Modeling Inflation Expectations: The Case of Iran

elektronisches dokument

ULB Sachsen-Anhalt

Dissertation

zur Erlangung des Grades

Doktor der Wirtschaftswissenschaft (Dr. rer. pol.)

der Juristischen und Wirtschaftswissenschaftlichen Fakultät

der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von

Shahram Fattahi Gakieh M.A.

aus Iran

Halle (Saale)

2008

Gutachter der Dissertation:

1. Gutachter: Prof. Dr. Heinz P. Galler

2. Gutachter: Prof. Dr. Gunter Steinmann

Tag des öffentlichen PromotionsKolloquiums: 17. Juli 2008

To my family

Table of Contents

1. Introduction

Expectations are central to our understanding of the behavior of the economy and any explanation of inflation dynamics needs to examine the process of expectation formation.

Economists have recognized that expectations play a determining role in economic theories. For example, Keynesian believe that the IS curve is volatile because firm's expectations about the future probability of their investment projects are themselves highly volatile; they are subject to "animal spirits". In his permanent income hypothesis, Friedman (1957) stressed the role of expected future incomes in determination of consumption expenditure. In fact, many important macroeconomic relationships include element of expectations. When such relationships are combined to build a full macroeconomic model, policy implications of that model will depend on how expectations are being specified.

The economic outcomes that agents can expect from economic policy are affected by the way expectations are formed and how they vary over time. It matters whether agents form their expectations by looking at the past or by looking forward by either trusting economic policymakers' promises or forecasting economic conditions. On the other hand, policy makers need to take expectation of economic agents into account when deciding on policy actions. For that purpose, an understanding of expectation formation is needed. Therefore, failure to investigate these issues fully could lead to flawed economic policy.

Public expectations about the central bank's objectives are important for price stability. If private agents are not sure that the central bank prefers lower to higher inflation, expectations about future policy actions and future inflation will highly become sensitive to the central bank's inflation target and thus result in economic instability (Bernanke, 2003). Furthermore, the reputation of a central bank has an impact on inflationary expectations. The change of the central bank's regime gives rise to a change in the level of inflation expectations. Changing patterns of inflation expectations formation may be resulting from learning process about new monetary regimes.

Any unfavorable economic shock raises actual inflation and causes private agents to raise forecasts of future inflation. Higher inflation expectations will in turn increase inflation. In this situation, policymakers need to have policy tools to anchor

expectations. Some economists believe that if the central bank announces an explicit target for inflation and credibly demonstrates that it will take actions to return inflation to the target when economic shocks occur, firms and households are less likely to increase their long run inflation expectations even if a shock increases inflation for a couple of months. The result is that with inflation expectations well-anchored, any given shock, whether it is from aggregate demand or supply, will not lead to increase in inflation but only to a change in relative prices.

Inflation expectations are very unstable in Iran's economy because the Central Bank is unable to adhere to an inflation target in practice. Thus, inflation expectations are not well-anchored and any oil price increase, which seems apparently to be a favorable shock, results in money creation, fueled by government spending out of oil revenues, and inflation and causes private agents to raise inflation expectations. This in turn will increase inflation. As a result, poor anchored inflation expectations make price stability much more difficult to achieve in the long run and decrease the Central Bank's ability to stabilize output and employment in the short run.

This research tries to examine how market participants form their inflation expectations in the Iranian economy over the period 1959-2003. The Iranian economy depends largely on oil revenues so that any change in oil prices can directly affect all economic sectors. An increase in oil prices will result in money creation and inflation. Furthermore, the large number of government-controlled enterprises, benefited from subsidies, which increase budget deficit through borrowing from the Central Bank and thus have increased the monetary base. During this period, money supply has become 10127 fold while real GNP recorded only a 10 fold increase, resulting in a relatively high inflation with an average inflation rate of about 15 percent. With such very high liquidity, any decision or news announced by the government or the Central Bank could severely change distribution of resources in the economy. In such circumstances, it matters for the Central Bank to know how private agents form their expectations. Moreover, optimal monetary policy depends considerably on the assumed nature of expectations formation process.

Empirical analyses on the formation of expectations can be divided into two categories: first, those studies that have been done by asking people about the future values of inflation (survey studies). Second, those studies that have tried to extract expectations from past data, on the assumption that people look to past experience as a guide to the future. This study will go the latter way.

This study compares two approaches to modeling inflation expectations: simple forecast and a multi-equation model. In the first case, parametric and nonparametric methods are applied and then it is evaluated whether nonparametric models yield better estimates of inflationary expectations than do parametric alternatives. The agents are assumed to use an optimal parametric autoregressive moving average (ARMA) model or nonparametric models including Additive, Multiple Adaptive Regression Splines, Projection-Pursuit Regression, and Neural Networks for forecasting. In fact, out-of-sample estimates of inflation generated by the parametric and nonparametric models will be compared.

In the case of a multi-equation model, this study will focus on the structural model of Phillips curve. The expected inflation generated by the rational, near rational and learning schemes will be examined in the augmented Philips curve equation.

The main focus of this study is on the following general questions:

(I) Do inflation expectations play a main role in determining the wages?

(II) How do private agents form their expectations? Are they rational, near rational, or do they use a learning mechanism?

(III) Are neural networks better suited for modeling expectations than nonparametric alternatives?

(IV) What implications arise from (II)?

(V) What conclusions can be drawn based on the findings above?

This thesis is organized as follows. Following this introduction in chapter one, an overview of the theoretical concepts of expectation formation including adaptive expectations, rational expectations and learning approach will be given in chapter two. The merits and demerits of the each approach are discussed in details.

In chapter three, expectation formation using statistical predictors is examined. Parametric models including autoregressive moving average (ARMA) models, state-space model, the Kalman filter, and nonparametric models including the additive model (AD), multiple adaptive regression splines (MARS), and projection-pursuit regression (PPR) will be discussed.

An innovation based on computational intelligence has been the use of neural networks as a semi parametric approach to describe learning procedures. This is presented in chapter four. The basics of neural networks are first explained. Then the process of learning in these models using the backpropagation algorithm is demonstrated. The interest is in examining whether rational expectations are learnable by use of neural networks.

Chapter five presents the results of an empirical analysis. The data as well as a background of Iranian economy are described. In this chapter, first simple statistical predictors will be used for forecasting and a then multi-equation model including the augmented Phillips curve equation will be used to examine inflation expectations generated by the rational, near rational and learning approaches. Finally, chapter six presents a brief summary, conclusions and policy implications.

2 Modeling expectation formation

In this chapter, different approaches to modeling inflation expectations are presented. First, theoretical concepts of adaptive expectations are demonstrated. Then, the rational expectations hypothesis is discussed in details. The merits and demerits of rational expectations as well as different versions and different tests of this hypothesis are also considered. Finally, the learning approach and its role in macroeconomics are explained. Approaches to learning including eductive learning, adaptive learning, and rational learning are also illustrated.

2.1 Theoretical concepts

2.1.1 Adaptive expectations

One of the most familiar traditional models of expectation formation is adaptive expectations. This model can be stated using the following equation, where P_t^e is this period's expected inflation; P_{t-1}^e is last period's expected inflation; and P_{t-1} last period's actual inflation:

$$P_t^e = P_{t-1}^e + \lambda (P_{t-1} - P_{t-1}^e) \tag{1}$$

with λ being a value between 0 and 1. According to this hypothesis, current expectations of inflation reflect past expectations and an "error-adjustment" term. The parameter value of λ depends on what we think about the likely source of last period's error. If it was a permanent shift in the process forming P, then we set $\lambda = 1$ so that $P_t^e = P_{t-1}$. This is static expectations: this year's inflation is expected to be the same as last year's. If last period's error was just due to a random event, we set $\lambda = 0$, so there is no adjustment, and we should not change expectations at all ($P_t^e = P_{t-1}^e$). People will change expected inflation if there is a difference between what they were expecting it to be last period and what it actually was last period. In fact, expected inflation is revised by some fraction of most recent forecast errors. If the expected inflation was, say 5 percent, but the actual inflation 10 percent, people raise their expectations by some fraction λ of the difference between 5 and 10. Using the Koyck transformation, the equation (1) can be transformed into

$$P_t^e = (1-\lambda)P_{t-1} + \lambda(1-\lambda)P_{t-2} + \lambda^2(1-\lambda)P_{t-3} + \lambda^3(1-\lambda)P_{t-4}... \qquad (2)$$

Now we can examine the relationship between P_t^e and P_t. Suppose that P_t has been constant for a long time at P_0. Then, suppose that at time period T, the inflation jumps up to P_1 and stays there indefinitely. At T, all the terms on the right-hand side of equation (2) are equal to P_0, so the expected inflation for T is given by P_0, that is $P_T^e = P_0$:

$$P_T^e = (1-\lambda)P_0 + \lambda(1-\lambda)P_0 + \lambda^2(1-\lambda)P_0 + \lambda^3(1-\lambda)P_0... = P_0$$

