Five Applied Theoretic and Time Series Econometric Essays with Applications to Accounting and Economics

A thesis submitted for the degree of Doctor of Philosophy

by
Stavros Dafnos

College of Business, Arts and Social Sciences
Brunel University

October 2016

CHAPTER 1

Introduction

Five Applied Theoretic and Time Series Econometric Essays with Applications to Accounting and Economics

Abstract

We employ some of the modern tools of economic theory and time series econometrics to consider a number of economic problems. The communication and coordination problems we study are relevant in accounting, business, economics and finance.

1 Introduction

Until the early 1980's the prevailing definition of the subject matter of economics was the one given by Lionel Robbins and popularised in the principles textbook by Paul Samuelson: economics is how society makes the best use of its scarce resources. But modern economics has developed into a large arsenal of tools which we can employ to understand socioeconomic problems. One definition that emphasizes tools of inquiry rather than context is provided by David Kreps (2004, p.8):

"Economics is concerned with modeling the behavior of individuals and organizations - firms, nonprofit organizations, and so on - in market and nonmarket settings. Its models almost always assume that behavior is *purposeful* - directed at some clear goal - and it usually studies how diverse behaviors that have conflicting objectives are brought into equilibrium by market and nonmarket institutions."

The present thesis begins with chapters two and three by examining the behaviour of people and organisations, who are supposed to share a common goal. The tools employed in these two chapters are standard optimisation techniques and graph theory, which is the methodology of modern network economics.

Then in chapter four it considers the equilibriating mechanisms of behaviour by groups of economic agents, who usually have conflicting interests. Here, we apply the tools of non-cooperative game theory, which constitutes a large part of modern economic theory.

Chapter five addresses the question of why people behave the way they do in their economic affairs. Peoples' economic behaviour is mirrrored in the aggregates of macroeconomics. We propose a Time Varying Autoregressive model to study the movements in the five main macroeconomic variables: Gross Domestic Product, Unemployment, Inflation, Interest rates and Exchage rates. The methods in this highly original chapter come from standard Time Series Analysis, but we do introduce some innovative time series techniques.

Chapter six is an empirical investigation of the movements in one of the five main macroeconomic variables, the rate of inflation. Among the econometric tools employed are standard Autoregressive models (AR), Autoregressive Distributed Lag models (ADL) and the more recent Generalised Autoregressive Conditional Heteroskedasticity (GARCH) methodology.

Chapter seven concludes and offers suggestions for future research. Below we briefly discuss the contents of each chapter and their contributions.

1.1 Chapter 2. The Development of Budgets with Bounded Rational Accountants

The first essay is related to the most basic building block of economic theory: the theory of decision making under certainty and more importantly under *uncertainty*. The current orthodox theory is based on two works. The first is the paper of Daniel Bernoulli originally published in Latin (1738) and published in English translation by Econometrica (1954). The second is the book by John Von Neumann and Oskar Morgenstern which established Game Theory as a separate discipline. In an appendix to the second and third editions (1953), they produced an axiomatic treatment of utility. The synthesis of all the work done by mathematical economists, mathematicians and statisticians was carried out by Leonard Savage in 1954. It constitutes the core of the prevailing approach as taught in leading PhD programmes of economics, finance and accounting; it is popularised in modern Microeconomic textbooks.

Contrary to the claims made by Modern Behavioural Economics, which criticises economic theory for representing people as ultra rational decision makers, our work as exposed in chapter two, shows how orthodox economic theory has injected realism into its model of decision making under uncertainty by incorporating peoples' bounded rationality. Since the 1950's leading economists, like the 1978 Nobel prize winner Herbert Simon, and Richard Cyert and James March (1963) have gone into great strengths to demonstrate the bounded rationality of individuals and especially organisations.

The contribution of Chapter two is to develop a model of budgeting *without* the fully rationality assumption. What accountants do is collect information, process information, compute aggregate - useful - numbers and communicate them. The paper demonstrates that budgeting is a costly activity. Taking into consideration costly information processing is one way of modelling the bounded rationality of accountants.

1.2 Chapter 3. Understanding a Group Resource: Information Processing in Teams

A challenge for economists, strategy scholars and business practitioners is to understand the causes and durability of an organisation's competitive advantage. The contribution of Chapter three is to illustrate the versatility of the modified orthodox theory of decision making under uncertainty by illuminating an internal group resource: information processing in teams. The techniques, first proposed by Roy Radner (1993), are employed to determine whether one group of people is more efficient than another in processing a certain amount of information.

1.3 Chapter 4. Auditing the Auditors

The first two papers are based on the assumption that the members of an organisation, which follows a particular purposeful behaviour, share a common goal. Accordingly, in the last two chapters we focused on communication problems. Next, we move to discuss the coordination of behavioural responses amongst people and organisations that have diverse objectives. The chapter employs the tools of noncooperative game theory to address the question: Can outsiders, like shareholders and bankers rely on the accounting information reported by an organisation?

Whereas the first two essays in the previous two chapters are related to the foundational building block of economic theory, the present paper goes to the heart of modern economic theory, the non-cooperative Nash equilibrium of an interactive situation among economic agents. A major weakness of Nash equilibrium is its inability to rule out incredible threats and promises as was demonstrated in two path breaking papers written in the 1960's by Reichard Selten, for which he shared the 1993 Nobel prize. The questioning by Selten of the appropriateness of Nash equilibrium in dynamic situations, represented by extensive form games, has directed the attention of the economics profession to the importance for careful study of dynamic interactions. Selten proposed a refinement of Nash equilibrium, called the Subgame Perfect Equilibrium, which has been applied in numerous economic settings, including the credibility of monetary policy, which is relevant to chapter six.

Our work differs from existing papers on auditing in a number of ways; one difference is that we do *not* treat the auditors as exogenous parties that automatically practice the prescribed accounting and auditing standards. Instead, we examine how the actions prescribed by the standards can become incentive compatible and thus self-enforcing for the auditors.

1.4 Chapter 5. The Fundamental properties of the Time Varying-AR(2) model

In chapter 2, where we study decision making under uncertainty, we assume that peoples' preferences are given. In the current chapter we develop a theory, which among a number of applications it can be employed to model how individuals' preferences and values evolve in response to their experiences.

The fourth paper studies the mechanisms that lead economic agents to behave in a particular way. It adresses two interelated questions. How do players form their *beliefs* on what is likely to happen in a game? What are their beliefs on how the other players will behave? David Levine (2012, p.6) says this is the critical part of examining what happens in the game and the formation of beliefs is at the heart of modern economic theory.

Since beliefs are formed under uncertainty we represent them with probability distributions. Once we assign probability to an outcome, then beliefs about the consequences of an outcome lead to the prediction that players will choose the strategy - course of action - that will maximize their expected utility given their beliefs.

The natural follow up question is: Where beliefs come from? For example, how can I decide whether to buy or sell shares of particular companies? How do I form my beliefs on whether the market will go up or down? In the present paper, which is a joint work with Professor Menelaos Karanasos (PhD Financial Time Series Econometrics, Birkbeck College, University of London) and Dr Alexandros Paraskevopoulos (PhD Pure Mathematics, Imperial College, University of London) we argue that one way to form beliefs is from what we learn from the past, from history.

Economic decisions in both Savage's small worlds and in Savage's large worlds (the difference is explained in chapter two) are realistically assumed in economic theory to be made by experienced decision makers. For example, many of the decisions involved in budgeting (the subject matter of chapter two) have a periodic or seasonal component. In the present chapter, we develop a dynamic theory in which the economic agents learn from the past and adapt to the changing circumstances. We call it a Time Varying Autoregressive Model, $TV-AR$. Players and game theorists may employ it to predict the Nash outcome, where no further learning is possible.

The particular autoregressive model we specify and then examine is of order two, denoted as $TV-AR(2)$. Such a low order specification, at first sight might seem unrealistic, but it has the power to model a variety of real world phenomena. It was an $AR(2)$ model that Yule (1927) developed in a paper that played a catalytic role in the advancement of time series analysis as a separate field.

1.5 Chapter 6. Econometric Inquiry of Inflation with AR, ADL and GARCH in Mean Models

Macroeconomists and policy makers try to understand past movements in the inflation rate and forecast its likely future values. To investigate the intertemporal properties of the inflation rate we apply theories of linear time series, including dynamic dependence, the autocorrelation function and stationarity. To represent the characteristics of our inflation data we specify, estimate and test a variety of econometric models, including AR, Unit-root nonstationarity, ADL and GARCH type of models.

References

[1] Bernulli, Daniel (1954). "Exposition of a New theory on the Measurement of Risk" Originally published 1738 in Latin by the Imperial Academy of Sciences in Petersburg, *Econometrica*, pp.23-36.

[2] Cyert, Richard and James March (1963). *A Behavioral Theory of the Firm*, Prentice Hall.

[3] Kreps, David (2004). *Microeconomics for Managers*, Norton.

[4] Levine, David (2012). *Is Behavioral Economics Doomed? The Ordinary versus the Extraordinary*, OpenBook Publishers.

[5] Radner, Roy (1993). "The Organization of Decentralized Information Processing", *Econometrica*, vol.61, no.5, pp.1109-1146.

[6] Savage, Leonard (1972). *The Foundations of Statistics*, Dover.

[7] Von Neumann, John and Morgenstern, Oskar (1953). *Theory of Games and Economic Behavior*, 3rd ed., Princeton University Press.

[8] Yule, Udny (1927). "On a Method of Investigating periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers", *Philosophical Transactions of the Royal Society of London. Series A*, Vol.226, pp.267-298.

CHAPTER 2

The Development of Budgets with Bounded Rational Accountants: Modeling Costly Accounting Information

Abstract

Budgeting is as a problematic area for most corporations. Budgeting is a resource-consuming practice. The analytical research on budgeting does not incorporate the costs of processing and communicating information. Explicitly or implicitly, the analytical models assume that management accountants and decision makers are fully rational. Taking into account the costly nature of information is one way of capturing the bounded rationality of people. Up to around 1990, economists remained silent about such costs. In the 1990's, new models have been developed that incorporate these costs. The present paper employs these tools to incorporate costly information processing into budgeting. I consider the processing of accounting information within a firm in order to find out how bounded rational accountants develop budgets and accounting information systems. This examination will help management accountants understand better the processing of accounting information and its costs, so that they can develop and implement more efficient budgets. In addition, the bounded rationality of accountants and other members of an organization provides a justification for the decentralized nature of most modern corporations.

1 Budgeting as an Example of Decision Making and Control Under Uncertainty

Acknowledgement 1 [1]

[1]

Acknowledgement 2 *I would like to thank the participants of the following conferences for their helpful comments: 2nd AIMA 2005, Monterey, CA; EAA Annual Congress 2005, Gotenborg; and GMARS 2006, Copenhagen. I wish to thank Marc Epstein and John Young for offering an opportunity to an outside practitioner. I am grateful to Mike Bromwich, my discussant at GMARS 2006, for his seven critical comments, which were instrumental in improving the paper. I remain responsible for the contents. Last but not least, I thank Mike*

Budgeting has attracted criticism from academics and practitioners. The main problems identified in the literature arise from games that accountants and managers sometimes play with the budget targets.[2] Such games result in agency costs, which are due to conflicts of interest and incomplete information among the members of an organization. The mitigation of agency costs has been the subject of intense research effort in the past thirty years by the theory of incentives, also known as contract theory.

The accounting theorists' almost exclusive preoccupation with the principal - agent interaction, and the design of incentive schemes has left an aspect of budgeting less well understood: How different units of the firm collect and send budgeting information, which ends up as few aggregate numbers in the hands of the budget manager? This is the main question addressed in the present paper. Why management accountants should care about how organizations aggregate large amounts of dispersed information? A better understanding of how information flows within the firm will help us to take prompt corrective actions. In other words, it will help the control role of budgets. Besides, as Demski (1994) argues, the accounting systems of an organization is an economic resource producing benefits for a cost. We must find a way to estimate the cost of providing a certain amount of accounting information.

The paper faces head on the two issues discussed in the previous paragraph. First, budgeting information is an economic resource and as such we must estimate its opportunity cost. Second, it should be provided on time, without delay, ideally on real time. Late information cannot play a control role; it can only be used for record keeping. If budgets help to coordinate activities and provide incentives within firms, then the present study aims for a better understanding of coordination. An important facet of coordination is the problem of information processing. The characterization of honest information is left to the voluminous literature of contract theory.

Budgeting, pro-forma financial statements, and business plans are examples of decision making and control under uncertainty. Budgeting concerns the following period; the future is unnkown and therefore uncertain. Arrow (1974, pp.33-34) defines uncertainty as

> "Uncertainty means that we do not have a complete description of the world which we fully believe to be true. Instead, we consider the world to be in one or another of a range of states".

Arrow's definition of uncertainty coincides with the management accountant's definition of budgets. A budget is a quantitative prediction of the operating and financial state of an entity for the next financial period; it is the management accountant's estimate that the firm will be in one particular state, but the actual state might turn out to be different. How certain we are that the

Shields for his lovely words of encouragement.

[2]Some of the deficiencies have been pointed out by Jensen (2001), and by Hansen et al. (2003).

budget numbers we produce are sound? We need a theory of decision making and control under uncertainty to guide us in modeling budgeting.

Theoretical research on budgeting has employed the maximization of expected utility hypothesis as its foundation. The expected utility model is a theory of rational decision making under uncertainty. The required information for maximizing expected utility is assumed to be freely available, and the required calculations are assumed costless to make. More to the point of this paper, the cost of optimizing is ignored or assumed to be zero; the resources required to make optimizing decisions are treated as if they are not scarce (See Demsetz, 1988). The canonical - economics oriented- management accounting model is based on this Walrasian, often called the neoclassical, view of the firm.

One of the drawbacks of the analytical research on budgeting is that it does not incorporate the time and costs of processing and communicating accounting information in the budgeting process.[3] Atkinson et al. (2004, p.405), in their authoritative textbook on management accounting say

> "Some organizations invest thousands of hours over many months to prepare the master budget documents".

The theoretical models ignore these costs and thus, explicitly or implicitly, assume that financial accountants, management accountants and other decision makers are fully rational. This is the drawback of the expected utility hypothesis that the present paper addresses. Taking into consideration the costly nature of information, is one way of capturing the *bounded rationality* of accountants.

The activities of information processing use resources: people, machines, and materials, and therefore are costly. Radner (1992, section 2) provides convincing evidence that in the majority of large firms, more than one third or even half, of all employees are engaged in information processing activities, or in jobs that support such activities. Accordingly, managing the firm affects its profitability, which means that managing may have an impact on returns to scale, which are different to the traditional technological returns to scale (See Radner and Van Zandt, 1992). Despite the well documented presence of the costs of acquiring, processing, and communicating information within a firm, up to around 1990, theory had little to say. In an article published in 1988, Demsetz (1988, p.159), an authority on the theory of the firm, writes

> "A more complete theory of the firm must give greater weight to information cost than is given in Coase's theory or in theories based on shirking and opportunism"

Beginning in the early 1990's, several pioneering papers [Radner (1992, and 1993), Radner and Van Zand (1993), Bolton and Dewatripont (1994), and Van

[3] A recent excellent survey of theoretical approaches to budgeting is Covaleski et al. 2003 and its update 2006). They point out, footnote 14, p.12, that

> "analytical economic models assume that individuals' information processing is costless".

Zand (1999)] developed techniques for incorporating such costs into theoretical models. I employ the tools developed in these pioneering articles and subsequent research, to incorporate *costly information processing* into budgeting.[4] In particular, I show two things. First, I consider how a programmed network of individual processors, accountants and computers, executes the decentralized information processing activities involved in budgeting; and second, I demonstrate that this information processing is costly and therefore these costs should be taken into consideration in the evaluation of budgets. In short, budgeting is a costly activity performed not by a single representative accountant, but by a network of accountants and computers.

The framework employed for the mathematical modeling of information processing is *parallel computation.* Traditional computer programs are based on algorithms which perform one step at a time; such algorithms are called *serial*. The majoriy of Accounting Information Systems are based on serial algorithms. But accounting involves many computationally intense problems, which cannot be solved within deadlines using serial operations. *Parallel processing*, which uses computers made up of many separate processors, helps to overcome the limitations of serial computers. *Parallel algorithms* decompose a problem into subproblems, that can be solved concurrently, and rapidly solve problems using a computer with multiple processors.[5]

The reported work is an application of the theories collectively known as optimization with decision costs taken into account. Gigerenzer and Selten (2001,b) criticize these theories for not being theories of truly bounded rationality. Although this is a justified criticism, models which take into account the costs of decision making are more realistic than the traditional fully rationality models. This type of modeling will enhance our understanding of how budgeting information flows within an organization and the costs of budgeting. The problems associated with budgets and the ever rising costs of budgeting makes it an urgent matter for us to be able to estimate the costs of budgets and to make budgets more effective. The more we understand about budgeting, we will be in a better position to avoid budget games and reduce budget costs.

The research reported in this paper should be viewed as a response to two different calls. First, theorists working on developing theories of bounded rationality believe that the practical application of their theoretical models should be viewed as tests of their relevance (See Lipman, 1995). Second, accounting scholars (e.g. Zimmerman, 2001) propose that the development of managerial accounting models should be based on firm economic foundations. I follow the advice given in a major synthetic paper by Mark Covaleski et al (2003, p.4):

> "research within a theoretical perspective often advances by modifying its assumptions and addressing issues that were previously simplified away".

[4]Radner (1993), which is the seminal contribution of this literature, in section 6 of the paper on Applications, has a subsection 6.1, entitled "*Linear Operations in Accounting and Control*", where he suggests that his ideas and techniques can be applied to problems in accounting.

[5]Rosen (2003, pp.552-553) provides a bird's eye view of parallel processing.

The paper shows that we can built formal models of budgeting in particular, and management accounting models in general, without the fully rationality assumption. It demonstrates the applicability of the theorists' techniques to a practical problem.

2 Theories of Decision Making and Control Under Uncertainty

In this section following Radner's (1996, and 1997) systematic classification and examination of *rationality* and *bounded rationality*, I distinguish between two types of bounded rationality: the first is *costly rationality*, and the second is *truly bounded rationality*. As I said in the introduction, the contribution of the paper is to show that accountants are bounded rational in the sense that the information they produce is costly. Before I consider the topic of bounded rationality and multiperson decision making, in order to have a point of reference I provide a critical commentary of the theory of individual - rational - decision making under uncertainty. It underlies the bulk of the theoretical research in management accounting (See Baiman's, 1990, professional survey).

2.1 Rationality in Small Worlds

The expected utility theory has taken its most developed form in the hands of the statistician Savage (1954), and is known as the *Savage Paradigm*.[6] The theory has three main aspects:

First, it proposes seven postulates that consistent-rational decision making must satisfy.

Second, these principles of rationality imply that the decision maker's choices among alternatives could be calculated as a function, called the Von-Neumman Morgenstern Utility function,[7] of two scales: i) a numerical scale of *probabilities of events*, and ii) a numerical scale of the *utilities of outcomes*. The decision maker prefers an action that gives the highest mathematical expectation of the utility of the outcome. The expectation is calculated with respect to the decision maker's scale of probabilities. A rational decision maker is characterized solely by his *beliefs* and *tastes*. The probability and utility scales are *subjective (idiosyncratic)* to the decision maker in that different decision makers can differ in their scales. Savage did not claim that the decision maker knows these scales, or makes the calculations, but only that these scales could be inferred by a modeler. Savage shows that a person who satisfies the postulates will make decisions as *though* is maximizing a Von Neumman - Morgenstern utility function.

Third, the theory implies how a decision maker should modify his decisions as new information becomes available; or how actions are chosen in a sequential

[6] My references are from the revised republication, reprinted with corrections in 1972, by Dover.

[7] Because it appeared for the first time in an appendix of the second edition of Von Neumman and Morgenstern's (1947) *Theory of Games and Economic Behavior.*

decision problem as new information is coming in. Updating is important in revising budgets. The theory assumes that the decision maker updates his beliefs by applying Bayes' rule in the light of new information. A decision maker who does not violate Savage's postulates when making decisions, and updates his beliefs using Bayes' rule, is said to be Bayesian rational. The Savage Paradigm is taken to be synonymous with *Bayesian rationality*.

There have been noumerous critics of this theory, the most famous of whom is the Nobel laureate Allais with his "paradox".[8] This is not the place to discuss the paradoxes that confront the Savage paradigm.[9] I just want to point out that Savage responded to the critics by arguing that the Von-Neumman-Morgenstern Utility theory is a good representation of how real people make decisions only in *"small worlds"*.[10]

By the term "small world" Savage meant that the decision maker is faced with a small number of alternatives, and that he knows the alternatives open to him and the consequences associated with each one of the alternatives. In such small worlds, Savage's axioms are likely to be satisfied. In Binmore's (1992, p.119)words, only then it is

> "practical for someone to evaluate in advance the implications of anything that might conceivably happen".

This assumption is called "logical omniscience" (see for example Arrow, 2004); it never holds in budgeting for real firms. In budgets, management accountants attempt to predict the operational and financial figures of a firm for a future period, but they have no way of knowing in advance what will happen and their implications. They just approximate next period's operating and financial figures using the available information in the current period. Results derived by analytical management accounting models apply only in small worlds. Next, I consider how the Savage paradigm has been transformed to live in more realistic habitats than the "small worlds".

2.2 Costly Rationality

The Savage paradigm, upon which the large majority of principal - agent models, contract theories, and incomplete contract models are based, does not take into account the resources used in the process of decision making. This is unsatisfactory as most decisions, and budgeting in particular, involve the employment of resources, especially human resources. Following Radner (1996, and 1997), we classify the *costly* (resource-used) activities of decision making into four groups:

[8]The ancestor of the Von-Neumman Morgenstern Utility function for representing reasonable behavior under uncertainty, namely the maximization of expected value, run against another famous problem, the St. Petersborough Paradox. The famous 18th century mathematician Daniel Bernulli responded by changing the theory of reasonable behavior to one that maximizes expected utility, see Gingensberger and Selten (2001).

[9]Interested readers should consult the excellent survey by Machina (1987).

[10]Savage (1954), covers in section 5 of chapter 5 on "Utility", the conditions under which his rationality postulates are likely to hold.

- Observation, or the *gathering of information.*
- Memory, or the *storage of information*; in measuring transactions, accountants gather large sets of data. These data sets require a method of storage. Stakeholders should be able to retrieve transaction data quickly and in usable form. The accounting storage system of recording transactions is made up of *accounts*, which culminate into the *general ledger.*
- Computation, or the *manipulation of information.*

When we consider groups of decision makers, as we usually do in accounting systems of organizations, we have to take into consideration the costs of

- Communication, or the *transmission of information.*

Anyone who has been involved with the practice of financial or management accounting will readily recognize that all the above four activities are the bread and butter of every day's life of accountants. Out of the costs of these activities, the first, second, and fourth can be accomodated by the Savage paradigm. But once we consider the costs of computation, we are lead to question the realism of the Savage paradigm as a theory of human decision making.

In 1950, the statistician Wald incorporated the *cost of gathering information* into his model of statistical decision making.[11] Wald showed how costly sampling can be eliminated by a sequential procedure. Under certain conditions, Wald's optimal sequential testing procedure results in savings. Radner (1996, p.1364) argues that for the Wald procedure to be fitted into the Savage system, it is necessary for the decision maker to remember the results of previous observations. If the observations are complicated, the memory required to do this might be costly, or infeasible in large problems. The *costs of keeping records of information* are analogous to the cost of observation, but not identical; they are modeled in a similar fashion.

Managerial decisions may be classified into two categories: decisions that are relatively *routine* or *periodic*, and decisions that are relatively *unique* or *"one time only"*. Budgeting involves mainly periodic decisions; I spend more time considering periodic decisions. For example, firms periodically update their production plans on the basis of observed sales, market conditions, and prospects of the (international) macroeconomy. In large firms, such *decision cycles* involve the collection and processing of large amounts of information, and the calculation of hundreds of individual decisions. Such computational activity is beyond the capability of any single human, even if he was to use the most powerful computer. Accordingly, the computational task involved in corporate decision making is divided among many people and machines. The activities of information processing for decision making are *decentralized.*[12]

[11] My telegraphic account of Wald's contribution comes from the systematic accounts by Radner (1996, and 1997); for more details see Radner's papers.

[12] We model decentralized information processing in section 4.

There is an additional "cost" of information processing, the *cost of delay*. This is crucial for "timely" accounting information; it is not an obvious pecuniary cost, like salaries, machine maintenance, etc., but a loss of accounting profit, due to the degradation of the decision. The cost of delay is important for budgetary control as "timely" accounting information enables corrective actions. Such delays can be reduced by employing more information processors - accountants and machines. It is analogous to replacing one "serial" computer with a "parallel machine" made up of many serial ones.

In addition, information processing require that primary data and intermediate results be *stored in* some *memory*. It adds to the cost of information processing, and to delays, since it takes time to read data into memory and to access the stored information. In general, machine memory is relatively cheap compared to computation. Finally, *communication* among individual processors requires additional resources and causes additional delays. Radner (1996) argues that *information transmission* is relatively cheap, compared to computation.

The costs and delays of information processing have an important implication for an organization: it is *not* efficient (except for very small firms) for every decision to use all of the information avaialble to the organization as a whole. The question of how a manager selects the relevant information set arises. In a large firm, only a small portion of the available information will be used on any single decision; ideally, the choice of information will be influenced by its cost and its relevance to the decision under study. In practice this choice is made rather arbitrary; often is based on rules of thumb. They lead to the *inevitability of decentralized decision making*, in which different decisions or sets of decisions, are made by different decision makers on the basis of different information.

2.3 Bounded Rationality

> "Bounded rationality needs to be, but it is not yet, understood." (Gigerenzer and Selten, 2001, p.1).

The considerations of the previous subsections imply that practical (real) problems of decision making under uncertainty are more complicated than the simple examples of textbooks. The assumption of rationality emerged in economics and related fields, like accounting, in the 1950's and 1960's, usually referring to the optimization (maximization or minimization) of some function. Up to then, economists assumed that people were motivated by self-interest. Soon after the idea of rationality as optimization took its most developed form by Savage (1954), the competing idea of bounded rationality emerged. It is noteworthy to point out that Herbert Simon introduced the idea of bounded rationality in the mid-1950's in the context of budgeting. In his words:

> "The idea, by the way, emerged not from speculation but from some very concrete observations I made on budgeting processes in the city government of Milwaukee." (Simon, 1998, p.189).

Simon emphasized two aspects of bounded rationality: cognitive limitations of people, and structure of the environment. In subsequent sections I attempt to model these two facets.

Decision makers are not merely uncertain about "empirical" events, they are also uncertain about logical inferences. In this kind of environment the Savage model is declimatized and finds it hard to survive. To conclude, costly rationality can be handled by extending the Savage paradigm. Truly bounded rationality *cannot* be handled by the Savage system. The paper incorporates the costs of information processing into budgeting. The modeling of truly bounded rationality is beyond the purpose of the present paper, and is left for future work.[13]

3 Two Critical Remarks on Budgeting Models with Rational Accountants

The progression of formal modeling leading to our current understanding of budgeting is well represented by Demski and Feltham (1978), Baiman and Evans (1983), Penno (1984), and Kanodia (1993). I refrain from discussing them since they are presented very well in Covaleski et al (2003-2006).[14]

With no intend to underestimate the contributions of this branch of accounting research, especially the insights of the seminal paper by Demski and Feltham (1978), I would like to make two related critical comments. First, the papers raise issues about the roles of incentives and private information, in relation to budgets; they say very little about how the budget numbers are reached. Second, these accounting theories, originated from economics, are based on an unrealistic model of human decision making: people are fully rational Bayesian maximizers of subjective utility.

In the remainder of the paper, I attempt to rectify the two shortcomings in the following way. The canonical accounting model has as its main player, a mythical accountant who knows the solutions to all problems and can immediately do all the required computations. But in real companies, accountants are different: their cognitive capabilities are limited. I model such limitations using Herbert Simon's idea of bounded rationality. The budgeting models described in the following sections demonstrate how bounded rational accountants reach a judgement or decision rather than only the outcome of the decision.

[13]The edited volume by Gigerenzer and Selten (2001) contains useful insights towards this direction.

[14]Previous drafts include a review of this literature.

4 Decentralization of Information Processing

Financial and management accountants spend most of their time processing information.[15] As argued persuasively by Radner (1992, pp.1392-1393), in modern large firms, the information processing activities are decentralized. Modern corporations are decentralized in the sense that they are decomposed into numerous responsibility centres, which make different decisions based on different information. The large scale of modern enterprises, makes the information processing task impossible for a single person. The limited capacity of people for information processing implies that it uses scarce resources, including people.

In the Walrasian paradigm, the resources required to make decisions, optimizing or not, are treated as if they are not scarce. (See Demsetz, 1988). Accordingly, up to the late 1980's, economic and accounting theory, dominated by the Walrasian paradigm, remained silent about the costs of processing and communicating information. In the 1990's, a series of pioneering papers began to incorporate such costs into formal models of the workings of the firm, and by doing so they have modeled one important aspect of the bounded rationality of people.

The research program, initiated by Radner (1993), is addressing the question of how to model the decentralization of information processing. In the Walrasian paradigm, accounting and management tasks are performed without error and costlessly, as if by a *free and perfect computer* (Demsetz, 1988). A more realistic alternative approach is to use the following metaphor: an accountant, who is working in information processing, can be represented by a *computer of limited capacity*. This is the route taken by Radner and other researchers working within this research paradigm.

4.1 A Framework for modeling Decentralized Information Processing

As I said above, the starting point of this approach is that an accountant is represented by a computer of limited capacity; limited in the sense that the individual computers (accountants) can process a maximum number of items per unit of time. The accountants / computers are linked to an accounting information system, together with machines / computers. The modeling of information processing using parallel computation assumes away any divergence of interest among processors. The individual members (accountants or machines) are called *processors*, and the system *a network*. Accounting systems are represented by a network.

I consider general networks, but following Radner (1993) I introduce this type of modeling with the special case of *a hierarchical network* with which we are all familiar. This (hierarchical) network will be asked to solve the following problem: Given N items to be processed, and P processors, arrange and

[15] The current and the remaining sections of the paper are in the spirit of the theories of costly rationality, that were briefly reviewed in subsection 2.2.

program the processors to process the N items in minimum time. Processing in our examples will be the operation of addition. Addition, subtraction, and multiplication are the most common operations in accounting.

A *processor* has 3 characteristics:

- an *"in-box"*, which can hold any number of unread data;
- a *"register"*, which can hold a single information item that has been read; i.e. the register is where the accountant processor keeps the running total, and
- a *"Central Processing Unit (CPU)"* or a *"clock"*, that can in 1 unit of time, read one item of information from the in-box, add it to the one in the register, and put their sum back in the register. We also assume that it takes 1 unit of time for a processor to read an item and put it in an empty register. We measure time in *"cycles"*; in 1 cycle, a processor can take 1 item from its in-box and add it to its register.

There can be one or more *one-way communication links* among the processors. A processor can also send information contained in its register to the in-boxes of other processors to which it is directly linked, and then re-set its own register to zero; this can be done in any cycle without additional time; i.e. this communication takes no additional time. There is, also, a particular processor that, at a designated time, sends out the contents of its register as the result of the computation; i.e. the grand total. The set of processors and links is called a *network*.

Finally, we need an *algorithm* (i.e. *a program*) that determines which processors send items to other processors and when, and which processor calculates and announces the final answer. That is, a program describes:

- the original assignment of information items to the in-boxes of specific processors;
- the times at which each individual processor sends the contents of its register to the processor(s) to which it is directly linked; and
- the time at which one designated processor sends out the final answer.

The combination of a network and a program is called a *programmed network*. The programmed network contains 3 elements:

- a specification of *the number of individual processors* (in this paper, we assume that they have identical capacities);
- an *initial assignment of items* to the in-boxes of processors; and
- an *algorithm* that determines how the network operates.

4.2 Costly and Efficient Networks

In each period, a processor can take an item from its in-box, add it to the current register, and send the value of the register to another processor. When it sends the value of its register to another processor, this information goes into the receiving processor's in-box, and the sending processor's register is set to zero. In each period, information from outside the network flows into the in-boxes of certain prespecified processors. The choice variables in a network are:

- the number of processors;
- which processors receive how much incoming information;
- when procesors send their register to other processors; and
- which processors they send the information to.

Two properties of the network are considered costly:

- the number of processors; processors are costly because they have to be paid, bought, or rented.
- the number of units of time it takes to deliver the answer. The *number of cycles* needed to perform the computation is called the *delay.*

A network is *efficient*, if it is not possible to decrease the delay C without increasing the number of processors, and vice versa. The following are examples of activities involved in information processing, which are costly, and should be "economized":

- Observation of information;
- The capabilities and number of individual processors (humans and machines);
- The "communication network" (i.e. the accounting systems) that transmits and switches information (both primary and partly processed) among the processors; and
- The "delay" between the observation of information and the implementation of decision(s). Delay is costly to the extent that delayed decisions are not timely. The "cost of delay" is not a pecuniary cost, like payroll or machine maintenance, but a loss of profit due to an outdated decision.

Computer scientists have, in the main, examined "delay"; that is, they have tried to estimate the time it takes to compute a particular function. Radner (1992 and 1993) pays equal attention to economizing the number of processors. Radner does not consider the "cost of communication"; he justifies this choice on the grounds that in human organizations, the presence of "information overload" is a testimony of the fact that with modern technology, information transmission is relatively cheap compared to computation.[16]

[16] In his 1993, Econometrica paper, Radner works with the assumption that the amount

5 Modeling Information Processing in Budgeting

Management accountants spend most of their time processing information. We can think of the information procesing part of the firm as a decision making machine, which takes signals and transforms them into actions, implemented by the employees. Every employee working on the front line makes many decisions that are not dictated by management. Radner (1992, and 1993) shows that hierarchical structures, which are usually thought as serving the centralization of authority are effective in decentralizing the activities of information processing. Decentralization of information processing is due to the large scale of modern enterprises, which renders the task impossible for a single person. The limited capacity of people for information processing implies that this activity uses *scarce resources*, including people.

Following Bolton and Dewatripont (1994), I place the organization in an environment, in which a steady flow of information arrives over time. It is a realistic assumption for the majority of firms. This flow of information is too large to be processed by one group of employees; this is called *information overload*. The organization's problem is to design a communication network to process this information most effectively. I adopt an important aspect of Bolton and Dewatripont's (1994) model, namely, the idea of *returns to specialization* in processing: by repeatedly processing the same type of information item an accountant can lower his unit time of processing that type of item. This is how most financial accounting and management accounting departments in large firms operate. Because of this reason, a group of several accountants want to work and process information as a *team* within the firm. Each accountant works on a different type of information and the different pieces of information are put together by the information network. At the end, one person (e.g. the budget manager, the CFO, or the CEO) or a group of people (e.g. the Board of Directors) receives all the processed information and makes a decision. In the sequel, I show with the help of simple examples the superiority of parallel computation, compared to the tradititonal serial processing of information.

