institutions, culture, and distance on patterns of trade.*
372. A. HOPFENSITZ, *The role of affect in reciprocity and risk taking: Experimental studies of economic behavior.*
373. R.A. SPARROW, *Health, education and economic crisis: Protecting the poor in Indonesia.*
374. M.J. KOETSE, *Determinants of investment behaviour: Methods and applications of meta-analysis.*
375. G. MÜLLER, *On the role of personality traits and social skills in adult economic attainment*.
376. E.H.B. FEIJEN, *The influence of powerful firms on financial markets.*
377. J.W. GROSSER, *Voting in the laboratory.*
378. M.R.E. BRONS, M*eta-analytical studies in transport economics: Methodology and applications*.
379. L.F. HOOGERHEIDE, *Essays on neural network sampling methods and instrumental variables.*
380. M. DE GRAAF-ZIJL, *Economic and social consequences of temporary employment.*
381. O.A.C. VAN HEMERT, *Dynamic investor decisions.*
382. Z. ŠAŠOVOVÁ, *Liking and disliking: The dynamic effects of social networks during a large-scale information system implementation.*