Once T is over, expectations are formed by equation (2) with t set equal to T+1. Therefore, the first term on the right-hand side for period T+1 is P_1:

$$P_{T+1}^e = (1-\lambda)P_1 + \lambda(1-\lambda)P_0 + \lambda^2(1-\lambda)P_0 + \lambda^3(1-\lambda)P_0...$$

Since $P_1 > P_0$, it is easy to verify that $P_1 > P_{T+1}^e > P_T^e = P_0$. There is some correction in T+1 for the error made at T, but is not complete. At the start of following period, two of the right-hand terms of equation (2) include P_1. The remaining error is again partly corrected but the absolute value of correction is less. This process continues until the second term on the right-hand side of equation (1) diminishes to make the difference ($P_t - P_t^e$) arbitrarily small

There are merits and demerits of the adaptive expectations hypothesis (AEH). On the one hand, the hypothesis has the advantages of being simple to operate as a "rule of thumb". It is at the best appropriate in a stable environment where the price level moves up and down in a fairly random fashion, with the possibility of somewhat more permanent shifts in the background. On the other hand, however, it has two disadvantages: first, it is a backward-looking approach (no account of fully-announced future policies). Second, it has systematic errors based on the previous forecast with some correction for previous forecast errors. Individuals do not systematically learn from previous forecast errors, they do ignore information that would help them improve the accuracy of their forecasts. Thus, the AEH assumes suboptimal behavior on the part of economic agents. For example, consider the Phillips curve equation:

$$P_t = P_{t-1} - (U_{t-1} - U^*) + \varepsilon$$

P_t = Actual inflation at time t

$U*$ = Natural of unemployment

Assume that (for simplicity):

$$U* = U_{t-1} = U_{t-2} = U_{t-3}\ldots$$

then

$$P_t = P_{t-1} + \varepsilon_t$$

with adaptive expectations:

$$P_t^e = \lambda P_{t-1} + (1 - \lambda) P_{t-1}^e$$

$$= \lambda P_{t-1} + (1 - \lambda)(\lambda P_{t-2} + (1 - \lambda) P_{t-2}^e) + \ldots$$

If $\lambda = 0.5$

$$P_t^e = 0.5 P_{t-1} + 0.25 P_{t-2} + 0.125 P_{t-3} + \ldots$$

$$= 0.5 P_{t-1} + 0.25[P_{t-1} - \varepsilon_{t-1}] + 0.125[P_{t-1} - \varepsilon_{t-1} - \varepsilon_{t-2}] + \ldots \quad (3)$$

Equation (3) shows that the AEH ignore past forecast errors in forming expectations. Under adaptive expectations, if the economy suffers from constantly rising inflation rates, people would be assumed to sequentially underestimate inflation. This may be regarded unrealistic- surely rational people would sooner or later realize the trend and take it into account in forming their expectations. Moreover, models of adaptive expectations never reach an equilibrium; instead they only move toward it asymptotically.

2.1.2 Rational expectations

The rise of Rational Expectations

The rational expectations hypothesis responds to this criticism by assuming that individuals use all information available in forming expectations. During the late 1960s, rational expectations economics started changing the face of macroeconomics. Robert Lucas, Tomas Sargent, and Neil Wallace started to dominate the macroeconomic discussion. Notions such as the Lucas critique, the Lucas supply curve, and the Sargent-Wallace policy irrelevance proposition became integral parts of the macroeconomics discourse.

There are different reasons behind the rise of rational expectations (RE). Sent (1998) argues that the main factors are as follows:

1. Expiration of the Phillips curve: in the late 1960s to early 1970s, policy makers used a trade-off between inflation and unemployment to lower unemployment. However, they faced high inflation rates accompanied by high unemployment rates in the 1970s. In other words, the result of policy making was higher inflation with no benefits in terms of lower unemployment. Rational expectations economists were able to explain the expiration of the Phillips curve. They, using the rational expectation hypothesis, demonstrated that government actions caused an adverse shift of the Phillips curve.

2. Policy irrelevance: orthodox prescriptions of economic policy crumbled, since much of the effectiveness of these policies were based on the government's ability to fool people. Rational expectations economists asserted that people can foil government policies by learning their mistakes. They justified the ineffectiveness of government intervention in the context of the failure of traditional Keynesian policies in the 1970s. Also, they recognized the limitations of their profession maintaining that the economy would basically be stable if it were not subjected to the shocks administered by the government.

3. Using available techniques: rational expectation economists used sophisticated mathematical techniques in order to predict. They learned and used the techniques of intertemporal optimization developed by mathematicians and control scientists. They also improved the tools of optimal prediction and filtering of stochastic processes. Some of these techniques such as classical linear prediction theory[1] was developed in 1940s to1950s but did not immediately become part of economists' toolkits. However, Peter Whittle made more accessible to economists this theory that was heavily used by rational expectation economists. This delay explains the lagged effect of Muth's contributions. Thus rational expectation economists were able to calculate rational expectation equilibria using new techniques.

4. Restoring symmetry: the hypothesis of adaptive expectations had been used heavily up until the late 1960s. According to this hypothesis, individuals used forecasting errors in revising their expectations. Econometricians were presumed to be fully knowledgeable whereas the agents were assumed to make systematic

----

1. The mathematical theory for interpreting distributed lags in terms of economic parameters and incorporating the rational expectations hypothesis in economic models.

forecasting errors period after period. Thus there was an asymmetry among economists or econometricians and the agents in that econometricians fit models that forecast better than agents. Rational expectations hypothesis (REH) removed this asymmetry making the econometricians part of the agents' behavior. Therefore, rational expectation economists placed econometricians and agents on an equal footing by postulating that forecasts made by the agents within the model were no worse than those the econometricians who had the model.

5. Optimizing over information: according to REH, optimization over perceptions implied agents did the best they could and formed their views of future using available information, including their understanding of how the economy works. Rational expectation theorists extended expectation theory into the optimizing behaviors theory. If perceptions were not optimally chosen, unexploited utility or profit-generating possibilities would exist within the system. Hence, these economists insisted on the disappearance of all such unexploited possibilities.

6. Endogenizing expectations: Keynes (1936) doubted that expectations could be modeled accurately. So he considered expectations as given. Also, Keynes followers assumed that people made guesses about the future by looking exclusively backward. In fact, the hypothesis of adaptive expectations is backward-looking in that it allows the possibility of systematic forecasting errors for many periods in succession. This is a suboptimal use of available information and is not consistent with the idea of optimization. Even though people used adaptive expectations, no widely accepted economic theory was offered to explain the amount of the adjustment parameter. The mechanism of rational expectations' formation is endogenously motivated and expectations or forecasts are correct on average if errors individuals remain satisfied with their mechanism. This hypothesis asserted that the resulting predictions might still be wrong, but the errors would be random. If errors follow a pattern, they contain information that could be used to make more accurate forecasts. Therefore errors were presumed to cancel out when all individual expectations are added together.

7. Making public predictions: some authors believed that the rise of rational expectations could fight the threat of indeterminacy of economic outcomes. This indeterminacy resulted from this fact that making both self-falsifying and self-fulfilling predictions about people was possible. Since outcomes depended partly on what people expected those outcomes to be if people's behavior depended on their

perceptions, economic systems were thought to be self-referential. This led some economists to despair that economic models could produce so many outcomes that they were useless as instruments for generating predictions. Rational expectations ,however, was a powerful hypothesis for restricting the range of possible outcomes since it focused only on outcomes and systems of beliefs that were consistent with one another. Under rational expectations, correct public predictions could be made because rational expectations predictions were presumed to be essentially the same as the predictions of the relevant economic theory. Also, the hypothesis consisted of expectational response of the agents and the influence of predictions on behavior of the agents.

8. Countering bounded rationality: rational expectations theory was born at the same time in the same situation as the concept of bounded rationality, namely, in the 1960s at the Graduate School of Industrial Administration (GSIA) at Carnegie Mellon University. Holt, Modigliani, Muth, and Simon were colleagues and worked on the Planning of Control of Industrial Operation Project, which consisted of developing and applying mathematical techniques to business decision making. Though Simon and Muth had both participated in the project, Simon saw the strong assumption underlying this project as an instance of satisfying, whereas Muth saw this special case as a paradigm for rational behavior under uncertainty. Some argue that Muth, in his announcement of rational expectations, explicity labeled this theory as a reply to the doctrine of Simon's bounded rationality.