5.1 A Hierachical processing of budgeting information with parallel processors

In this subsection I show how a hierarchical network of accountants can be used for parallel computation. In particular, I show how seven (7) accountants can be organized to process 8 numbers in 3 steps.[17] Suppose a firm has operations

of environmental information is given , but he is quick to point out that it should be an endogeneous variable, determined by the balance between its cost and its value. Radner considers it in his 1992 paper, section 5. Geanakoplos and Milgrom (1991) study hierarchical networks in the spirit of team theory, in which the acquisition of information by managers is time-consuming; Geanakoplos and Milgrom (1991) should be seen as comlementary to Radner (1993); they emphasize the costly acquisition of information by organizations.

[17] The numbers of the example come from Rosen (2003), only the little story is mine.

in two countries, and in each country it has two business units (for the shake of the argument suppose that all the business units are profit centers). There are 4 budget coordinators that correspond to the 4 profit centres. The four budget coordinators are labeled with the symbols P_4, P_5, P_6, P_7. The 8 numbers are indicated by $x_1, x_2, ..., x_8$. The odd numbers represent estimated revenues and the even numbers represent estimated costs. The 3 stages of parallel budgeting are: In the first stage, P_4 calculates the difference between x_1 and x_2 ,$(x_1 - x_2)$; P_5 calclulates $(x_3 - x_4)$; P_6 calculates $(x_5 - x_6)$; and P_7 calculates $(x_7 - x_8)$. The four budget coordinators, who are at the lower level of the budgeting hierachy, are sending their "partial totals" to the two accountants P_2 and P_3; P_2 is budget manager for the operations in one country, and P_3 is budget manager for the second country. P_2 adds $(x_1 - x_2) + (x_3 - x_4)$, and P_3 calculates $(x_5 - x_6) + (x_7 - x_8)$. In the third and final step, the accountant at the higher level in the hierarchy of the budgeting process (the budget manager of the whole firm) calculates $(x_1 - x_2) + (x_3 - x_4) + (x_5 - x_6) + (x_7 - x_8)$, which is the "grand total".

The three stages used to process the 8 financial numbers constitutes an improvement compared to the 7 steps required to add 8 numbers serially. The above parallel processing of accounting information is a crude - schematic - representation of how real companies prepare their budgets. In this case, theory follows the practice. Where the theory of parallel processing becomes useful is when we move away from the simple examples like the one above and consider accounting information processing carried by tens or hundreds of accountants. Then, the issues of accurate, timely, and costly infomation can only be considered with an organized theoretical framework. In the next subsection I show how the idea of *efficient networks* can help us to design better and less costly accounting information systems.

5.2 Efficient Accounting Information Systems using Parallel Computation

Consider the following hierachical processing of information that comes from Radner (1992 and 1993). Suppose information comes into the network only in one period (i.e. we exclude Bayesian updating and budget revisions), and consists of a vector of 40 numbers. The objective of the hierarchical network is to compute the grand sum of these forty numbers. The cost of the information system is increasing in both the number of processors and the number of periods required to compute the sum. One network that produces this partition is the following.[18] At the lowest level of the hierarchy there are 8 processors. Each of the 8 processors receives 5 of the incoming numbers. In the first 5 periods, the processors add the 5 numbers. (Remember that it takes 1 unit of time to process 1 item). At the end of period 5, the 8 processors send their totals to 4 higher level processors, each of which receives information from 2 processors.

[18]It might help the readers to draw their own trees to keep track of information processing by the hierarchies.

In periods 6 and 7, the $four$ processors add the incoming numbers. At the end of period 7, they send the totals to 2 other processors, each of which receives information from 2 processors. In periods 8 and 9, the *two* processors sum their incoming numbers and forward the totals to a single processor at the end of period 9. In periods 10 and 11, the last processor computes the grand sum. This network requires 15 processors and 11 periods to compute the sum.

There is a redundancy in this network, since higher up processors are idle while they are waiting for the lower processors to add. This is a characteristic of information systems in which knowledge is assembled *vertically*. But as Zimmerman (2003, p.266) argues

> "budgeting is also an important device for assembling specialized knowledge *horizontally* within the firm."

Now I demonstrate that once we allow accountants to share their knowledge among peers in other parts of the firm, the 40 numbers can be summed by a network of 8 processors in 8 periods. This more efficient network operates as follows. Each of the "lower level" processors receives 5 numbers. As in the previous network, the processors spend periods 1 through 5 adding their numbers. $Four$ of these 8 processors send their totals to the other 4, each processor receiving one number. This is added to the processor's previous total in period 6. At the end of period 6, *two* of the 4 processors send their totals to the other 2. These numbers are added to previous totals in period 7, after which one processor sends its partial total to the other. The grand total is computed in period 8. Radner (1993) demonstrates that this network is *efficient*: a network cannot compute this sum with both fewer processors and less delay. As there is a trade off between processors and periods, this is not necessarily the minimum cost information network. The relative cost of processors versus delay make one of the information systems, the optimal.

6 Related Literature and Suggestions for Further Research

In this section I describe some issues regarding the modeling of budgeting, which deserve closer attention. A cornerstone of every budget is the estimation of demand. An interesting extension of this paper would be to incorporate forecasting demand, using the techniques of Radner and Van Zand (1993) and Van Zand (1999 and 2003). Another possible research direction is to explore differences between batch posting and real-time posting. Van Zand has looked at the differences on a theoretical level in a series of papers; see for example, Van Zand (2003).

There are also several issues that need to be addressed in the general area of transferring the graph theoretic computations into matrix forms.[19] This work

[19] Textbooks of Graph Theory or on Discrete and Combinatorial Mathematics (e.g. Rosen, 2003) show how this transformation is done.

will produce useful results that will help in the practical processing of large amounts of accounting information. Also, a natural extension is to apply this modeling to complex organizational structures.

The spirit of the "aspiration-adaptation" literature (see the volume by Gigerenzer and Selten, 2001) is quite different from that of the current work. Here I try to model how the costly processing of accounting information makes the assumption of unbounded rationality unrealistic. Instead, the paper tried to show that the costly rationality assumption seems to be more appropriate for real-life company budgeting practises. The aspiration-adaptation theories of bounded rationality may be shown to be useful in modeling budgeting, especially, recursive budgets. This will be a critical step towards constructing applied models of truly bounded rationality.

Aware that I repeat myself, all the work reported in the present paper is done under the team theoretic assumption of no conflict of interests among the members of the organization. This is an unrealistic assumption. The next step is to introduce incentive considerations into the costly information framework. A first attempt has been made by Dewatripont (2006), who injects incentive considerations into a model of costly information (specifically, costly communication of information) using the framework developed by Dessein and Santos (2006). This merging explains the coexistence of various forms of communication and is a promising area for future research.

7 Concluding Remarks

In line with the bulk of the management accounting and accounting information systems literature, I view an accountant as an information processor. In the paper, I described the interaction of accountants who are aware of how they are processing information. Where I depart from existing literature is when I drop the assumption that accountants are fully rational. Instead, I believe that accountants are bounded rational. By relaxing the assumption (made usually in the economics literature) of unbounded rationality, and by introducing elements from the phychology and sociology perspectives to budgeting, the paper has produced a better understanding of budgeting information.

Part of bounded rationality of accountants is the result of their limited capacity of information processing. My concern is with optimal processing networks made up of bounded rational accountants and computers. The literature on processing networks examines how infomation is processed. Parallel processing is the technique of performing sequences of tasks simultaneously. Its effect is to decrease the time needed to process accounting information using computers. The message of our naive application of parallel computing to managerial accounting is that it can be used in the design of accounting information systems to increase the accuracy and timeliness of information at the lowest possible costs.

The present paper constitutes an attempt to apply some techniques that model aspects of bounded rational behavior to the practical area of budgeting. The state of the different research programs that try to model bounded rationality is summarized by the following quotation:

> "A comprehensive theory of bounded rationality is not available. This is a task for the future. At the moment we must be content with models of limited scope." (Selten, 2001, p.14).

Hansen et al (2003) classify the views of practitioners concerning budgeting into two categories: one that supports improving the budgeting process, the other abandoning it. The present work belongs to the improving camp. With our present state of knowledge, I cannot envisage an alternative and better way to allocate and control resources within an organization.

References

[1] Arrow, K. (1974). *The Limits of Organization*, Norton.

[2] Arrow, K. (2004). "Is Bounded Rationality Unboundedly Rational? Some Ruminations" in *Models of Man*, edited by Mie Augier and James G. March, The MIT Press.

[3] Atkinson, A., Kaplan, R., and Young, M.S. (2004). *Management Accounting*, 4rth edition, Pearson - Prentice Hall.

[4] Baiman, S. (1990). "Agency Research in Managerial Accounting: A Second Look", *Accounting, Organizations and Society*, vol.15, no.4, pp.341-371.

[5] Baiman, S. and Evans, J. III, (1983). "Pre-Decision Information and Participative Management Control Systems", *Journal of Accounting Research*, vol.21, no.2, pp.371-395.

[6] Binmore, K. (1992). *Fun and Games: A Text on Game Theory*, D.C.Heath.

[7] Bolton, P. and Dewatripont, M. (1994). "The Firm as a Communication Network", *Quarterly Journal of Economics*, vol.CIX, Non., issue 4, pp.809-839.

[8] Covaleski, M, Evans III, J., Luft, J., and Shields, M. (2003). "Budgeting Research: Three Theoretical Perspectives and Criteria for Selective Integration", *Journal of Management Accounting Research*, vol.15, pp.3-49.

[9] Demsetz, H. (1988). "The Theory of the Firm Revisited", *The Journal of Law, Economics, and Organization*, 4, Spring; reprinted as chapter 10, pp. 159-178 in *The Nature of the Firm*, ed. by O. Williamson and S.Winter, Oxford University Press, 1991.

[10] Demski, J. (1994). *Managerial Uses of Accounting Information*, Kluwer.

[11] Demski, J., and Feltham, G. (1978). "Economic Incentives in Budgetary Control Systems", *The Accounting Review*, vol.LIII, no.2, pp.336-359.

[12] Dessein, W., and Santos, T. (2006). "Adaptive Organizations", *Journal of Political Economy*, vol.114, no.5, pp.956-995.

[13] Dewatripont, M. (2006). "Costly Communication and Incentives", *Journal of the European Economic Association*, vol.4, no.2-3, April-May, pp.253-268.

[14] Geanakoplos, J. and Milgrom, P. (1991). "A Theory of Hierarchies Based on Limited Managerial Attention", *Journal of Japanese and International Economies, vol.5, pp.205-225.*

[15] Gigerenzer, G., and Selten, R. (2001). "Rethinking Rationality", chapter 1, in Gigerenzer, G.and Selten, R. (Edit.) *Bounded Rationality; The Adaptive Toolbox*, (pp.1-12), The MIT Press.

[16] Gigerenzer, G. and Selten, R. (2001). *Bounded Rationality; The Adaptive Toolbox*, The MIT Press.

[17] Hansen, S., Otley, D., and Van der Stede, W. (2003) "Practice Developments in Budgeting: An Overview and Research Perspective", *Journal of Management Accounting Research*, vol.15, pp.95-116.

[18] Jensen, M. (2001). "Corporate Budgeting Is Broken - Let's Fix It", *Harvard Business Review*, Nov., pp.94-101.

[19] Kanodia, C. (1993). "Participative Budgets as Coordination and Motivational Devices", *Journal of Accounting Research*, vol.31, no.2, pp.172-189.

[20] Kaplan, R. and Atkinson, A. (1989). *Advanced Management Accounting*, 2nd edition, Prentice Hall.

[21] Lipman, B. (1995). "Information processing and bounded rationality: a survey", *Canadian Journal of Economics*, XXVIII, No1, pp.42-67.

[22] Machina, M. (1987). "Choice Under Uncertainty: Problems Solved and Unsolved", *Journal of Economic Perspectives*, vol.1, no.1, pp.121-154.

[23] Penno, M. (1984). "Asymmetry of Pre-Decision Information and Managerial Accounting", *Journal of Accounting Research*, vol.22, no.1, pp.177-191.

[24] Radner, R. (1992). "Hierarchy: The Economics of Managing", *Journal of Economic Literature*, vol.XXX, September, pp.1382-1415.

[25] Radner, R. (1993). "The Organization of Decentralized Information Processing", *Econometrica*, vol.61, no.5, September, pp.1109-1146.

[26] Radner, R. (1996). "Bounded Rationality, Indeterminacy, and the Theory of the Firm", *The Economic Journal*, vol.106, Sept., pp.1360-1373.

[27] Radner, R. (1997). "Bounded Rationality, Indeterminacy, and the Managerial Theory of the Firm" chapter 15, pp.324-352, in *Organizational Decision Making*, edited by Shapira, Z., Cambridge University Press.

[28] Radner, R., and Van Zandt, T. (1992). "Information Processing in Firms and Returns to Scale", *Annales D'Economie et Statistique*, no25/26, pp.265-298.

[29] Rosen, K. (2003). *Discrete Mathematics and Its Applications*, Fifth edition, McGraw-Hill.

[30] Rubinstein, A. (1998). *Modeling Bounded Rationality*, MIT press.

[31] Savage, L. (1954). *The Foundations of Statistics*, Revised Republication in 1972 by Dover.

[32] Selten, R. (2001) "What Is Bounded Rationality?", chapter 2 in Gigerenzer, G.and Selten, R. (edited) *Bounded Rationality; The Adaptive Toolbox*, (pp.13-36), The MIT Press.

[33] Simon, H. (1959). "Theories of Decision-Making in Economics and Behavioral Science", *American Economic Review*, vol.49, pp.253-283.

[34] Simon, H. (1998). "Letter (excerpts) to Ariel Rubinstein", in A. Rubinstein, *Modeling Bounded Rationality*, (pp.187-190), MIT Press.

[35] Van Zandt, T. (1999a) "Real-Time Decentralized Information Processing as a Model of Organizations with Bounded Rational Agents", *Review of Economic Studies*, vol.66, pp.633-658.

[36] Van Zandt, T. (1999b). "Decentralized Information Processing in the Theory of Organizations", Chapter 7 in *Contemporary Economic Issues*, Vol.4: Economic Behavior and Design, Edited by Murat Sertel, pp.125-160, Macmillan Pres.

[37] Van Zandt, T. (2003) "Real-Time Hierarchical Resource Allocation", Unpublished manuscript, Insead.

[38] Zimmerman, J. (2001). "Conjectures regarding empirical managerial accounting research", *Journal of Accounting and Economics*, vol.32, pp.411-427.

[39] Zimmerman, J. (2003). *Accounting for Decision Making and Control*, 4rth edition, McGraw Hill.

CHAPTER 3

UNDERSTANDING A GROUP RESOURCE: INFORMATION PROCESSING IN TEAMS

Abstract

Firms succeed when they enjoy competitive advantage. A critical challenge is to explain the causes and durability of competitive advantage. The research paradigm, known as the resource-based view (RBV), suggests that the foundation for the success of a business firm lies in its resources. Despite its enthusiastic reception by academics, the RBV has had limited applicability in the real world. The thesis of the present work is that this lack of wide applicability is due to our limited understanding of what constitutes valuable resources and their role in gaining and sustaining competitive advantage. This is hardly a new idea; it is a rejoinder to strategy scholars, who believe that many resources are mystical. The note's contribution is to illuminate a particular type of group resource. Applying ideas and techniques developed by economists in the 1990's, I examine how a group of employees process information within a firm. A better understanding of the notion of group resources will enable us to examine how they contribute to the causes and sustainability of competitive advantage. By doing so, it will make the RBV a more useful tool in the practice of strategic management.

THE ROLE OF RESOURCES IN GAINING AND SUSTAINING COMPETITIVE ADVANTAGE

A critical challenge for strategy scholars, managers and consultants is to explain the causes of competitive advantage. Beginning in the 1980's and culminating in the early 1990's, a series of writings published that contain the elements and implications of what is known as the *resource-based view (RBV)* of the firm. This view suggests that the foundation for the success of a business firm lies in its *resources*. Turning these resources into *distinctive capabilities* enables the firm to achieve competitive advantage and earn *rents*.

The existence of rents generates greater competition from incumbent rivals and from new entrants to the sector. An important aspect of strategic management is how to keep a competitive advantage once it is obtained.

> A competitive advantage is *sustainable* when the advantage persists despite efforts by competitors or potential entrants to duplicate or neutralize it.[1]

[1]This definition of *durable competitive advantage* is due to Professor Jay Barney and was brought to my attention by a quotation provided in Besanko et al. (1996, p.543,),

This will be the case when *ex-post limits to competition* exist: after a firm has gained a competitive advantage and earns rents, there must be forces, which limit the competition for rents.

The Resource Based View has been received with enthusiasm by academics but so far has had a limited applicability in the business world. It is this armchair observation that motivated my inquiry into the causes of this lack of wide applicability and its possible remedies. The thesis of the present work is to attribute this lack to our limited understanding of what makes resources valuable. This is hardly a new idea; it is a rejoinder to strategy scholars who believe that we do not fully understand how resources contribute to the attainment of competitive advantage. As Professor Birger Wernerfelt (1997) says

> "many resources remain mystical".

My response to this justified criticism is to examine a specific example of *group resources*. Wernerfelt (1997) speculates that this class of resources contains most of the critical ones. Specifically, this note takes seriously Harold Demsetz's long standing concern with the undergrading of the role of information within the firm as well as a barrier to entry. (See for example,

---

and Besanko et al. (2004, p.426).

Demsetz, 1988). In particular, I explore how a group of people process information within a firm. This type of resource comes under the label *efficient procedures*, in Wernerfelt's (1984) list. Suprisingly, such a group resource has eluded theoretical treatment until the 1990's.

Beginning in the early 1990's, following the seminal work of Roy Radner (1993), several pioneering papers (Radner, 1992; Radner and Van Zand, 1993; Bolton and Dewatripont, 1994; and Van Zand, 1999;), developed techniques for examining how information flows within a firm. This research program is known as *Costly Information Processing in Teams.* What is important in the context of the present work, is that these techniques enable us to determine whether one group of people is more efficient than another in processing a certain amount of information. The contribution of the note is to take the first steps toward demystification of such an important group resource. Information and ideas are critical inputs in the production process of most firms, especially in the new economy firms. As Bowles et al (2005, Ch.20) argue, *Brains, Information, and Reputation* have replaced Land, Labor, and Capital as factors of production in today's weightless economy.

The next section provides an example of a group capability that has until recently escaped theoretical modeling. Using the tools developed by researchers working within the Costly Information Processing in Teams paradigm, I show how we can determine the efficiency of one group over another. The last section draws few tentative conclusions. It also suggests some avenues for future research, which might produce results that can open the doors of the boardrooms to the resource based theory.

EXPLORING A GROUP RESOURCE: INFORMATION PROCESING IN TEAMS

In planning, controlling and day-to-day management of a firm, managers are faced with a host of *coordination* issues that can be classified into various categories: Information Processing; Resource Allocation; Monitoring; Problem Solving; Allocation of Decision Rights.

In this note, I concentrate on information processing, which includes processing sales data, and keeping the books. The specific research question I ask is: How firms aggregate large amounts of information that is widely dispersed. Why this question is important? Neo-classical theory, on which many strategy papers are based, has no role for the time and costs required to process information. This unrealistic assumption is relaxed in the Costly

Information Processing in Teams literature I employ.

Explicitly, or implicitly, neo-classical based theories assume that employees are fully rational. Taking into consideration the costly nature of information is one way of capturing the bounded rationality of people. In many large firms, more than one third or even half of all employees are engaged in information processing activities, or in jobs that support such activities (See Radner, 1992, section 2). The activities of information processing use resources: people, machines and materials, and therefore are costly.

Up to around 1990, theory had little to say about such costs. In the 1990's, Radner's (1993) seminal contribution has stimulated a number of important papers, which look at how information flows within a firm. In the sequel, I employ the tools developed in these pioneering papers to explore one group resource: how efficiently a team of employees process a certain amount of information.

The Model

The framework employed for the mathematical modeling of information processing is *parallel computation.* Traditional computer programs are based on algorithms, which perform one step at a time; such algorithms are called serial. The majority of Management Information Systems (MIS) are based on serial algorithms. But the management of large enterprices involves many computationally intense problems, which cannot be solved within deadlines, using serial operations. *Parallel processing*, which uses computers made up of many separate processors, helps to overcome the limitations of serial computers. Parallel algorithms decompose a problem into subproblems, that can be solved concurrently, are designed to rapidly solve problems using a computer with multiple processors.[2]

In the neo-classical paradigm, management tasks are performed without error and costlessly, as if by a free and perfect computer (See Demsetz, 1988). A more realistic alternative approach is to use the following metaphor: an employee is represented by a computer of limited capacity. Computers (employees) can process a maximum number of items per unit of time. This is the route taken by Radner and other researchers working within this research paradigm.

The employees / computers are linked to a MIS with machines / computers. Individual members (employees or machines) are called *processors* and the system is called *network*. Following Radner (1993), I introduce this type

[2]Kenneth Rosen (2003, pp.552-553) provides a bird's eye view of parallel processing.

of modeling with the special case of a *hierarchical network*, with which we are all familiar. This hierarchical network will be asked to solve the following problem: Given N items to be processed, and P processors, arrange and program the processors to process the N items in minimum time.

Time is measured in *cycles*; in one cycle, a processor can process one item of information and add it to the running total. There can be one or more communication links among the processors. Finally, we need an algorithm (i.e. a program) that determines which processors send items to other processors and when, and which processor calculates and announces the final answer.

Parallel versus Serial Computation

Two properties of the network are costly: The number of processors and the number of units of time it takes to deliver the answer. The number of cycles needed to perform the computation is called the *delay*. A network is efficient, if it is not possible to decrease the delay without increasing the number of processors and vice-versa.

Example 1 *In this example I show how a hierarchical network can be used for parallel computation. In particular, I show how* 7 *accountants can be orga-*

nized to process 8 *numbers in* 3 *steps.*[3] *Suppose a firm operates in two countries, and in each country it has two business units (Assume that all the business units are profir centres). There are* 4 *budget coordinators that correspond to the* 4 *profit centres, who are labeled with the symbols* P_1, P_2, P_3, P_4*. Let the* 8 *numbers be* $x_1, x_2, ...x_8$*. Odd numbers represent estimated revenue, and even numbers represent estimated costs. The* 3 *stages of parallel computation are: In the 1st stage,* P_1 *calculates the difference between* x_1 *and* x_2*,* $(x_1 - x_2)$*;* P_2 *calculates* $(x_3 - x_4)$*;* P_3 *calculates* $(x_5 - x_6)$*; and* P_4 *calculates* $(x_7 - x_8)$*. The* 4 *budget coordinators are sending their partial totals to the* 2 *budget managers of the two countries. In the 2nd stage,* P_5 *adds* $(x_1 - x_2) + (x_3 - x_4)$*, and* P_6 *adds* $(x_5 - x_6) + (x_7 - x_8)$*. In the 3rd and final step, the budget manager of the entire firm calculates* $(x_1 - x_2) + (x_3 - x_4) + (x_5 - x_6) + (x_7 - x_8)$*, which is the grand total. The* 3 *stages used to process the* 8 *numbers is an improvement, compared to the* 7 *steps required to add* 7 *numbers serially.*

Where the theory of parallel processing becomes useful is when we move away from the simple examples, like the one above, and consider information processing carried by tens or hundreds of employees. Then, the issue

[3]The numbers of the example come from Rosen (2003), only the business scenario is mine.

of accurate, timely, and costly information can only be considered with an organized theoretical framework. In the next subsection, I show how the idea of *efficient networks* can help us to determine when a group resource is efficient.

When a Group Resource is Efficient

Consider the following hierarchical processing of information that comes from Radner (1992 and 1993). Suppose information comes into the network only in one period, and consists of a vector of 40 numbers. The objective of the network is to compute the grand total of these 40 numbers.

One network that can compute this sum is the following. At the lowest level of the hierarchy there are 8 processors. Each of the 8 processors receives 5 of the incoming numbers (Remember from subsection 5.1 that it takes 1 unit of time to process 1 item). At the end of period 5, the 8 processors send their partial totals to 4 higher level processors, each of which receives information from 2 processors.

In periods 6 and 7, the 4 processors add the incoming numbers. At the end of period 7, they send their totals to 2 other processors, each of which receives information from 2 processors.

In periods 8 and 9, the two processors sum their incoming numbers and forward the totals to a single processor at the end of period 9.

In periods 10 and 11, the highest standing procesor computes the grand total. This network requires 15 processors and 11 periods to compute the sum.

There is a redundancy in this network, since the higher up processors stay idle while they are waiting for the lower processors to add. This is how information flows within groups in which knowledge is communicated vertically. But once we allow team members to share their knowledge among peers in other parts of the firm, the 40 numbers can be summed by a network of 8 processors in 8 periods. This more efficient network operates as follows. Each of the 8 lower level processors receives 5 numbers. As in the previous network, they spend periods 1 to 5 adding numbers.

Four (4) of these 8 processors send their partial totals to the other 4, each processor receiving 1 number. This is added to the processor's previous total in period 6. At the end of period 6, two (2) of the 4 processors send their totals to the other 2.

These numbers are added to the previous total in period 7, after which 1 processor sends its partial sum to the other.

The grand total is computed in period 8. Radner (1993) demonstrates that this network is *efficient* in the sense that we cannot obtain this result with both fewer processors and less delay.

CONCLUDING REMARKS

This note suggests that the main contribution of the Resource Based View is the emphasis it places on factors internal to the firm when searching for the source of heterogeneous firm performance. But as Saloner et al. (2001), in their authoritative strategy book, warn,

> "using the RBV to justify a focus on the firm's assets in and of themeselves is a mistake. The resources are a source of competitive advantage only if they create positional advantage or advantageous capabilities".

This mistake is not committed by most of the writers of the RBV; according to the RBV, the strategy process begins from an examination of the firm's resources, but the business environment is not forgotten. As in Porter's five-forces framework, or better in its extension *the value-net* by Brandenburger and Nalebuff (1996), value is subject to bargaining amongst employees, suppliers, distributors, complementors, customers, and owners (See Foss, 1997, p.11).

The Resource Based View of the firm emphasizes the firm's internal resources as a source of competitive advantage. As long as the firm's capabilities are based on its resources (e.g. its organization), the RBV encompasses capabilities. To the extent that a firm's advantage is based on tangible and intangible assets, the RBV can be viewed as a call for a deeper understanding of the internal sources of competitive advantage. Beginning in the 1990's, economists and other social scientists have responded to this call by developing formal models about the *inner workings of organizations.* [4]

An important class of models are the ones that attempt to formally model the *bounded rationality* of economic agents within organizations. A subset of this class has made considerable progress in modeling the processing of information by a group of people within a firm, by emphasizing that this

[4] Robert Gibbons is a leading figure in this movement and his articles and forthcoming graduate book on Organizational Economics are excellent representatives of this trend. (e.g. Gibbons, 2000 and 2003). Forthcoming is also a Hanbook of Organizational Economics by Gibbons and Roberts that will contain articles for research workers in the field.

is a costly business.[5] Economists have taken important steps in the formal incorporation of bounded rational behaviour, and it will not be very long before the insights of the models help us to demystify Wernerfelt's "group" resources.

Another important strand of the new organizational economics, and particularly relevant to the design and implementation of strategy are the formal models of *leadership*.[6]

The writings of the Resource Based View have created a more balanced view of competitive advantage arising from both internal and external factors. More research effort towards a better understanding of the role that ideas, information, and reputation play is likely to produce the kind of additions that the RBV needs to become a full fledged tool for practitioners. The contribution of the present note was to show how the RBV once energized by the addition of how a group capability works, can become a useful tool in the practice of strategic management.

[5] Roy Radner and Timothy Van Zandt have been among the protagonists in this strand of the literature (see Radner 1992 and Van Zandt 1998).

[6] Benjamin Hermalin has been particularly active in this literature (see Hermalin, 1998).

References

[1] Barney J. 1986. Strategic factor markets: expectations, luck, and business strategy. *Management Science* 32(10): 1231-1241.

[2] Barney J. 1991. Firm resources and sustained competitive advantage. *Journal of Management* 17(1): 99-120.

[3] Besanko D, Dranove D, Shanley M. 1996. *Economics of Strategy*. Wiley: New York.

[4] Besanko D, Dranove D, Shanley M, Schaefer S. 2004. *Economics of Strategy* (3rd edn). Wiley.

[5] Bowles S, Edwards R, Roosevelt F. 2005. *Understanding Capitalism*, (3rd edn). Oxford.

[6] Bolton P, Dewatripont M. 1994. The firm as a communication network. *Quarterly Journal of Economics* CIX(4): 809-839.

[7] Brandenburger A, Nalebuff B. 1996. *Co-opetition*. Harper Collins Business: New York.

[8] Demsetz H. 1988. The theory of the firm revisited. *The Journal of Law, Economics, and Organization*, 4, Spring; reprinted as chapter 10, pp.

159-178 in *The Nature of the Firm*, ed. by O. Williamson and S.Winter. 1991.Oxford University Press.

[9] Foss. N. 1997. (ed.) *Resources, Firms and Strategies*. Oxford university Press.

[10] Gibbons R. 2000. Why organizsations are such a mess (and what an economist might do about it. Unpublished manuscript, MIT.

[11] Gibbons R. 2003. Team theory, garbage cans, and real organizations: some history and prospects of economic research on decision-making in organizations, *Industrial and Corporate Change*. 12: 753-787.

[12] Hermalin B. 1998. Toward an economic theory of leadership: leading by example, *American Economic Review*, 88: 1188-1206.

[13] Radner R. 1992. Hierarchy: the economics of managing. *Journal of Economics Literature* 30(3): 1382-1415.

[14] Radner R. 1993. The organization of decentralized information processing. *Econometrica* 61(5): 1109-1146.

[15] Radner R, Van Zandt T. 1992. Information processing in firms and returns to scale. *Annales D'Economie et Statistique*, no25/26: 265-298.

[16] Rosen K. 2003. *Discrete Mathematics and Its Applications* (5th edn). McGraw-Hill.

[17] Saloner G, Shepard A, Podolny J. 2001. *Strategic Management.* Wiley.

[18] VanZandt T. 1999 Real-time decentralized information processing as a model of organizations with bounded rational agents. *Review of Economic Studies* 66: 633-658.

[19] Wernefelt B. 1984. A resource-based view of the firm. *Strategic Management Journal* 5: 171-180.

[20] Wernerfelt B. 1995. The resource-based view of the firm: ten years after. *Strategic Management Journal*. 16: 171-174.

[21] Wernerfelt B. 1997. Foreword to the Foss N. 1997 volume.

CHAPTER 4

Auditing the Auditors: Reducing Inefficiencies in a Trilateral Relationship using the Subgame Perfect Folk Theorem

Abstract

This is *a theory-based case study*, in which facts about the recent accounting-auditing scandals are integrated with game theoretic examination to obtain a better understanding of what has happened, why, how we can reduce the likelihood of repitition and whether the regulatory responses that followed are likely to produce their intentional aims. We consider the strategic problem of financial reporting and auditing as a repeated game. Auditors are called in to provide external enforcement, but we then face the question of who audits the auditors. The answer given by Game Theory's Subgame Perfect Folk Theorem is that we should rely on mutual enforcement: each party with stakes in the financial disclosure game must watch the behavior of the other parties. The importance of the game theoretic answer lies on the implication that regulators should design appropriate intertemporal incentives for the parties involved. The design and application of proper incentive schemes will produce better audits and avoid conditions like the accounting and auditing scandals we witnessed in the dawn of the twenty first century.

1 Financial Reporting and Auditing: A Trilateral Relationship

Measuring and recording transactions of separate entities is one of the main tasks of accounting; it is mirrored in the reported financial statements of firms. The owners of a firm, investors, *delegate* the preparation of financial reports to managers.[1] Numbers reported in financial statements reflect the consequences of all the contracts, complete and incomplete, formal and implicit, the entity is engaged with the parties it interacts. The accurate reflection of transactions into published financial statements is audited and verified by independent certified public accountants (CPA's). Once the second agent, a certified public

[1] ■ classic example of a Principal-Agent interaction examined by the theory of incentives.

accountant, is hired to monitor the actions of the first agent, we are getting a *trilateral interaction*.[2] The auditors are liable to shareholders and other stakeholders if they prove to be negligent in issuing a certification on the basis of misinformation.[3]

How we can be confident that the CPA's perform appropriate audits? Why can a firm's stakeholders rely on the reported accounting information and the audits of the CPA's? These are the questions we will try to answer. Providing convincing answers to such questions has become increasingly relevant and pressing in light of the recent accounting - auditing scandals and the failure of some giant business, financial and accounting firms.

The scandals made investors hesitant to channel funds to investment opportunities; by doing so, some funds that could go to productive activities are wasted. Economists have a rhetorical phrase for it: they say that corporate governance and financial reporting are at a *Pareto inefficient* state. Such inefficiencies are the result of *agency problems*; they are conflicts that arise because of the *incongruence of interests and incomplete information* among the parties involved in the relationships described above. The aim of the Sarbanes-Oxley Act of 2002 and other regulatory reforms introduced in the aftermath of the scandals, is to try and reduce the agency problems and thus move corporate governance towards a more efficient state.

The problem before us is a tantalizing one because, while the regulators impose new measures to correct existing deficiencies, the majority of reporting entities comply but some try to circumvent them. It is the desire to understand this practical problem, by providing an appropriate framework, tools of examination, and method of study that motivated our investigation. In the remaining of this introductory section, we develop a skeleton of the framework, and describe some of the tools that will be employed in this analytical narrative. The paper is a *theory-based case study*, in which facts about the recent accounting - auditing scandals and subsequent regulatory responses, are integrated with game theoretic explanation to obtain a clearer understanding of what has happened, why, how it can be avoided in the future, and whether the regulatory reforms will be successful.[4]

[2] Here we have an example of a multiagent interaction: one principal is interacting with two agents. Multi-agent contracting problems are formulated with models of mechanism design, common agency, collusion, contract externalities, and others. Interested readers may consult the graduate texts on contract theory by Patrick Bolton and Mathias Dewatripont (2005) and Bernard Salanie (2005).

[3] The institutional framework for financial reporting and auditing is described in Krishna Palepou, Paul Healy and Victor Bernard (2004, Chapters 1, 3, and 13). The theory of financial reporting is covered by William Beaver (1999). The theoretical background to auditing is covered in among other places by Shyam Sunder (1997 - Chapter 7, and 2003), and its practical aspects, in the context of the new regulatory environment, by Timothy Bell, Mark Peecher and Ira Solomon (2005). This is not to suggest that they are the best sources; they are just part of our reading on the topic.