9. Restricting distributed lags: in the late 1960s, rational expectation economists were confronted with theoritical models that analyzed individual behavior in a context without uncertainty and randomness. At the same time, since they treated their data probabilistically, they had to incorporate uncertainty and randomness in optimizing economic theory and using the outcome to understand, interpret, and restrict the distributed lags that abounded in the decision rules of dynamic macroeconomic models. They promised to tighten the link between theory and estimation.

10. Incorporating vector autoregression: the final causal background of rational expectations is related to the belief that it created a connection between vector autoregressions and economic theory. Some argue the REH was able to revive theory by showing that vector autoregressions was not necessarily atheoritical and could provide a statistical setting within which the restrictions implied by theoretical models could be imposed. In particular, rational expectation theorists exploited cross-

equation restrictions to connect the vector autoregressive parameters of decision rules with theoretical parameters of taste, technology, and other stochastic environments.

Rational expectations and processes

The rational expectation hypothesis (REH) assumes economic variables are generated by recurring processes (Attfield et al, 1991). Over time, economic agents learn the process determining a variable and they will use this knowledge and all information available (that is related to the variable) to form expectations of that variable. As a result, the agents' subjective probability distribution coincides with the objective probability distribution of events[1]. In other words, the expectations of agents will be the same as the conditional mathematical expectations based on the true probability model of the economy. For example, suppose the value of variable Y in period t is determined by its own lagged value and by lagged values of other variables X and Z in the following way:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_{t-1} + \alpha_3 Z_{t-1} \tag{4}$$

where α_0, α_1, α_2 and α_3 are constant coefficients. Consider a person who, at the end of period t-1, is trying to form an expectation about the value that Y is going to take in period t. She knows that the process determining Y is given by equation (4): knowledge of this process is said to be part of her information set at the end of period t-1. She also knows the values of all lagged variables of X, Y, Z, that also are part of her information set at the end of period t-1. If she is rational, her expectation of what Y is going to be in period t, on the basis of her information set at the end of period t-1, will be formed as follows:

$$E_{t-1}Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_{t-1} + \alpha_3 Z_{t-1} \tag{5}$$

where E_{t-1} is the expectation of Y_t formed on the basis of the information available at the end of period t-1. The rational expectation of Y_t formed at period t-1 (denoted as

---

1. This is the strong version of the rational expectations hypothesis, due to Muth, (Pesaran, 1987).

$E[Y_t \mid I_{t-1}]$ is the mathematical expectation of Y_t conditional on the available information at the end of period t-1 (I_{t-1}). If Y does indeed continue to follow the process shown in equation (5) then this person's expectation will be perfectly accurate, the person's forecasting or expectational error is zero. This result is not general because in this case we assumed the process determining Y is deterministic. However, most processes in real world are stochastic; that is, they include an unpredictable element of randomness in human responses. One way to incorporate this element in equation (4) is to add to it a random term (v_t):

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_{t-1} + \alpha_3 Z_{t-1} + v_t \qquad (6)$$

v_t may be positive or negative. Since this variable is seen as the result of a large number of random factors affecting human behavior, it is natural to think of small values of v_t rather than large values. In fact, we assume that variable v_t has a probability distribution centered at zero and a constant, finite variance $\{\sigma_v^2\}$. The value of v in period t is unknown at the end of period t-1; it is not part of the information set at period t-1. But it is clear that a rational forecaster has to form some expectation of the value that v is going to take in period t. The rational expectation of Y in accordance with equation (6) is as follows:

$$E_{t-1}Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_{t-1} + \alpha_3 Z_{t-1} + E_{t-1}v_t \qquad (7)$$

where $E_{t-1}v_t$ is the expectation of v_t formed on the basis of all the information available at the end of period t-1. The best guess a rational agent can make of v_t is that it will equal its mean value $E_{t-1}v_t = 0$. Thus, the rational expectation of Y in period t, based on information available at the end of period t-1, can be written as:

$$E_{t-1}Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_{t-1} + \alpha_3 Z_{t-1} \qquad (8)$$

Thus the rational expectation of the variable Y in period t is its mathematical expectation given the available information. Rational expectations, as Muth (1961) explained, should be generated by the same stochastic process that generates the variable to be forecast.

In equation (8), if the process determining Y remains unchanged, it follows that the expectational error will be the random component v of Y:

$$Y_t - E_{t-1}Y_t = v_t \tag{9}$$

The general characteristics of Rational Expectations

A number important implications follow from the fact that, if the process determining Y is understood, the error of rational expectation of Y is the same as the random component of the process determining Y. They are as follows:

(a) The errors of rational expectations are on average zero

It is clear from equation (9) that once the process determining Y is allowed to be stochastic the rational expectation of Y will not always be perfectly accurate, for the random component v is inherently unpredictable. The best a rational forecaster could do is expect the mean value of v and that is defined to be zero. In fact, the error may be positive, negative or zero. But on average or over a large number of periods the negative errors will cancel out with the positive ones, leaving an average error of zero.

(b) The errors of rational expectations exhibit no pattern

If expectations are rationally formed, the forecasting error will equal the random element in the process being forecast. This random variable, and hence forecasting error, may be surprises or news in the system. If it exhibits no pattern, then the forecasting error does not exhibit any pattern either. But what happen if v exhibits a pattern in the following way:

$$v_t = \beta_1 v_{t-1} + \varepsilon_t \tag{10}$$

The current value of v is linked to the previous period's value of v. ε_t is a random error with zero mean which can not be predicted on the basis of any information available at the end of period t-1; β_1 is a coefficient, the value of which lies between -1 and +1. If v is being determined according to equation (10) then rational people will form their expectation of current period's value of v in accordance with that process.

And since the value of v in the previous period, t-1, will be part of the available information at the end of period t-1, it follows that the forecast of v will diverge from the actual value of v by an unknown, unpredictable element ε_t. The error term (ε_t) exhibits no pattern and has a mean value of zero. Thus even if v does exhibit a pattern, the rational forecast of Y would, on average, still be correct and the forecasting error would exhibit no pattern. As for the timing of a change in the method of forming expectations, the rational expectations hypothesis suggests that as long as there is no change in the process determining a variable, the method of forming expectations will not change. But if the actual process determining a variable is known to have changed, then the method by which expectations are formed will change in line with it.

(c) Rational expectations are the most accurate expectations

Rational expectations is the most efficient method of forecasting in that the variance of the forecasting errors will be lower under rational expectations than under any other method of forecasting or forming expectations. Because forecasts of a variable on the basis of rational expectations hypothesis will use all available information on the process determining that variable. In other words, as expectations are formed the unpredictable part of Y can not regularly be predicted. So any method of expectation formation will be inaccurate to a degree determined by the likely range of values that v can take. But it is possible to be even more inaccurate by forecasting without reference or with only partial reference to the process determining the variable.

General critique of the rational expectations hypothesis

Criticisms of the REH are as follows (Attfield et al, 1991):

(a) The plausibility of rationality

REH assumes people to use all the information about the process determining a variable when forming expectations. Is it really plausible? Can we really assume that all decision-makers are intelligent enough to use and fully understand all the available information? In reality people often ignore economic matters. This criticism is that a major assumption behind rational expectations is implausible.

The advocates of REH respond to this criticism in this way: first of all the idea that the typical individual is capable of making the best of opportunities open to him is a common one in economics. For example, in demand theory it is assumed that the typical person chooses to consume goods at a point given by the tangency of an indifference curve and a budget constraint. The mathematics behind this choice strategy is highly sophisticated for most people. Yet it is assumed that people act as if they understand it. If such assumption leads to a theory which makes accurate predictions, then the assumption of mathematical awareness is thereby shown to be a useful one. People forming expectations use firms- who specialize in or provide the service of making economic forecasts- or government bodies-who make forecast public.

Some economists also criticize the role of rationality in REH. Advocates of the hypothesis state that the role of rationality has been used in REH in that the process of acquiring information has been carried out up to the point where the marginal cost of acquiring more information equals the marginal benefit of making more accurate forecasts. But this point does not necessarily correspond to the point at which the forecasting error is equal to the purely random component of the determining process. It may be that knowledge about some determining variable could be obtained and extra accuracy thereby achieved, but only at a price which it is not worth paying. In that case the forecasting error will tend to be absolutely greater than the random element in the determining process. Advocates of REH accept this criticism but they assert that for most purposes it is not of great significance. The reason for this is that forecasting errors themselves are observed at no cost. For example, any error in your forecast about the level of prices is observed as a costless side-effect of shopping. In other words, it must be worthwhile to exploit this information fully until its marginal benefit is zero.