[4] Such theory-based case studies appeared in the 1990's in other fields, like strategic management, political science, and economic history. Pankaj Ghemawat (1997) is a collection of such cases in the relative field to Management Accounting, that of Strategic Management. An alternative name given to this type of methodology is *analytic narratives*. The distinguished

1.1 The Institutional Framework

The paper considers the issue of accurate reporting and verification of accounting information, within the framework of the new institutional economics. Oliver Williamson (2000), one of the protagonists of this school and corecipient of the 2009 Nobel prize in economics, put forward a four-level classification scheme for the examination of institutions and transactions (See Figure 1). At the first level we have *informal institutions*, such as social customs and norms; they evolve over centuries. Double-entry and ethics of the accounting profession can be classified in this category. Sudipta Basu and Gregory Waymire (2006) provide an informal but convincing evolutionary explanation for the emergence of accounting recordkeeping. Shyam Sunder has repeatedly emphasized the development and adherence to social norms. He argues for a better balance between evolved norms and formal standards to be found in the following level, rather than an exclusive reliance on standards (e.g. Sunder, 2005 and 2006).

The second level is the *institutional environment*, made up of rules and laws; their development takes decades. Accounting and auditing standards belong to this category. The structures of this level are partly the result of evolution and partly the result of authoritative design. For example, the Securities and Exchange Commission (SEC) was established by Federal decree; it continues, even after the Sarbanes - Oxley Act, to oversee public companies. An example of evolved rule is the conservatism bias embodied in about a third of standards. Williamson (2000) believes that at this level, the government has opportunities for first-order economizing: get the rules of the game, such as standards and regulations, right. It is at this level, that recent reforms on the regulation of accounting information are introduced. Disclosure rules determine how players involved in financial reporting share information with each other. The post-2002 regulatory reforms concentrate on information and how it is communicated.

The *play of the financial disclosure game* takes place at the third level, and is where we find the modes of governance, which are relevant to prudent corporate governance and accurate financial reporting. The fourth level contains *economic activities* such as operating, investment, and financing activities; it has being traditionally the subject matter of neoclassical economics, but increasingly other approaches are applied to illuminate transactions and relationships in these spheres of economic engagement.[5] Neoclassical economics is an appropriate approach for the lowest level, because it assumes that interactions take the exclusive form of contractual exchanges. A firm's financial statements summarize the consequences of its activities. In the language of game theory, the numbers reported in the financial statements, measure the consequences of

economic historian Avner Greif is a champion of this methodology.

[5]Probably, the most influential of such metaneoclassical treatments is the *multi-tasking* extension of the standard agency theory, pioneered by Bengt Holmstrom and Paul Milgrom (1991) and developed further by Gerald Feltham and Jim Xie (1994) and Srikant Datar, Susan Kulp, and Richard Lambert (2001). Cutting edge research on operations from a contract theoretic perspective is to be found in Bolton and Dewatripont (2005) and Salanie (2005). The book by Jean Tirole (2006) contains the state of the art on the strategic perspective to investment and financing.

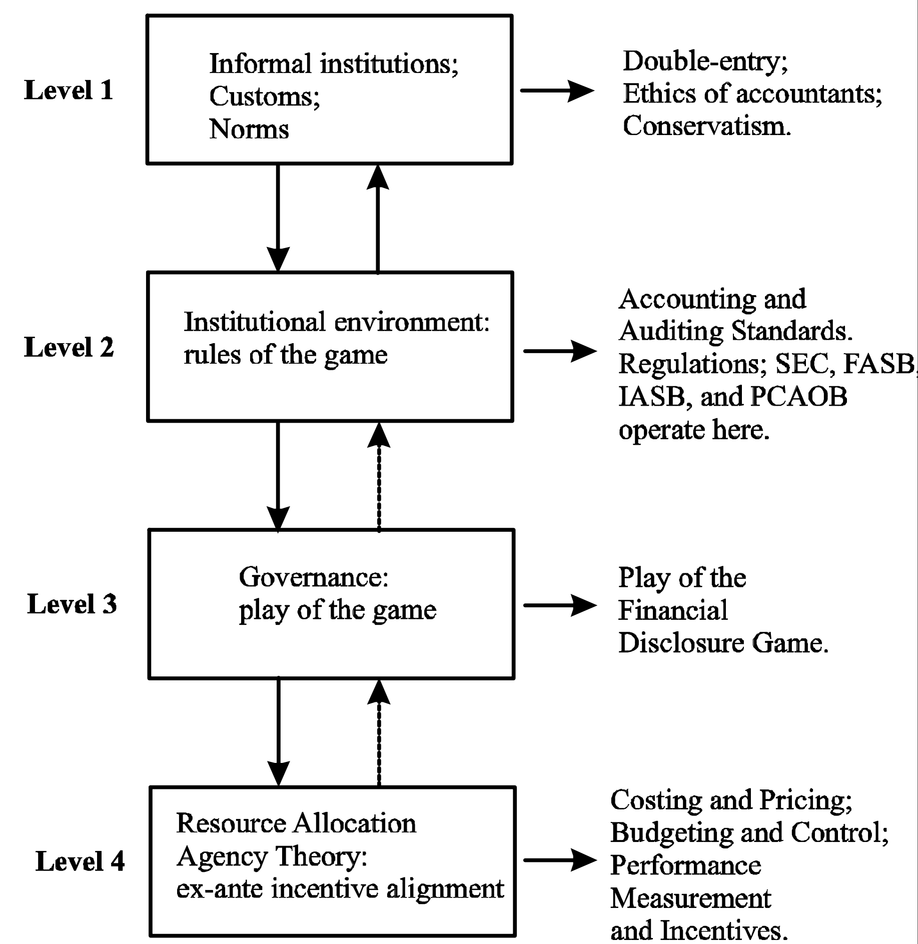

Figure 1. Williamson's 4-level Schema applied to Accounting and Auditing.

Figure 1:

decisions and actions taken by the entity and by all the other entities it interacts.

Why we should be concerned with the institutional environment and not just proceed directly to consider how accountants and other decision makers are playing the financial reporting game? The reason is that governance of modern corporations involves the interaction of hundreds or even thousands of people. As a result, we cannot solely rely on accountants' professional ethics for good governance. We also need to investigate the evolution and design of institutions that promote accurate financial reporting given the absence of perfect ethics.[6] As Rick Antle and Stanley Garstka (2004, p.4) point out, the institutional environment disciplines both the constructors and users of financial statements.

1.2 Subgame Perfect Nash Equilibrium Analysis

> "The strength of a control system resides in the threat of what might be reported or what might take place were off-equilibrium play to occur." (Joel Demski, 2003, p.56).

The specific research question we ask is: Given that the government's law apparatus, situated at level two, has proved ineffective to prevent accounting - auditing scandals from happening, what alternative institutions may provide appropriate rules of the game?[7] A number of distinguished accounting scholars and economists (e.g. George Benston, Michael Bromwich, Robert Litan, and Alfred Wagenhofer, 2003; and Preston McAffe, 2004) seem to agree that the question of *trust* is at the heart of scandals.

The paper investigates *whether and how trust is established* by the interaction, play, of people, involved in the preparation - reporting of financial information, the verification of its accuracy (supply side), and its use (demand). The purpose of the paper is to predict the path of play. That is, the paper is mainly concerned with the third level of Williamson's schema and the feedbacks between level three and level two. The choice to concentrate on level three is based on the empirical regularity that in financial disclosure, the management of transactions and dispute resolutions tend to be resolved directly by the players, through private ordering. (See Antle and Barry Nalebuff, 1991; Williamson, 2000). In the terminology of Dixit (2004), private ordering operates in the shadow of the law.

Masahiko Aoki (2001, pp.60-61) calls the mechanisms that facilitate honest interaction, the *governance mechanisms* of interaction. Since most interactions are some type of contract, formal or relational, the mechanism of governance is also called *contract enforcement mechanism*. One of the objectives of the new institutional economics, and of the present study, is to understand the conditions under which various governance mechanisms can become *self-enforcing*, and sustain honest relationships.

[6] This is an adaptation of Samuel Bowles (2004, p.26) justification for modern economics increasing preoccupation with the workings of institutions, using a game theoretic approach.

[7] Here, we apply Avinash Dixit's research agenda in his Lawlessness and Economics (2004).

There is not a general consensus about what constitutes a self-enforcing agreement, but most game theorists believe that the requirement of the play being a *Subgame Perfect (Nash) Equilibrium, SPE*, is reasonable (See David Pearce, 1992, p.134, and pp.161-167). It is a refiniment of Nash equilibrium proposed by Reichard Selten.[8] The Subgame Perfect Equilibrium is based on the requirement that Nash Equilibrium play (which means best response play by all the participants, with no participant having an incentive unilaterally to change this behavior) occurs even at out of equilibrium subgames. It formalizes the idea contained in Demski's scholium, quoted at the beginning of the current subsection. In everyday English, it must be to the continuing advantage of all players to honor such agreements (See Robert Aumann and Lloyd Shapley, 1976, p.2). It is the definition of self-enforcing behavior employed in the present study. The modern theory of *repeated games* is the appropriate model, in which self-enforcing arrangements can be explored.

The fundamental problem every country's regulators face is to design institutions were the conditions that ensure Subgame Perfect Nash Equilibrium play apply to arrangements that promote the public firms' stakeholders interests. Ken Binmore (2010) provides an authoritative discussion of the interplay between institutional economics and game theory in the same spirit with the present work.

1.3 The Players

> "reality is multiple conflicts among multiple players, in the context of an enlarged, interactive web of controls." (Demski, 2003, p.53)

Up to this point, we have discussed the requirements concerning the behavior of the participants involved in the financial disclosure game, that will ensure honest interaction between the suppliers and users of accounting information. To put it in a slightly different way, we have described the standard of reasonable play. But, who are the participants, players, in the game?

Demski (2003) lists the following players as being involved in the corporate governance game: auditors, board of directors, analysts, investment bankers, regulators, management, and others such as attorneys and investors.[9] To keep the formalization tractable, following a suggestion made by David Kreps (1990a, ch.21, and 1996, p.573), we model the auditing problem as a trilateral relationship.[10] The three players are: investors (which includes analysts and investment bankers); managers (which includes board of directors); and auditors;

[8] John Nash, who proposed *Nash Equilibrium* (and hero of the film Beautiful Mind), and Reihard Selten are two of the three scholars, who shared the 1994 Nobel prize in economics, mainly for these equilibrium notions. The third was John Harsanyi, who developed the methodology for solving games under incomplete information. Harsanyi's methods are relevant to section five.

[9] Palepou et al (2004, chapter 13, p.3) provide a schematic summary, which portrays the interdependencies of players, in the interaction of supply and demand of financial information. We employ it later to illustrate the main message of the paper.

[10] See also James Baron and David Kreps (1999, ch.4 and appendix A), and Kreps (2004,

they are the three protagonists in Sunder's (1997) Theory of Accounting and Control. When the two players, the managers who prepare the financial statements and the investors who use them, cannot identify ex-ante their common Pareto-efficient outcome, a third party, the auditors, may become necessary to govern such a relationship (see Aoki, 2001, p.61). Once the auditing firm, who is supposed to act with independence, is added, we move into the world of what Williamson (1979) calls *trilateral governance*.

The auditing contract does not specify what kind of adaptation will be made in various contingencies, but prescribes a third party, who will determine appropriate adaptation (see Kreps, 1990a). One can say that auditing has evolved to deal with unforseen contingencies that characterize the incomplete contracts of bounded rational contracting parties. The role of unexpected events and changes in conditions throughout the audit engagement is emphasized in the revised International Standard on Auditing (ISA) 300 by the International Auditing and Assurance Standards Board (IAASB) (see IFAC, 2004). In practice, adaptation follows some convention or specified procedure, familiar to the auditing profession.

The present work does not treat the auditors as exogenous neutral parties that automatically follow and observe the prescribed accounting and auditing standards. Instead, the paper examines whether standards for auditors actions, can be voluntarily and credibly chosen; it investigates how the actions prescribed by the standards can become *incentive compatible* (self-enforcing) for the auditors (See Aoki, 2001,p.61).

Despite the increasing complexity that the consideration of interacting multiple players brings, we will see below that the subgame perfect equilibrium is telling us that control becomes easier as one player watches the behavior of the others. We develop this seemingly paradoxical result in the main part of the paper.

But what is it that auditors should verify?

1.4 The Subject Matter of Auditing

Shyam Sunder (1997, and 2003) takes a broader view of auditing than the traditional one, arguing that the auditors' main contribution to the firm is verification of the accounting system. In Sunder (2003) the audit requirement together with a mechanism to control its quality, is one of the four main elements of the U.S. corporate governance system; the other three being the accounting rules, organization to set accounting rules, and the participation of the board of directors in audit and reporting. This broader view is in line with the position taken by Douglas Carmichael, the first Chief Auditor of the Public Company Accounting Oversight Board (PCAOB) established by the Sarbanes-

ch.24). Robert Wilson (1983, p.312, footnote 7) also suggests that auditing can be examined as a 3-player repeated game. The idea of modeling auditing as a repeated game comes from the writings of Ken Binmore. Binmore (1992, 1998, and 2005) discusses the more general question of who guards the guardians as a repeated game and he suggests that the answer is given by game theory's perfect folk theorem.

Oxley Act of 2002. Carmichael (2004) argues that following the standards set by PCAOB, the audit of a public company becomes an audit of the company's financial reporting process.

The PCAOB is another semi-public institution established by the authorities, belonging to the second level of the Williamsonian schema. It was created to oversee and regulate auditing. John C. Coates IV (2007) provides an illuminating discussion of its design and role.

One of the first standards proposed by PCAOB requires *audits of internal control* over financial reporting. Such requirements aim to alleviate moral hazard, which is due to insufficient effort or attention to the oversight of subordinates; scandals in the 1990's, involving losses caused by traders who were subject to isufficient internal control, are textbook cases of this form of moral hazard (e.g. Metallgesellsaft, and Barings). The January 2008 huge loss caused by a trader at the second largest French bank, Sociate Generale, is a recent example. As we will see below, this stance is consistent with the overriding principle of the present study, namely that each party involved in the controling, reporting - auditing of public firms should watch the behavior of others.

1.5 Tools and How the Narrative will Unfold

The paper employs the tools of non-cooperative game theory, which are appropriate in settings where markets are peripherally relevant, such as the relationship between a shareholder and a manager or a regulator and a firm. We examine financial reporting, governance, and auditing in the context of *long-run relationships*. The modern theory of *repeated games* provide us with the tools to consider values such as *trust*, *credibility* and *reputation*, as well as the role of *intertemporal incentives* in such relationships. The theory of repeated games opposes opportunistic behaviour by punishing it. Since the paper employs many of the basic ideas and tools of modern noncooperative game theory, it can also be used as a quick refresher course in the field for graduate accounting students, especially on the theory of informal contracts.

Having described the institutional framework, the standard of reasonable play, the players, and the subject matter of auditing, we provide an outline of the rest of the paper in the language of the principal-agent model, since it has been applied widely in the accounting literature. The two approaches are isomorphic, because the principal–agent model is a leader-follower, Stackelberg game (See Salanie, 2005, p.5). The study is a partial equilibrium one as it isolates the market of accounting information from the rest of the economy; it examines the interactions of a small number of players. We pose the questions in the context of simplified hypothetical examples, in which many of the real world details are ignored. Whereever possible, we supress technical details or simplify to improve readability.

Section two considers the simplest 2-player principal-agent model: the first player is investors and the second is managers. We make the conventional assumption that managers are the informed player; they possess some private information. It is the familiar setting that gives rise to adverse selection and moral hazard problems. Moral hazard may come in many forms (See Tirole, 2006, Chapter 1), but in this paper we concentrate on accounting manipulations. The uninformed party, the investors, moves first and is imperfectly informed of the actions of the informed party. The model represents the constraints imposed by the institutional setting with a contract. The Sarbanes-Oxley act has made the contract between investors and managers formal as it requires the signing of reported financial statements by managers. The certifications requirement is critically discussed by Marshall Geiger and Porcher Taylor (2003). The contract retains its implicit characteristics because the managers have flexibility over the numbers they report, and are supposed to behave according to ethical norms. Such managerial discretion can be abused; it turns out that abuse is the theoretical prediction of the 2-player game when it is played only once. In short, this section shows that often in one-shot games, a player has an incentive to deviate from an efficient outcome. In the real world, we observe sustained long-term relationships between investors and managers. This setting is modeled by a repeated game.

Section three investigates whether in a repeated game setting, where the two players play the financial disclosure game (the stage game) repeatedly, managers have incentives to honor the trust of investors. We will see that in the case of a finite repeated game, with a known terminal date, theory predicts that the managers will betray the trust of stakeholders. In the infinite repeated case, we get one of the most celebrated results of game theory, the *folk theorem*, which says that many outcomes are possible, including the cooperative, honored, efficient outcome. What the folk theorem does not say is that the cooperative outcome will happen. One way to find out that the explicit contract is honored in a long-run relationship, is by the desire of the managers to develop and maintain a reputation for honest behavior.

In section four, the explicit contract is guaranteed by a third party, the auditors. Both explicit and implicit contracts between investors and managers, and investors, managers and auditors are sustained by equilibria that result from the interactions of the involved parties. The *subgame perfect folk theorem* produces a powerful insight to the theory of corporate governance: it internalizes the concerns of a corporation's stakeholders. By doing so, it provides a resolution to the debate of shareholder value maximization versus stakeholder society.

Section five suggests few avenues for future research. Finally, section six concludes by providing a critical evaluation of the usefulness of the game theoretic approach.

2 Trust or Do Not Trust the Managers?

> "... the crucial step in solving a real-life strategic problem nearly always consists of locating a toy game that lies at its heart" (Ken Binmore, 2005, p.58).

This section employs a toy game, called the Trust-Abuse Game, or the Trust Game for short, to portray the relationship between a shareholder and a manager in their interaction concerning financial disclosure. This simple game will help us to explore the question of how credible the manager's promise is to honor the trust of shareholders, and report the accurate position of the entity. The Trust Game captures the *incentives* faced by the players involved in financial disclosure, and will be the workhorse for discussion in later sections.[11]

2.1 The 2-player Trust Abuse Game

> "Enron and WorldCom thus became both examples and symbols of a broken system. Investors turned skittish because no longer had full and complete trust in all the financial information they were being given" Michael G. Oxley (2007, p.C1).

One way of formulating an interactive problem as a game is the *extensive form*.[12] The extensive form is a dynamic structure that specifies the rules of the game: who moves when, what actions are available to each player at his decision points, and what each player knows when it is his turn to move. A class of extensive form games consists of games in which players are perfectly informed of all previous actions taken, whenever it is their turn to move; they are called *games of perfect information*. The Trust-Abuse Game is an example of a 2-player game with perfect information.[13] The following figure, called a tree, portrays its extensive form.[14]

There are two players, represented by I and II. The paper follows the convention that odd numbered players are male and even numbered are female. The game has two decision nodes: the initial node, also known as the root, at which player I decides, and a second decision node at which is player II turn to move. And it has three terminal nodes corresponding to the three possible plays of the game, to which are assigned three 2-dimensional payoff vectors. The first element of the vector is the payoff of player I and the second is the payoff of player II.

The game begins with the initial decision node for player I, who must choose either to trust (denoted by T) or not trust (denoted by N) player II. If player

[11] We dedicate the paper to the memory of the nice accounting and experimental economics scholar John Dickhaut who with his coauthors Berg and McCabe first proposed the Trust game in 1995.

[12] The exposition in this section is intentionally at a slower pace than the rest of the paper as it conveys ideas and techniques, especially the mechanics of the backward induction technique, that will be used throughout the paper.

[13] The choice of the Trust Game for capturing the structure of the Financial Reporting interaction is based on the writings of David Kreps. It is also suggested by Eric Rasmussen (1994, section 5.3, pp.129-131; and 2007, section 5.3, pp.136-7). The discussion draws from Kreps (1996, and 2004), and Robert Gibbons (1997, and 2003). Payoff numbers come from Binmore (2005). The Trust Game is often employed to illustrate the *hold up* problem in Incomplete Contract theories of the firm.

[14] Certified Management Accountants (CMA's) have studied one-person decision trees to get certified. Trees are also used to portray the value of Real Options in the Investment Analysis part of the CMA curriculum. The difference here is that different players decide at various decision nodes.

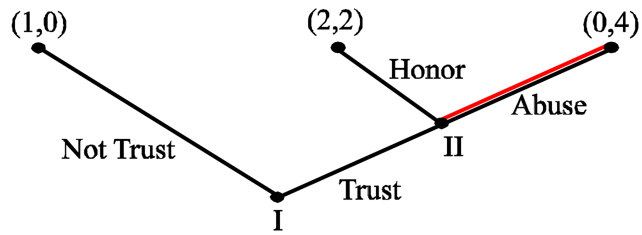

Figure 2. The Extensive Form of the Trust Game.

Figure 2:

I chooses not to trust II, then the game ends with player I geting 1 and II get nothing, 0. If I chooses to trust II, II is made aware of this and at her decision node has the choice either to honor that trust (denoted by H) or to abuse it (denoted by A). If I trusts II, and II chooses to honor that trust, both get 2. But if I trusts II and II chooses to abuse that trust (i.e. player II engages in opportunistic behavior), II gets 4 and I gets nothing, 0. In the scenario of the present paper, investors use firm disclosures to judge whether managers have governed the firm in line with investors interests, or have abused the authority and control over firm resources. Manipulations of accounting numbers (one in a range of managers' entrenchment strategies) is one of several forms that hidden action (moral hazard) may happen in this relationship.

Having described financial disclosure as an extensive form game, we proceed to solve it using backward induction; that is, by working backward through the game tree, starting at the last decision node. One of the advantages of applying game theory to strategic problems is that the specification of a game is forcing us to describe the institutional setting, in which the interaction takes place. Outcomes, usually depend on the institutional details: having specified that investors move first, determines the solution of the game.

2.2 An Inefficient Subgame Perfect Nash Equilibrium and the Problem of Credibility

> "What (investors) want is a transparent system so they can make up their minds as to whether the investment is credible" Michael G. Oxley (2007, p.C2).

Suppose that investors have decided to delegate the preparation of financial statements to managers; that is, player I has decided to trust II. Consequently, the game has moved to the second decision node. What is best for player II at this point in the game? Player II can receive a payoff of 2 by choosing to honor I's trust or a payoff of 4 by choosing to abuse I's trust. Since 4 is greater than 2, player II's best response is to betray I's trust because by doing so, she receives a payoff of 4 rather than 2. From player I's point of view, the game reduces to

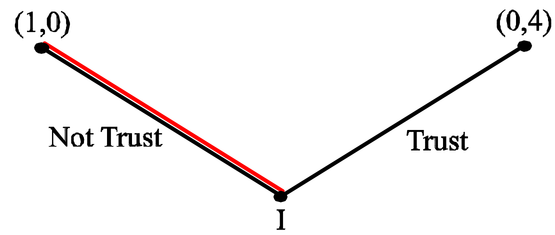

Figure 3. The 1-player game obtained after replacing the subgame with its value.

Figure 3:

that given in the following figure, where we have replaced player II's decision node and the subgame that follows with the payoff vector that will result once her decision node is reached.

At the initial node, player I will choose not to trust II because this yields a payoff of 1 rather than 0. We have arrived at a pair of strategies, one strategy for each player by solving the game using backward induction. The strategies are: player I does not trust II and player II abuses I's trust. Embodied in backward induction strategies is the idea that choices made early in the game ought to take into account the optimal play of future players. These backward induction strategies are portrayed by the bold lines in the game tree. A pair of strategies that constitutes a Nash equilibrium in every subgame, like the two strategies described above, is called *Subgame Perfect Equilibrium (SPE)*.

We can confirm the robustness of the solution by formulating the Trust Game in *strategic form* (the older, still in use, synonymous term is normal form) and find its *Nash Equilibrium*.

The strategic form with ordinal preferences[15] specifies three elements:

1. The two players: Player I , the shareholder, and player II, the manager.
2. The strategies available to each player. Player I has two strategies: he can either trust (T) or not trust (N) the manager. Player II has also two strategies: she can either honor (H) or abuse (A) the shareholder's trust.
3. For each player, preferences over the set of strategy profiles. It is convenient to specify players' preference orderings by giving payoff functions

[15] For an excellent description of the strategic form as a modeling tool and for an introduction to game theory in general, see Martin Osborne (2004).

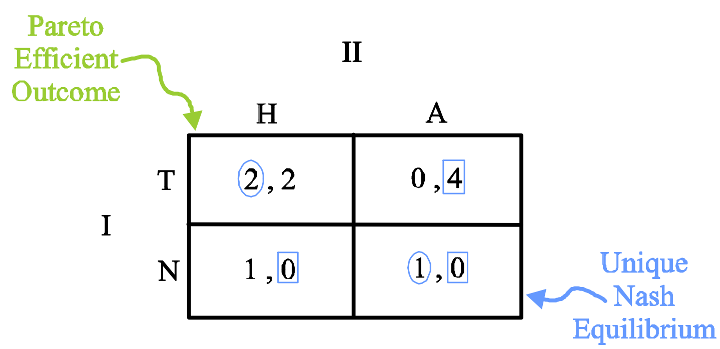

Figure 4. The Strategic Form of the Trust Game.

Figure 4:

> that represent them. For each combination of strategies, also called *strategy profiles*, one strategy for each player, corresponds a payoff to each player.

Applying the best response solution method to the strategic form of the Trust Game, we find that it has a unique Nash equilibrium, which calls for player I not to trust and for player II to abuse. In the bimatrix, best responses of player I are shown by circles and best responses of player II by squares. The cell that has both, is the payoff vector that corresponds to Nash equilibrium strategies. This is a particular instance of a theorem due to one of the early pioneers of game theory, Harold Kuhn. It states that if s is a backward induction strategy for a perfect information finite extensive form game, then s is also a Nash equilibrium of this game.[16]

The important feature of this game is that when it is played once and once only (that is, the *one-shot* Trust Game), player I would not willingly trust player II and the resulting payoff vector would be $(1, 0)$. It makes both players worse off than they would be if I had chosen trust and II honor; it is an inefficient outcome. But, it is the unique, self-enforcing outcome of the game. In any other strategy profile, one of the players wants to deviate to a different strategy. The Trust Game is like a one-sided Prisoner's Dilemma. In this methaphor, we do not claim that financial reporting necessarily has the structure of the One-Sided Prisoner's Dilemma. We only say that people involved in financial reporting may have the same preferences as in the One-Sided Prisoner's Dilemma. The dilemma part is that player I ends up with payoffs of 1 and player II with zero, (0), while payoffs of 2 to each are available. It is one sided because only player II has the opportunity to make a private gain. It rerpesents the archetypal transaction with some element of moral hazard (See Kreps, 1986). This theo-

[16] Geoffrey Jehle and Philip Reny (2001, p.296) provide a straightforward proof of this theorem.

retical prediction is in line with the public's widespread concern that managers are not accountable (See Tirole, 2006, p.16).

Player II, the managers, could promise to honor the trust, but as Dixit (2004, p.15) says

> "... in the absence of some form of governance, the promise is not credible".

The lack of credibility due to the empty promise of managers can cause a breakdown in the relationship between the shareholder and the manager. Anticipation of opportunism provides a disincentive to making valuable investments. The unique Pareto inefficient outcome means that some funds that could potentially get invested in wealth creating activities are wasted. This is precisely economists's definition of *Pareto inefficiency*. How can the two parties avoid the waste and attain the first-best, cooperative outcome $(2, 2)$? Tirole (2006, p.15) argues that insider moral hazard can be reduced in two ways: The first, is by trying to align insiders' incentives with investors interests through performance-based incentive schemes. The second is by monitoring insiders. The present work is mainly concerned with monitoring and punishing deviant behavior, although it repeatedly emphasizes the importance of providing appropriate intertemporal incentives.

In the one-shot Trust Game, a method can be found for penalizing Player II, if she deviates (Dixit, 2004, p.16). The two players agree before the play that (Trust, Honor) is the Pareto efficient outcome. They can sign a formal contract, which binds them to comply with their part in the (Trust, Honor) strategy profile. Courts enforce the contracts. The Sarbanes - Oxley Act requirement that managers must sign the financial statements, has to some extent made the financial reporting contract formal. Enforcement requires that deviations can be proved before court, or are verifiable. But, managers retain a certain amount of discretion in what they report; the contract between investors and managers still exhibits informal aspects. In the presence of such grey areas, how can we explain enduring relationships between shareholders and managers?

Two real world phenomena ensure that shareholders trust their investment funds in the hands of managers. First, the shareholder and the manager do not meet only once but repeatedly: they are not engaged in a one-shot game, they play a repeated game. Second, the auditing institution has evolved to oversee that trust is not abused. Modern noncooperative game theory has developed tools that model these two aspects. In the following section, we model repeated interactions. Auditors come into play afterwards.

3 Will Trust Prevail in Long-Run Relationships?

Game theorists employ the model of a repeated game to formalize long-run relationships. Drew Fudenberg and Jean Tirole (1991, p.145), in their authoritative game theory book, say that repeated games are a good approximation for some long-term relationships, particularly those were "trust" and "social pressure" are important.

Can the inefficient outcome of the one-shot Trust Game be avoided when the game is repeated many times? The model of a repeated game help us to focus attention on the variables that determine when players are guided by long rather than short-run considerations. The following subsection considers the simplest model of repeated games, the twice repeated simultaneous move case.

3.1 The Twice Repeated Trust Game

Suppose the Trust Game (T.G.) is played twice by the same players.[17] Then the T.G. is said to be the *stage-game* of the two-period repeated game, denoted $(\text{T.G})^2$. The following assumptions are standard; we will question them in later sections:

- The two players are observing the outcome of the first round before the second begins.
- The payoffs in the repeated game $(\text{T.G.})^2$ are obtained by summing the payoffs in each stage game: there is no discounting.[18] For example, if the action profile (T, H) is played in the first stage and the action profile (T, A) is played at the second stage, then player I gets $2 + 0 = 2$ and player II gets $2 + 4 = 6$ in the repeated game $(\text{T.G.})^2$.

Implicitly, we are also assuming that

- The stage game is a simultaneous-move game;

and it is also useful to have explicitly stated the assumption made in the first sentence of the subsection, namely

- At each stage, the game is played by the same players.

Caution:

Pure strategies in the repeated Trust Game and in general repeated games are not the pairs of strategies that result from combining the available actions in the two stages. The reason is that the pure strategies that result from those combinations ignore that players in the repeated game, will wish to make their behavior at the second stage contingent on what happened at the first stage. Each player can condition his action at the second stage on the other player's previous action. Players in repeated games follow contingent strategies that depend on behavior in previous rounds of the game. Instead of formulating this principle in general terms, we look at the heuristic example of our Trust Game.

Let $S = \{T, N\}$ be the set of pure strategies for player I in the stage game. We call these pure strategies *actions* to distinguish them from pure strategies

[17] The discussion in the present subsection draws on Binmore (1992) and Gardner (1995).

[18] This simplifying assumption is relaxed later in the paper.

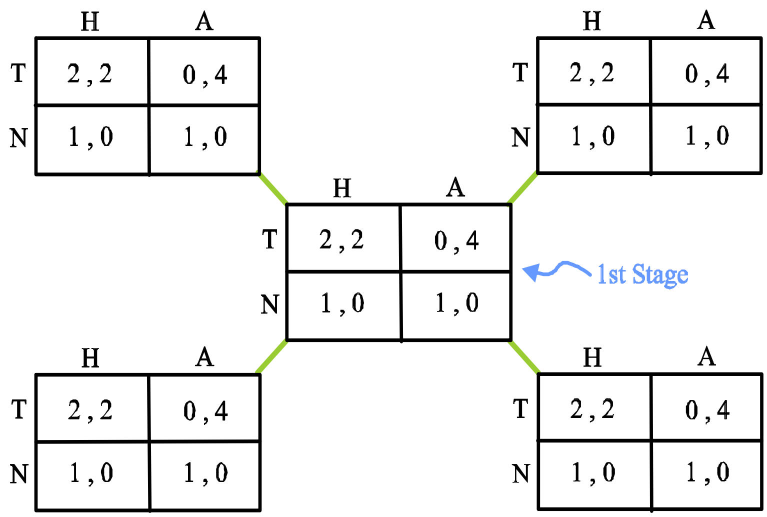

Figure 5. The Twice Repeated Trust Game: the actions in the 2nd stage are functions of the 4 possible outcomes, histories, of the 1st stage.

Figure 5:

in (T.G.)^2. The set of actions for player II in the stage game is $T = \{H, A\}$. Assume that

- At the second stage, the players remember the play of the first stage.

There are four possible outcomes of the first round in the Trust Game. Each of these outcomes is a possible *history* that has to be taken into consideration for the second stage of the game. We can illustrate the setting schematically in the following figure.[19]

The bimatrix at the centre represents the first stage of the repeated game. Each of the four peripheral bimatrices represents a possible play at the second stage of the twice repeated Trust Game. Each line from a cell of the first stage bimatrix to the second stage bimatrix represents a history.

Four possible *histories* of the game must be considered; they are the four elements of the set $H = S \times T$.

$H = S \times T = (T, H), (T, A), (N, H), (N, A)$.

For example, the history $h_{TA} = (T, A)$ means that player I used action T and player II used action A at the first stage.

[19] The schematic device for portraying a twice repeated 2 by 2 (2 players and 2 strategies for each player) game is due to Roy Gardner (1995). Only its application to the Trust Game is ours.

A *strategy* is defined as a complete plan of action for (T.G.)^2. A strategy specifies what a player does in the first stage and what this player does in the second stage following each history leading to it. In symbols: A (*pure*) *strategy* for player I in (T.G.)^2 is a pair (s, f) in which s is an action in S to be used in the first stage, and f is a function such that

$$f : H \rightarrow S$$

The strategy of player I in (T.G.)^2 specifies that at the first stage he either decides T or N; he has two choices at the first round. And it specifies what player I does at the second stage, following each history; there are four (4) histories and two (2) choices for each history: player I has 2^4 choices at the second stage. Therefore, player I has $2* \, 2^4 = 32$ strategies in the twice repeated Trust Game.

Similar is the definition of strategy for player II, who also has 32 strategies. The strategic form of (T.G.)^2 involves a $32 * 32$ matrix. Further repetitions make the matrices bigger and bigger. But the **sequential** nature of the relationship explored by repeated games suggests there is a convenient way to examine repeated games: the extensive form. The reason is that repetition creates subgames. Repeated games have a subgame structure, and thus we can employ subgame perfection in their solution. Using backward induction, we will show that the (T.G.)^2 has a unique subgame perfect equilibrium, in which player I always plays Not Trust (N) and player II always plays Abuse (A).

Following the backward induction method we examine the first stage of the 2-stage T.G., (T.G.)^2, by taking into account that the outcome of the game remaining in the second stage will be the Nash equilibrium of that remaining game: (N, A) with payoff $(1, 0)$. That is, we start with the four final subgames: the four peripheral matrices of the picture. Each of the four subgames has the same unique Nash equilibrium: player I plays N and player II plays A. The players' first stage interaction in the (T.G.)^2 is equivalent to the one-shot game portrayed in the following matrix, in which the payoff pair $(1, 0)$ for the second stage has been added to each first stage payoff vector.

Again the game has a unique Nash equilibrium: (N, A). We conclude that there is a unique Subgame Perfect Equilibrium of the (T.G.)^2 in which player I in both stages chooses Not Trust (N) and player II in both stages chooses Abuse (A).

Lack of credibility by the managers prevents the players from achieving a Pareto improvement outcome over the one-shot equilibrium. A manager's promise to honor the trust of shareholders in the second stage, cannot by itself stand the credibility criterion.

3.2 Finitely Repeated Trust Game

The backward induction argument employed above to establish the Subgame Perfect Equilibrium of the twice repeated Trust Game, (T.G.)^2, holds for the

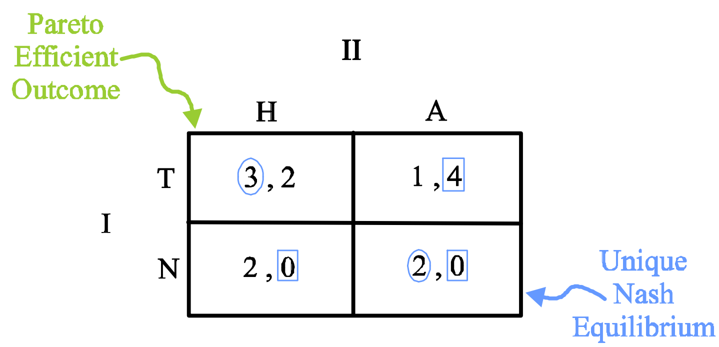

Figure 6. The first stage of the twice repeated Trust Game, after backward induction.

Figure 6:

general finitely repeated Trust Game, $(\text{T.G.})^n$, where n is a fixed integer. It amounts to folding the finitely repeated Trust Game, one subgame at a time, beginning from the end. We present it as a proposition that mimics the theorem given by Binmore (1992, pp.354-355) for the Prisoners' Dilemma.

Proposition 1 *The **finitely** repeated Trust Game has a unique Subgame Perfect Equilibrium in which player I always plays Not Trust and player II always Abuses.*

Sketch Proof

To demonstrate the proposition we use a backward induction argument and thus we start from the smaller subgame, which is the last round of the game.

Each player thinks about the play of the last stage. In the last round, no player cares about how an action in this last round will affect future play, because the game will end after this period. Accordingly, in the last stage, both players act as if they are playing a one-shot game. Each will therefore play its part of the unique one-shot Nash equilibrium.

Given the outcome of the last round, the players consider the penultimate round of the game. Since it is known that in the last stage, the players will play their one-shot Nash strategies, no player has any concern about how his action in the penultimate stage will affect the last stage of the game. Therefore, both players will behave in the next to last stage as if it were a one-shot game and thus they will play their one-shot Nash strategies.

Next, let us examine what happens in the pre-penultimate stage. Since no player needs a good reputation in the penultimate round, they treat this stage as a one-shot interaction and thus again play the one-shot Nash responses.

Following backward induction until the first round, suggests that in the finitely repeated Trust Game, each player in every round, will play his-her unique

Nash equilibrium action. Therefore, finite repetion of the interaction cannot build trust between the parties. We have demonstrated that the Finitely Repeated Trust Game has a Unique Subgame Perfect Equilibrium, in which in every stage, each player plays his/her one-shot Nash Equilibrium action.

So far, we have shown that in all three cases of:

- the One-shot Trust Game;
- the Twice repeated Trust Game; and
- the Finitely repeated Trust Game,

the two players cannot avoid the inefficient, but unique, subgame perfect Nash equilibrium. Accordingly, the answer to the question-heading of the present section is, thus far, no: trust will not prevail if the trust game is repeated a finite number of times. How then we can explain the presence of ongoing shareholders-managers relationships that we observe in the real world? The key to the resolution of this apparent divergence between theoretical predictions and real world phenomena, is to understand that the argument we employed to find the Subgame Perfect Equilibrium in the finitely repeated Trust Game is based on four strong assumptions:

1. The final stage of the game (i.e. the number of repetitions) is common knowledge;
2. Players are rational;
3. It is common knowledge that they are rational; and
4. Nomatter what happens in the game, it remains common knowledge that they are rational.

The assumptions on which the Subgame Perfect Equilibrium is based are not realistic. Game theorists have developed models with more realistic assumptions, which can produce cooperative equilibria and thus explain the long-run relationship between shareholders and managers. The most popular way of making the model more realistic is to consider repetitions of the stage game (in our case, the Trust Game) that may continue for ever. It is the subject of the next subsection.

3.3 Infinitely Repeated Trust Game and the Folk Theorem

> "...ethical behavior does not come from within the individual but must be motivated extrinsically" David Kreps (2004, p.614)

Infinite repeated games are relevant to corporate interactions, because the corporation is a going concern: it has an infinite life (Gardner, 1995, p.177).

Let the Trust Game to be repeated infinitely and assume that, for each period t, the outcomes of the $t-1$ preceeding plays of the stage game are observed before the t stage begins. Subsection 3.1 on the twice repeated Trust Game, has shown that a strategy of each player in the repeated game is a function of the history of play thus far. We have shown that finite repetition of the Trust Game cannot produce a Pareto improvement over the inefficient one-shot unique equilibrium. Next, we will demonstrate that once we allow the Trust Game to be repeated infinitely, we get a lot of other equilibria. The characterization of all possible Nash equilibrium outcomes of any infinitely repeated game, is provided by a celebrated result of game theory, called the Folk Theorem.

Before providing a proof of the folk theorem, let us be clear about what it says. The Folk Theorem specifies all possible behaviors that can arise as equilibria of an infinitely repeated game. It says that if we allow a game to be repeated infinitely, then we can get many outcomes as equilibria, which were not possible in the one-shot play, or, its finite repetition. The question is what are the characteristics of the many Nash equilibria that can prevail in the infinitely repeated game. We cannot do better than quote from Sergiu Hart's talk at the 2005 Nobel sumposium. The **Folk theorem** states:

> "The set of Nash Equilibrium outcomes of the *repeated game* equals the set of feasible and individually rational outcomes of the one-shot game" Sergiu Hart (2005, p.8).

A Heuristic Proof of the Folk Theorem:

To give the theorem and the terms in its statement substance, we provide a heuristic proof of the folk theorem for the infinitelty repeated Trust Game. The idea of the proof is to have the players coordinate on a feasible *master plan*, which is supported by the *threat of punishment* in case of deviation from equilibrium play. It is what the regulators of the accounting and auditing industry have in mind when they introduce guidelines for accurate financial reporting.

The three payoff pairs that correspond to all the four possible outcomes of the one shot Trust Game are shown in the following 2-dimensional coordinate plane.

If the two players were to bargain about how to play, they might agree on any of these 3 points, but they have many other options as possible agreements. For example, they might agree by tossing a coin or take turns to settle any dispute that may arise about which alternative should be adopted. To take all such posssibilities into account, it is necessary to expand the two players' set of feasible agreements to the shaded region of the diagram. This is the smallest convex set, called the convex hull, that contains all three payoff pairs. A payoff pair (π_I, π_{II}) is *feasible* in the stage game, T.G., if it is a convex combination of the pure strategy payoffs of T.G.. In other words, a payoff pair is feasible if it is a weighted average, where the weights are all nonnegative and sum to 1. The shaded area represents all the feasible payoff pairs.

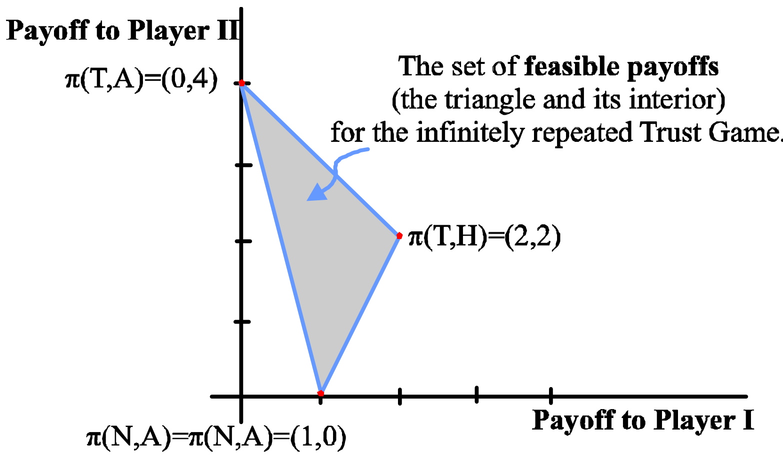

Figure 7. The Maximum Cooperative Payoff Region.

Figure 7:

To place the folk theorem in the context of the paper, suppose the two players play a financial disclosure game every quarter, and agree that player II will always Honor the trust, and player I will play Trust only every other quarter; then they will implement the payoff pair $(2, 2)$ one half of the quarters and the payoff pair $(1, 0)$ the other half of the quarters. Player I then expects 1.5 on average, and player II expects 1. It is a Pareto improvement on the unique subgame perfect equilibrium of the one-shot game. The point $(1.5, 1)$ lies halfway between $(1, 0)$ and $(2, 2)$, and is therefore in the shaded set. The remaining payoff pairs in the interior of the shaded region are weighted averages of more that two pure strategy payoffs. To achieve a weighted average of pure strategy payoffs, the players could use a public randomizing device: by playing (T, H) or (N, A) depending on a toss of a coin, for example, they achieve the expected payoffs $(2.5, 1)$.

The problem with the bargaining scenario is that it works only when an *external agency* is willing and able to enforce any contracts that the two players write. Without an external agency, any agreements that the two players make must be *self-enforcing*: only equilibria are viable agreements.

In the one-shot Trust Game, the absence of external enforcement implies the breakdown of the relationship, because the only equilibrium is the inefficient outcome. But the infinitely repeated Trust Game gives more equilibria. Every outcome on which players might agree in the presence of external enforcement is available as an equilibrium in the repeated game.

The argument runs as follows. Arbitrary choose a point P in the feasible set of the Trust Game. This is made an equilibrium outcome by punishing anyone who deviates from the strategy necessary for the two players to get P

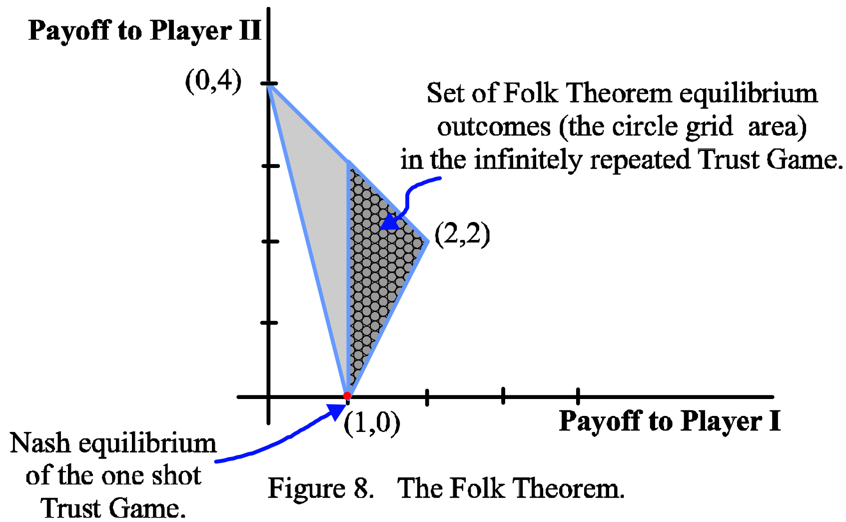

Figure 8:

each time the Trust Game is played. To simplify the exposition, the unrelenting punishment of the GRIM strategy is applied, in which any deviation is punished forever. In the diagram, the worst payoffs that each player can inflict (impose) on the other is indicated by the letter M. In the T.G., the worst that player I can do to player II is to play Not Trust (N). The worst that player II can do to player I is to play Abuse: $M = (1, 0)$.

The worst punishment that can be inflicted on player I is his minmax payoff. A player will never agree to a deal that pays off less than his maximin payoff, because he has a strategy that guarantees at least this much whatever the opponent may do. A pair of payoffs (and in general, a payoff vector) is called *individually rational*, if each player receives at least his minmax payoff, which is the payoff below which he cannot be forced by the remaining players. Since the minimax payoff is $(1, 0)$, the set of all individually rational payoff pairs is indicated by the circle grid region. Note that the first best outcome appears as the payoff to a Nash equilibrium in the $(\text{T.G.})^{\infty}$.

The Folk Theorem says that the set of payoff pairs associated with equilibria in the repeated game is indicated by the circle grid region. In Ken Binmore's words, the Folk theorem says that every contract on which players agree in the presence of external enforcement is available as an equilibrium outcome in an infinitely repeated game. The Folk theorem seems to give us the paradoxical result that because the two players, investors and managers, play the T.G. repeatedly, they do not need auditors to enforce equilibria. The outcomes of a game enforced by auditors are the same as the outcomes of the repeated game played by the two players.

The discussion of the Folk theorem suggests that investors *may* trust the managers, as it happens in the real world. To put it in a slightly different way, the theory of repeated games and the folk theorem in particular, can explain the ongoing cooperation between the two parties. But, we also get the seemingly paradoxical result that in an ongoing relationship, the two players do not need the auditors to achieve the first best outcome. How can we then justify theoretically the existence of auditors?

Two strong assumptions drive the result expressed by the Folk theorem: First, in the version of the Folk Theorem presented, the payoffs must be measured in utils that we can add. In this way, we evaluate both lotteries and income streams in terms of their average payoffs. The latter implies that the players are infinitely patient. This unrealistic assumption is relaxed in two ways: One way is to replace infinitely repeated games by idefinitely repeated games. The accounting - economic interpretation is more solid for indefinite repeated games. The manager, player II, knows that she will get fired, once she is caught to cook the books. A second way is to assume that the players discount the future at a fixed rate of interest. The next subsection pursues these two routes and shows that an approximate version of the folk theorem is valid for the case when both the interest rate and the probability that any repetition is the last are small.

The second assumption is more serious:

> "the version of the Folk theorem presented assumes that any deviations from equilibrium will be observed by the other players" Binmore (2005, p.82).

This assumption might hold in interactions amongst small group of players, but in the scenario of financial disclosure, the assumption of perfect monitoring is not realistic: investors cannot readily detect deviations by managers to deter them from cooking the books. As soon as we relax the assumption of perfect monitoring, we get the need for auditors. We undertake this task in section following the next.

3.4 Indefinitely Repeated Trust Game

When a game has no identifiable end, that is when N is infinite, we cannot add up payoffs because we run into problems. The problems are related to comparing infinite quantities, which makes no sense.[20]

To incorporate time-value considerations in the calculations of the players, we introduce the discount factor $\delta = \frac{1}{1+r}$; δ represents today's value of 1 dollar to be received one period later, where r is the interest rate. The players evaluate each sequence of outcomes in the repeated game by the discount sum of the associated sequence of payoffs. Taking account of the discount factor allows us to compute the present value of the infinite sequence of payoffs $\pi_1, \pi_2, ..., \pi_3, ...$ as

[20] A concise discussion of the problems can be found in Prajit Dutta (1999, p.228).

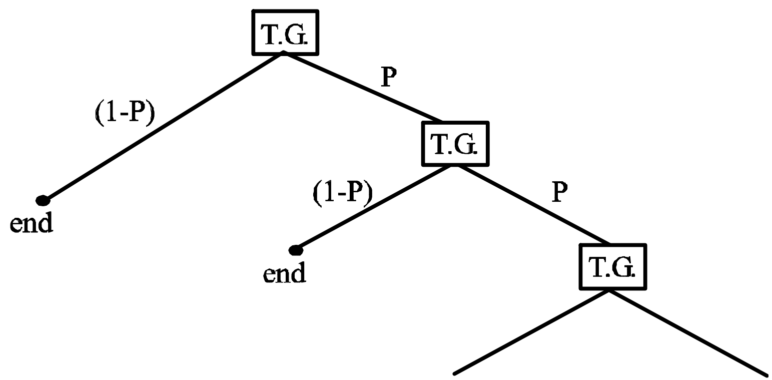

Figure 9.

Figure 9:

$$\pi_1 + \delta\pi_2 + \delta^2\pi_3 + ... = \sum_{t=1}^{\infty} \delta^{t-1}\pi_t$$

The common objection is to criticize the infinite repeated model as being unrealistic, since in real life, games among the same players, do not go on forever. But, people involved in long-run relationships, although they know that have finite lifes, they rarely know for certain the exact date of their last interaction.

The most popular method of modeling such uncertainty is by introducing a chance move after each stage-game of the play, that resolves whether or not the game will continue. In other words, we reinterpret an infinitely repeated game as a repeated game that ends after an unknown number of repetitions.

One way to incorporate the chance move is to suppose that after each stage is played a coin is tossed or a dice is rolled, to decide whether the game will end. Let the probability that the game continues at least one more time be p and $(1-p)$ the probability that the game ends. For example, a player may believe that there is a positive probability $(1-p)$ that the current interaction is the last one. In the case of a manager, this belief might come from his expectation that she may get fired, or the governmant may introduce new regulations. The payoff for the next period is an expected value estimated by multiplying it by the probability p. To fix ideas, we show this interpretation of the interest rate in the following schema:

The payoff, π, to be received in the next stage (if it takes place) is valued only

$$\frac{p*\pi}{(1+r)}$$

before the stage's random event occurs. Similarly, a payoff π, to be received two stages from now is valued only

$$\frac{p^2*\pi}{(1+r)^2}$$

before this stage random event happens. If

$$\delta = \frac{p}{(1+r)},$$

the present value

$$\pi_1 + \delta\pi_2 + \delta^2\pi_3 + ...$$

represents both the time-value of money and the probability that the game will end.

To be specific, consider the case in which the probability is $\frac{1}{5}$ that any stage reached is the last. That is, after each round of play there is a 0.2 chance that it was the last encounter and a 0.8 chance that they interact at least one more time, nomatter what happened in the past.

Since each player is uncertain about what payoffs she will get, say because of uncertainty of how long the game will last, the player tries to maximize the expected value or probability - weighted average of his summed payoff.

Repeating the game in the above way allows for many other outcomes as part of equilibria. For example, suppose that the two players adopt the following pair of strategies:

- Player I trusts player II in the first round and continues to trust II as long as II respects that trust. But, if II ever abuses I's trust, player I grimly refuses to offer trust ever again.
- Player II treats player I fairly in the first round and for as long as he has done so in the past. But if he ever - say, by mistake - abuses I, he will abuse her in all subsequent rounds, given the chance. Such a strategy is called grim-trigger strategy.

Following Kreps (2004), we consider the case when the probability that the game will continue from any stage to the next is always $\frac{4}{5}$. At the beginning of the game, the probability that the Nth stage will be reached is $(\frac{4}{5})^{N-1}$.

Consider the grim-trigger strategy described above. In this strategy, any deviation will be heavily punished. But, if both players use the grim strategy, there will never be a need for punishment. Computation shows that the pair of grim strategies constitutes a Nash equilibrium for the repeated game. The crux of the calculation is, will II honor I's trust? She can do so and continue to do so, accumulating payoffs of 1 in each round, for as long as the game lasts. This gives her an expected payoff of

$$E(\pi_{II}) = 1 + 1\left(\frac{4}{5}\right) + 1\left(\frac{4}{5}\right)^2 + ... = \frac{1}{0.2} = 5$$

Her other option is to abuse I's trust, getting an immediate payoff of 2. But then she would never be trusted again, getting 0 in all subsequent rounds, for an expected payoff of

$$E(\pi_{II}) = 2 + 0\left(\frac{4}{5}\right) + 0\left(\frac{4}{5}\right)^2 + ... = 2$$

A defection by player II triggers relentless, grim, defection, which can be viewed as retaliatory punishment. Since $5\rangle 2$ she prefers to stick with her part of the equilibrium profile. Player II is deterred from exploiting her short-term advantage by the threat of punishment, which reduces her long-term payoff.

Let us rephrase those two strategies:

- Player I will trust II in any round if II has a reputation of a trustworthy person. But I will not trust II if II's reputation is that she is untrustworthy.
- Player II lives up to her current reputation. She treats I fairly if her (II's) reputation is that of a trustworthy person. She behaves abusively if her reputation is that she is not trustworthy.

  In the words of Kreps : "II's reputation begins as a trustworthy person and stays that way as long as she never abuses I's trust. If she does abuse I, she gains a reputation for being untrustworthy, a reputation that can never shed" Kreps (2004, p.570).

The important point of the discussion is that the two players' strategies are described implicitly: the actions of I and II both depend on a mysterious property, II's reputation. Then, the rule by which II's reputation evolves is specified. In other words, II's reputation is a product of her past.

Finally, we mention a warning raised by Fudenberg (1992, p.90). Repeated game models explain how cooperation and trust might emerge, they do not predict that cooperation will occur. Folk theorems demonstrate that repeated games exhibit many equilibria; among the possible equilibria, there might exist one or few with desired properties. But, the folk theorems do not prove that the good equilibrium will occur. What then explains the ongoing relationship that often exists between stakeholders and managers.

Following Fudenberg (1992), we can classify the explanations for the emergence of certain long-run equlibria into two classes. The first is based on reputational effects; the second resorts to evolutionary arguments. In the remaining of the paper, we rely on reputation. The evolutionary explanation deserves more time and space, and is left for future work.

3.5 The Repeated Trust Game with Imperfect Detection

Up to now our discussion of repeated games assumed that each player observes the other player's action perfectly. In some real world environments this assumption is not realistic. In particular, the folk theorem requires that the players observe each other's behavior. This enables them to punish defection. But, what happens if players cannot observe the other players' actions, but observe only imperfect signals about the actions taken by the other players.

How an infinitely repeated game with discounting unfolds when players have information about others' play, but not perfect information. Strategic problems under such information structures are called *games of imperfect monitoring*. There are two different information structures that give rise to two categories of games with imperfect monitoring.

The first is games with *imperfect public monitoring*, in which players observe only a probabilistic signal about actions taken in each period; they observe the signal jointly, and thus all players know what their rivals have observed.

The second category is games with *private monitoring*, in which players observe possibly different signals and thus they might not know their rivals' signals. Both categories are the subject of current research. The outcome of recent research has established Folk theorem outcomes only in limited cases of imperfect monitoring.

4 The Subgame Perfect Folk Theorem: Auditing the Auditors

> "The equilibrium is held together by an infinite regress of threatened punishments, in which a player who does not punish a defector as he should is in turn punished by the defector for not punishing him. Thus a motorist stopped by the highway patrol may refrain from offering the patrolman a bribe for fear of being turned in by him; and the patrolman would probably indeed turn him in, for fear of being himself turned in by the motorist otherwise. Much of whatever stability society may possess is perhaps traceable to this kind of perfect equilibrium" Robert Aumann (1986, p.214).

Our proof and discussion of the folk theorem begged three questions: First, would the folk theorem still be correct when the Nash equilibrium is replaced by Subgame Perfect Equilibrium? Second, does the folk theorem applies to games played by more than two players? Third, does the folk theorem holds under imperfect monitoring of the players' behavior?

The question we pursue in this section is whether the Folk theorem is true not only for Nash equilibria, as we discussed in the previous section, but also for Subgame Perfect Nash Equilibria. A Subgame Perfect Equilibrium is a strategy profile that requires the play of Nash Equilibrium in all subgames, whether these are on the equilibrium path or not.

Binmore (2005, Chapter 4) points out that in modern game theory:

> "perfect equilibria are no more rational than other Nash equilibria".

But, the inventor of the notion of perfect equilibrium, Selten, showed that this parity is not valid "if we change the game slightly by assuming that there is always a chance that players will make a mistake with some small probability. Nash equilibria of this new game are then approximately perfect equilibria of the old game".

People, accountants included, tend to make mistakes. If one player deviates, and this leads to one of the subgames off the equilibrium path, it remains optimal in a perfect equilibrium to play your original strategy provided everybody else does: If the strategy of an auditor tells him to punish the deviant manager, for inaccurate reporting, at some cost to him, it is optimal for the auditor to execute the punishment. If the auditor deviates by escaping his duty to punish, it will take us to yet another subgame where it is optimal for some other player, say, a financial analyst, to punish the auditors for their dereliction of duty. If the financial analyst fails in his duty, we go to another subgame, and so on for ever. The financial disclosure game has a finite number of players and hence these chains of responsibility are by definition closed.

It is the *fear* on the part of each player that he will get punished, if he doesn't play his part of the equilibrium strategy profile, that sustains the perfect equilibrium.

Immanuel Kant's answer to the ancient question of who guards the guardians involved an infinite regress, but we can apply the Perfect Folk theorem to show that the chains of responsibility are bent back on each other (see Binmore, 2005, pp.85-86). A critic of the Perfect Folk Theorem might say that a spiral of self-confirming beliefs is too fragile to support equilibrium play. Beliefs may go round in a circle, but hte Folk theorem shows that behavior generated by the beliefs constitutes a perfect equilibrium.

Binmore (2005, p.86) warn us that we should not allow the institutional framework of the financial reporting game to blurr the insight that it all rests on the *self-confirming beliefs* of the participants. Regulatory laws and accounting - auditing standards are devices that help players coordinate on an equilibrium (see also Wilson, 1982, on this point). For example, a regulator is supposedly an instrument of the law but this does not exclude him from the financial reporting contract. Policing the regulator might seem inconceivable in the U.S., but in another country we had a recent episode, in which the stock exchange regulatory body allegeddly turned a blind eye to market manipulation (pumping-up the index) by state-owned companies, in a pre-election period.

Binmore (2005) believes that to coordinate efficiently, we must retain the power lend to the leaders in our collective hands. Because people with responsibility cannot be trusted not to abuse their privileges, a well designed financial reporting contract should make provision for the collective power to get on moving when corruption threatens the corporate governance system. Binmore (2005, p.86) concludes that the roles assigned to officers of the bodies that overlook the system must be compatible with their incentives: the bigger the guardians, the more they need to be guarded.

There are three players involved in the game of auditing: Investors, Managers, and Auditors (Public Certified Accountants). We model auditing as a

3-player game. We treat the auditor as a single player and not as an organization of different members. Brickley, Smith and Zimmerman (2004) treat auditing firms as multiagent organizations. In particular, they contribute the fall of Arthur Andersen to its deficient internal organization.

Auditors obey both accounting and auditing standards (See Demski, 2003). But these standards rarely cover and govern all possible contingencies for transactions. If disagreements arise, the players first attempt to resolve them by negotiation. In practice, this happens in the majority of cases. But, if a private settlement fails, then the law is available; this last option usually is identified with the end of an ongoing relationship (See Dixit, 2004, p.25). Dixit calls this empirical regularity, private ordering in the shadow of the law.

The introduction of a third player into a 2-player interaction, not only changes the relationship between the two, but gives rise to two new relationships and the associated agency costs that did not exist before. In a shareholder-auditor-manager relationship, the shareholder does not know what the auditor actually discovered or how thorough the audit was. Both managers and auditors have an information advantage over the shareholders. Also, managers have an information advantage over the auditors. In short, the market for auditing services is affected by strategic considerations, which are due to differences in information among the players (See Wilson, 1983). Accordingly, this 3-player game is a game of incomplete information.

The trilateral relationship must promise better prospects to each of the three players, compared to what they could get alone, or by forming smaller coalitions. Shareholders are protected against manager-auditor collusion by engaging large, reputable audit firms.[21]

The three players play this game repeatedly. As Demski (2003) points out, long-term relationships between accounting and client firms are common in practice. The repeated nature of the game is clear from the recent ISA 300 (Revised) issued by the International Federation of Accountants (see IFAC, 2004). The auditors are supposed to provide independent reports about the reliability of the financial statements produced by the management of the firm. As we discussed above, the question is whether the investors can rely on the auditors reports, or whether the auditors collude with management and verify misinformation? Put it in another way: How can this trilateral relationship be controled?

This game is a special case of the general question of who guards the guardians, in the context of the theme of what keeps society together. Ken Binmore (1992, and 1998) suggests that the answer is given by Game Theory's Subgame Perfect Folk Theorem. For auditing in particular, David Kreps (1990 and 1996) suggests that one can construct a perfect folk theorem for the trilateral relationship.

The supposed advantage of trilateral governance (bringing in an outside, independent expert) is that the third party does not have a direct stake on the business and, so, can be equitable. It is an ideal condition. Auditors might be

[21] The paper by Baiman, Evans and Nagarajan (1991) considers collusion, and the papers by Wilson (1983) and Datar and Alles (1996) examine the role of reputation.

corrupted by one side or the other, usually by managers, and so the best and most successful auditors are those that have a reputation stake in appearing to be fair and neutral. To put it in a slightly different way, since the auditors income depends on having other parties perceive that he is fair and neutral, he has a reputation stake in financial statements he is asked to verify.

5 Directions for Future Research

We believe that it is worth pursuing the following possibility, pointed out in Baron and Kreps (1999, ch.4), which as far as we know remains unexplored in the existing literature. Often the auditor is less well informed about the firm under examination than are either of the two parties directly involved. This is the case when the two parties are professional managers and institutional investors, who usually employ financial analysts that follow the operation of the firm closely. Under such conditions, the less-informed party, who has a reputation stake, has the final word.

In the games of the previous sections, the ability to observe what each player does in each round is total. There is neither noise in observables nor ambiquity. But, in the real world, both noise and ambiquity are present, and both can destroy reputation equilibria. We can employ a version of the Trust game and extend it to three players. The trusting party in the auditing game is the community of investors in the companies being audited. Investors must trust the auditor to put the hard work needed to unravel what is being going on at the firm being audited. If the auditing firm works hard and honestly, it is being fair. If it slacks off or shades its report because of, say, the consulting work it might get from the audited firm, it abuses the trust of the investing community. But even if the auditor tries to do the best audit that it can, it might miss something. If that something is uncovered, the auditing firm may appear to have abused the public's trust. In other words, investors would not be able to tell if the undiscovered facts were the result of abusive behaviour or honest error.

Kreps (2004) makes two observations, which we think deserve further research. First, public accounting firms protect their reputations to collect economic benefits, which come in the form of continuing audit engagements, based on a reputation for trustworthiness. In recent years, audit fees have decreased, as competition in the audit business has increased. This has lowered the value of good reputation, which means less incentive to behave. Second, an auditor that seems to have missed something defends itself, ex post, by showing that it followed standard auditing procedures. But, to verify, it requires that the standard auditing procedures have to be narrowly defined, with less room for subjective judgement by the auditor about what to do at a particular engagement.

Following the suggestion of David Levine (2001), one can attempt to answer the question of who audits the auditors with reference to the literature on random matching and information systems.

An empirical question: Do the auditors enhance the credibility of financial reports?

Healy and Palepou (2001) point out that although theory suggests that auditors enhance the credibility of financial reports, empirical research has provided little evidence to support this claim. Little is kown about why financial reporting and disclosure is regulated. Empirical work can address questions like: Is there a significant market imperfection or externality that regulation attempts to resolve? If so, how effective is disclosure regulation in resolving this problem?

6 Concluding Remarks: What Can We Gain from a Game Theoretic Approach

Financial scandals involving the manipulation of accounting information moved corporate governance and financial reporting into an inefficient state. Inefficient states are likely to be Nash equilibria of dynamic games that violate optimality by specifying irrational behavior at out of equilibrium decision nodes.

In the second part of the twenty century, economists have established three ways to pursue efficiency (See Michihiro Kandori, 2008). The first is *free market competition*, which is irrelevant to our study, because accounting information is not a standardized commodity. It is the non-homogeneity of accounting numbers the issue under study. Accordingly, we cannot invoke neither of the two fundamental theorems of General Equilibrium Theory.

The other two ways of achieving Pareto improvements work by providing incentives, rewards or penalties, with the aim to align individual incentives with common goals. Two routes can be used to provide incentives: One is by signing optimal contracts, and the other is through long-run relationships. Beginning with the seminal paper by Demski and Feltham (1978), *contract theory*, and to a lesser extent *mechanism design*, has been applied widely by accounting theorists to explore the second way of pursuing efficiency. Since it relies mainly on formal contracts, it is more appropriate for modeling problems of managerial accounting. The *theory of repeated games* explores the third method, namely *informal contracts*. The theory of repeated games has been the toolbox employed in the present study to investigate recent historical episodes and their lessons for the future.

We shaw that repeated game strategies can be interpreted as *explicit or implicit agreements* among the participants of the financial disclosure interaction to cooperate on an efficient outcome. The stability of such agreements is based on whether one or more of the participants has an incentive to deviate from their agreed part in the equilibrium strategy profile. The Perfect Folk Theorem is teaching us that for the players to stick to the equilibrium strategies, they have to be appropriately incentivized and must watch the behavior of the others. Each player must know that if he deviates, he will be caught and be punished accordingly. Only if the threat of punishment is credible or self-enforcing will the outcome of the financial disclosure game is likely to be efficient.

Credibility is formalized using the Subgame Perfect Equilibrium notion. Subgame Perfect Equilibrium refines Nash Equilibrium by imposing the sequential rationality requirement that the threat of punishment is credible if a deviation by a single player is never profitable. The practical message from this game theoretic prediction for regulators is: try to incentivize players in a way that makes unilateral deviations from the social welfare goal unprofitable. The regulators must introduce penalties such that the incentive constraint, penalty $\geq$ gain from deviation, is satisfied. In this way, repeated games internalize the benefits or costs of players' actions on their opponents.

The policy implication is that we cannot rely solely on federal-state law to govern the communication of accounting information; we need supporting institutions. The present work has stressed the role of social norms and punishments for contract enforcement. It has provided a framework, in which to conduct research and debate on how various institutions of governance work, how they interact with each other and with an imperfect state law.

The paper is a contribution in what Shyam Sunder (1997) calls the microtheory of accounting and control. It provides a theoretical framework in which to do theoretical and experimental research, as well as debate on these burning issues. Its contribution is to craft a balanced view between the agency theory view of ex ante incentive allignment and the Transaction Cost Economics emphasis on ex post governance. The message of the present study is that intertemporal incentives is what matters. This is where the regulatory authorities and participants with influence must concentrate; they have to get the balance right.

Another major point that comes out and should not be overlooked in light of the almost universal convergence towards the International Financial Accounting Standards is the institutional variations in the accounting - auditing standards across countries. Proposed regulatory policies based on institution-free arguments may produce adverse consequences. The accounting sector's instituttional structure should de taken seriously into consideration before policy reforms go into implementation.