(b) The availability of information

REH assumes that the process Y is known and that the values of variables in that process are known at the end of period t-1. But what happens if we do not know the process determining the variable (Y) and if we are not able to acquire the necessary information? Advocates of the REH state that it is true that people cannot automatically know which variables are important in the process determining Y but it is also true that the REH doesn't claim that they do. What the hypothesis argues is

that on average and after a period of time, economic agents will learn from past experience what the process is. They will combine this developed knowledge with current available information to form their expectations[1]. For example, if, at the end of period t-1 the rational agent does not know the true value of X in period t-1, and if the value of X in period t-1 determines the value Y in this period, the agent will have to form expectations of the value X in period t-1. Suppose the process determining Y is as follows:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_{t-1} + \alpha_3 Z_{t-1} + v_t \tag{11}$$

Suppose that the value of X_{t-1} is unknown at the end of period t-1. And let the process determining X in any period t as follows:

$$X_t = \beta_0 + \beta_1 V_{t-1} + \beta_2 W_{t-1} + \varepsilon_t \tag{12}$$

where V and W are other variables, the β's are coefficients, and ε is a random error term with mean zero. The rational forecast of the unknown value of X in period t-1 will be as follows:

$$E_{t-1} X_{t-1} = \beta_0 + \beta_1 V_{t-2} + \beta_2 W_{t-2} \tag{13}$$

$E_{t-1} X_{t-1}$ will be used in place of X_{t-1} in equation (11). Thus if X_{t-1} is unknown the rational expectations of Y in period t will be:

$$E_{t-1} Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 (\beta_0 + \beta_1 V_{t-2} + \beta_2 W_{t-2}) + \alpha_3 Z_{t-1} \tag{14}$$

The forecasting error will therefore be given by:

$$Y_t - E_{t-1} Y_t = v_t + \alpha_2 \varepsilon_{t-1} \tag{15}$$

Since v_t and ε_{t-1} are random errors with means of zero, neither of which can be even partly predicted on the basis of any information available at the end of period t-1. The rational forecast or expectation of Y in equation (14) is, in general, the most accurate forecast.

1.Friedman(1979), criticizing the REH, asserted that what is typically missing in rational expectations models is a clear outline of the way in which economic agents derive the knowledge which then they use to formulate expectations meeting requirement.

(c) Limits to the applicability of rational expectation

Many important economic events can be seen as unique, or at least exceptional or unusual due to the particular political circumstances of day. In what sense can the REH be said to apply to these exceptional cases? The advocates of rational expectations assert that the REH can best be applied to variables or events which can be seen as a part of recurring process. However, this class of events may be a larger one than is commonly thought. For example, governments desire to have a high level of economic activity at the time of general elections and may switch some policies. Such switches of policy could be seen as part of a fairly regular and reasonably predictable process. So an event which could be portrayed as unique from another viewpoint may well be part of an underlying recurring process.

(d) Testability of REH

Some economists have criticized that REH is not testable. Rational expectations theorists state that there are several layers to this criticism. First, if REH is taken rather loosely to imply that people make the best of their available information, then it may always be possible to define the available information so that the hypothesis becomes immune to falsification. This criticism is valid if tests of REH tended to employ the loose form of the hypothesis. But if they tend to employ strong versions of the hypothesis in which people's knowledge of the process determining a variable is assumed to be the same as the best estimate that can be made of that process by econometric techniques then this criticism is hardly a strong one. Because this assumption leads to predictions which are both clear and different from the predictions derived from other theories about expectations.

An important criticism is that expectations about a variable are almost always only part of a model. Thus there are joint tests of the REH itself and the rest of the model. If the model fails the tests to which it is subjected one can always 'rescue' the REH by arguing that it is the rest of the model which is wrong. It is at times possible to distinguish between the restrictions imposed on the data by REH itself and the restrictions imposed by the rest of the model. However, the usefulness of the REH, in this way, can be tested informally and less satisfactory. If, time after time, this kind of models were rejected then we can reject the REH.

The final type of criticism of testability of REH is what is known as 'observational equivalence'. For many rational expectations models which 'fits the data' there will always be a non-rational expectations model which fits the data equally well. The data themselves cannot discriminate between two theories, which are therefore said to be observationally equivalent. The implication of this is that, even if a rational expectations model 'passes' conventional empirical tests, this does not necessarily imply that one should accept the hypothesis. Whether you do or do not, depends on whether you find it more 'plausible' than the non-rational expectations model on some other unspecified grounds.

(e) Multi rational expectations equilibria

The models of Muth and Lucas assume that at any specific time, a market or the economy has only one equilibrium (which was determined ahead of time), so that people form their expectations around this unique equilibrium. If there is more than one possible equilibrium at any time then the more interesting implications of the theory of rational expectations do not apply. In fact, expectations would determine the nature of the equilibrium attained, reversing the line of causation posited by rational expectations theorists.

(f) Ability of agents in action

In many cases, working people and business executives are unable to act on their expectations of the future. For example, they may lack the bargaining power to raise nominal wages or prices. Alternatively, wages or prices may have been set in the past by contracts that cannot easily be modified. (In sum, the setting of wages and prices of goods and services is not as simple or as flexible as in financial markets.). This means that even if they have rational expectations, wages and prices are set as if people had adaptive expectations, slowly adjusting to economic conditions.

Different versions of RE

Many definitions of rational expectations (RE) have been proposed since Muth (1961) published his seminal article on this concept. In its stronger forms, RE operates as a coordination devise that permits the construction of a "representative agent" having "representative expectations." Generally, two definitions for RE is used are applied research: the weak form and the strong form.

Weak-form RE

The weak version of RE is independent of the content of the agent's information set. Suppose there are N agents (i=1,...,N) in an economy and $E_{t,i} Y_{t+k}$ denote agent i's subjective (personal) expectation formed at the end of period t of Y_{t+k} (k≥1). Also let E $[Y_{t+k}|I_{t,i}]$ denote the objectively true expectation for Y_{t+k} conditional on the information available to agent i at the end of period t ($I_{t,I}$). The agents are said to have weak-form rational expectations for variable Y_{t+k} if the following condition holds:

For each i = 1,..., N, $E_{t,i} Y_{t+k} = E[Y_{t+k}|I_{t,i}] + \epsilon_{t,i}$ *where* $\epsilon_{t,i}$ *are serially and mutually independent finite-variance error terms that satisfy* $E[\epsilon_{t,i}|I_{t,i}] = 0$.

Weak-form RE has some features. First, it is applicable only if there are "objectively true" conditional expectations. Weak-form RE assumes that agents make optimal use of all available information. Second, it is consistent with the idea of "economically rational expectations", proposed by Feige and Pearce (1976), in which agent's information sets are the result of cost-benefit calculations by the agents regarding how much information to obtain. Finally, many economists are willing to use this version, as a useful benchmark assumption consistent with the idea that agents are arbitrageurs who make optimal use of information.

Strong-form RE

Muth (1961) used a stronger version of RE in that he placed a restriction on the information sets of agents in theoretical economic models. This version guarantees the existence of "objectively true" conditional expectations but at the cost of transforming RE into an incredible concept in relation to the form of expectations that real economic agents could reasonably be supposed to have.

Agents in a theoretical model of a multi-agent economy will be said to have strong-form RE if they have weak-form RE and, in addition, their information sets at the end of period t contain the following information:

(a)   The true structural equations and classification of variables for the model, including the actual decision rules used by each private and public (government) agent to generate actions and/or expectations;

(b)   The true values for all deterministic exogenous variables of the model;

(c)   The true probability distributions governing all exogenous stochastic terms;

(d)   Realized values for all endogenous variables observed by the modeler through the end of period t.

Strong-form RE has some interesting features. First, it is assumed that agents are smart and as well informed about the economy. The issue that agents know a priori the actual decision rules used by each other agent is incredible. This version can therefore be interpreted as an idealized Nash equilibrium[1] benchmark for agents' expectations that agents may (or may not) eventually arrive at through some process of reasoning and/or learning.

Second, in practice theorists modeling economic systems assume that they have an extraordinary amount of information about the true working of the economy. As a result, under strong-form RE, economic agents are presumed to have a great deal more information than would actually be available to any econometrician who attempted to test these models against data (Sargent, 1993).

Third, many economists are uncomfortable with the more common assumption in the strong-form RE. Nevertheless, this version becomes more acceptable if it is viewed as a possible ideal limiting point for the expectations of boundedly rational agents with limited information who engage in learning in successive time periods.

Finally, considering perfect foresight[2] RE is interesting. Agents in a theoretical model of a multi-agent economy will be said to have perfect foresight RE if the following two conditions hold:

----

1. If there is a set of strategies with the property that no player can benefit by changing her strategy while the other players keep their strategies unchanged, then that set of strategies and the corresponding payoffs constitute the Nash Equilibrium.

2. It must be noted that perfect-foresight RE differs from the perfect foresight assumption used in "Walrasian equilibrium models." In the latter kind of models, perfect foresight is the assumption that households and firms correctly foresee the market-clearing levels and solve their optimization problems conditional on these levels.