Despite the insights we gain from the game theoretic treatment, we are left with three major unsolved issues: First, is the *multiplicity of equilibria* (i.e. lack of prediction). Second, is the *problem of renegotiation proofness*; and third, is the *lack of Folk theorem for Imperfect Private Monitoring*.

These are challenging problems and remain open for theorists to solve. Until we get reports on progress on these three fronts we must live with our inadequate understanding of the strategic problem of credible financial disclosure. Having said that, the debate that is going on among accounting scholars, economists, and practitioners about the causes of the Enron-Arthur Andersen etc failures and the effects of the regulatory responses, will enhance our understanding on how to reduce the likelihood of similar episodes. The paper has provided a game theoretic platform for the debate.

References

[1] Antle, R., and Garstka, S. (2004). *Financial Accounting*, 2nd edition, South Western - Thomson.

[2] Antle, R., and Nalebuff, B. (1991). "Conservatism and Auditor - Client Negotiations", *Journal of Accounting Research*, Autumn, vol. 29, Supplement, pp.31-54.

[3] Aoki, M. (2001) *Toward a Comparative Institutional Analysis*, MIT Press.

[4] Arya, A. and Glover, J. (1996). "Verification of Historical Cost Reports", *The Accounting Review*, April, pp.255-269.

[5] Aumann, R. (1986) "Repeated Games", in G. Feiwell, edition, *Issues in Contemporary Microeconomics and Welfare*. London: Macmillan, pp.209-242.

[6] Aumann, R., and Shapley, L. (1976) "Long-Term Competition; A Game-Theoretic Analysis", published in *Essays in Game Theory*, Essay 1, pp.1-15, edited by N. Megiddo (1994), Springer-Verlag.

[7] Baiman, S., Evans, J. and Nagarajan, N. (1991). "Collusion in auditing", *Journal of Accounting Research*, vol.29, Spring, pp.1-18.

[8] Baird, D., Gertner, R., and Picker, R. (1994). *Game Theory and the Law*, Harvard University Press.

[9] Baron, J. and Kreps, D. (1999). *Strategic Human Resources: Frameworks for General Managers*, Wiley.

[10] Basu, S., and Waymire, G. (2006). "Recordkeeping and Human Evolution", *Accounting Horizons*, Vol.20, No.3, pp.201-229.

[11] Bell, T., Peecher, M, and Solomon, I. (2005) *The 21st Century Public Company Audit,* KPMG International.

[12] Benston, G., Bromwich, M., Litan, R., and Wagenhofer, A. (2003). *Following the Money; The Enron Failure and the State of Corporate Disclosure*, AEI-Brooking Joint Center for Regulating Studies.

[13] Binmore, K. (1992). *Fun and Games; A Text on Game Theory*, D.C.Heath.

[14] Binmore, K. (1998). *Game Theory and the Social Contract; vol II: Just Playing*, MIT Press.

[15] Binmore, K. (2005). *Natural Justice*, Oxford University Press.

[16] Binmore, K. (2007). *Playing for Real; A Text on Game Theory*, Oxford University Press.

[17] Binmore, K. (2010). "Game Theory and Institutions", *Journal of Comparative Economics,vol.38, issue 3, Sept.,pp.245-252.*

[18] Bowles, S. (2004). *Microeconomics; Behavior, Institutions, and Evolution*, Princeton University Press.

[19] Brickley, J., Smith, Jr. C., and Zimmerman, J. (2004). *Managerial Economics and Organizational Architecture*, 3rd edition, McGraw-Hill / Irwin.

[20] Carmichael, D. (2004). "The Public Company Accounting Oversight Board (PCAOB) and the Social Responsibility of the Independent Auditor", *Accounting Horizons*, vol.18, no.2, pp.127-133.

[21] Coates IV, J.C. (2007). "The Goals and Promise of the Sarbanes - Oxley Act", *Journal of Economic Perspectives*, Vol.21, No.1, pp.91-116.

[22] Christensen, J., and Demski, J. (2003). *Accounting Theory; An Information Content Perspective*, McGraw - Hill.

[23] Dasgupta, P. (2007). *Economics; A Very Short introduction*, Oxford University Press.

[24] Datar, S., and Alles, M. (199). "The Formation and Role of Reputation and Litigation in the Auditor-Manager Relationship", *Journal of Accounting, Auditing & Finance*, pp.401-428.

[25] Datar, S., Feltham, G., and Hughes, J. (1991). "The role of audits and audit quality in valuing new issues", *Journal of Accounting and Economics*, vol.14, pp.3-49.

[26] Datar, S., Kulp, S., and Lambert, R. (2001). "Balancing Performance Measures", *Journal of Accounting Research*, vol.39, 75-92.

[27] Demski, J. (2003). "Corporate Conflicts of Interest", *Journal of Economic Perspectives*, Vol.17, No.2, Spring, pp.51-72.

[28] Dixit, A. (2004) *Lawlessness and Economics; Alternative Modes of Governance*, Princeton University Press.

[29] Dutta, P. (1999). *Strategies and Games*, MIT Press.

[30] Fellingham, J., and Neuman, P., (1985). "Strategic Considerations in Auditing", *The Accounting Review*, October, pp.634-650.

[31] Feltham, G., and Xie, J. (1994). "Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations", *The Accounting Review*, vol.69, pp.429-453.

[32] Fudenberg, D. (1992) "Explaining cooperation and commitment in repeated games", Chapter 3 in *Advances in Economic Theory-Sixth World Congress*, Vol.1, edited by J-J Laffont, Cambridge University Press.

[33] Fudenberg, D. and Maskin, E. (1986) "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information", *Econometrica*, 54, pp.533-554.

[34] Fudenberg, D. and Tirole, J. (1991) *Game Theory*, MIT Press.

[35] Gardner, R. (1995). Games for Business and Economics, Wiley.

[36] Geiger, A. M. and Taylor III, L. P. (2003). "CEO and CFO Certifications of Financial Information", *Accounting Horizons*, Vol.17, No.4, pp.357-368.

[37] Ghemawat, P. (1997). *Games Businesses Play; Cases and Models*, The MIT Press.

[38] Gibbons, R. (1992). *A Primer in Game Theory*, Harvester Wheatsheaf. Note: It is the non-American edition of Game Theory for Applied Economists.

[39] Gibbons, R. (1997). "An Introduction to Applicable Game Theory", *Journal of Economic Perspectives*, vol.11, no.1, pp.127-149.

[40] Gibbons, R. (2003). "Relational Contracts", mimeo MIT.

[41] Hart, S. (2005). "Robert Aumann's Game and Economic Theory", Dec. 9 2005, Stockholm School of Economics.

[42] Healy, P., and Palepou, K. (2001). "Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature", *Journal of Accounting and Economics*, vol.31, pp.405-440.

[43] Healy, P., and Palepou, K. (2003). "The Fall of Enron", *Journal of Economic Perspectives*, Vol.17, No.2, Spring, pp.3-26.

[44] Holmstrom, B., and Milgrom, P. (1991). "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics, and Organization*, 7, pp.24-52.

[45] Jehle, G. and Reny, P. (2001). *Advanced Microeconomic Theory*, 2nd edition, Addison Welsey Longman.

[46] International Auditing and Assurance Standards Board (2004). "Planning an Audit of Financial Statements", International Standard on Auditing (ISA) 300 (Revised), International Federation of Accountants (IFAC).

[47] Kandori, M. (2008). "Repeated Games", Entry in *The New Palgrave Dictionary of Economics*, 2nd edition, edited by Blume and Duerlof.

[48] Kreps, D. (1986). "Corporate Culture and Economic Theory", Reprinted in the volume edited by Buckley, P. and Michie, J. (1996) *Firms, Organizations and Contracts*, Oxford.

[49] Kreps, D. (1990a). *A Course in Microeconomic Theory*, Princeton University Press.

[50] Kreps, D. (1990b). *Game Theory and Economic Modeling*, Oxford University Press.

[51] Kreps, D. (1996). "Markets and Hierarchies and (Mathematical) Economic Theory", *Industrial and Corporate Change*, vol.5, no.2, pp.561-595.

[52] Kreps, D. (2004). *Microeconomics for Managers*, Norton.

[53] McAfee, P. (2004) "The Real Lesson of Enron's Implosion: Market Makers Are In the Trust Business", *The Economists' Voice*, vol.1, issue 2, article 4.

[54] McAfee, P., Vakkur, N., and Kipperman, F. (2008). "The Real Costs of Sarbanes Oxley", Mimeo, 1/14/2008.

[55] McLennan, A. and Park, I-U. (2003). "The Market for Liars: Reputation and Auditor Honesty", The Institute of Social and Economic Research, Osaka University, Discussion Paper No. 587, June.

[56] Osborne, M. (2004) *An Introduction to Game Theory*, Oxford University Press.

[57] Oxley, M.G. (2007) "The Sarbanes - Oxley Act of 2002 - Restoring Investor Confidence", *Current Issues in Auditing*, Vol.One, pp.C1-C2.

[58] Palepou, K., Healy, P., and Bernard, V. (2004) *Business Analysis and Valuation*, Third edition, Thomson / South-Western.

[59] Pearce, D. (1992) "Repeated games: cooperation and rationality" Chapter 4 in *Advances in Economic Theory-Sixth World Congress*, Vol.1, edited by J-J Laffont, Cambridge University Press.

[60] Rasmusen, E. (1994) *Games and Information*, 2nd edition, Blackwell.

[61] Rasmusen, E. (2007) *Games and Information*, 4rth edition, Blackwell.

[62] Salanie, B. (2005). *The Theory of Contracts*, 2nd edition, MIT Press.

[63] Samuelson, L. (2006). "The Economics of Relationships". Chapter 4, pp.136-185, in Advances in Economics and Econometrics, edited by Blundell, R., Newey, W., and Persson, T., Cambridge University Press.

[64] Sunder, S. (1997). *Theory of Accounting and Control*, South-Western College Publishing.

[65] Sunder, S. (2003). "Rethinking the Structure of Accounting and Auditing", *Indian Accounting Review*, vol.7, no.1, June, pp.1-15.

[66] Sunder, S. (2004) "Competition in the Market for Audit Services and Corporate Governance", *DowJones*, August 4.

[67] Sunder, S. (2005) "Minding our manners: Accounting as social norms", *The British Accounting Review*, vol.37, pp.367-387.

[68] Sunder, S. (2006). "President's Message: Imagined Worlds of Accounting" *Accounting Education News*, Fall Issue, p.3 and p.6.

[69] Tirole, J. (2006). *The Theory of Corporate Finance*, Princeton University Press.

[70] Weigelt, K. and Camerer, C. (1988). "Reputation and Corporate Strategy: A Review of Recent Theory and Applications", *Strategic Management Journal*, Vol. 9, pp.443-454.

[71] Williamson, O. (1979). "Transaction-Cost Economics: The Governance of Contractual Relations", *Journal of Law and Economics*, vol.22, no.2, October, pp.233-261.

[72] Williamson, O. (2000). "The New Institutional Economics: Taking Stock, Looking Ahead", *Journal of Economic Literature*, Vol.XXXVIII, September, pp.595-613.

[73] Wilson, R. (1983). "Auditing: Perspectives from Multi-Person Decision Theory", *The Accounting Review*, April, pp.305-318.

CHAPTER 5

The fundamental properties of time varying AR models with non stochastic coefficients

M. Karanasos†, A. G. Paraskevopoulos and S. Dafnos
Brunel University, London, UK

Abstract

The paper examines the problem of representing the dynamics of low order autoregressive (AR) models with time varying (TV) coefficients. The existing literature computes the forecasts of the series from a recursion relation. Instead, we provide the linearly independent solutions to TV-AR models. Our solution formulas enable us to derive the fundamental properties of these processes, and obtain explicit expressions for the optimal predictors. We illustrate our methodology and results with a few classic examples amenable to time varying treatment, e.g, periodic, cyclical, and AR models subject to multiple structural breaks.

Keywords: abrupt breaks, covariance structure, cyclical processes, homogeneous and particular solutions, optimal predictors, periodic AR models.

We gratefully acknowledge the helpful conversations we had with L. Giraitis, G. Kapetanios and A. Magdalinos in the preparation of the paper. We would also like to thank R. Baillie, L. Bauwens, M. Brennan, D. van Dijk, W. Distaso, C. Francq, C. Gourieroux, E. Guerre, M. Guidolin, A. Harvey, C. Hommes, S. Leybourne, P. Minford, A. Monfort, C. Robotti, W. Semmler, R. Smith, T. Teräsvirta, P. Zaffaroni, and J-M Zakoian for suggestions and comments on a closely related work (see Paraskevopoulos et al., 2013) which greatly improved many aspects of the current paper as well. We are grateful to seminar participants at CREST, Erasmus University, London School of Economics, Queen Mary University of London, Imperial College, University of Essex, Birkbeck College University of London, University of Nottingham, Cardiff University, University of Manchester, Athens University of Economics and Business, and University of Piraeus. We have also benefited from the comments given by participants (on the closely related work) at the 3rd Humboldt-Copenhagen Conference on Financial Econometrics (Humboldt University, Berlin, March 2013), the SNDE 21st Annual Symposium (University of Milan-Bicocca, March 2013), the 8th and 9th BMRC-QASS conferences on Macro and Financial Economics (Brunel University, London, May 2013), the 7th CFE Conference (Senate House, University of London, December 2013), and the 1st RASTANEWS Conference (University of Milan-Bicocca, January 2014).

†Address for correspondence: Menelaos Karanasos, Economics and Finance, Brunel University, West London, UB3 3PH, UK; email: menelaos.karanasos@brunel.ac.uk, tel: +44(0)1895265284, fax: +44 (0)1895269770.

1 Introduction

The constancy of the parameters assumption made in the specification of time series econometric models has been the subject of criticism for a long time. It is argued that the assumption is inappropriate in the face of changing institutions and a dynamically responding economic policy. These evolving factors cause the parameter values characterizing economic relationships to change over time. Partly to respond to the criticism and partly motivated by the desire to construct dynamic models, econometricians have developed an arsenal of powerful methods that attempt to capture the evolving nature of our economy. Such frameworks include AR processes which contain multiple abrupt breaks, and periodic and cyclical autoregressive models.

A methodology is presented in this paper for analyzing time varying systems which is also applicable to the three aforementioned processes. A technique is set forth for examining the periodic AR model, which overcomes the usual requirement of expressing the periodic process in a vector AR (VAR) form.

The first attempts to develop theories for time varying models, made in the 1960's, were based on a recursive approach (Whittle, 1965) and on evolutionary spectral representations (Abdrabbo and Priestley, 1967). Rao (1970) used the method of weighted least squares to estimate an autoregressive model with time dependent coefficients. Despite nearly half a century of research work, the great advances, and the widely recognized importance of time varying structures, the bulk of econometric models have constant coefficients. There is a lack of a general theory that can be employed to systematically explore their time series properties. Granger in some of his last contributions highlighted the importance of the topic (see, Granger 2007, and 2008).

There is a general agreement that the main obstacle to progress is the lack of a universally applicable method yielding a closed form solution to stochastic time varying difference equations. The present paper is part of a research program aiming to produce and utilize closed form solutions to AR processes with non stochastic time dependent coefficients. Our methodology attempts to trace the path of these changing coefficients. To be specific, in the time series literature, there is no method for finding the p linearly independent solutions that we need to obtain the general solution of the TV-AR model of order p. To keep the exposition tractable and reveal its practical significance we work with low order specifications.

The main part of the paper begins with subsection 2.2, where we state the second order difference equation with time variable coefficients, which is our main object of inquiry. We start by writing this equation in a more efficient way as an infinite linear system. The next step is to define the matrix of coefficients, called the fundamental solution matrix, associated with the system representation. This matrix is a workhorse of our research and it is derived step by step from the time varying coefficients of the difference equation.

The reader will have noticed that we have moved the goalposts, paradoxically against us, from obtaining a solution for a time varying (low order) difference equation, to solving an infinite linear system. The reason is that the solution of such infinite systems has been made possible recently, due to an extension of the standard Gaussian elimination, called the infinite Gaussian elimination (see Paraskevopoulos, 2012). Applying this infinite extension algorithm, we obtain the fundamental solutions, which take explicit forms in terms of the determinants of the fundamental solution matrix.

Subsection 2.3 contains the main theoretical result of the paper. Pursuing the conventional route followed by the differential and difference equations literature, we construct the general solution by finding its two parts, the homogeneous one and a particular part. It is expressed as Theorem 1 and its proof is in Appendix A. The coefficients in these solutions are expressed as determinants of tridiagonal matrices. The second order properties of the TV-AR process can easily be deduced from these solutions. An additional benefit of these solutions is the facility with which linear prediction can be produced. This allows us to provide a thorough description of time varying models by deriving: first, multistep ahead forecasts, the associated forecast error and the mean square error; second, the first two unconditional moments of the process and its covariance structure. In related works we provide results for the p order and the more general ascending order (see, for example, Paraskevopoulos et al., 2013). Our method is a natural extension of the first order solution formula. It also includes the linear difference equation with constant coefficients (see, for example, Karanasos, 2001) as a special case.

The next two Sections of the paper, 3 and 4, apply our theoretical framework to a few classic time series models, which are obvious candidates for a time varying treatment. Linear systems with time dependent coefficients are not only of interest in their own right, but, because of their connection with periodic models and time series data which are subject to structural breaks. They also provide insight into these processes as well. Viewing a periodic AR (PAR) formulation as a TV model clearly obviates the need for VAR analysis. For surveys and a review of some important aspects of PAR processes see Franses (1996b), Franses and Paap (2004), Ghysels and Osborn (2001), and Hurd and Miamee (2007). The authoritative studies by Osborn (1988), Birchenhall et al. (1989), and Osborn and Smith (1989) applied these models to consumption. del Barrio Castro and Osborn (2008) pointed out that "*despite the attraction of PAR models from the perspective of economic decision making in a seasonal context, the more prominent approach of empirical workers is to assume that the AR coefficients, except for the intercept, are constant over the seasons of the year*".[1]

[1] del Barrio Castro and Osborn (2008, 2012) (see the references therein for this stream of important research; see also

Despite the recognized importance of periodic processes for economics there have been few attempts to investigate their time series properties (see, among others, Franses, 1994, Franses, 1996a, Lund and Basawa, 2000, Franses and Paap, 2005). Tiao and Grupe (1980) and Osborn (1991) analyzed these models by converting them into a VAR process with constant coefficients. In this paper we develop a general theory that can be employed to systematically explore the fundamental properties of the periodic formulation. We remain within the univariate framework and we look upon the PAR model as a stochastic difference equation with time varying (albeit periodically varying) parameters.

Although some theoretical analysis of periodic specifications was carried out by the aforementioned studies the investigation of their fundamental properties appears to have been limited to date. Cipra and Tlustý (1987), Anderson and Vecchia (1993), Adams and Goodwin (1995), Shao (2008), and Tesfaye et al. (2011) discuss parameter estimation and asymptotic properties of periodic AR moving average (PARMA) specifications. Bentarzi and Hallin (1994) and McLeod (1994) derive invertibility conditions and diagnostic checks for such processes. Lund and Basawa (2000) develop a recursive scheme for computing one-step ahead predictors for PARMA specifications, and compute multi-step-ahead predictors recursively from the one-step-ahead predictions. Anderson et al. (2013) develop a recursive forecasting algorithm for periodic models. We derive explicit formulas that allow the analytic calculation of the multi-step-ahead predictors.

We begin Section 3 with a PAR(2) model. We limit our analysis to a low order to save space and also since Franses (1996a) has documented that low order PAR specifications often emerge in practice. First, we formulate it as a TV model; then, we express its fundamental solution matrix as a block Toeplitz matrix. This representation enables us to establish an explicit formula for the general solution in terms of the determinant of such a block matrix. The result is presented in Proposition 1, which is the equivalent to Theorem 1 with the incorporation of the seasonal effects. That is, by taking account of seasons and periodicities, we obtain the general solution, by constructing its homogeneous and particular parts and then adding them up. In subsection 3.1, we turn our attention to a different type of seasonality, namely the cyclical AR (CAR) model and we provide its solution.

Section 4 is an application of the time varying framework to time series subject to multiple structural breaks. We employ a technique analogous to the one used in Section 3 on the PAR formulation. In particular, we express the fundamental solutions of the AR(2) model with r abrupt breaks, as determinants of block tridiagonal matrices. Again, we are able to obtain the general solution by finding and adding the homogeneous and particular solutions.

One of the advantages of our time varying framework is that we can trace the entire path of the

Taylor, 2002, 2003 and 2005) test for seasonal unit roots in integrated PAR models.

series under consideration. In Section 5, we employ this information feature to derive the fundamental properties of the various TV-AR processes. For example, simplified closed-form expressions of the multi-step forecast error variances are derived for time series when low order PAR models adequately describe the data. These formulae allow a fast computation of the multi-step-ahead predictors. Finally, Section 6 concludes.

2 Time Varying AR Models

2.1 Preliminaries and Purpose of Analysis

2.1.1 Notation

Throughout the paper we adhere to the following conventions: ($\mathbb{Z}^+$) $\mathbb{Z}$, and ($\mathbb{R}^+$) $\mathbb{R}$ stand for the sets of (positive) integers, and (positive) real numbers, respectively. Matrices and vectors are denoted by upper and lower case boldface symbols, respectively. For square matrices $\mathbf{X} = [x_{ij}]_{i,j=1,\ldots,k} \in \mathbb{R}^{kxk}$ using standard notation, $\det(\mathbf{X})$ or $|\mathbf{X}|$ denotes the determinant of matrix $\mathbf{X}$ and $adj(\mathbf{X})$ its adjoint matrix. To simplify our exposition we also introduce the following notation: $t \in \mathbb{Z}$, $(n, l) \in \mathbb{Z}^+ \times \mathbb{Z}^+$; $T = 0, \ldots, n$ denotes the 'periods' (i.e., years); $s = 1, \ldots, l$, denotes the 'seasons' (i.e, quarters in a year: $l = 4$). The t represents the present time and $k \in \mathbb{Z}^+$ the number of seasons such that at time $\tau_k = t - k$ information is given.

Let the triple $(\Omega, \{\mathcal{F}_t, t \in \mathbb{Z}\}, P)$ denote a complete probability space with a filtration, $\{\mathcal{F}_t\}$, which is a non-decreasing sequence of σ-fields $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t \subseteq \mathcal{F}$, $t \in \mathbb{Z}$. The space of P-equivalence classes of finite complex random variables with finite p-order is indicated by L_p. Finally, $H = L_2(\Omega, \mathcal{F}_t, P)$ stands for a Hilbert space of random variables with finite first and second moments.

2.1.2 The Problem

The solution of the second order linear difference equation with non constant coefficients is the building block for the extension of the well known closed form solution of the first order to the pth order time varying equation. As noted by Sydsaeter et al. (2008), in their classic text (Further Mathematics for Economic Analysis, p. 403), in the case of second order homogeneous linear difference equations with variable coefficients:

"*There is no universally applicable method of discovering the two linearly independent solutions that we need in order to find the general solution of the equation.*"

We can identify two lines of inquiry that can be pursued to solve linear difference equations with time varying coefficients. Searching for a solution, one can follow either of the following two paths. The first is to develop an analogous method to the standard one that exists for the linear p order difference equation with constant coefficients: find the eigenvalues, solve the characteristic equation, and obtain the closed form. The second line of research searches for the generalization of the closed form formula that exists for first order time varying difference equations. Here, the way to proceed is to make up a conjecture and try to prove it by induction. The two strands of the literature have taken important steps, but have not provided us with a general solution method that we can apply; the existing results lack generality and applicability. To be more specific, the research problem we face is that there is a lack of a universally applicable method yielding a closed form solution to stochastic higher order difference equations with time dependent coefficients.

A general method for solving infinite linear systems with row-finite coefficient matrices has recently been established by Paraskevopoulos (2012). It is a modified version of the standard Gauss-Jordan elimination method implemented under a right pivot strategy, called infinite Gauss-Jordan elimination. Expressing the linear difference equation of second order with time dependent coefficients as an infinite linear system, the Gaussian elimination part of the method is directly applicable. It generates two linearly independent homogeneous solution sequences. The general term of each solution sequence turns out to be a continuant determinant. The general solutions of the homogeneous and nonhomogeneous difference equation are expressible as a single Hessenbergian, that is, a determinant of a lower Hessenberg matrix (see Karanasos, Paraskevopoulos and Dafnos 2013). Theorem 3 in Paraskevopoulos et al. (2013) affords an easy means of finding, for a given lower Hessenberg matrix, its ordinary expansion in non-determinant form (see also Paraskevopoulos and Karanasos, 2013). These results are extendible to the solution of the pth and ascending order time varying linear difference equations in terms of a single Hessenbergian (see Paraskevopoulos et al., 2013). This makes it possible to introduce, in the above cited reference, a unified theory for time varying models.

2.2 Fundamental Solution Matrices

The main theoretical contribution of this Section is the development of a method that provides the closed form of the general solution to a TV-AR(2) model.

Next we give the main definition that we will use in the rest of the paper. Consider a second order stochastic difference equation with time dependent coefficients, which is equivalent to the time varying AR(2) process, given by

$$y_t = \phi_0(t) + \phi_1(t)y_{t-1} + \phi_2(t)y_{t-2} + \varepsilon_t, \tag{1}$$

where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of zero mean serially uncorrelated random variables defined on $L_2(\Omega, \mathcal{F}_t, P)$ with $\mathbb{E}[\varepsilon_t \,|\mathcal{F}_{t-1}] = 0$ a.s., and finite variance: $0 < M_l < \sigma_t^2 < M < \infty$, $\forall\ t$, $(M_l, M) \in \mathbb{R}^+ \times \mathbb{R}^+$.

Remark 1 *We have relaxed the assumption of homoscedasticity (see also, among others, Paraskevopoulos et al., 2013 and Karanasos et al., 2013), which is likely to be violated in practice and allow ε_t to follow, for example, a periodical GARCH type of process (see, Bollerslev and Ghysels, 1996).*

The fundamental solution sequence, and in general all the solution sequences, must necessarily be functions of the independent variable t, so as to satisfy eq. (1). Our intermediate objective is to obtain the fundamental solution matrix, denoted below by $\mathbf{\Phi}_{t,k}$, which is associated with our stochastic difference equation (1); the $\mathbf{\Phi}_{t,k}$ matrix will be derived from the time varying coefficients of eq. (1). The best way to appreciate the representation of the fundamental solution matrix is to view the stochastic difference equation as a linear system. We carry out this construction below. Once we have this stepping stone in place, then we can pursue our ultimate objective, by computing the determinants of the $\mathbf{\Phi}_{t,k}$, which will give us the linearly independent solutions sequences to the difference equation.

Equation (1) written as

$$\phi_2(t)y_{t-2} + \phi_1(t)y_{t-1} - y_t = -[\phi_0(t) + \varepsilon_t], \tag{2}$$

takes the infinite row (and column)-finite system form

$$\mathbf{\Phi} \cdot \mathbf{y} = -\boldsymbol{\phi} - \boldsymbol{\varepsilon}, \tag{3}$$

where

$$\mathbf{\Phi} = \begin{pmatrix} \phi_2(\tau_k+1) & \phi_1(\tau_k+1) & -1 & 0 & 0 & 0 & \ldots \\ 0 & \phi_2(\tau_k+2) & \phi_1(\tau_k+2) & -1 & 0 & 0 & \ldots \\ 0 & 0 & \phi_2(\tau_k+3) & \phi_1(\tau_k+3) & -1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\vdots\vdots \end{pmatrix},$$

(row-finite is an infinite matrix whose rows have finite non zero elements) and

$$\mathbf{y} = \begin{pmatrix} y_{\tau_k-1} \\ y_{\tau_k} \\ y_{\tau_k+1} \\ y_{\tau_n+2} \\ y_{\tau_k+3} \\ y_{\tau_k+4} \\ \vdots \end{pmatrix}, \ \boldsymbol{\phi}= \begin{pmatrix} \phi_0(\tau_k+1) \\ \phi_0(\tau_k+2) \\ \phi_0(\tau_k+3) \\ \vdots \end{pmatrix}, \ \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{\tau_k+1} \\ \varepsilon_{\tau_k+2} \\ \varepsilon_{\tau_k+3} \\ \vdots \end{pmatrix}$$

(recall that $\tau_k = t - k$). The system representation results from the values that the coefficients take in successive time periods. The equivalence of (2) and (3) follows from the fact that the ith equation in (3), as a result of the multiplication of the ith row of $\mathbf{\Phi}$ by the column of ys equated to $-[\phi_0(\tau_k+i)+\varepsilon_{\tau_k+i}]$, is equivalent to eq. (2), as of time τ_k+i. The $\mathbf{\Phi}$ matrix in eq. (3) can be partitioned as

$$\mathbf{\Phi} = \left(\begin{array}{c|c} \mathbf{P} & \mathbf{C} \end{array} \right),$$

where

$$\mathbf{P} = \begin{pmatrix} \phi_2(\tau_k+1) & \phi_1(\tau_k+1) \\ 0 & \phi_2(\tau_k+2) \\ 0 & 0 \\ \vdots & \vdots \end{pmatrix}, \ \mathbf{C} = \begin{pmatrix} -1 & 0 & 0 & 0 & \ldots \\ \phi_1(\tau_k+2) & -1 & 0 & 0 & \ldots \\ \phi_2(\tau_k+3) & \phi_1(\tau_k+3) & -1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

That is, $\mathbf{P}$ consists of the first 2 columns of $\mathbf{\Phi}$ and the jth column of $\mathbf{C}$, $j = 1, 2, \ldots$, is the $(2+j)$th column of $\mathbf{\Phi}$. We will denote the 2nd column of the $k \times 2$ top submatrix of the matrix $\mathbf{P}$ by $\boldsymbol{\phi}_{t,k}$:

$$(\boldsymbol{\phi}_{t,k})' = \begin{pmatrix} \phi_1(\tau_k{+}1), & \phi_2(\tau_k{+}2), & 0, & \ldots & ,0 \end{pmatrix}.$$

The $k \times (k-1)$ top submatrix of matrix $\mathbf{C}$ is called the core solution matrix and is denoted as

$$\mathbf{C}_{t,k} = \begin{pmatrix} -1 & & & & \\ \phi_1(\tau_k+2) & -1 & & & \\ \phi_2(\tau_k+3) & \phi_1(\tau_k+3) & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & \phi_2(t-1) & \phi_1(t-1) & -1 \\ & & & \phi_2(t) & \phi_1(t) \end{pmatrix} \tag{4}$$

(here and in what follows empty spaces in a matrix have to be replaced by zeros). The fundamental solution matrix is obtained from the core solution matrix $\mathbf{C}_{t,k}$, augmented on the left by the $\boldsymbol{\phi}_{t,k}$ column. That is,

$$\begin{aligned} \boldsymbol{\Phi}_{t,k} &= \left(\boldsymbol{\phi}_{t,k} \,\middle|\, \mathbf{C}_{t,k} \right) = \\ &\begin{pmatrix} \phi_1(\tau_k+1) & -1 & & & & \\ \phi_2(\tau_k+2) & \phi_1(\tau_k+2) & -1 & & & \\ & \phi_2(\tau_k+3) & \phi_1(\tau_k+3) & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \phi_2(t-1) & \phi_1(t-1) & -1 \\ & & & & \phi_2(t) & \phi_1(t) \end{pmatrix}, \end{aligned} \tag{5}$$

(recall that $\tau_k = t-k$). Formally $\boldsymbol{\Phi}_{t,k}$ is a square $k \times k$ matrix whose (i,j) entry $1 \le i,j \le k$ is given by

$$\begin{cases} -1 & \text{if} \quad i = j-1, \quad \text{and} \quad 2 \le j \le k, \\ \phi_{1+m}(t-k+i) & \text{if} \quad m = 0,1, \quad i = j+m, \quad \text{and} \quad 1 \le j \le k-m, \\ 0 & \text{otherwise.} \end{cases}$$

It is a continuant or tridiagonal matrix, that is a matrix that is both an upper and lower Hessenberg matrix. We may also characterize it as a 'time varying' Toeplitz matrix, because its time invariant version is a Toeplitz matrix of bandwidth 3. We next define the bivariate function $\xi : \mathbb{Z} \times \mathbb{Z}^+ \longmapsto \mathbb{R}$ by

$$\xi_{t,k} = \det(\boldsymbol{\Phi}_{t,k}) \tag{6}$$

coupled with the initial values $\xi_{t,0} = 1$, and $\xi_{t,-1} = 0$. That is, $\xi_{t,k}$ for $k \ge 2$, is a determinant of a $k \times k$

matrix; each of the two nonzero diagonals (below the superdiagonal) of this matrix consists of the time varying coefficients $\phi_m(\cdot)$, $m = 1, 2$, from $t - k + m$ to t. In other words, $\xi_{t,k}$ is a kth-order tridiagonal determinant. Paraskevopoulos and Karanasos (2013) give its ordinary expansion in non-determinant form (a closed form solution).

2.3 Main Theorem

This short section contains the statement of our main theorem.

Theorem 1 *The general solution of eq. (1) with free constants (initial condition values)* y_{t-k}, y_{t-k-1} *is given by*

$$y_{t,k}^{gen} = y_{t,k}^{hom} + y_{t,k}^{par}, \tag{7}$$

where

$$\begin{aligned} y_{t,k}^{hom} &= \xi_{t,k} y_{t-k} + \phi_2(t-k+1)\xi_{t,k-1} y_{t-k-1}, \\ y_{t,k}^{par} &= \sum_{i=0}^{k-1} \xi_{t,i}\phi_0(t-i) + \sum_{i=0}^{k-1} \xi_{t,i}\varepsilon_{t-i}. \end{aligned}$$

In the above Theorem $y_{t,k}^{gen}$ is decomposed into two parts: first, the $y_{t,k}^{hom}$ part, which is written in terms of the two free constants (y_{t-k-m}, $m = 0, 1$), and, second, the $y_{t,k}^{par}$ part, which contains the time varying drift terms and the error terms from time $t - k + 1$ to time t.

Notice that the 'coefficients' of eq. (7), that is, the ξ's are expressed as continuant determinants. Moreover, for '$k = 0$' (for $i > j$ we use the convention $\sum_{q=i}^{j}(\cdot) = 0$), since $\xi_{t,0} = 1$ and $\xi_{t,-1} = 0$ (see eq. (6)), eq. (7) becomes an 'identity': $y_{t,0}^{gen} = y_t$. Similarly, when '$k = 1$' eq. (7), since $\xi_{t,1} = \phi_1(t)$ and $\xi_{t,0} = 1$, reduces to $y_{t,1}^{gen} = \phi_1(t)y_{t-1} + \phi_2(t)y_{t-2} + \phi_0(t) + \epsilon_t$.