(a)   Agents have strong-form RE;

(b)   There are no exogenous shock terms affecting the economy, so that all expectations are correct without error, e.g. $E_{t,i}Y_{t+k} = Y_{t+k}$

There are some implications of strong-form RE. First, if there is a change in the way a variable moves, then the way in which expectations of this variable are formed also changes. For example, a change in the government's monetary policy rule leads to a change in the movements of the Fed Funds rate. Second, forecasts are not always exactly correct, but forecast errors are not predictable in advance and they average out to zero. Third, two reasons why expectations can fail to be rational in the strong-form sense: (a) agents fail to use all available relevant information and (b) agents fail to make optimal use of all available relevant information.

An example of strong-form RE

Suppose an economy is described by the Lucas Model (Caplan, 2000):

(IS)	$y_t = -ar_t + u_t$	(1)
(LM)	$m_t - p_t = by_t - ci_t + v_t$	(2)
(Fisher equation)	$i_t = r_t + E_t p_{t+1} - p_t$	(3)
(AS)	$y_t = y^* + \alpha(p_t - E_{t-1}p_t)$	(4)
(Monetary Policy Rule)	$m_{t+1} = m_t + \varphi_{t+1}$	(5)
(Strong-Form RE)	$E_t p_{t+1} = E[p_{t+1} \mid I_t)$	(6)

Where y_t = output, p_t = price level, m_t = money supply, r_t = real interest rate, i_t = nominal rate, u_t, v_t, and φ_t = random variables with mean 0, y^* = potential output, $E_t p_{t+1}$ = the subjective forward-looking expectation of representative agent at time t regarding the price level in period t+1, $E[p_{t+1} \mid I_t)$ = the objectively true conditional expectation, I_t = information set that is available to the representative agent at the end of period t whose contents are assumed to be consistent with strong-form RE.

All variables are logs of their level values. In the period t predetermined variables are m_t and $E_{t-1}p_t$ for $t > 1$. The exogenous variables are: y^*, u_t, v_t and φ_t ; the positive exogenous constants a, b, c, and α ; an initial value $m_1 = m_0 + \varphi_1$ for the period 1 money supply m_1, where m_0 is exogenously given, and initial value for $E_0 p_1$.

Model equation (6) is incomplete as it stands, in that the "true conditional expectation" on the right hand side needs to be determined in a manner consistent with strong-form RE. That is, given this expectation, the subsequent way in which the price level for period t+1 is actually determined by the model equations must conform to this expectation in the sense that the objectively true I_t-conditioned expectation of the model-generated solution for the price level in period t+1 must coincide with the expectation assumed for this price level in model equation (6). To complete this model with strong-form RE, we must solve a fixed point problem of the form f(x) = x, where $x = E\ [p_{t+1}|I_t)$[1]. To determine the needed expectational form, $E\ [p_{t+1}\ |I_t)$, the method of undetermined coefficients is used.

Conjecture a possible solution form for p_t as a parameterized function of other variables, where the parameter values are unknown. Then, determine values for these unknown parameters that ensure strong-form RE. For simplicity assume that $y^* = 0$. Combining model equations (1) through (4) plus (6) leads to

----

1. It must be noted that there is a problem for the RE solution, it is not unique. In fact, multi rational expectations are likely to exist for models that include equations that are nonlinear in the endogenous variables. This spreads some doubts about the "rationality" of these RE solutions. For example, consider the following model of an economy:

$$y_t = a + b\ E_{t-1}y_t + \varepsilon_t\ ,$$ (1)
$$t \geq 1,\ a > 0,\ 0 < b < 1,\ E\ [\varepsilon_t\ |\ I_{t-1}] = 0$$

If a representative agent forms his expectations for y_t in period t-1 in accordance with strong-form RE, that is, $\quad E_{t-1}y_t = E\ [y_t\ |\ I_{t-1}]$ (2)

In this case the y_t generating process in (1) takes the form
$$y_t = a + b\ E\ [y_t\ |\ I_{t-1}]\ + \varepsilon_t\ ,\ t \geq 1$$ (3)

The right side of equation (3) can be expressed as a function M(x) of x, where $x = E\ [y_t\ |\ I_{t-1}]$. Taking conditional expectations of both sides of (3), one can obtains a relation of the form
$$x = E\ [M(x)\ |\ I_{t-1}]$$
$$\equiv f_t\ (x)\ ,\quad t \geq 1$$ (4)

Suppose that the RE solution for output of a model economy in period t satisfies a fixed point problem having form (4) and that two distinct solutions x' and x" exist- that is, $f_t\ (x') = x'$ and $f_t\ (x") = x"$. Thus, if all agents in the economy at the end of period t-1 anticipate output level x' for period t, the objectively true expected output level for the economy in period t will be x'; and if instead, all agents in the economy at the end of period t-1 anticipate output level x" for period t, the objectively true expected output level for the economy in period t will be x".

$$p_t = (1/1+c) \, m_t + (c/1+c) \, E\,[p_{t+1}\,|\,I_t] - \beta\,[p_t - E\,[p_t\,|\,I_{t-1}]] + w_t \qquad (7)$$

where $\quad \beta = \alpha[(b+c/a)/(1+c)]$; $w_t = (1/1+c)[(c_t/a)u_t - v_t]$ $\qquad\qquad (8)$

Suppose it is conjectured that the solution for p_t takes the form

$$p_t = q_1 m_t + q_2 w_t + q_3 \varphi_t \qquad , t \geq 1, \qquad (9)$$

Lead equation (9) one period and taking conditional expectation of both sides:

$$E[p_{t+1}\,|\,I_t] = q_1 E[m_{t+1}\,|\,I_t] \quad , t \geq 0, \qquad (10)$$

Taking conditional expectation of both sides of equation (5) leads to $E[m_{t+1}\,|\,I_t] = m_t$, hence

$$E[p_{t+1}\,|\,I_t] = q_1 m_t \quad , t \geq 0 \qquad (11)$$

Now lag equation (11) one period and lag equation (5) one period to substitute $m_t - \varphi_t$ in for m_{t-1}, thus obtaining

$$E[p_t\,|\,I_{t-1}] = q_1[m_t - \varphi_t] \, , t \geq 1 \qquad (12)$$

Combining equations (9) and (12), one then has

$$p_t - E[p_t\,|\,I_{t-1}] = [q_1 + q_3] \, \varphi_t + q_2 w_t, \, t \geq 1 \qquad (13)$$

Using equations (11) and (13) to substitute out for the expectations in the price equation (7) and combining terms leads to

$$p_t = [(1/1+c) + (c/1+c)q_1] \, m_t + [1 - \beta q_2]w_t - \beta[q_1 + q_3] \, \varphi_t \, , t \geq 1 \qquad (14)$$

Now we have two distinct equations-equations (9) and (14) that state p_t as linear function of m_t, w_t, and φ_t. To make these equations consistent, set the three coefficients in (9) equal to the three coefficients in (14). It yields:

$$q_1 = 1; \qquad (15)$$

$$q_2 = (1/1+\beta); \qquad (16)$$

$$q_3 = -(\beta/1+\beta); \qquad (17)$$

Thus it follows that one possible solution for p_t consistent with strong-form RE is:

$$p_t = m_t + (1/1+\beta)w_t - (\beta/1+\beta)\varphi_t \tag{18}$$

Equation (18) shows that the price level is directly proportional to the money supply, a positive function of investment shocks, a negative function of money demand shocks, and a negative function of unexpected money supply increases. The corresponding strong-form RE for p_t, to be substituted in on the right hand side of model equation(6), is then found by taking the I_t-conditional expectation of each side of equation (18) bumped up one period, which yields

$$E[p_{t+1} \mid I_t] = E[m_{t+1} \mid I_t] = m_t , t \geq 1 \tag{19}$$

Combining model equation (4) (with $y^* = 0$) with (18) and (19), it follows that the solution for period t output consistent with strong-form RE is given by

$$y_t = \alpha[(1/1+\beta) \varphi_t + (1/1+\beta) w_t] \tag{20}$$

Output is an increasing function not of money, but of unexpected money shocks as well as of shocks u_t and v_t to the IS and LM curves.

Equation (20) has some conclusions for economic policymaking. From Lucas' point of view, if the Central Bank decides to lower the unemployment rate by an expansionary monetary policy, then according to the REH the policy will be ineffective. People will see what the Central Bank is doing and raise their expectations of future inflation. This is in turn will counteract the expansionary effect of the increased money supply. All that the Central Bank can do is to raise the inflation rate, with at most temporary decreases in unemployment.

Different tests of REH

Following Sargent (1993), four different tests of Muthian rationality may be distinguished. Letting ${}_{t-k}x_t^e$ signify the expectation reported in the survey for a variable X_t made at time t-k.