In the next Section, we illustrate the above claims in the context of a simple seasonal process with fixed periodicity, and a cyclical model as well.

3 Periodic AR(2) Model

Periodic regularities are phenomena occurring at the same season every year, so analogous to each other that we can view them as recurrences of the same event. Many economic time series are periodic in this sense. In the present Section we express them in a mathematical model, so that we can then employ it for forecasting and control. Gladyshev (1961) introduced a technique which still dominates the literature. He begins by decomposing the series into subperiods; then he treats each point within a subperiod as one part of a multivariate process. In this way he transforms a univariate non-stationary formulation into a multivariate stationary one. Following Gladyshev, Tiao and Grupe (1980) and Osborn (1991) treated periodic autoregressions as conventional nonperiodic VAR processes. But, as pointed out by Lund et al. (2006), even low order specifications can have an inordinately large numbers of parameters. A PAR(1) model for daily data, for example, has 365 autoregressive parameters. Its time invariant VAR form will contain 365 variables, and this is a handicap, especially for forecasting.

The most common case is the modeling in one dimensional time repetition at equal intervals. In this Section we present a re-examination of the periodic modeling problem. Our approach differs from most of the existing literature in that we stay within the univariate framework.

A periodic AR model of order 2 with l seasons, PAR$(2;l)$, is defined as

$$y_{t_s} = \phi_{0,s} + \phi_{1,s} y_{t_s-1} + \phi_{2,s} y_{t_s-2} + \varepsilon_{t_s} \tag{8}$$

where $t_s = Tl + s$, $s = 1, \ldots, l$, that is time t_s is at the sth season and $\phi_{m,s}$, $m = 1, 2$, are the periodically (or seasonally) varying autoregressive coefficients. For example, if $s = l$ (that is, we are at the lth season) then the periodically varying coefficients are $\phi_{m,l}$ whereas if $s = 1$ (that is, we are at the 1st season) then the periodically varying coefficients are $\phi_{m,1}$; $\phi_{0,s}$ is a periodically varying drift. The above process nests the AR(2) model as a special case if we assume that the drift and all the AR parameters are constant, that is: $\phi_{m,s} = \varphi_m$, $m = 0, 1, 2$, for all t.

The PAR$(2;l)$ model can be expressed as the time varying AR(2) model in eq.(1):

$$y_t = \phi_0(t) + \phi_1(t) y_{t-1} + \phi_2(t) y_{t-2} + \varepsilon_t,$$

where $\phi_m(t) = \phi_m(\tau_{Tl})$, $m = 0, 1, 2$, $\tau_{Tl} = t - Tl$, are the periodically (or seasonally) varying autoregressive coefficients: $\phi_{m,s} \triangleq \phi_m(Tl + s)$, $s = 1, \ldots, l$.

For the PAR$(2;l)$ model the continuant matrix $\mathbf{\Phi}_{t,nl}$ in eq. (5) (we assume that information is given at time $\tau_{nl} = t - nl$ for ease of exposition; it can of course be given at any time $\tau_{nl+s} = t - nl - s$) can be expressed as a block Toeplitz matrix. Thus, we have

$$\xi_{t,nl} = \left| \mathbf{\Phi}_{t,nl} \right|, \tag{9}$$

with

$$
\mathbf{\Phi}_{t,nl} = \begin{pmatrix}
\mathbf{\Phi}_{\tau_{(n-1)l},l} & \overline{\mathbf{0}} & & & \\
\widetilde{\mathbf{0}}_{\tau_{n-2}} & \mathbf{\Phi}_{\tau_{(n-2)l},l} & \overline{\mathbf{0}} & & \\
 & \ddots & \ddots & \ddots & \\
 & & \widetilde{\mathbf{0}}_{\tau_1} & \mathbf{\Phi}_{\tau_l,l} & \overline{\mathbf{0}} \\
 & & & \widetilde{\mathbf{0}}_t & \mathbf{\Phi}_{t,l}
\end{pmatrix},
$$

where $\overline{\mathbf{0}}$ is an $l \times l$ matrix of zeros except for -1 in its $(l,1)$ entry; $\widetilde{\mathbf{0}}_t$ is an $l \times l$ matrix of zeros except $\phi_2(t-l+1)$, in its $(1,l)$ entry. Since $\phi_m(\tau_{Tl}) = \phi_m(t)$: $\widetilde{\mathbf{0}}_{\tau_{Tl}} = \widetilde{\mathbf{0}}_t$ and $\mathbf{\Phi}_{\tau_{Tl},l} = \mathbf{\Phi}_{t,l}$. Thus the block diagonal matrix $\mathbf{\Phi}_{t,nl}$ can be written as

$$
\mathbf{\Phi}_{t,nl} = \begin{pmatrix}
\mathbf{\Phi}_{t,l} & \overline{\mathbf{0}} & & & \\
\widetilde{\mathbf{0}}_t & \mathbf{\Phi}_{t,l} & \overline{\mathbf{0}} & & \\
 & \ddots & \ddots & \ddots & \\
 & & \widetilde{\mathbf{0}}_t & \mathbf{\Phi}_{t,l} & \overline{\mathbf{0}} \\
 & & & \widetilde{\mathbf{0}}_t & \mathbf{\Phi}_{t,l}
\end{pmatrix}, \tag{10}
$$

where $\mathbf{\Phi}_{t,l}$ is the continuant or tridiagonal matrix $\mathbf{\Phi}_{t,k}$ matrix defined in eq. (5) when $k = l$. That is

$$
\mathbf{\Phi}_{t,l} = \begin{pmatrix}
\phi_1(\tau_l+1) & -1 & & & & \\
\phi_2(\tau_l+2) & \phi_1(\tau_l+2) & -1 & & & \\
 & \phi_2(\tau_l+3) & \phi_1(\tau_l+3) & -1 & & \\
 & & \ddots & \ddots & \ddots & \\
 & & & \phi_2(t-1) & \phi_1(t-1) & -1 \\
 & & & & \phi_2(t) & \phi_1(t)
\end{pmatrix}. \tag{11}
$$

Proposition 1 *The general solution of eq. (8) with free constants (initial condition values)* y_{t-nl}, y_{t-nl-1} *is given by*

$$
y_{t,nl}^{gen} = y_{t,nl}^{hom} + y_{t,nl}^{par}, \tag{12}
$$

where

$$
\begin{aligned}
y_{t,nl}^{hom} &= \xi_{t,nl} y_{t-nl} + \phi_2(t-nl+1)\xi_{t,nl-1} y_{t-nl-1}, \\
y_{t,nl}^{par} &= \sum_{s=0}^{l-1} \sum_{T=0}^{n-1} \xi_{t,Tl+s}\phi_0(t-s) + \sum_{i=0}^{nl-1} \xi_{t,i}\varepsilon_{t-i},
\end{aligned}
$$

and $\xi_{t,nl}$ is given either in eq. (9) or in Proposition (6) in Appendix B.

The proof of the above Proposition follows immediately from Theorem 1 and the definition of the periodic model (8).

3.1 Cyclical AR(2) process

Some economic series exhibit oscillations which are not associated with the same fixed period every year. Despite their lack of fixed periodicity, such time series are predictable to a certain degree.

Rather than setting up a general model from first principles, we re-interpret the periodic model with some modifications. In particular, we now assume that we have $d+1$ cycles, with $0 \leq d \leq l-1$. Then, $s_j = l_{j-1}+1, \ldots, l_j$, $j = 1, \ldots, d+1$ (with $0 = l_0 < l_1 < \ldots < l_d < l_{d+1} = l$) are the seasons in cycle j. Thus we can write $\phi_{m,s_j} \triangleq \phi_m(t_{s_j})$, $m = 0, 1, 2$, $t_{s_j} = Tl + s_j$. A CAR(2) model with l seasons and $d+1$ cycles (CAR$(2; l; d)$) is defined as

$$y_{t_{s_j}} = \phi_{0,s_j} + \phi_{1,s_j} y_{t_{s_j}-1} + \phi_{2,s_j} y_{t_{s_j}-2} + \varepsilon_{t_{s_j}}. \tag{13}$$

For the above process, $\mathbf{\Phi}_{t,l}$ in eq. (11) can be written as

$$\mathbf{\Phi}_{t,l} = \begin{bmatrix} \mathbf{\Phi}_{t-l_d, l_{d+1}-l_d} & \overline{\mathbf{0}}_d & & & \\ \widetilde{\mathbf{0}}_d & \mathbf{\Phi}_{t-l_{d-1}, l_d - l_{d-1}} & \overline{\mathbf{0}}_{d-1} & & \\ & \ddots & \ddots & \ddots & \\ & & \widetilde{\mathbf{0}}_2 & \mathbf{\Phi}_{t-l_1, l_2-l_1} & \overline{\mathbf{0}}_1 \\ & & & \widetilde{\mathbf{0}}_1 & \mathbf{\Phi}_{t,l_1} \end{bmatrix}, \tag{14}$$

where first, the j ($j = 1, \ldots, d+1$) block of the main diagonal is $\mathbf{\Phi}_{t-l_{j-1}, l_j - l_{j-1}}$, that is a $(l_j - l_{j-1}) \times (l_j - l_{j-1})$ banded 'time varying' Toeplitz matrix of bandwidth 3:

$$\mathbf{\Phi}_{t-l_{j-1}, l_j-l_{j-1}} = \begin{pmatrix} \phi_1(\tau_{l_j}+1) & -1 & & & & \\ \phi_2(\tau_{l_j}+2) & \phi_1(\tau_{l_j}+2) & -1 & & & \\ & \phi_2(\tau_{l_j}+3) & \phi_1(\tau_{l_j}+3) & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \phi_2(\tau_{l_{j-1}}-1) & \phi_1(\tau_{l_{j-1}}-1) & -1 \\ & & & & \phi_2(\tau_{l_{j-1}}) & \phi_1(\tau_{l_{j-1}}) \end{pmatrix},$$

second, the j ($j = 1, \ldots, d$) block of the subdiagonal, $\widetilde{\mathbf{0}}_j$, is a $(l_j - l_{j-1}) \times (l_{j+1} - l_j)$ matrix of zeros except for $\phi_2(\tau_{l_j}+1)$ in its $1\times(l_{j+1}-l_j)$ entry, and third, the j block of the superdiagonal $\overline{\mathbf{0}}_j$, is a $(l_{j+1}-l_j)\times(l_j-l_{j-1})$ matrix of zeros except for -1 in its $(l_{j+1}-l_j)\times 1$ entry, and iv) there are zeros elsewhere.

4 Abrupt Breaks

Our general result has been presented in Section 2.3. In the current Section, we discuss still another example in order to both make our analysis clearer and to demonstrate its applicability. One important case is that of r, $0\le r\le k-1$, abrupt breaks at times $t-k_1$, $t-k_2$, $\ldots$, $t-k_r$, where $0=k_0<k_1<k_2<\cdots<k_r<k_{r+1}=k$, $k_r\in\mathbb{Z}^+$, and k_r is finite. That is, between $t-k=t-k_{r+1}$ and the present time $t=t-k_0$ the AR(2) process contains r structural breaks and the switch from one set of parameters to another is abrupt. In particular

$$y_\tau=\phi_{0,j}+\phi_{1,j}y_{\tau-1}+\phi_{2,j}y_{\tau-2}+\sigma_j^2e_{\tau,j}, \tag{15}$$

for $\tau=t-k_{j-1},\ldots,t-k_j+1, j=1,\ldots,r+1$ and $e_{t,j}\sim$ i.i.d $(0,1)$ $\forall\ t,j$. Within the class of AR(2) processes, this specification is quite general and allows for intercept and slope shifts as well as changes in the error variances (see also Pesaran et al., 2006). Each regime j is characterized by a vector of autoregressive coefficients: $\phi_{0,j}$, $\boldsymbol{\phi}_j'=(\phi_{1,j},\phi_{2,j})$, and an error term variance, $0<\sigma_j^2<M_j<\infty\ \forall\ j$, $M_j\in\mathbb{R}^+$. We term this model abrupt breaks AR process of order $(2;r)$ (ABAR$(2;r)$).

For the AR(2) model with r abrupt breaks, $\xi_{t,k}$ in eq. (6) can be written as the determinant of a partitioned (or a block) tridiagonal matrix

$$\xi_{t,k}=\det(\mathbf{\Phi}_{t,k})=\begin{vmatrix}\mathbf{\Phi}_{t-k_r,k_{r+1}-k_r} & \overline{\mathbf{0}}_r & & & \\ \widetilde{\mathbf{0}}_r & \mathbf{\Phi}_{t-k_{r-1},k_r-k_{r-1}} & \overline{\mathbf{0}}_{r-1} & & \\ & \ddots & \ddots & \ddots & \\ & & \widetilde{\mathbf{0}}_2 & \mathbf{\Phi}_{t-k_1,k_2-k_1} & \overline{\mathbf{0}}_1 \\ & & & \widetilde{\mathbf{0}}_1 & \mathbf{\Phi}_{t,k_1}\end{vmatrix}, \tag{16}$$

where first, the j $(j=1,\ldots,r+1)$ block of the main diagonal is $\mathbf{\Phi}_{t-k_{j-1},k_j-k_{j-1}}$, that is a $(k_j - k_{j-1}) \times (k_j - k_{j-1})$ banded Toeplitz matrix of bandwidth 3:

$$\mathbf{\Phi}_{t-k_{j-1},k_j-k_{j-1}} = \begin{pmatrix} \phi_{1,j} & -1 & & & \\ \phi_{2,j} & \phi_{1,j} & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & \phi_{2,j} & \phi_{1,j} & -1 \\ & & & \phi_{2,j} & \phi_{1,j} \end{pmatrix},$$

with $\xi_{t-k_{j-1},k_j-k_{j-1}} = \left|\mathbf{\Phi}_{t-k_{j-1},k_j-k_{j-1}}\right| = \frac{1}{\lambda_{1,j}-\lambda_{2,j}}(\lambda_{1,j}^{k_j-k_{j-1}+1} - \lambda_{2,j}^{k_j-k_{j-1}+1})$, and the second equality holds if and only if $\lambda_{1,j} \neq \lambda_{2,j}$ (where $1 - \phi_{1,j}B - \phi_{2,j}B^2 = (1-\lambda_{1,j}B)(1-\lambda_{2,j}B)$), second, the j ($j = 1, \ldots, r$) block of the subdiagonal, $\widetilde{\mathbf{0}}_j$, is a $(k_j - k_{j-1}) \times (k_{j+1} - k_j)$ matrix of zeros except for $\phi_{2,j}$ in its $1 \times (k_{j+1} - k_j)$ entry, and third, the j block of the superdiagonal $\overline{\mathbf{0}}_j$, is a $(k_{j+1} - k_j) \times (k_j - k_{j-1})$ matrix of zeros except for -1 in its $(k_{j+1} - k_j) \times 1$ entry, and iv) there are zeros elsewhere.

Corollary 1 *The general solution of the ABAR$(2; r)$ model in eq. (15) with free constants (initial condition values) y_{t-k}, y_{t-k-1}, is given by*

$$y_{t,k}^{gen} = y_{t,k}^{hom} + y_{t,k}^{par},$$

where

$$\begin{aligned} y_{t,k}^{hom} &= \xi_{t,k}y_{t-k} + \phi_2(t-k+1)\xi_{t,k-1}y_{t-k-1}, \\ y_{t,k}^{par} &= \sum_{j=1}^{r+1}\phi_{0,j}\sum_{i=k_{j-1}}^{k_j-1}\xi_{t,i} + \sum_{j=1}^{r+1}\sigma_j^2\sum_{i=k_{j-1}}^{k_j-1}\xi_{t,i}e_{t-i,j}, \end{aligned}$$

and $\xi_{t,k}$ is given either in eq. (16) or in Proposition 7 (see Appendix B).

The proof of the above Corollary follows immediately from Theorem 1 and the definition of the ABAR$(2; r)$ model in eq. (15).

5 Prediction and Moment Structure

We turn our attention to the fundamental properties of the various TV-AR(2) processes. Armed with a powerful technique for manipulating time varying models we may now provide a thorough description of the processes (1) by deriving, first, its multistep ahead predictor, the associated forecast error and the mean square error; second, the first two unconditional moments of this process, and third, its covariance structure.

5.1 Multi Step Forecasts

Taking the conditional expectation of eq. (7) with respect to the σ field $\mathcal{F}_{\tau_k}$ ($\tau_k = t - k$) yields the following Proposition.

Proposition 2 *For the TV-AR(2) model the k-step-ahead optimal (in L_2-sense) linear predictor of y_t, $\mathbb{E}(y_t \,|\mathcal{F}_{\tau_k})$, is readily seen to be*

$$\mathbb{E}(y_t \,|\mathcal{F}_{\tau_k}) = \sum_{i=0}^{k-1} \xi_{t,i}\phi_0(t-i) + \xi_{t,k} y_{t-k} + \phi_2(t-k+1)\xi_{t,k-1} y_{t-k-1}. \tag{17}$$

In addition, the forecast error for the above k-step-ahead predictor, $\mathbb{FE}(y_t \,|\mathcal{F}_{\tau_k}) = y_t - \mathbb{E}[y_t \,|\mathcal{F}_{\tau_k}]$, is given by

$$\mathbb{FE}(y_t \,|\mathcal{F}_{\tau_k}) = \Xi_{t,k}(B)\varepsilon_t = \sum_{i=0}^{k-1} \xi_{t,i} B^i \varepsilon_t, \tag{18}$$

and it is expressed in terms of k error terms from time $t-k+1$ to time t; the coefficient of the error term at time $t-i$, $\xi_{t,i}$, is the determinant of an $i \times i$ matrix ($\Phi_{t,i}$), each nonzero variable diagonal of which consists of the AR time varying coefficients $\phi_m(\cdot)$, $m = 1, 2$ from time $t-i+m$ to t.
The mean square error is

$$\mathbb{V}ar[\mathbb{FE}(y_t \,|\mathcal{F}_{\tau_k})] = \Xi_{t,k}^{(2)}(B)\sigma_t^2 = \sum_{i=0}^{k-1} \xi_{t,i}^2 B^i \sigma_t^2, \tag{19}$$

which is expressed in terms of k variances from time $t-k+1$ to time t, with time varying coefficients (the squared ξs).

The following Corollary presents results for the forecasts from PAR and CAR processes.

Corollary 2 *For the PAR(2; l) and the CAR(2; l; d) models (see eqs. (8) and (13) respectively) the nl-step-ahead optimal linear predictor is given by eq. (17) (with $k = nl$) in Proposition (2) where*

$$\begin{aligned}
\sum_{i=0}^{nl-1} \xi_{t,i}\phi_0(t-i) &= \sum_{s=0}^{l-1}\sum_{T=0}^{n-1} \xi_{t,Tl+s}\phi_0(t-s) \quad \textit{(PAR model)},\\
\sum_{i=0}^{nl-1} \xi_{t,i}\phi_0(t-i) &= \sum_{j=1}^{d+1}\sum_{s_j=l_{j-1}+1}^{l_j}\sum_{T=0}^{n-1} \xi_{t,Tl+s_j}\phi_0(t-s_j) \quad \textit{(CAR model)},
\end{aligned}$$

and $\xi_{t,Tl+s}$, $\xi_{t,Tl+s_j}$ are given in Proposition (6) and Corollary (3) respectively (see Appendix B).
Finally, for the ABAR(2; r) model in eq. (15) the k-step-ahead optimal linear predictor is given by eq.

(17) where

$$\sum_{i=0}^{k-1} \xi_{t,i}\phi_0(t-i) = \sum_{j=1}^{r+1} \phi_{0,j} \sum_{i=k_{j-1}}^{kj-1} \xi_{t,i},$$

and $\xi_{t,i}$ *is given either in eq. (16) or in Proposition (7) (see Appendix B).*

Franses and Paap (2005) employ the vector season representation to compute forecasts and forecast error variances for a PAR(1; 4) process. In this way forecasts can be generated along the same lines with quadrivariate VAR(1) models. Franses (1996a) derives multi-step forecast error variances for low-order PAR models with $l = 4$, using the VS representation. But, if l is large even low order specifications will have large VAR representations and this is a handicap especially for forecasting. In contrast, our formulae using the univariate framework allow a fast computation of the multi-step-ahead predictors even if l is large.

In what follows we give conditions for the first and second unconditional moments of model (1) to exist.

5.2 Wold Representation

First, we need an assumption.

Assumption A.1. $\sum_{i=0}^{k} \xi_{t,i}\phi_0(t-i)$ as $k \to \infty$ converges $\forall\, t$ and $\sum_{i=0}^{\infty} \sup_t(\xi_{t,i}^2\sigma_{t-i}^2) < M_u < \infty$ ($M_u \in \mathbb{R}^+$).

Assumption A.1 is a sufficient condition for the model in eq. (1) to admit a second-order MA(∞) representation. A necessary but not sufficient condition for $\sum_{i=0}^{k} \xi_{t,i}\phi_0(t-i)$ to converge is $\lim_{k\to\infty}[\xi_{t,k}\phi_0(t-k)] = 0\ \forall\, t$. A sufficient condition for this limit to be zero is: $\lim_{k\to\infty} \xi_{t,k} = 0$ and $\phi_0(t-k)$ is bounded.

Another immediate consequence of Theorem 1 is the following Proposition, where we state an expression for the first unconditional moment of y_t.

Proposition 3 *Let Assumption A.1 hold. Then for the TV-AR(2) model we have:*

$$y_t = \lim_{k\to\infty} y_{t,k}^{gen} \stackrel{L_2}{=} \lim_{k\to\infty} y_{t,k}^{par} \stackrel{L_2}{=} \Xi_{t,\infty}(B)[\phi_0(t) + \varepsilon_t] = \sum_{i=0}^{\infty} \xi_{t,i}B^i[\phi_0(t) + \varepsilon_t], \tag{20}$$

is a unique solution of the TV-AR(2) model in eq. (1). The above expression states that $\{y_{t,k}^{par}, t \in \mathbb{Z}\}$ *(defined in eq. (7))* L_2 *converges as* $k \to \infty$ *if and only if* $\sum_{i=0}^{k} \xi_{t,i}\phi_0(t-i)$ *converges and* $\sum_{i=0}^{k} \xi_{t,i}\varepsilon_{t-i}$ *converges a.s., and thus under assumption A.1* $\lim_{k\to\infty} y_{t,k}^{gen} \stackrel{L_2}{=} \lim_{n\to\infty} y_{t,k}^{par}$ *satisfies eq. (1).*

In other words $\lim_{k\to\infty} y_{t,k}^{gen}$ *is decomposed into a non random part and a zero mean random part. In* *particular,*

$$\mathbb{E}(y_t) = \lim_{k\to\infty} \mathbb{E}(y_t\,|\mathcal{F}_{\tau_k}) = \Xi_{t,\infty}(B)\phi_0(t) = \sum_{i=0}^{\infty} \xi_{t,i} B^i \phi_0(t), \tag{21}$$

is the non random part of y_t *and it is an infinite sum of the periodical drifts where the time varying coefficients are expressed as determinants of continuant matrices (the* ξs*), while* $\lim_{k\to\infty} \mathbb{FE}(y_t\,|\mathcal{F}_{\tau_k}) = \sum_{i=0}^{\infty} \xi_{t,i}\varepsilon_{t-i}$ *is the zero mean random part. Therefore the* $\xi_{t,i}$ *as defined in eq. (6) are the Green functions associated with the second order time varying AR polynomial:* $\Phi_t(B) = 1 - \phi_1(t)B - \phi_2(t)B^2$.

5.3 Second Moments

In this subsection we state as a Proposition the result for the second moment structure.

Proposition 4 *Let Assumption A.1 hold. Then the second unconditional moment of* y_t *exists and it is given by*

$$\mathbb{E}(y_t^2) = [\mathbb{E}(y_t)]^2 + \Xi_{t,\infty}^{(2)}(B)\sigma_t^2 = [\mathbb{E}(y_t)]^2 + \sum_{i=0}^{\infty} \xi_{t,i}^2 B^i \sigma_t^2. \tag{22}$$

That is, the time varying variance of y_t *is an infinite sum of the time varying variances of the errors with time varying coefficients (the squared values of the* ξs*).*
In addition, the time varying autocovariance function $\gamma_{t,k}$ *is given by*

$$\begin{aligned} \gamma_{t,k} &= \mathbb{C}ov(y_t, y_{\tau_k}) = \sum_{i=0}^{\infty} \xi_{t,k+i}\xi_{\tau_k,i}\sigma_{\tau_k - i}^2 = \xi_{t,k}\mathbb{V}ar(y_{\tau_k}) + \\ & \quad \phi_2(\tau_k + 1)\xi_{t,k-1}\mathbb{C}ov(y_{\tau_k}, y_{\tau_k - 1}), \end{aligned} \tag{23}$$

where the second equality follows from the MA(∞*) representation of* y_t *in eq. (20) and the third one from eq. (7) in Theorem 1. For any fixed* t*,* $\lim_{k\to\infty} \gamma_{t,k} = 0$ *when* $\lim_{k\to\infty} \xi_{t,k} = 0\ \forall\ t$*. Finally, recall that for the PAR and ABAR models the* ξs *are given either in eqs. (9) and (16) respectively, or in Propositions (6) and (7) respectively.*

Although it may be difficult to compute the covariance structure of $\{y_t\}$ explicitly, for numerical work, one can always calculate it by computing the Green functions (that is, the continuant determinants ξs) with eqs. (5) and (6) and summing these with eq. (17).

6 Conclusions

We have provided the general solutions to low order TV-AR models in terms of their homogeneous and particular parts. Our first step was to find the fundamental set of solutions by computing the determinants of the matrix of coefficients associated with the infinite linear system that represents the difference equation.

The framework developed in Section 2, proved itself to be a general time varying theory, encompassing a number of seemingly unrelated models, discussed in Sections 3 and 4. We have identified common properties (throughout the paper and in particular in Section 5), which are basic to each of the particular application.

We believe that time varying models should take center stage in the time series literature; this is why we have labored to develop a theory with rigorous foundations that can encompass a variety of dynamic systems, i.e., periodic and cyclical processes, and AR models which contain multiple structural breaks. Work that remains to be done by us and fellow researchers is on estimation and testing (for one application on this front see the paper by Karanasos et al., 2013) to demonstrate the usefulness of time varying models. In the long run, a sound mathematical theory has to be cointegrated with its applicability.

REFERENCES

Abdrabbo, N. A. and Priestley, M. B. (1967) On the prediction of non-stationary processes. *Journal of the Royal Statistical Society*, Series B **29**, 570-85.

Adams, G. J. and Goodwin, G. C. (1995) Parameter estimation for periodic ARMA models. *Journal of Time Series Analysis* **16**, 127-45.

Anderson, P. L., Meerschaert, M. M. and Zhang, K. (2013) Forecasting with prediction intervals for periodic autoregressive moving average models. *Journal of Time Series Analysis* **34**, 187-93.

Anderson, P. L. and Vecchia, A. V. (1993) Asymptotic results for periodic autoregressive moving-average processes. *Journal of Time Series Analysis* **14**, 1-18.

Bentarzi, M. and Hallin, M. (1994) On the invertibility of periodic moving-average models. *Journal of Time Series Analysis* **15**, 263-68.

Birchenhall, C. R., Bladen-Hovell, R. C., Chui, A. P. L., Osborn, D. R. and Smith, J. P. (1989) A seasonal model of consumption. *The Economic Journal* **99**, 837-43.

Bollerslev, T. and Ghysels, E. (1996) Periodic autoregressive conditional heteroscedasticity. *Journal of Business & Economic Statistics* **14**, 139-51.

Cipra, T. and Tlustý, P. (1987) Estimation in multiple autoregressive-moving average models using periodicity. *Journal of Time Series Analysis* **8**, 293-300.

del Barrio Castro, T. and Osborn, D. R. (2008) Testing for seasonal unit roots in periodic integrated autoregressive processes. *Econometric Theory* **24**, 1093–129 del Barrio Castro, T. and Osborn, D. R. (2012) Non-parametric testing for seasonally and periodically integrated processes. *Journal of Time Series Analysis* **33**, 424-37.

Franses, P. H. (1994) A multivariate approach to modeling univariate seasonal time series. *Journal of Econometrics* **63**, 133–51.

Franses, P. H. (1996a) Multi-step forecast error variances for periodically integrated time series. *Journal of Forecasting* **15**, 83-95.

Franses, P. H. (1996b) Periodicity and Stochastic Trends in Economic Time Series. *Oxford University Press.*

Franses, P. H. and Paap, R. (2004) Periodic Time Series models. Oxford. *Oxford University Press.*

Franses, P. H. and Paap, R. (2005) Forecasting with periodic autoregressive time-series models, in A Companion to Economic Forecasting, Eds. Clements, M. P. and Hendry, D. F. Oxford. *Wiley-Blackwell*, pp. 432-52.

Ghysels, E. and Osborn, D. R. (2001) The Econometric Analysis of Seasonal Time Series. *Cambridge University Press.*

Gladyshev, E. G. (1961) On periodically correlated random sequences. *Soviet Mathematics* **2**, 385-88.

Granger, C. W. J. (2007) Forecasting - looking back and forward: Paper to celebrate the 50th anniversary of the econometrics institute at the Erasmus University, Rotterdam. *Journal of Econometrics* **138**, 3-13.

Granger, C. W. J. (2008) Non-linear models: Where do we go next - time varying parameter models? *Studies in Nonlinear Dynamics and Econometrics* **12**, 1-9.

Hurd, H. L. and Miamee, A. (2007) Periodically Correlated Random Sequences: Spectral Theory and Practice. Hoboken, New Jersey. *Wiley-Blackwell.*

Karanasos, M. (2001) Prediction in ARMA models with GARCH in mean effects. *Journal of Time Series Analysis* **22**, 555-76.

Karanasos, M., Paraskevopoulos, A. G. and Dafnos, S. (2013) A univariate time varying analysis of periodic ARMA processes. *Unpublished Paper.*

Karanasos, M., Paraskevopoulos, A. G., Menla Ali, F., Karoglou, M. and Yfanti, S. (2013) Modelling returns and volatilities during financial crises: a time varying coefficient approach. *Unpublished Paper.*

Lund, R. and Basawa, I. V. (2000) Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis* **21**, 75-93.

Lund, R., Shao, Q. and Basawa, I. (2006) Parsimonious periodic time series modeling. *Australian & New Zealand Journal of Statistics* **48**, 33-47.

McLeod, A. I. (1994) Diagnostic checking of periodic autoregression models with application. *Journal* *of Time Series Analysis* **15**, 221-33.

Osborn, D. R. (1988) Seasonality and habit persistence in a life cycle model of consumption. *Journal of Applied Econometrics* **3**, 255–66.

Osborn, D. R. (1991) The implications of periodically varying coefficients for seasonal time-series processes. *Journal of Econometrics* **48**, 373–84.

Osborn, D. R. and Smith, J. P. (1989) The performance of periodic autoregressive models in forecasting seasonal U.K. consumption. *Journal of Business & Economic Statistics* **7**, 117-27.

Pagano, M. (1978) On periodic and multiple autoregressions. *The Annals of Statistics* **6**, 1310-317.

Paraskevopoulos, A. G. (2012) The Infinite Gauss-Jordan elimination on row-finite $\omega \times \omega$ matrices. arXiv: 1201.2950.

Paraskevopoulos, A. G. and Karanasos, M. (2013) Closed form solutions for linear difference equations with time dependent coefficients. *Unpublished paper.*

Paraskevopoulos, A. G., Karanasos, M. and Dafnos, S. (2013) A unified theory for time varying models: foundations with applications in the presence of breaks and heteroskedasticity (and some results on companion and Hessenberg matrices). *Unpublished Paper.*

Pesaran, M. H., Pettenuzzo, D. and Timmermann, A. (2006) Forecasting time series subject to multiple structural breaks. *Review of Economic Studies* 73, 1057-084.

Rao, S. T. (1970) The fitting of non-stationary time-series models with time-dependent parameters. *Journal of the Royal Statistical Society,* Series B **32**, 312-22.

Shao, Q. (2008) Robust estimation for periodic autoregressive time series. *Journal of Time Series Analysis* **29**, 251-63.

Sydsaeter, K., Hammond, P. J., Seierstad, A. and Strom, A. (2008) Further Mathematics for Economic Analysis. New York. *Prentice Hall,* 2nd edition.

Taylor, A. M. R. (2002) Regression-based unit root tests with recursive mean adjustment for seasonal and nonseasonal time series. *Journal of Business & Economic Statistics* **20**, 269–81.

Taylor, A. M. R. (2003) On the asymptotic properties of some seasonal unit root tests. *Econometric Theory* **19**, 311–21.

Taylor, A. M. R. (2005) Variance ratio tests of the seasonal unit root hypothesis. *Journal of Econometrics* **124**, 33–54.

Tesfaye Y. G., Anderson, P. L. and Meerschaert, M. M. (2011) Asymptotic results for Fourier-PARMA time series. *Journal of Time Series Analysis* **32**, 157–74.

Tiao, G. C. and Grupe, M. R. (1980) Hidden periodic autoregressive-moving average models in time series data. *Biometrika* **67**, 365–73.

Tiao, G. C. and Guttman, I. (1980) Forecasting contemporal aggregates of multiple time series. *Journal of Econometrics* **12**, 219–30.

Whittle, P. (1965) Recursive relations for predictors of non-stationary processes. *Journal of the Royal Statistical Society,* Series B **27**, 523-32.

A APPENDIX

In this appendix we prove Theorem 1. Before proceeding with the main body of the proof, we present two essential tools for carrying it out.

The Infinite Gaussian Elimination. Following Paraskevopoulos (2012), we apply the infinite Gaussian elimination algorithm implemented under a rightmost pivot strategy to the coefficient matrix $\mathbf{\Phi}$ of (3). The process is briefly described below.

Call $\mathbf{h}^{(1)} = \mathbf{H}^{(1)} = (-\phi_2(\tau_k+1), -\phi_1(\tau_k+1), 1, 0, ...)$ the opposite-sign first row of $\mathbf{\Phi}$. Insert the second row of $\mathbf{\Phi}$ below $\mathbf{H}^{(1)}$ to build the matrix $\mathbf{B}^{(2)}$:

$$\mathbf{B}^{(2)} = \begin{pmatrix} -\phi_2(\tau_k+1) & -\phi_1(\tau_k+1) & 1 & 0 & \dots \\ 0 & \phi_2(\tau_k+2) & \phi_1(\tau_k+2) & -1 & \dots \end{pmatrix}.$$

Use as pivot the rightmost one of $\mathbf{H}^{(1)}$ to clear the element $\phi_1(\tau_k+2)$ in the second row of $\mathbf{B}^{(2)}$. After normalization it yields the matrix:

$$\mathbf{H}^{(2)} = \begin{pmatrix} -\phi_2(\tau_k+1) & -\phi_1(\tau_k+1) & 1 & 0 & \dots \\ -\phi_2(\tau_k+1)\phi_1(\tau_k+2) & -\phi_2(\tau_k+2) - \phi_1(\tau_k+1)\phi_1(\tau_k+2) & 0 & 1 & \dots \end{pmatrix}.$$

Insert the third row of $\mathbf{\Phi}$ below $\mathbf{H}^{(2)}$ to build the matrix $\mathbf{B}^{(3)}$:

$$\begin{pmatrix} -\phi_2(\tau_k+1) & -\phi_1(\tau_k+1) & 1 & 0 & 0 & \dots \\ -\phi_2(\tau_k+1)\phi_1(\tau_k+2) & -\phi_2(\tau_k+2)-\phi_1(\tau_k+1)\phi_1(\tau_k+2) & 0 & 1 & 0 & \dots \\ 0 & 0 & \phi_2(\tau_k+3) & \phi_1(\tau_k+3) & -1 & \dots \end{pmatrix}.$$

Use the first two rows of $\mathbf{B}^{(3)}$ as pivot rows and their rightmost 1s as pivot elements to clear the entries $\phi_2(\tau_k+3)$ and $\phi_1(\tau_k+3)$ of $\mathbf{B}^{(3)}$, producing the matrix $\mathbf{H}^{(3)}$:

$$\mathbf{H}^{(3)} = \begin{pmatrix} h_{11} & h_{12} & 1 & 0 & 0 & 0 & ... \\ h_{21} & h_{22} & 0 & 1 & 0 & 0 & ... \\ h_{31} & h_{32} & 0 & 0 & 1 & 0 & ... \end{pmatrix}.$$

where the entries of the first column of $\mathbf{H}^{(3)}$ are given by

$$h_{11} = -\phi_2(\tau_k + 1),\ h_{21} = -\phi_2(\tau_k + 1)\phi_1(\tau_k + 2),$$
$$h_{31} = -\phi_2(\tau_k + 1)\phi_1(\tau_k + 2)\phi_1(\tau_k + 3) - \phi_2(\tau_k + 1)\phi_2(\tau_k + 3), ...$$

and the entries of the second column are given by

$$h_{12} = -\phi_1(\tau_k + 1),\ h_{22} = -\phi_2(\tau_k + 2) - \phi_1(\tau_k + 1)\phi_1(\tau_k + 2),$$
$$h_{32} = -\phi_1(\tau_k+1)\phi_1(\tau_k+2)\phi_1(\tau_k+3) - \phi_2(\tau_k+2)\phi_1(\tau_k+3) - \phi_2(\tau_k+3)\phi_1(\tau_k+1).$$

This process continues ad infinitum, generating an infinite chain of submatrices

$$\mathbf{H}^{(1)} \sqsubset \mathbf{H}^{(2)} \sqsubset \mathbf{H}^{(3)} \sqsubset \ldots \sqsubset \mathbf{H}$$

whose limit row-finite matrix $\mathbf{H}$ is the Hermite Form (HF) of $\mathbf{\Phi}$. The ith row of $\mathbf{H}$ is defined to be the last row of $\mathbf{H}^{(i)}$.

Two Fundamental Solutions. The opposite-sign two first columns of $\mathbf{H}$ augmented at the top by $(1, 0)$ and $(0, 1)$, respectively, that is

$$\begin{aligned}\boldsymbol{\xi}^{(2)}_{\tau_k} = (1,\ \ 0,\ \ &\phi_2(\tau_k + 1),\ \ \phi_2(\tau_k + 1)\phi_1(\tau_k + 2),\\ &\phi_2(\tau_k + 1)\phi_1(\tau_k + 2)\phi_1(\tau_k + 3) + \phi_2(\tau_k + 1)\phi_2(\tau_k + 3), ...)',\\ \boldsymbol{\xi}^{(1)}_{\tau_k} = (0,\ \ 1,\ \ &\phi_1(\tau_k + 1),\ \ \phi_2(\tau_k + 2) + \phi_1(\tau_k + 1)\phi_1(\tau_k + 2),\\ &\phi_1(\tau_k+1)\phi_1(\tau_k+2)\phi_1(\tau_k+3) + \phi_2(\tau_k+2)\phi_1(\tau_k+3) + \phi_2(\tau_k+3)\phi_1(\tau_k+1), ...)'\end{aligned}$$

are the two linearly independent solution sequences of the space of homogeneous solutions of eq. (2). The linear independence of $\boldsymbol{\xi}^{(1)}_{\tau_k}, \boldsymbol{\xi}^{(2)}_{\tau_k}$ follows from the fact that they possess the Casoratian:

$$\det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \neq 0.$$

We observe that the terms of the sequences $\boldsymbol{\xi}^{(1)}_{\tau_k}, \boldsymbol{\xi}^{(2)}_{\tau_k}$ are expansions of the following determinants

$$\boldsymbol{\xi}^{(2)}_{\tau_k} = \begin{cases} 1 \\ 0 \\ \phi_2(\tau_k+1) \\ \det\begin{pmatrix} \phi_2(\tau_k+1) & -1 \\ 0 & \phi_1(\tau_k+2) \end{pmatrix} \\ \det\begin{pmatrix} \phi_2(\tau_k+1) & -1 & 0 \\ 0 & \phi_1(\tau_k+2) & -1 \\ 0 & \phi_2(\tau_k+3) & \phi_1(\tau_k+3) \end{pmatrix}, \\ . \\ . \\ . \end{cases} \tag{A.1}$$

$$\boldsymbol{\xi}^{(1)}_{\tau_k} = \begin{cases} 0 \\ 1 \\ \phi_1(\tau_k+1) \\ \det\begin{pmatrix} \phi_1(\tau_k+1) & -1 \\ \phi_2(\tau_k+2) & \phi_1(\tau_k+2) \end{pmatrix} \\ \det\begin{pmatrix} \phi_1(\tau_k+1) & -1 & 0 \\ \phi_2(\tau_k+2) & \phi_1(\tau_k+2) & -1 \\ 0 & \phi_2(\tau_k+3) & \phi_1(\tau_k+3) \end{pmatrix}. \\ . \\ . \\ . \end{cases} \tag{A.2}$$

The first few terms of the homogeneous solution sequences, as shown above, suggest that the general terms of $\boldsymbol{\xi}^{(1)}_{\tau_k}, \boldsymbol{\xi}^{(2)}_{\tau_k}$ are

$$\xi^{(m)}_{t,k} = \det(\boldsymbol{\Phi}^{(m)}_{t,k}), \quad m = 1, 2, \tag{A.3}$$

where $\boldsymbol{\Phi}^{(1)}_{t,k} = \boldsymbol{\Phi}_{t,k}$ and $\xi^{(1)}_{t,k} = \xi_{t,k}$ (we drop the superscript 1 for notational convenience), as introduced in eqs. (5) and (6), and

$$\mathbf{\Phi}^{(2)}_{t,k} = \begin{pmatrix} \phi_2(\tau_k+1) & -1 & & & & \\ & \phi_1(\tau_k+2) & -1 & & & \\ & \phi_2(\tau_k+3) & \phi_1(\tau_k+3) & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \phi_2(t-1) & \phi_1(t-1) & -1 \\ & & & & \phi_2(t) & \phi_1(t) \end{pmatrix}.$$

In the following Proposition we use mathematical induction to verify the above generalization formally.

Proposition 5 *The general terms of the fundamental solution sequences* $\boldsymbol{\xi}^{(m)}_{\tau_k}$, $m = 1, 2$, *are given by eq.* (A.3), *that is*

$$\xi^{(2)}_{t,k} = \det \begin{pmatrix} \phi_2(\tau_k+1) & -1 & & & & \\ & \phi_1(\tau_k+2) & -1 & & & \\ & \phi_2(\tau_k+3) & \phi_1(\tau_k+3) & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \phi_2(t-1) & \phi_1(t-1) & -1 \\ & & & & \phi_2(t) & \phi_1(t) \end{pmatrix} \tag{A.4}$$

and

$$\xi_{t,k} = \det \begin{pmatrix} \phi_1(\tau_k+1) & -1 & & & & \\ \phi_2(\tau_k+2) & \phi_1(\tau_k+2) & -1 & & & \\ & \phi_2(\tau_k+3) & \phi_1(\tau_k+3) & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \phi_2(t-1) & \phi_1(t-1) & -1 \\ & & & & \phi_2(t) & \phi_1(t) \end{pmatrix}. \tag{A.5}$$

Proof. If $t = \tau_k + 1$ and $t = \tau_k + 2$ then $\xi_{\tau_k+1,1}$ and $\xi_{\tau_k+2,2}$ is the third term and fourth term of the sequences as directly verified by eq. (A.2). We assume that $\xi_{t-2,k-2}$ and $\xi_{t-1,k-1}$ are terms of $\boldsymbol{\xi}^{(1)}_{\tau_k}$. We show that $\xi_{t,k}$ is also a term of $\boldsymbol{\xi}^{(1)}_{\tau_k}$. Expanding $\xi_{t,k}$ along the last row and taking into account that $\mathbf{\Phi}_{t,k}$

is a $k \times k$ matrix, we have:

$$\xi_{t,k} = (-1)^{2k}\phi_1(t)\det\begin{pmatrix} \phi_1(\tau_k+1) & -1 & & & & \\ \phi_2(\tau_k+2) & \phi_1(\tau_k+2) & -1 & & & \\ & \phi_2(\tau_k+3) & \phi_1(\tau_k+3) & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \phi_2(t-2) & \phi_1(t-2) & -1 \\ & & & & \phi_2(t-1) & \phi_1(t-1) \end{pmatrix} +$$

$$(-1)^{2k-1}(-1)\phi_2(t)\det\begin{pmatrix} \phi_1(\tau_k+1) & -1 & & & & \\ \phi_2(\tau_k+2) & \phi_1(\tau_k+2) & -1 & & & \\ & \phi_2(\tau_k+3) & \phi_1(\tau_k+3) & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \phi_2(t-3) & \phi_1(t-3) & -1 \\ & & & & \phi_2(t-2) & \phi_1(t-2) \end{pmatrix}.$$

Using the induction hypothesis, the above result can be written as

$$\xi_{t,k} = \phi_1(t)\xi_{t-1,k-1} + \phi_2(t)\xi_{t-2,k-2},$$

which shows that $\xi_{t,k}$ is a homogeneous solution of (2). Thus $\xi_{t,k}$ in (A.5) is a term of the solution sequence and the induction is complete. By analogy, we can show (A.4) and the proof is complete. ■

The fundamental solution $\xi_{t,k}$ (respectively $\xi^{(2)}_{t,k}$) can be obtained by augmenting the core solution matrix $\mathbf{C}_{t,k}$ (see eq. (4) in the main body of the paper) on the left by a $k \times 1$ column consisting of the first k entries of the second column (respectively of the first column) of $\mathbf{P}$ or $\mathbf{\Phi}$.

Proof. (**of Theorem 1**) As a direct consequence of Proposition 1, the general homogeneous solution of eq. (2) is the linear combination of the fundamental solutions as given below:

$$y^{hom}_{t,k} = \xi_{t,k}y_{\tau_k} + \xi^{(2)}_{t,k}y_{\tau_k-1}. \tag{A.6}$$

By expanding $\xi^{(2)}_{t,k}$ along the first column we obtain

$$\xi^{(2)}_{t,k} = \phi_2(\tau_k+1)\xi_{t,k-1}$$

and therefore (A.6) takes the form

$$y_{t,k}^{hom} = \xi_{t,k} y_{\tau_k} + \phi_2(\tau_k + 1)\xi_{t,k-1} y_{\tau_k - 1},$$

which coincides with the general homogeneous solution employed in eq. (7).

Next we show that $y_{t,k}^{par}$, employed in eq. (7), is a particular solution of eq. (2). Using the same arguments as in the proof of Proposition 5 we can show that

$$y_{t,k}^{par} = \det \begin{pmatrix} \phi_0(\tau_k+1) + \epsilon_{\tau_k+1} & -1 & & & & & \\ \phi_0(\tau_k+2) + \epsilon_{\tau_k+2} & \phi_1(\tau_k+2) & -1 & & & & \\ \phi_0(\tau_k+3) + \epsilon_{\tau_k+3} & \phi_2(\tau_k+3) & \phi_1(\tau_k+3) & -1 & & & \\ \vdots & & \ddots & \ddots & \ddots & & \\ \phi_0(t-1) + \epsilon_{t-1} & & & \phi_2(t-1) & \phi_1(t-1) & -1 \\ \phi_0(t) + \epsilon_t & & & & \phi_2(t) & \phi_1(t) \end{pmatrix}, \qquad \text{(A.7)}$$

is the solution of the initial value problem determined by eq. (2) subject to the initial values $y_{-1} = y_0 = 0$. This is the determinant of the core solution matrix $\mathbf{C}_{t,k}$ augmented on the left by a $k \times 1$ column consisting of the opposite sign first k entries of the right-hand side sequence of eq. (2).

Now expanding the determinant in eq. (A.7) along the first column we obtain $y_{t,k}^{par}$ in terms of $\xi_{t,i}$ and $\phi_0(t-i), \epsilon_{t-i}$ for $i = 0, 1, ..., k-1$, as used in eq. (7). Therefore the general solution in eq. (7), as the sum of the general homogeneous solution plus a particular solution, has been established. This completes the proof of Theorem 1. ■

B APPENDIX

In this Appendix we will make use of the block Toeplitz matrix in eq. (10) to obtain an explicit formula of $\xi_{t,nl}$ in which we decompose it into tridiagonal determinants, $\xi_{t,l}$. To prepare the reader, before we present the main result we consider the case where $n = 2$, that is we go from time t back to time $t - 2l$. The tridiagonal determinant $\xi_{t,2l}$ can be written as the sum of two terms

$$\begin{aligned} \xi_{t,2l} &= \begin{vmatrix} \mathbf{\Phi}_{t,l} & \overline{\mathbf{0}} \\ \widetilde{\mathbf{0}}_t & \mathbf{\Phi}_{t,l} \end{vmatrix} = \\ &= \xi_{t,l}^2 + \phi_2(\tau_l + 1)\xi_{t,l-1}\xi_{t-1,l-1}, \end{aligned} \qquad \text{(B.1)}$$

where each term is the product of two continuant (or tridiagonal) determinants.

Next let $i_j \in \{0,1\}$, $j = 1, \ldots, n-1$, and define

$$\varphi_{2,j} = \begin{cases} 1 & \text{if} \quad i_j = 0, \\ \phi_2(\tau_{jl}+1) & \text{if} \quad i_j = 1. \end{cases} \tag{B.2}$$

Proposition 6 *For the PAR(2; l) process in eq. (8), $\xi_{t,nl}$ is the determinant of $\mathbf{\Phi}_{t,nl}$ in eq. (9), and therefore can be written as*

$$\xi_{t,nl} = \sum_{i_1=0}^{1} \cdots \sum_{i_{n-1}=0}^{1} \{\xi_{t,l-i_1}(\prod_{T=2}^{n-1} \varphi_{2,T-1}\xi_{t-i_{T-1},l-i_T-i_{T-1}})\varphi_{2,n-1}\xi_{t-i_{n-1},l-i_{n-1}}\}, \tag{B.3}$$

where $\sum \cdots \sum$ stands for a multiple but finite summation, and recall that $\xi_{t,l} = |\mathbf{\Phi}_{t,l}|$ and $\mathbf{\Phi}_{t,l}$ is given by eq. (11).

In the above Proposition $\xi_{t,nl}$ is expressed as the sum of $\sum_{j=0}^{n-1} \binom{n-1}{j} = 2^{n-1}$ terms, each of which is the product of n terms. In other words, it is decomposed into determinants of $(l-m) \times (l-m)$ continuant matrices, $m = 0, 1, 2$: $\Phi_{t-i_{T-1},l-i_T-i_{T-1}}$.

When $n = 3$, eq. (B.3) reduces to:

$$\begin{aligned} \xi_{t,nl} &= \xi_{t,l}^3 + \phi_2(\tau_l+1)\xi_{t,l-1}\xi_{t-1,l-1}\xi_{t,l} \\ &\quad +\xi_{t,l}\phi_2(\tau_l+1)\xi_{t,l-1}\xi_{t-1,l-1} \\ &\quad +\phi_2^2(\tau_l+1)\xi_{t,l-1}\xi_{t-1,l-2}\xi_{t-1,l-1} \\ &= \xi_{t,l}^3 + 2\phi_2(\tau_l+1)\xi_{t,l-1}\xi_{t-1,l-1}\xi_{t,l} + \phi_2^2(\tau_l+1)\xi_{t,l-1}\xi_{t-1,l-2}\xi_{t-1,l-1}, \end{aligned}$$

that is, $\xi_{t,nl}$ is equal to the sum of four ($p^{n-1} = 2^2$; $i_1 = i_2 = 0$, $i_1 = i_2 = 1$, $i_1 = 0$ and $i_2 = 1$, and $i_1 = 1$ and $i_2 = 0$) terms, each of which is the product of three ($n = 3$) ξ's (continuant determinants).

Next we will prove Proposition 6 by mathematical induction. For $n = 2$ the result has been proved in eq. (B.1). If we assume that eq. (B.3) holds for n then it will be sufficient to prove that it holds for $n+1$ as well.

Proof. (**Proposition** 6) Assume that

$$\xi_{t,nl} = |\Phi_{t,nl}| = \sum_{i_1=0}^{1} \cdots \sum_{i_{n-1}=0}^{1} \{\xi_{t,l-i_1}(\prod_{l=2}^{n-1} \varphi_{2,T-1}\xi_{t-i_{T-1},l-i_T-i_{T-1}})\varphi_{2,n-1}\xi_{t-i_{n-1},l-i_{n-1}}\}. \tag{B.4}$$

Similarly to eq. (B.1) we can express $\xi_{t,(n+1)l}$ as the determinant of a 2×2 block matrix:

$$\begin{aligned}\xi_{t,(n+1)l} &= \begin{vmatrix} \mathbf{\Phi}_{t,l} & \overline{\mathbf{0}} \\ \widetilde{\mathbf{0}}_t & \mathbf{\Phi}_{t,nl} \end{vmatrix} = |\mathbf{\Phi}_{t,nl}|\,|\mathbf{\Phi}_{t,l}| + \phi_2(t-nl+1)\,|\mathbf{\Phi}_{t,nl-1}|\,|\mathbf{\Phi}_{t-1,l-1}| \\ &= \xi_{t,nl}\xi_{t,l} + \phi_2(t-nl+1)\xi_{t,nl-1}\xi_{t-1,l-1}, \end{aligned} \tag{B.5}$$

where $\widetilde{\mathbf{0}}_t$, is an $nl \times l$ matrix of zeros except for $\phi_2(t-nl+1)$ in its $1 \times l$ entry and the second equality follows from eq. (B.1). Combining eqs. (B.4) and (B.5) yields

$$\begin{aligned}\xi_{t,(n+1)l} &= \sum_{i_1=0}^{1}\cdots\sum_{i_{n-1}=0}^{1}\{\xi_{t,l-i_1}(\prod_{T=2}^{n-1}\varphi_{2,T-1}\xi_{t-i_{T-1},l-i_T-i_{T-1}})\varphi_{2,n-1}\xi_{t-i_{n-1},l-i_{n-1}}\}\xi_{t,l} + \\ &\sum_{i_1=0}^{1}\cdots\sum_{i_{n-1}=0}^{1}\{\xi_{t,l-i_1}(\prod_{T=2}^{n-1}\varphi_{2,T-1}\xi_{t-i_{T-1},l-i_T-i_{T-1}})\varphi_{2,n-1}\xi_{t-i_{n-1},l-1-i_{n-1}}\}\phi_2(t-nl+1)\xi_{t-1,l-1} \\ &= \sum_{i_1=0}^{1}\cdots\sum_{i_n=0}^{1}\{\xi_{t,l-i_1}(\prod_{T=2}^{n}\varphi_{2,T-1}\xi_{t-i_{T-1},l-i_T-i_{T-1}})\varphi_{2,n}\xi_{t-i_n,l-i_n}\}, \end{aligned} \tag{B.6}$$

which completes the proof. ■

Corollary 3 *For the CAR(2; l; d) process, in eq. (13), with* $0 \leq d \leq l-1$, $\xi_{t,l} = |\mathbf{\Phi}_{t,l}|$ *(see eq. (14)) can be written as*

$$\xi_{t,l} = \sum_{i_1=0}^{1}\cdots\sum_{i_d=0}^{1}\{\xi_{t,l_1-i_1}(\prod_{j=2}^{d}\varphi_{2,j-1}\xi_{t-l_{j-1}-i_{j-1},l_j-l_{j-1}-i_j-i_{j-1}})\varphi_{2,d}\xi_{t-l_d-i_d,l-l_d-i_d}\}, \tag{B.7}$$

where $\varphi_{2,j}$ *is defined similarly to the one in eq. (B.2), i.e.,* $\varphi_{2,j} = \phi_2(t-(l_j-l_{j-1})+1)$ *if* $i_j = 1$ *(the proof is along the lines of that of Proposition (6) above).*

Proposition 7 *For the ABAR(2; r) process in eq. (15) with* r, $0 \leq r \leq k-1$, *abrupt breaks at times* $t-k_1$, $t-k_2$, ..., $t-k_r$, $\xi_{t,k}$ *in eq. (16) can be written as*

$$\xi_{t,k} = \sum_{i_1=0}^{1}\cdots\sum_{i_r=0}^{1}\{\xi_{t,k_1-i_1}(\prod_{j=2}^{r}\varphi_{2,j-1}\xi_{t-k_{j-1}-i_{j-1},k_j-k_{j-1}-i_j-i_{j-1}})\varphi_{2,r}\xi_{t-k_r-i_r,k-k_r-i_r}\}. \tag{B.8}$$

where $\varphi_{2,j}$ *is defined similarly to the one in eq. (B.2) (the proof is similar to that of Proposition (6) above).*

C APPENDIX

Vector Seasons Representation

For the benefit of the reader this Appendix reviews some results on PARMA models. Recall that the drift and the autoregressive coefficients are periodically varying: $\phi_m(t) = \phi_m(\tau_n)$, $m = 0, 1, 2$, $\tau_n = \tau - nl$. Recall also that t_s denotes time at the sth season: $t_s = Tl + s$, $s = 1, \ldots, l$, and that we can write $\phi_m(Tl + s) = \phi_{m,s}$ (see eq. (8)).

We assume without loss of generality that time t is at the lth season (e.g., $t = t_l = (T + 1)l$). Thus our $\mathbf{\Phi}_{t,l}$ matrix in eq. (11) becomes:

$$\mathbf{\Phi}_{t,l} = \mathbf{\Phi}(l) = \begin{pmatrix} \phi_{1,1} & -1 & & & & \\ \phi_{2,2} & \phi_{1,2} & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & \phi_{2,l-1} & \phi_{1,l-1} & -1 \\ & & & \phi_{2,l} & \phi_{1,l} \end{pmatrix}.$$

A convenient representation of the PAR$(2;l)$ model (8) is the VAR(1) representation- hereafter we will refer to it as the vector of seasons (VS) representation (see, for example, Tiao and Guttman, 1980; Osborn, 1991; Franses, 1994, 1996a,b; del Barrio Castro and Osborn, 2008).

The corresponding VS representation of the PAR$(2;l)$ model (ignoring the drifts) is given by

$$\mathbf{\Phi}_0 \mathbf{y}_T = \mathbf{\Phi}_1 \mathbf{y}_{T-1} + \boldsymbol{\varepsilon}_T, \tag{C.1}$$

with $\mathbf{y}_T = (y_{1T}, \ldots, y_{lT})'$, $\boldsymbol{\varepsilon}_T = (\varepsilon_{1T}, \ldots, \varepsilon_{lT})'$, where the first subscript refers to the season (s) and the second one to the 'period' (T). Moreover, $\mathbf{\Phi}_0$ is an $l \times l$ parameter matrix whose (i, j) entry is:

$$\begin{cases} 1 & \text{if} \quad i = j, \\ 0 & \text{if} \quad j > i, \\ -\phi_{i-j,i} & if \quad j < i, \end{cases}$$

and $\mathbf{\Phi}_1$ is an $l \times l$ parameter matrix with (i, j) elements $\phi_{i+l-j,i}$, (see, for example, Lund and Basawa, 2000, and Franses and Paap, 2005).

As pointed out by Franses (1994), the idea of stacking was introduced by Gladyshev (1961) and is also considered in e.g., Pagano (1978). Tiao and Guttman (1980), Osborn (1991) and Franses (1994) used it in the AR setting. The dynamic system in eq. (C.1) can be written in a compact form

$$\mathbf{\Phi}(B)\mathbf{y}_T = \boldsymbol{\varepsilon}_T \text{ or } |\mathbf{\Phi}(B)|\, \mathbf{y}_T = adj[\mathbf{\Phi}(B)]\boldsymbol{\varepsilon}_T$$

where $\mathbf{\Phi}(B) = \mathbf{\Phi}_0 - \mathbf{\Phi}_1(B)$. Stationarity of $\mathbf{y}_T$ requires the roots of $\left|\mathbf{\Phi}(z^{-1})\right| = 0$ to lie strictly inside the unit circle (see, among others, Tiao and Guttman, 1980, Osborn, 1991; Franses, 1994, 1996a; Franses and Paap, 2005; del Barrio Castro and Osborn, 2008).

As an example, consider the PAR$(2;4)$ model

$$y_{t_s} = \phi_{1,s} y_{t_s-1} + \phi_{2,s} y_{t_s-2} + \varepsilon_{t_s},$$

which can be written as

$$\mathbf{\Phi}_0 \mathbf{y}_T = \mathbf{\Phi}_1 \mathbf{y}_{T-1} + \boldsymbol{\varepsilon}_T,$$

for which the characteristic equation is

$$|\mathbf{\Phi}_0 - \mathbf{\Phi}_1 z| = \begin{vmatrix} 1 & 0 & -\phi_{2,1}z & -\phi_{1,1}z \\ -\phi_{1,2} & 1 & 0 & -\phi_{2,2}z \\ -\phi_{2,3} & -\phi_{1,3} & 1 & 0 \\ 0 & -\phi_{2,4} & -\phi_{1,4} & 1 \end{vmatrix} = 0.$$

Hence, when the nonlinear parameter restriction

$$\begin{aligned} &\left|\phi_{2,2}\phi_{1,3}\phi_{1,4} + \phi_{2,2}\phi_{2,4} + \phi_{2,1}\phi_{1,2}\phi_{1,3} + \phi_{2,1}\phi_{2,3} + \phi_{1,1}\phi_{1,2}\phi_{1,3}\phi_{1,4}\right. \\ &\left.+\phi_{1,1}\phi_{1,2}\phi_{2,4} + \phi_{1,1}\phi_{1,4}\phi_{2,3} - \phi_{2,1}\phi_{2,2}\phi_{2,3}\phi_{2,4}\right| < 1, \end{aligned}$$

is imposed on the parameters, the VS representation of the PAR$(2;4)$ model is stationary (see Franses and Paap, 2005). When $\phi_{2,s} = 0$ for all s, that is we have the PAR$(1;4)$ model, then the stationarity condition reduces to: $\left|\phi_{1,1}\phi_{1,2}\phi_{1,3}\phi_{1,4}\right| < 1$, which is equivalent to our condition $\left|\xi_{t,l}\right| < 1$, or in other words, that the absolute value of $|\mathbf{\Phi}(l)|$ is less than one.

CHAPTER 6

An Econometric Investigation of Inflation and its Persistence using AR, ADL and GARCH in Mean models

1 Introduction

On the inflation mechanism, scholars and practitioners are interested in: Its causes; its effects, both benefits and costs; its reduction, methods of reduction; and its control, once reduced to the target rate. In recent times control has been the major policy objective and thus the preocuppation of central bankers. Following the great recession, the actual inflation rate has been persistently below the target rate, set by the central banks. Rates of inflation that have stayed persistently below the 2% targets but above zero have been called *lowflation*, adding a new term to the economists' lexicon. There are both domestic and international factors that keep inflation low, to which the present paper tries to shed light on. For example, the sharp decrease in the price of oil in 2015 and its pass through to the price indices.

Inflation is the process of prices increasing over an extended period; is not a once for all change in the price level, it is a sustained increase in the price level. Utilizing the definition we can break the study of the inflation rate into two steps: first, what causes inflation and second what perpetuates a specific inflation level. In the present chapter we focus in the second part, namely the persistence of inflation.

The more persistent inflation is, the hardest it becomes for the monetary authorities to manipulate it. Lack of understanding the effects of inflation persistence may steer the monetary authorities to pursue inappropriate monetary policy.

2 Are there any Stylised Facts on Inflation?

There are three main empirical regularities concerning changes in inflation dynamics during the post World War II period.

2.1 Lower Persistence

The first is the question of whether inflation persistence has remained constant or whether it has decreased. Most empirical studies provide evidence of inflation becoming less persistent. This tendency has contributed to low inflation (see Ceccheti and Debelle, 2006). Shocks to the price level produce smaller and shorter effects on the inflation rate. The empirical findings that indicate lower inflation persistence call for a theoretical explanation.

2.2 Flatter Phillips Curve

The second is that the Phillips Curve has become flatter. Leading macroeconomists and central bankers (e.g. Olivier Blanchard, 2016 and Stanley Fisher, 2016) argue that the Phillips curve has become flatter moving closer to the horizontal Keynesian Phillips curve portrayed in intermediate textbooks.

2.3 Inflation Less influenced by Supply Shocks

Third, inflation has become less elastic to external shocks, like energy prices and import prices. In short, the pass through is smaller. Frederic Mishkin (2007) argues that these changes result from the anchoring of inflation expectations due to better monetary policy. In the present work we study the first two regularities but we also make references to the third.

2.4 A Puzzle:The deflation that did not appear after the Great Recession

In addition, in the Great Recession, inflation has not decreased as much as was predicted by all the econometric models. This short horizon phenomenon has acquired the status of a non-repeating puzzle that cries out for a sound explanation. In the spirit of Joan Robinson (1976) we believe that the power of multinational corporations to control prices - in the new era of monopoly (Joseph Stiglitz, 2016) - has saved the advanced economies from falling into deflation.

Corresponding to the regularities are three hypotheses that can be tested.

3 Preview of Data

3.1 The Primary Data

Our basic time series are monthly observations provided by FRED and it has been transformed to quarterly through averaging; in other words, the aggregation method is average. We have mainly worked with U.S. and U.K. data. We have also used French and Austrian data for GARCH in Mean estimations. We will illustrate characteristics and results with graphs for the U.S.A. data and only illustrate the corresponding graphs for the U.K. data in case it makes a different point. Exhibit 1 shows the quarterly Consumer Price Index (CPI) All Urban Consumers All Items for the U.S.A., denoted by p_t. So, the interval between observations, that is the period of time between observation t and observation $t+1$ is a quarter of a year. The number on the vertical axis is a CPI index, which represents the data of a price level of a basket of goods-services, purchased at different points in time. On the horizontal line we have the time in quarters running from from 1947-q1 to 2015q4, corresponding to the sample period. In the regressions we run, we point out our effective sample. That is, the figure shows 272 quarterly time series data on a particular price index. The series is indexed so that the average of the 12 quarterly values of CPI between 1982 to 1984 is 100. A CPI of 130 in 2014q3 means that we have paid \$130 to purchase the basket of goods that in 1982-1984 we purchased with \$100. ftbpF4.5792in3.1358in0ptFigure

Figure 1. First observation 1947Q1=21.700 and last observation 2015Q4=238.097

The series p_t displays exponential growth. In particular, the price index is increasing over time, which means that there has been a steady increase in consumer prices. We observe some variation; for example, there are several slight falls in the price index corresponding to the Volker disinflation of the early 1980's and to the Great Recession in 2008-2009. But, in general, the time series plot is a curve with a positive slope. It dispalys a sustained upward movement, that is, an upward trend.

3.2 Transformations of the Data to obtain Time series of Inflation

To obtain a series for inflation we transform the series :

$$y = \frac{[p-p(-1)]}{p(-1)} \times 400.$$

The number of observations is 275.

Exhibit 2 plots the quarterly inflation rate.

ftbpF4.5515in3.1358in0ptFigure

ftbpF8.4354in3.7421in0ptFigure

Figure 2 The US Inflation Rate from 1947 to 2015

The line graph of the inflation series looks different to the plot of the price index. The trend behaviour exhibited in the price index figure has disappeared.

It shows that the inflation rate has varied over the years, which does not show in the plot of the price index. Does the plot of the inflation series look independent and identically distributed? From the line graph, we cannot say whether the series is independent. On the contrary, we observe that the value of inflation in the current quarter is not far from its value in the previous quarter. The inflation rate shows some degree of persistence. It experienced two upward surges in the 1970's and remained at high levels both in the early 1970's and in the late 1970's. Following a sharp decline in the early 1980's associated with the Volker disinflation, it has remained below 3%-4% for over three decades.

3.3 The Inflation Rate as a Stochastic Process

Let y_t denote the value of inflation in time period t. We cannot predict inflation perfectly; accordingly, y_t is a random variable. When we learn the value of inflation in period t, then y_t is one of the realized values from a stochastic process. Since we have 275 quarterly observations for the inflation rate, we have a series of 275 random variables.

From the time plot we observe that the distribution of inflation rates varies over time; it is time varying. The questions we face: Are there any empirical evidence of time-varying means, variances and covariances? If they are time-varying, we cannot employ the standard multiple regression model, which regards them as consatnt. We need to specify models, which allow them to vary over time. Next, we provide a skeleton of such a prototype model.

Let the information available to us at time t, which might be useful in predicting the variation in the future distribution at time $t+1$, be denoted by I_t. The information set I_t may include past inflation rates, Fed announcements and other macroeconomic information, such as recent unemployment rates - a la Phillips curve. In the current work we develop time series models, which mainly utilize past inflation rates. In general, I_t includes all available information relevant to predicting inflation.

Given the I_t, we can put forward a general time-varying setting where the distribution of inflation rates changes over time. The conditional distribution of inflation rates varies as a function of I_t:

$$f(y_{t+1} \mid I_t)$$

In the sequel, we study the link between I_t and the future distribution of inflation rates, y_{t+1}. We assume that inflation rates are conditionally normal. The number of observations, denoted $T = 275$, can be represented as:

$$y_{1947q1},\, y_{1947q2}, \ldots,\, y_t, \ldots, y_{2015q4}$$

It is our sample. The question we try to answer is: what is our best prediction for next quarter's, y_{2016q1}, inflation rate?