1. Unbiasedness: the survey expectation should be an unbiased predictor of the variable. That is, a regression of the form

$$x_t = a + b \, {}_{t-k}x_t^e + \varepsilon_t$$

Should yield coefficient estimates a=0 and b=1. This is necessary condition. A sufficient condition is as follows

$$x_t - {}_{t-k}x_t^e = E_t = \mu_t + \varepsilon_t$$

The hypothesis to test is $\mu = 0$

2. Efficiency: the survey expectation should use information about the past history of the variable in the same way that the variable actually evolves through time. That is, in the two regressions,

$$_{t-k}x_t^e = a_1X_{t-1} + a_2X_{t-2} + ... + a_nX_{t-n} + \epsilon_t$$

$$X_t = b_1X_{t-1} + b_2X_{t-2} + ... + b_nX_{t-n} + u_t$$

It must be true that $a_i = b_i$ for all i. This test is called orthogonality test. Another possibility for examining the efficiency property is that the forecast error is tested for serial correlation.

3. Forecast error unpredictability: The forecast error, that is, the difference between the survey expectation and the actual realization of the variable, should be uncorrelated with any information available at the time the forecast is made.

4. Consistency: when forecasts are given for the same variable at different times in the future, the forecasts should be consistent with one another. For example, in the regressions,

$$_{t-2}x_t^e = c_1 \, {}_{t-2}x_{t-1}^e + c_2X_{t-2} + ... + c_nX_{t-n} + \epsilon_t$$

$$_{t-1}x_t^e = a_1X_{t-1} + a_2X_{t-2} + \ldots + a_nX_{t-n} + u_t$$

It must be true that $c_i = a_i$ for all i.

These tests are different ways of testing properties of conditional expectations in that whether the reported survey expectations are consistent with being conditional expectations. For example, consider the efficiency test and suppose that $a_1 \neq b_1$. Substracting the first equation from the second yields the expression

$$X_t - {}_{t-1}x_t^e = \text{forecast error} = (a_1 - b_1)X_{t-1}$$

Since, by hypothesis $a_1 \neq b_1$, this implies that the forecast error is correlated with X_{t-1}, which violates the orthogonality of conditional expectations as long as X_{t-1} is contained in the information set. Although it would be desirable for any expectation mechanism to satisfy at least some of these four properties, conditional expectations must satisfy all of them.

2.1.3 Learning processes

Role of learning in macroeconomics

Learning in macroeconomics refers to models of expectation formation in which agents revise their forecasting rules over time as new data becomes available. Learning plays a key rule in macroeconomics. Rational expectations can be assessed for stability under different kinds of learning such as least squares learning. Learning can be useful when there is a structural change in economy. Suppose a new government appears. Agents need to learn about the new regime. Besides, learning can be used as a selection criterion when a model has more than one equilibrium solution. (Bullard, 1991) Let us illustrate this point using a model of hyperinflation. Assume a government prints money to finance a constant budget deficit, then

$$P_tG_t = M_t - M_{t-1} \tag{1}$$

where P_t is the price level, $G_t = \overline{G}_t$ is the constant real deficit, and M_t is the money stock. Suppose the demand function for money is as follows

$$\frac{M_t}{P_t} = f(E_t p_{t+1}) \tag{2}$$

$$\frac{\partial \dfrac{M_t}{P_t}}{\partial E_t p_{t+1}} < 0$$

where $E_t p_{t+1} = E_t(\log(\frac{P_{t+1}}{P_t}))$ is the expected rate of inflation and real output has

been assumed constant. Considering equilibrium in the money market and substituting (2) into (1) will result in

$$\bar{G} = f(E_t p_{t+1}) - f(E_{t-1} p_t)e^{-p_t} \tag{3}$$

(since $\log\dfrac{P_t}{P_{t-1}} = p_t, \dfrac{P_t}{P_{t-1}} = e^{p_t}$)

It can be shown that equation (3) has two RE equilibria: the high inflation equilibrium and the low inflation equilibrium. If we assume rational expectations, the high inflation equilibrium is locally stable and the lower one is unstable. These rankings will be reversed if we assume adaptive expectations. If it is considered that stability is not the appropriate selection criteria in a rational expectations model then there is no mechanism to choose between the two equilibrium solutions. In such cases, learning provides a selection criterion.

Researchers have frequently faced the issue of multiplicity of RE equilibria in nonlinear models. Assume a nonlinear model $y_t = F(y_{t+1}^e)$ has the S-shape shown below
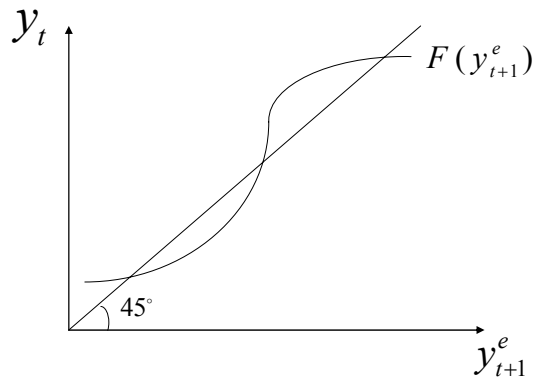
Figure 1.1: Multiplicity of solutions in nonlinear models

The multiple steady states $\bar{y} = F(\bar{y})$ occur at the intersection of the graph of $F(.)$ and 45-degree line. This possibility can appear in models with monopolistic competition, increasing returns to scale production or externalities. Other specifications of this model can present multiple perfect foresight equilibia taking the form of regular cycles in addition to a steady state or sunspot eqilibria, taking the form of a finite state Markov process (Evans and Honkapohja, 2001). An interesting question may now be posed: which of the steady states are stable under learning.

Approaches to learning

Following Evans and Honkapohja (1999, 2001), the approaches to learning can be categorized into three groups: eductive learning, adaptive learning, and rational learning.

2.1.3.1 Eductive learning

In the eductive approach, we examine whether expectations converge to the rational expectations equilibrium through a process of reasoning. Consider the following example based on Decanio(1979)

Consider the demand and supply in a market are given by

$$q_t = a - bp_t + w_t \tag{4}$$

$$q_t = c + dp_t^e + v_t \tag{5}$$

Here q_t and p_t are the actual quantity and price level, w_t and v_t are random disturbances which are assumed to be white noise and a, b, c, and d are constant. Demand is downward-sloping linear function of the market price and supply depends positively and linearly on expected price due to a production lag. p_t^e denotes the expectations of the representative supplier (average expectations). The good is assumed to perishable and markets clear. The reduced form for the prices is given by

$$p_t = A - Bp_t^e + u_t \tag{6}$$

where $A = \dfrac{a-c}{b}, B = \dfrac{d}{b}$, and $u_t = \dfrac{w_t - v_t}{b}$.

First we examine the model under RE. The RE hypothesis can be formally stated as

$$p_t^e = E(p_t | \mathrm{I}_{t\text{-}1}) = E_{t-1} p_t \tag{7}$$

So that expectations are the true mathematical conditional expectations, conditional on available information at the end of period t-1. The information set includes past data $\{u_{t-1}, u_{t-2}, \ldots, \mathrm{P}_{t\text{-}1}, \mathrm{P}_{t\text{-}2,\ldots}\} = \mathrm{I}_{t\text{-}1}$ and knowledge of the model. We can compute RE by substituting (7) in (6) and obtain

$$p_t = A - BE_{t-1} p_t + u_t \tag{8}$$

Taking conditional expectations $\mathrm{E}_{t\text{-}1}$ of both sides yields $E_{t-1} p_t = A - BE_{t-1} p_t$ so that expectations are given by

$$E_{t-1} p_t = \frac{A}{1+B}$$

And the unique RE solution is of this form: $p_t = \dfrac{A}{1+B} + v_t$.

The RE equilibrium for the model is a random variable that is of the form constant plus noise. Under RE the appropriate way to form expectations depends on the stochastic process followed by the exogenous variables, v_t in this case.