3.4 Characterizing the Inflation Data: The Autocorrelation Function

One typical propery of many time series data is the presence of correlation across adjacent observations. A positive correlation between two adjacent inflation rates means that when the inflation rate in period t is below the sample average, it is likely that the inflation rate in period $t+1$ will also be below the sample average. The correlations between the current value of inflation y_t and its lagged values $y_{t-1}, y_{t-2}, ..., y_{t-k}$, for k lags, are called autocorrelations. We obtain the autocorrelations by dividing the autocovariances by the variance of y_t. The autocorrelations are numbers independent of the units of measurement of the underlying process; viewed as a function of integer valued lags k, are known as the **autocorrelation function (ACF), or the correlogram** of the inflation series. That is, for any lag k, the ACF assigns a number for the correlation between y_t and the value of y at that specific lag. It assigns ρ_k for $k = ..., -1, 0, 1, ...$ The sequence ρ_k for $k = ..., -1, 0, 1, ...,$ indicates the extent to which one value of the process is correlated with previous ones. The ACF is employed to model the dependencies among our time series inflation data. In particular, we use the ACF to characterize the process describing the evolution of inflation during the sample period. As Furher (2011) points out, most of the various measures of inflation persistence, that we will meet in the sequel, can be viewed as efforts to quantify the rate at which inflation's autocorrelations decline.

To compute the correlogram of our inflation series we have to specify two things. First, whether we want to obtain the correlogram for the level of the series or for its first or second difference. We have chosen to work with the level of the inflation series. Second, we must specify the order of the lag. It is a common practice to specify multiples of the seasonal period; since we are working with quarterly data, we use multiples of 4, in particular 16 lags. The resulting correlogram is portrayed in Figure 3.

ftbpF4.5117in3.1349in0ptFigure

The $\rho_1 = 0.732$, is pretty high and its rate of decay is slow; it goes down to $\rho_8 = 0.233$. Then is starts to increase again and hovers around 0.3 in the next 6 lags with the $\rho_{16} \simeq 0.25$. Since $\rho_1 = 0.732 > 0$, we have an indication that the inflation series is first order serially correlated. The value of ρ decays gradually with increasing k, which indicates that the series might well be represented by a low order autoregressive (AR) process. Since the value of ρ does not go down to zero after several lags, we exclude the possibility of the series following a low order moving average (MA) process.

One of the reasons that we get autocorellations in the inflation series is that the information set relevant to inflation is likely to be similar in neighbouring periods. Before we proceed to represent the autocorrelated inflation series with a low order autoregressive model , we test to decide whether our inflation series is stationary.

3.5 Unit Root Testing to check whether the Inflation Rate is Stationary

Looking at the graph of the inflation data we observe that the inflation rate is not growing. Utilizing this information, as suggested in the text by James Hamilton

(1994, p.501) and the survey by James Stock (1994, p.2829), we proceed to consider the following equation:

$$y_t = y_{t-1} + \alpha + \varepsilon_t$$

where α is a constant.

Rearranging we get

$$\triangle\, y_t = (\rho - 1)y_{t-1} + \alpha + \varepsilon_t$$

There are two possibilities: Either there is a unit root ($\rho = 1$) and a zero intercept ($\alpha = 0$), which means that there is no drift term to produce a trend; or there is no unit root ($\rho < 1$) and a nonzero intercept, so y_t will be stationary around the mean ($\frac{\alpha}{1-\rho)}$).

The unit root test we performed is an Augmented Dickey Fuller (ADF) test in levels, including only an intercept as a deterministic regressor. The critical value for the ADF test in our regression, which does not include a time trend, at the 5% significance level is -2.86. Since the ADF statistic we obtained is -4.192, we have that

$$-4.192 < -2.86,$$

and therefore the null hypothesis of a unit root is rejected. We conclude that the U.S. inflation is stationary. But if inflation is not mean stationary or close to being nonstationary, then a stationary modelling of the dynamics of inflation tends to understate the true persistence (see Furher, 1995, p.8).

We have also conducted an Augmented Dickey Fuller test on the U.K. inflation data with both a maximum lag 15 and a maximum lag of 4. The value obtained for the ADF test statistic is -2.80, which in the limit is smaller than the critical value of -2.87 at the 5% level: $-2.80 > -2.87$. Therefore, we do not reject the null hypothesis of a unit root. Inflation with our U.K. quarterly data is not stationary.

4 The Persistence of Inflation

One of the properties of inflation that has received attention, especially by empirical workers, is its persistence. Inflation is persistent if it remains in the neighborhood of its recent value, if no other forces cause it to move. To express it the other way around, it is the time it takes for the effects of a shock to dissapear. Furher (1995) gives a more accurate definition of inflation persistence as its propensity to remain different from its mean value. Furher's definition presupposes an existing mean, which is a sensible assumption.

A rudimentary measure of inflation persistence is its autocorrelation function; it provides the correlation of inflation with its own history. When an event cause inflation to move away from its average level and if it tends to stay away, then we observe positive autocorrelation in the rate of inflation. But when events cause inflation away from its norm and it comes back immediately to its norm, we observe low or no autocorrelation in the rate of inflation.

Insert the autocorrelation function of inflation. Then comment on the graph like: If inflation exhibits persistence, then its autocorrelation dies out slowly. (Fuhrer, 1995)

To evaluate the changes on inflation persistence, we must measure how long the impact of a shock to inflation will be present: whether inflation will rapidly return to its previous state (ex ante value) or whether the impact of the shock will lead to a changed (ex post value) level for a prolonged period. In the terminology of time series, inflation persistence is concerned with whether and how quickly inflation goes back to its trend rate, following a shock. In his professional survey on inflation persistence, Fuhrer (2011) distinguishes between reduced form persistence and structural persistence.

4.1 Metrics of Persistence

There are three (3) main measures of persistence in the context of univariate models:

First, we can measure the persistence as the largest AR root in the inflation rate (see Stock and Watson, 2007).

Second, we can measure the persistence as the sum of the AR parameters in a time-varying parameter model (see Pivetta and Reis, 2007). That is, we can measure inflation persistence by regressing inflation on its own lags and then compute the sum of the coefficients on lagged inflation. If the sum of the coefficients is close to one, then shocks to inflation have long lived effects on inflation; i.e. inflation is a random walk: when inflation goes up, it stays up. If the sum of coefficients is less than one, then a shock to inflation has a temporary effect, and inflation reverts to its trend.

Third, we can use the half-life response of inflation to a shock (see Pivetta and Reiss, 2007).

4.2 Explanations of Persistence

The first explanation of (inherent) persistence was advanced by Jordi Gali and Mark Gertler (1999). They base their thesis on the view that persistence arises when not all the price setters in the economy are forward looking. Accordingly, they decompose the price setters into two categories. The first, has rational expectations. The second sets prices by looking at past inflation. The backward looking pricing behaviour generates inflation persistence. Gali and Gertler model this dichotomous behaviour with "a hybrid" Phillips curve, in which current inflation is a function of both expected future inflation and past inflation."

The second explanation says that inflation persistence is due to monetary policy. It is developed by Dotsey (2002) and Sbordone (2006).

The third explanation places the emphasis on the characteristics of shocks. It has been articulated by Lubik and Schorfheide (2004).

5 Univariate A-theoretical models: Explaining Inflation using past Inflation Rates

Since both the informal and formal tests indicate that our observations are not independent, we proceed to formulate the time dependence with time series models. We focus on univariate models, that is models with a single variable, namely inflation. Univariate modelling is subject to omitted variable bias as it may leave out other causes of change in the inflation process. As a response to this problem in the next section we develop Autoregressive Distributed Lag, ADL, models.

5.1 An AR(1) model for Inflation

The AR models use the history of inflation to forecast its future. Viewing inflation as an AR process lead us to examine regressions where y_t is the outcome and where past values of the inflation process are the predictor variables.

When we studied our data above, we observed that the observations display a significant 1st order autocorrelation. The autocorrelation is an indication that the lagged value of inflation might be useful in predicting the current value of inflation. The simplest model that makes use of such predictive power is the $AR(1)$. The $AR(1)$ is in the same form as the Linear Regression Model in which y_t is the dependent variable and y_{t-1} is the regressor.

To predict the future of a time series that represents inflation, we can employ the information contained in its immediate past. For example, if we want to forecast the change in inflation from the current quarter to the next, we can examine whether inflation rose or fell last quarter. The $AR(1)$ model says that *conditional* on the immediate past value of inflation, y_{t-1}, the current value is centered around $\phi_0 + \phi_1 y_{t-1}$:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t.$$

The parameter ϕ_0 allows for the possibility of a nonzero mean. The $AR(1)$ is a first order stochastic linear equation, that is a straight line. It is based on the Markov property which says that conditional on y_{t-1}, the current value y_t is not correlated with y_{t-k} for $k > 1$. The Markovian property is represented by the conditional expectations operator which computes the predictor of y_t using the information available at time $t-1$:

$$E_{t-1}(y_t) = \phi_0 + \phi_1 y_{t-1}$$

The quantities ϕ_0 and ϕ_1 are unkown parameters and ε_t is a disturbance with zero mean and variance σ^2.

5.1.1 Estimation of the AR(1) Inflation Process

Estimation of the unkown parameters ϕ_0 and ϕ_1 is accomplished by applying ordinary least squares. We want to fit an $AR(1)$ model to the observations for the inflation series. In levels, the estimated equation is:

$$\underset{(0.51)}{y_t = 3.45} + \underset{(0.041)}{0.73} y_{t-1} + \varepsilon_t$$

In parantheses are the standard errors and ε_t is independent of $y_{t-1}, y_{t-2}, ...$ The parameter estimates are employed to find out the contribution of past information by implementing tests on the parameters. The estimated $AR(1)$ says that the current value of inflation is made up of two parts: the first, the quantity $3.45 + 0.73y_{t-1}$, is the part that depends on the past; the second is ε_t, which is not predictable from the past. The $AR(1)$ implies that knowing previous lagged values of inflation, $(y_{t-2}, y_{t-3}, ...)$, does not help us to predict the current inflation y_t if we know y_{t-1}.

Although the inflation data exhibit non-randomness, having fitted y_t against y_{t-1}, we expect to get random resisuals.

A problem with $AR(1)$ is that there are cases in which y_{t-1} alone cannot determine the conditional expectation of y_t and a more flexible model is needed. A generalization of the $AR(1)$ is the $AR(p)$ model to which we turn our attention in the next subsection.

5.2 AR(p) models

General AR processes allow for the possibility that y_t may not only depend on y_{t-1}, but also on y_{t-2}, y_{t-3} and on previous past values of y. The general dependencies are modelled with an $AR(p)$, where p is a non negative integer. To develop our $AR(p)$ inflation process, we must empirically identify the order p for the underying inflation series. The general $AR(p)$ is:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$$

where ϕ_0 is an intercept parameter related to the mean of y_t; the $\phi_i s$ are the unkown AR parameters and ε_t are errors assumed to be uncorrelated random variables with mean zero and variance σ^2_ε. Now we have to estimate $p + 1$ parameters: $\phi_0, \phi_1, \phi_2, ..., \phi_p$. The $AR(p)$ says that

5.2.1 Order Determination

The order p of an $AR(p)$ is unkown; it has to be found empirically. The process is called order determination of AR models. Two approaches can be followed to determine the value of p. The first approach employs the Partial Autocorrelation Function, denoted with the acronym $PACF$. The second approach employs some Information Criterion Function. We apply them in turn beginning with $PACF$. We obtain some information about the oscillatory behaviour of the sample ACF, but we can get more information from the $PACF$.

Estimating an $AR(4)$ for the inflation rate y_t we obtained the following equation.

$$y_t = 3.3 + 0.63y_{t-1} - 0.017y_{t-2} + 0.33y_{t-3} - 0.173y_{t-4} + \varepsilon_t$$

The $PACF$ method for estimating the order of the $AR(p)$ model says that the lag$-p$ value of the $PACF$ should be different from zero and the values for higher than p orders should be approximately equal to zero (see Ruey Tsay, 2005, pp.40-41).

6 Autoregressive Distributived Lag (ADL) Model: Forecasting the Inflation Rate Using Past Unemployment Rates

We turn our attention to time series models with exogenous variables as regressors. The question adressed is whether inflation forecasts made using lagged values of the rate of unemployment in addition to lagged inflation, that is, forecasts based on an empirical Phillips curve, improve on the AR inflation forecasts. In other words, whether ADL forecasts of inflation are more accurate than AR forecasts.

Keynes writing in 1940 on "How to Pay for the War" traced the roots of inflation to an excess of aggregate demand over real income. When an increase in government spending pushes the level of income above its equilibrium, it generates an *inflationary gap*. The Keynesian theory of the inflationary gap cried out for empirical support, which came in the form of the Phillips Curve.

6.1 The Phillips Curve

The Phillips Curve expresses relationships between inflation and real variables. The original Phillips curve (Phillips, 1958) depicted an inverse relationship between wage inflation and unemployment. As the power of trade unions progressively declined - one of the factors in operation - wage inflation was replaced by price inflation. So, its most common manifestation represents a negative relationship between price inflation and unemployment. The influence flows from unemployment to inflation. In a recent report, Olivier Blanchard (2016) concludes that current empirical evidence indicate that this type of causation is still valid. The introduction of the Phillipc curve shifted the attention from the root causes of inflation to the dynamics of the labour market.

For macroeconomic and monetary policy, the Phillips curve can be viewed as a vehicle of representing how money may not be neutral. Different stances on monetary neutrality can be resolved with references to the distinction between the short run and the long run. Monetary neutrality is relevant to the long run.

As Clive Granger and Yongil Jeon (2011) say one of the two main characteristics of the Phillips curve is that it is *non-linear*: successive reductions in the unemployment rate result in ever higher increases in the inflation rate.

For empirical testing, the inflationary gap theory has numerous shortcomings. The most serious is the difficulty of finding an appropriate measure which will indicate how close to full employment we actually are. In particular, we need a measurable indicator of how near to full capacity the economy is operating. One of the contributions of Phillips is his suggestion that a good indicator of capacity utilization is the unemployment rate. An increase in the level of capacity utilization, say, due to expansionary policy, will be shown in a decrease in the unemployment rate. Fluctuations in the unemployment rate can be viewed as observable manifestations of changes in the underlying aggregate demand.

Paul Samuelson and Robert Solow in 1960 using U.S. data from the beginning of the 20th century to 1958, found a pattern similar to Phillips. It was with the publication of Samuelson and Solow (1960) that economists and policy makers began to recognise the implications of a *stable* Phillips curve. If an inverse relationship existed, then the authorities could manage aggregate demand in such a way as to balance an acceptable unemployment rate with an acceptable rate of inflation. A positive rate of inflation was viewed as the price which had to be paid to achieve near full employment and vice versa. The Phillips curve was regarded as a trade-off relation. It was the constraint which policy makers should take into account when they are tempted to manipulate aggregate demand.

Phillips fitted a curve which represented the data well for both high and low unemployment rates. It is due, as Clive Granger and Yongil Jeon (2011) say, to one of the two main characteristics of the Phillips curve, namely it is *non-linear*: successive reductions in the unemployment rate result in ever higher increases in the inflation rate.

6.2 A canonical specification of the (short-run) Phillips Curve

Milton Friedman (1968) put forward a short-run theory of inflation. In Friedman's model, known in the literature as the accelerationist Phillips curve, inflation depends on two factors: first, on expected inflation, π^e, and second, on the gap between recorded unemployment, u, and its natural level, u^N. The specification for Friedman's theory is:

$$y_t = y_t^e + (u - u^N)_t + \varepsilon_t$$

where y is annualised quarterly inflation;
y^e is expected inflation;
u is actual observed unemployment;
u^N is the natural rate of unemployment; and
ε is an error assumed to be uncorrelated (orthogonal) with $(u - u^N)$.

Following Friedman, numerous authors have derived equations, similar to the one above, from models in which price setters have incomplete information or in which nominal prices are sticky.

In line with Ball and Mazumder (2011), we assume backward looking expectations: expected inflation is approximated by past inflation. In particular, expected inflation is the average of inflation in the past 4 quarters:

$$y_t = \frac{1}{4}(y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4}) + a(u - u^N)_t + \varepsilon_t$$

This Phillips curve type includes lags of inflation with unrestricted coefficients, except for the accelerationist assumption that the sum of coefficients is equal to 1.

7 Is Inflation a (G)ARCH?

7.1 Friedman's Conjecture

The stagflation of the 1970's has brought into question the empirical validity of the long-run trade off between inflation and unemployment, indicated by the empirical breakdown of the Phillips curve. The long run trade off was earlier criticized on theoretical grounds by the Phelps ( ) and Friedman (....). Friedman (1968) put forward the notion of the natural rate of unemployment.

Responding to the stagflation experience of the 1970's, Friedman (1977), in his Nobel lecture, recognized the non-constancy of the natural rate of unemployment. He argued that higher inflation can be held responsible for higher unemployment and lower growth in output. In other words, Friedman proposed a temporary positively sloped Phillips curve.

One mechanism that contributes to the change in the slope of the short-run Phillips curve from negative to positive is the role of inflation uncertainty. Friedman's thesis was that higher inflation uncertainty shortens the average duration of contracts and distorts the efficiency of the market signals concerning actual and expected relative prices. To put it in another way, to explain the coexistence of inflation and unemployment in the 1970's, Friedman (1977) hypothesized that higher inflation causes greater uncertainty about expected future prices (i.e. inflation uncertainty), which in turn implies a less efficient allocation of resources. The disruption of the market mechanism due to inflation uncertainty causes unemployment to increase above its natural rate.

7.2 Engle's ARCH and Bollerslev's GARCH modelling of Inflation

One of the economic motivations behind Engle's (1982) development of the Autoregressive Conditional Heteroscedasticity, ARCH, model was to investigate Friedman's (1977) conjecture. It was later generalised by Bollerslev (1986) into GARCH. How can be Friedman's conjecture be parametrized?

All the models we introduced up to now were interested in the rate of inflation itself. But as our discussion of Friedman's arguments made clear, in some cases we are not interested in the inflation rate itself but in its variability, measured by its variance. Until 1982, empirical workers modeled heteroskedasticity by introducing ad hoc explanatory variables and by performing data transformations. Engle (1982) broke with this tradition with the ARCH methodology, which simultaneously models the mean and the variance of the inflation series. General ARCH-type models are made up of two equations. The first is a regression to the mean equation. The second is a volatility equation, where volatility is defined as the time-varying variance of the regresssion equation.

7.3 Estimating GARCH-type Specifications

All the econometric models we applied thus far to our inflation data, namely the AR and ADL models, have a significant limitation: they assume a constant variance. The AR and ADL models are not appropriate for modeling the variance when is not constant. Up to now we ha studied Markovian lagged effects on the level

$$E(y_t \mid y_{t-1})$$

of the observed time series. There may exist lagged effects on the conditional variance

$$\sigma^2 = var(y_t \mid y_{t-1}).$$

We perform a joint estimation of the mean and variance equations.

7.3.1 Combined models for level and variance

In our formulation of the $AR(1)$ model we noticed that it has a nonconstant conditional mean and a constant conditional variance. The original $GARCH$ models we discussed in the previous subsection suppose that the inflation process has a zero conditional mean. Accordingly, in the pure $GARCH$ models the variance of inflation is predictable but the level of inflation is not predictable. If we want to capture the path dependency of both the conditional mean and the conditional variance of inflation, we have to combine features of both the AR and $GARCH$ models.

The first mixed model we specify is the $AR(1)$ for the level of the inflation rate, y_t, with $GARCH(1,1)$ for the variance of the innovations.

First, we specify the mean equation:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$$

Second, we need to specify the variance equation:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

where ε_{t-1}^2 is the $ARCH$ (moving average) term and σ_{t-1}^2 is the $GARCH$ (autoregressive) term.

That is, we study the inflation rate as a stochastic process with path dependent mean, whose errors follow a $GARCH$ model. Running the $GARCH$ regression using the default options for the variance equation, which means we do not enter variance regressors, results in the following fitted model:

$$y_t = 2.424 + 0.717 y_{t-1}$$

$$\sigma_t^2 = 0.116 + 0.274\varepsilon_{t-1}^2 + 0.754\sigma_{t-1}^2$$

The results show that the conditional variance coefficients, the one in front of the $ARCH$ effect (0.274) and the one in front of the $GARCH$ effect (0.754) are both positive and their sum is 1.028, which is approximately equal to one (1). The near unity sum of the two variance cofficients indicates that inflation shocks are persistent.

We found that the sum of the coefficients of the $ARCH$ and $GARCH$ terms is approximately one (1) indicating that inflation shocks are persistent. To put it in another way, is an indication that changes in the conditional variance are persistent.

Then we run an $AR(4)$ - $GARCH(1,1)$

ftbpF4.3422in3.1341in0ptFigure

The conditional standard deviation graph displays short-lived periods of high volatility.

7.4 AR(1) with GARCH (1,1) in Mean

When we examine financial and macroeconomic time series, we observe that in some series the level of the series under study depends on its variability. To capture this dependency, Robert Engle, David Lilien and Russel Robins (1987) extended the ARCH to allow the mean of a recorded time series to depend on its own conditional variance.

The first model we formulate is an $AR(1)$ with $GARCH(1,1)$ *in Mean*. It is given by:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \lambda\sigma_t^2 + \varepsilon_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$$
$$\text{where } \varepsilon_t \mid I_t \sim N(0, \sigma_{t-1}^2)$$

Let us digress for a moment to interpret the model in the context of assets markets, for which the extension is particularly suited. In financial applications, the parameter λ represents the risk premium that a risk averse investor requires to hold a risky asset instead of investing in a risk-free asset.

Given that an asset riskyness can be measured by the variance of returns, Engle et al. (1987) assume that the risk premium represented by the parameter λ, is an increasing function of the conditional variance of returns. A positive λ indicates that the return is positively correlated to its volatility: the greater the conditional variance of returns, the higher the premium necessary to incentivize the investor to hold the risky asset.

We apply the $GARCH - M$ to help us characterize the evolution of the mean and the variability of our inflation series, simultaneously. We employ the model to investigate whether as the variance of inflation changes over time, the value of inflation will change as well. Therefore, the $GARCH - M$ enable us to examine whether the relationship between the mean of the inflation series and its variance is positive but not constant (time varying).

By inserting the variance of inflation as a regressor into the equation that describes the average level of inflation, we hope to evaluate whether the variance of inflation can influence the mean of future inflation rates. Since it incorporates the conditional variance into the conditional mean equation, the process is called $GARCH$ *in mean* for GARCH effect in the mean. We have run regressions with three versions for the in mean effect: the standard deviation, the variance and the log of the variance. Below are the results for the standard deviation version; the results fot the other two are similar.

$$y_t = 1.69 + 0.76y_{t-1} + (-0.55)\sigma_{t-1} + \varepsilon_t$$

The coefficient of the conditional variance term that enters in the mean equation has a negative sign and is significant in all three variants of the general model. On the contrary, all the three coefficients: the $AR(1)$ coefficient and the ARCH-GARCH effects of the variance equation are positive and highly significant. Since the parameter λ is negative, we expect that higher volatility of inflation will cause the average level of inflation to fall.

8 Conclusion

We have examined the evolution of inflation and its persistence in the U.S. and the U.K.. As part of our pre-testing, we have conducted numerous unit root tests, based on the Dickey-Fuller work. It was revealed that the unit root tests have difficulty in differentiating on whether inflation is near to be a unit root process or not. Accordingly, we run several regressions both in levels and in differences.

The paper reports evidence that inflation is persistent, recently at below the 2% target set by the major central banks. The sluggish economic performance that followed the Great Recession did not produced the downward price pressure, anticipated by Phillips curve type of models. But the Phillips curve is still present in the data with a flatter slope. Our GARCH-in mean models suggests that the lower variance of the inflation rate, experienced in the past two decades seems to contribute to a lower mean.

A number of authors emphasize the stability of inflationary expectations in contributing to lower inflation rates. Whether the stable inflation expectations are linked to the credibility of monetary policy is controversial.

Other factors that have played a role in producing lowflation are the lower oil prices and the prices of other commodities. In the case of the US, the appreciation in the value of the dollar has also played a role. Therefore, it seems that a combination of factor are at work, which cannot all of them be represented by a single econometric model. Finally, we believe that more effort has to be put in empirically estimating the power of firms to control prices, especially on the international level.

References

[1] Atkeson, Andrew and Ohanian Lee (2001). "Are Phillips Curves Useful for Forecasting Inflation?", *Federal Reserve Bank of Minneapolis Quarterly Review*, vol.25, no.1,Winter, pp.2-11.

[2] Ball, Laurence and Mazumder, Sandeep (2011). "Inflation Dynamics and the Great Recession", *Brookings Papers on Economic Activity*, Spring, pp.337-405.

[3] Bassetto, Marco, Messer, Todd, and Ostrowski, Christine (2013). "Forecasting inflation and the Great Recesion", *Economic Perspectives*, Federal Reserve Bank of Chicago, vol. , 3rd quarter, pp.79-106.

[4] Beechey, Meredith and Osterholm, Par (2012). "The Rise and Fall of U.S. Inflation Persistence", *International Journal of Central Banking*, vol.8, No.3, pp.55- .

[5] Blanchard, Olivier (2016). "The US Phillipcs Curve: Back to the 60s?" Policy Brief, PB16-1, Peterson Institute for International Economics, Jan. 2016.

[6] Blanchard, Olivier and Katz, Laurence (1997). "What We Know and Do Not Know About the Natural Rate of Unemployment", *Journal of Economic Perspectives*, vol.11, no.1, Winter, pp.51-72.

[7] Bollerslev, Timothy (1986). "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31, pp.307-327.

[8] Brave, Scott and Fisher Jonas (2004). "In search for a robust inflation forecast", *Economic Perspectives*, Federal Reserve Bank of Chicago, vol.40, 3rd quarter, pp.12-31.

[9] Cecchetti, G. Stephen and Debelle Guy (2006). "Has the inflation process changed?", *Economic Policy,* April, pp.311-352.

[10] Cecchetti, Stephen, Peter Hooper, Bruce Kasman, Kermit Schoenholtz and Mark Watson (2007). "Understanding the Evolving Inflation Process", U.S. Monetary Policy Report.

[11] Cogley, Timothy., Primiceri, Giorgio., and Sargent Thomas (2010). "Inflation-Gap Persistence in the US", *American Economic Journal: Macroeconomics*, vol. 2, no.1, pp.43-69.

[12] Conrad, Christian and Karanasos Menelaos (2005). "On the inflation-uncertainty hypothesis in the U.S.A., Japan and the UK: a dual long memory approach", *Japan and the World Economy*, vol.17, pp.327-343.

[13] Cosimano, Thomas F. and Jansen, Dennis (1988). "Estimates of the Variance of U.S. Inflation Based upon the ARCH Model", *Journal of Money, Credit, and Banking*, Vol.20, No.3, August, pp.409-421.

[14] Engle Robert (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, Vol.50, No.4, July, pp.987-1007.

[15] Engle Robert (1983). "Estimates of the Variance of U.S. Inflation Based upon the ARCH Model", *Journal of Money, Credit, and Banking*, Vol.15, No.3, pp.286-301.

[16] Faust, Jon and Wright Jonathan (2014). "Forecasting Inflation", *Handbook of Economic Forecasting*, vol.2A.

[17] Fischer, Stanley (2016). "Reflections in macroeconomics then and now" Speech delivered at the "Policy Challenges in an Interconnected World", 32nd Annual NABE Conference, Washington DC, 7 March 2016.

[18] Fuhrer, Jeffrey (1995). "The Persistence of Inflation and the Cost of Disinflation", *New England Economic Review*, January/February, pp.3-16.

[19] Fuhrer, Jeffrey (2011). "Inflation Persistence", Chapter 9 in *Handbook of Monetary Economics*, vol.3A, edited by Friedman, B. and Woodford, M. eds, pp.423-486, North Holland.

[20] Fuhrer, Jeff and Moore, George (1995). "Inflation Persistence", *The Quarterly Journal of Economics*, Feb., pp.127-159.

[21] Gali Jordi and Gertler Mark (1999). "Inflation dynamics: A structural econometric analysis", *Journal of Monetary Economics*, vol.44, pp.195-222.

[22] Gordon Robert (1997). "The Time-Varying NAIRU and its Implications for Economic Policy", *Journal of Economic Perspectives*, vol.11, no.1, Winter, pp.11-32.

[23] Gordon Robert (1998). "Foundations of the Goldilocks Economy: Supply Shocks and the Time-Varying NAIRU", *Brookings Papers on Economic Activity*, vol.2, pp.297-346.

[24] Gordon Robert (2011). "The History of the Phillips Curve: Consesus and Bifurcation", *Economica*, vol.78, pp.10-50.

[25] Granger, Clive, and Yongil Jeon (2011). "The Evolution of the Phillips Curve: A Modern Time Series Viewpoint", *Economica*, vol.78, pp.51-66.

[26] Holland, Steven (1984). "Does Higher Inflation Lead to More Uncertain Inflation?" *Federal Reserve Bank of St.Louis,* Feb., pp.15-26.

[27] Kang Ho, Kyo, Kim, Chang-Jim and Morley, James (2009). "Changes in U.S. Inflation Persistence", *Studies in Nonlinear Dynamics & Econometrics*, vol. 13, issue 4, pp.1-21.

[28] Karanasos, Menelaos and Schurer, Stefanie (2008). "Is the Relationship between Inflation and Its Uncertainty Linear?" *German Economic Review*, vol.9, no.3., pp.265-286.

[29] King Robert, Stock James, and Watson, Mark (1995). "Temporal instability of the unemployment-inflation relationship", *Economic Perspectives of the Federal Reserve Bank of Chicago*, May/June, vol.19, no.3, pp.2-12.

[30] King, Robert and Watson Mark (1994). "The post-war U.S. Phillips curve: a revisionist econometric history", *Carnegie-Rochester Conference Series on Public Policy*, 41, pp.157-219.

[31] Mishkin, F. (2007). "Inflation Dynamics", *International Finance*, 10:3, pp.317-334.

[32] Nason, J. (2006). "Instability in U.S. Inflation: 1967-2005", *Economic Review, Federal Reserve Bank of Atlanda*, pp.39-59.

[33] Pivetta, Frederic and Reis Ricardo (2007). "The persistence of inflation in the United States", *Journal of Economic Dynamics & Control*, vol.31, pp.1326-1358.

[34] Robinson, Joan (1979). "Solving the Stagflation Puzzle", *Challenge*, Nov.-Dec., pp.40-46.

[35] Staiger Douglas, Stock James, and Watson Mark (1997). "The NAIRU, Unemployment and Monetary Policy", *Journal of Economic Perspectives*, vol.11, no.1, Winter, pp.33-49.

[36] Staiger Douglas, Stock James, and Watson Mark (2001). "Prices, Wages, and the U.S. NAIRU in the 1990's", in A. B. Krueger and R. M. Solow (eds), *The Roaring Nineties: Can Full Employment Be Sustained?*, Russell Sage Foundation and Century Foundation Press, pp.3-60.

[37] Stiglitz, Joseph (1997). "Reflections on the Natural Rate Hypothesis", *Journal of Economic Perspectives*, vol.11, no.1, Winter, pp.3-10.

[38] Stock James and Watson Mark (1999). "Forecasting Inflation", *Journal of Monetary Economics*, vol.44, pp.293-335.

[39] Stock James and Watson Mark (2007). "Why Has U.S. Inflation Become Harder to Forecast?", *Journal of Money, Credit and Banking,* Supplement to Vol. 39, 3-33.

[40] Stock James and Watson Mark (2009). "Phillips Curve Inflation Forecasts", in Jeffrey Fuhrer, Yolanda Kodrzycki, Jane Little, and Giovanni Olivei (eds) *Understanding Inflation and the Implications for Monetary Policy*, Cambridge: MIT Press, pp.99-202.

[41] Stock James and Watson Mark (2011). *Introduction to Econometrics*, 3rd ed., Pearson.

[42] Stock James and Watson Mark (2012). "Inflation Forecasting", NBER Reporter, Number 1, pp.13-16.

[43] Watson Mark (2014). "Inflation Persistence, the NAIRU, and the Great Recession", *American Economic Review:Papers & Proceedings"*, 104(5), pp.31-35.

CHAPTER 7

Conclusion

1 What have we learned

Employing some of the tools available to a modern economist we have laboured to show the unity of economic theory and econometrics in attacking a number of economic problems.

1.1 Our Approach

We have taken a three (3) dimensional approach: The first dimension studies relationships of *command*. The first two papers studied the communication of information, leaving in the background the exercise of power in hierarchical organisations. These two papers belong to the field of Organisational Economics (see the Handbook of Organisational Economics, edited by Robert Gibbons and John Roberts). The two papers employ basic Graph Theory, which is the main tool in Network Economics.

The exercise of power among the different groups of stakeholders is the subject matter of the third paper on auditing. The main tool applied to study the credibility of the reported information is Non-cooperative Game Theory. The schema proposed by the 2009 Nobel winner Oliver Williamson on Transaction Cost Economics, which is one of the foundations of Organizational Economics, is adapted and employed in this paper, as far as we know for the first time, to provide a framework for the examination of the credibility of accounting information, our research question.

The second dimension is our emphasis on processes of *historical change*. The innovative paper of chapter 5, written by Professor Menelaos Karanasos, Dr Alexandros Paraskevopoulos and myself, provides a rigorous explanation of how an economic variable changes over time. The specific stochastic process it considers is a Time Varying Autoregressive of order 2, $TV-AR(2)$. In contrast to the existing literature, the paper proposes a theory which allows the coefficients of the equations in time series models to vary. Accordingly, it can capture features of time series data, like the existence of trends and heteroscedasticity, which are typical characteristics of a nonstationary variable. In our time varying models, historical episodes do not repeat themselves as a farce, but economic agents do learn from the past and adapt their behaviour. The main tools are Linear Algebra and Time Series Analysis.

The third dimension of our approach is market *competition*. The applied econometrics of the last chapter, 6, explores the dynamics in the goods market and its interactions with the labour market. The tools come from Time Series Econometrics.

1.2 A Final Word

Notice that the final section of the last chapter is about the estimation of the Conditional Variance with the $GARCH$ models. We have come back, a full circle, where we started. The thesis began in Chapter 2 with exploring the modelling decision making under uncertainty. The thesis ends up with the best tool available to measure uncertainty. As the Nobel Laureate Kenneth Arrow (in Colander et al., 2004, p.295) points out $ARCH$ and $GARCH$ is an analysis based on pattern recognition. The remark made by Arrow incentivises us to finish the thesis on an optimistic tone. The $GARCH$ methodology may enlarge the territory of Savage's "small worlds" and make economics able to predict in more realistic environments.

References

[1] Colander, David, Holt, Richard and Barkley Rosser, Jr. (2004) *The Changing Face of Economics*, University of Michigan Press.

[2] Gibbons Robert and John Roberts (2012) *Handbook of Organizational Economics*, Princeton University Press