Now we consider the model under eductive learning. Suppose agents form their expectations initially in an arbitrary manner, for example, static expectations

$$E_{t-1}^0 p_t = p_{t-1} \tag{9}$$

The question is whether they can modify their behavior so that rational expectation equilibrium, given by $\dfrac{A}{1+B}$, would be attainable. Plugging (9) into (8) results in the actual evolution of prices

$$p_t = A - Bp_{t-1} + u_t \qquad (10)$$

It is assumed that after some passage of time agents realize (reason or deduce) that prices are evolving according to (10) and form new expectation

$$E_{t-1}^1 p_t = A - Bp_{t-1} \qquad (11)$$

The evolution of the system is changed by this new expectation

$$p_t = A - B(A - Bp_{t-1}) + u_t = A - BA + B^2 p_{t-1} + u_t$$

Observing the new evolution of prices, agents revise their expectations to

$$E_{t-1}^2 p_t = A - BA + B^2 p_{t-1} \qquad (12)$$

So that prices evolve as follows by plugging (12) into (8)

$$p_t = A - B(A - BA + B^2 p_{t-1}) + u_t = A - BA + B^2 A - B^3 p_{t-1} + u_t \qquad (13)$$

If we repeat this process, the expectations after n iterations will be

$$E_{t-1}^n p_t = A - BA + B^2 A - B^3 A + ... + B^n A + B^n p_{t-1} \qquad (14)$$

$$= A(1 - B + B^2 - B^3 + ... + B^n) + B^n p_{t-1}$$

Since $\left(1 - B + B^2 - B^3 + ... + B^n = \dfrac{1}{1+B}\right)$ and $B^n p_{t-1} \to 0$ for $|B|<1$ and large n, expectations will converge to rational expectations

$$E_{t-1}^n p_t = \frac{A}{1+B}$$

The rational expectations, in this case, is said to be iteratively E-stable. It is clear that convergence to rational expectations is not guaranteed if $|B|>1$. Guesnerie, 1992; Evans, 1985, 1986; Peel and Chappell, 1986; and Bullard and Mitra, 2000), employing the iterative expectations method in different models, examined convergence to rational expectations.

2.1.3.2 Adaptive learning

Agents would learn from data via regression about the model and the policy regime. Although this would produce expectations formation very similar to adaptive

expectations, it will not necessarily ever converge to rational expectations (Benjamin Friedman, 1979). Bray and Savin (1986) and Fourgeaud, Gourieroux and Pradel (1986) initially applied least-squares learning mechanism to see whether it would converge to rational expectations. Here, for simplicity, it will be assumed that the reduced form for prices is as follows

$$p_t = A - Bp_t^e + Cz_{t-1} + u_t \tag{15}$$

where z_{t-1} denotes observable exogenous variables. The rational expectations will be $E_{t-1}^n p_t = \dfrac{A + Cz_{t-1}}{1+B}$ and prices evolve as

$$p_t = \frac{A}{1+B} + \frac{C}{1+B} z_{t-1} + u_t = \bar{\alpha} + \bar{\beta} z_{t-1} + u_t \tag{16}$$

where $\bar{\alpha} = \dfrac{A}{1+B}$ and $\bar{\beta} = \dfrac{C}{1+B}$

It should be noted that this model has a unique RE since p_t does not depend on expected future prices. Now assume that agents know the true model but are unaware of the parameter values $\bar{\alpha}$ and $\bar{\beta}$. According to least-squares learning, agents are assumed to run least-squares regressions of p_t on z_{t-1} and an intercept. Rational forecast will be generated from the estimated model ($E_{t-1}^n p_t = \alpha_{t-1} + \beta_{t-1} z_{t-1}$). Agents revise the expectations by reestimating the model as more data becomes available. The coefficients (α_t, β_t) will converge to the unique RE ($\bar{\alpha}, \bar{\beta}$) if B<1. The conditions for convergence of recursive least-squares expectations (B<1) can be weaker than those under iterative expectations (|B|<1).

Agents perceive the reduced form as

$$y_t = \beta' x_t + e_t \tag{17}$$

Where the least squares estimated coefficients are given by

$$\beta_t = (\sum_{i=0}^{t-1} x_i y_i)(\sum_{i=0}^{t-1} x_i x_i')^{-1} \tag{18}$$

The recursive least-squares estimates can be shown to be

$$\beta_t = \beta_{t-1} + \gamma_t R_t^{-1} x_{t-1}(y_{t-1} - \beta_{t-1}' x_{t-1}) \tag{19}$$

and

$$R_t = R_{t-1} + \gamma_t(x_{t-1} x_{t-1}' - R_{t-1}) \tag{20}$$

with the gain $\gamma_t = \dfrac{1}{t}$, an important factor in determining the speed of convergence to the true parameter, and where R_t is an estimate of the moment matrix for x_t. For suitable initial conditions $R_t = t^{-1} \sum\limits_{i=0}^{t-1} x_i x_i'$.

Considering the recursive least-squares of the mean $Ez_t = \mu$ can help to understand the least-squares updating formula. The least-squares estimate is the sample mean $\bar{z}_t = \dfrac{1}{t} \sum\limits_{n=1}^{t} z_n$. If we subtract the sample mean at t-1 from both sides of $\bar{z}_t$ and rearrange, then

$$\bar{z}_t = \bar{z}_{t-1} + \frac{1}{t}(z_t - \bar{z}_{t-1}) \tag{21}$$

Since
$$
\begin{cases}
\bar{z}_t = \dfrac{1}{t} \sum\limits_{n=1}^{t} z_n \\[2mm]
\bar{z}_{t-1} = \dfrac{1}{t-1} \sum\limits_{n=1}^{t-1} z_n
\end{cases}
\qquad
\begin{cases}
t\bar{z}_t = \dfrac{1}{t} \sum\limits_{n=1}^{t} z_n = z_t + \sum\limits_{n=1}^{t-1} z_n \\[2mm]
(t-1)\bar{z}_{t-1} = \sum\limits_{n=1}^{t-1} z_n
\end{cases}
$$

$$t\bar{z}_t - (t-1)\bar{z}_{t-1} = z_t \rightarrow t(\bar{z}_t - \bar{z}_{t-1}) = z_t - \bar{z}_{t-1} \rightarrow \bar{z}_t - \bar{z}_{t-1} = \frac{1}{t}(z_t - \bar{z}_{t-1}) \rightarrow$$

$$\bar{z}_t = \bar{z}_{t-1} + \frac{1}{t}(z_t - \bar{z}_{t-1})$$

Adaptive methods of learning have the same structure which is given by

$$\theta_t = \theta_{t-1} + \lambda_t Q(\theta_t, \theta_{t-1}, X_t) \tag{22}$$

Where $\lambda_t = \dfrac{1}{t}$ in the case of least-squares, θ is a vector of parameters, Q is a function and X_t is the vector of variables in the structural model. Adaptive expectations is in fact a special case of least-square adaptive learning (21) if the gain parameter ($\lambda_t = \lambda$) is constant.

The evolution of X_t will depend on θ_{t-1}, in the case of a linear system

$$X_t = A(\theta_{t-1})X_{t-1} + B(\theta_{t-1})W_t \tag{23}$$

where W_t is a vector of disturbance term.

Stability results for linear and nonlinear systems have been derived by Marcet and Sargent (1989a, 1989b), Evans and Honkapohja (1998). Sergent (1999) asserts that if it is assumed that the US authorities employed constant-gain least-squares learning about the Phillips curve and maximized a social objective function to pick inflation, this fits US post-war data including the 'great inflation' well while rational expectations do not.

Stability under adaptive learning

When expectations are modeled by least-squares learning there is convergence to the rational expectation equilibrium (REE) as t→ ∞ provided that that a stability condition is met. This condition can usually be obtained by the expectational stability (*E-stability*) approach. Consider the agents' view of stochastic process for the market price as $p_t = \alpha + \beta z_{t-1} + u_t$ which is called the *perceived law of motion* (PLM). Expectations are based on the PLM and hence given by $p_t^e = \alpha + \beta z_{t-1}$, where ($\alpha,\beta$) may not be the REE values. Agents are boundedly rational because they do not initially know parameters (α,β) and they try to learn the REE solution over time. Inserting the PLM into the reduced form yields the corresponding *actual law of motion* (ALM):

$$p_t = (A - B\alpha) + (C - B\beta)z_{t-1} + u_t \tag{24}$$

This ALM has the same form as the PLM but with different values of the parameters. In fact, the above equation yields a mapping from the PLM parameters (α,β) into the ALM parameters T (α,β) = ($A - B\alpha, C - B\beta$).

Only at the REE values does one have T (α,β) = (α,β). Expectational stability looks at whether the REE is the stable outcome of a process that parameters of the PLM are adjusted slowly toward the parameters of the ALM that they induce. This adjustment is described by a differential equation and E-stability corresponds to local stability of the REE under these dynamics.

Consider a vector version of the model. Using the T mapping, E-stability is defined by the ordinary differential equation

$$\frac{d}{d\tau} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = T \begin{bmatrix} \alpha \\ \beta \end{bmatrix} - \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{25}$$

An REE is said to be E-stable if it is a locally asymptotically stable fixed point (or equilibrium point or steady state) of this differential equation. Here τ denotes virtual time (it is distinct from real time t and is measured in discrete periods). Plugging in the form of the mapping the system of differential equations will then be:

$$\frac{d\alpha}{d\tau} = A - (B+1)\alpha \ , \tag{26}$$

$$\frac{d\beta_i}{d\tau} = C_i - (B+1)\beta_i \ , \text{ for i = 1, 2,..., n,} \tag{27}$$

where n is the dimension of the vector of exogenous variables. Clearly, the unique fixed point is E-stable if B<1. Least-squares learning converges locally to an REE if and only if that REE is E-stable. Intuitively, a model is stable or learnable if the new data generated by one more observation under learning is on average closer to the REE than the current belief derived from past data.

2.1.3.3 Rational learning

The rational approach to learning recognizes the benefits and costs of more accurate forecasts for an agent so that rational expectations may not be achieved unless calculation costs are zero (Feige and Pierce, 1976; Evans and Ramsey, 1992). However the widely used method to model rational learning has been based on Bayes' theorem. It is a method of updating belief, implying that beliefs change by learning. Data or new facts only influence the posterior belief, P (A|B), through the likelihood function P (B|A)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{28}$$

where P(A) is prior belief. Many researchers used Bayes' rule to model learning in the economic literature including learning about a new regime (see Cyert and Degroot,1974; Backus and Driffill, 1985; Lewis, 1998; and Ellison and Valla, 2000). Consider the following example presented by Lewis (1988).

Assume the reduced form for the exchange rate is given by

$$s_t = m_t + \alpha(E_t s_{t+1} - s_t) \tag{29}$$

where m_t is the money supply at time t, s_t exchange rate and α is a positive constant. Also assume the money supply is as follows

$$m_t = \theta_0 + \varepsilon_t^0 \tag{30}$$

where θ_0 is a constant and $\varepsilon_t^0 \sim N(0, \sigma_0^2)$. Suppose at t=0 agents come to believe that the supply money process may have changed due to a new regime. The new process has the same form (30) except with different mean and variance:

$$m_t = \theta_1 + \varepsilon_t^1 \text{ for } t \geq 0 \tag{31}$$

We assume $\theta_1 < \theta_0$ and $\theta_1 = 0$, so that the process can be interpreted as going from 'loose' to 'tight' money. It is also assumed agents believe that if the policy has changed it will not be changed back and they also know the parameters of the potential new process. We can obtain the solution by solving (31) forward

$$s_t = (1 - \gamma) \sum_{i=0}^{\infty} \gamma^i E_t m_{t+i} \tag{32}$$

where $\gamma = \dfrac{\alpha}{1+\alpha}$

Expected money supply equals

$$E_t m_{t+i} = \theta_0 (1 - P_{1,t}) \text{ for any i>0 and } t \geq 0 \tag{33}$$

where $P_{1,t}$ is agents' assessed probability at time t that the process changed at time 0. Finally the exchange rate is obtained as

$$s_t = (1 - \gamma) m_t + \gamma (1 - P_{1,t}) \theta_0 \tag{34}$$

To obtain the best estimate of $P_{1,t}$, agents combine their prior beliefs about the probability together with their observations of money outcomes each period to update their posterior probabilities according to the Bayes' rule

$$P_{1,t} = \frac{P_{1,t-1} f(I_t | \theta_1)}{P_{1,t-1} f(I_t | \theta_1) + P_{0,t-1} f(I_t | \theta_0)} \tag{35}$$

where $P_{0,t}$ is the conditional probability of no change at t=0, $f(I_t | \theta_1)$ is the probability of observing the information set I_t given that m_t follows the ith process. The ratio of posterior probabilities of each process, the posterior odds, is given by

$$\frac{P_{1,t}}{P_{0,t}} = \frac{P_{1,t-1} f(m_t | \theta_1)}{P_{0,t-1} f(m_t | \theta_0)} = \left[\frac{P_{1,t-1}}{P_{0,t-1}} \right] \left[\frac{(1/\sigma_1) \exp(-(1/2)[m/\sigma_1]^2)}{(1/\sigma_0) \exp(-(1/2)[(m-\theta_0)/\sigma_0]^2)} \right] \tag{36}$$

The first term on the right-hand side of equation (36) indicates that the change from t-1 to t in the relative conditional probabilities depends on the observation of the current money supply at time t. For instance, for some observation of current money supply, say $\overline{m}$, the probability of being under either policy process is the same; i.e., $f(\overline{m}_t|\theta_1) = f(\overline{m}_t|\theta_0)$, so that the posterior probabilities, $\dfrac{P_{1,t}}{P_{0,t}}$, equal the prior probabilities, $\dfrac{P_{1,t-1}}{P_{0,t-1}}$, and therefore the conditional probabilities do not change.

However, observations of money different from $\overline{m}$ convey information, the last term on the right-hand side of (36), about the regimes causing probabilities to be revised. To analyze the behavior of the probabilities, equation (36) can be written as

$$\log(\frac{P_{1,t}}{P_{0,t}}) = \log(\frac{P_{1,t-1}}{P_{0,t-1}}) + \log(\frac{f(m_t|\theta_1)}{f(m_t|\theta_0)}) \tag{37}$$

Equation (37) is a linear difference equation in the dependent variable, $\log(\dfrac{P_{1,t}}{P_{0,t}})$. For simplicity assume that $\sigma_1 = \sigma_0 = \sigma$, then

$$\log\frac{f(m_k|\theta_1)}{f(m_k|\theta_0)} = \left[\frac{(m_k - \theta_0)^2 - m_k^2}{2\sigma^2}\right] \tag{38}$$

Given the initial probabilities $P_{1,0}$ and $P_{0,0}$, and plugging (38) into (37), we obtain the solution to the difference equation as

$$\log(\frac{P_{1,t}}{P_{0,t}}) = \log(\frac{P_{1,0}}{P_{0,0}}) + \sum_{k=1}^{t}\left[\frac{\theta_0^2 - 2m_k\theta_0}{2\sigma^2}\right] \tag{39}$$

Equation (39) indicates that the behavior of the probabilities depends on the actual observations of the process. For example, when the money supply observed today is strongly negative, agents think it is more likely that policy has changed.

Taking expectations of (39) and defining θ_i as the true θ gives

$$E\log(\frac{P_{1,t}}{P_{0,t}}) = \log(\frac{P_{1,0}}{P_{0,0}}) + t\left[\frac{\theta_0^2 - 2\theta_i\theta_0}{2\sigma^2}\right] \tag{40}$$

Equation (40) shows that the expected value of the 'true' process rises over time. For example, if policy has changed so that $\theta_i = \theta_1 = 0$, then the log probability increases

to infinity due to the term $\dfrac{t\theta_0^2}{2\sigma^2}$ as t goes to infinity. Similarly, when policy has not

changed so that $\theta_i = \theta_0$ the log probability goes to negative infinity due to the

term $\dfrac{-t\theta_0^2}{2\sigma^2}$ as t goes to infinity. Also, it can be demonstrated that the expected value

of the log ratio of probabilities converges and its speed depends positively on $\dfrac{\theta_0^2}{\sigma^2}$.

Therefore, the speed of market learning depends upon the squared signal-to-noise ratio.

Using the above analysis, Lewis (1988) investigates the effects of the probability behavior on the exchange rate and forecast errors. Taking expectation of the exchange rate at t-1 and subtracting from (34) we obtain the forecast errors of exchange rate corresponding to each potential process

$$s_t - E_{t-1}s_t = (1-\gamma)\varepsilon_t^0 + \theta_0(P_{1,t-1} - \gamma P_{1,t}) \text{ if } \theta_i = \theta_0 \tag{41}$$

$$s_t - E_{t-1}s_t = (1-\gamma)\varepsilon_t^1 + \theta_0(P_{0,t-1} - \gamma P_{0,t}) \text{ if } \theta_i = \theta_1 \tag{42}$$

The expected value of the last component of the equations (41) and (42) shows dependence on the conditional probabilities. Whilst agents are learning, the evolution of theses probabilities depends on the random observations of the money process, and does not equal the true values.

Taking expectations of forecast errors in equation (42) conditional upon a change in policy to θ_1 and initial probabilities, gives the expected evolution of forecast errors, for a large number of m_k as

$$E(s_t - E_{t-1}s_t|\theta_1) = -\theta_0[E(P_{0,t-1}|\theta_1) - \gamma E(P_{0,t}|\theta_1)] < 0 \tag{43}$$

The inequality is negative since the discount rate, γ, is less than one and $E(P_{0,t}|\theta_1) < E(P_{0,t-1}|\theta_1)$. Hence, if agents do not completely realize that the policy has changed to a 'tighter' money supply process, the exchange rate will be expected to be weaker than subsequently occurs. Lewis' model shows well how learning about a regime change using Bayes' rule can imitate the outcomes of the Peso problem[1].

----

1. The peso problem, which was initially examined by Milton Friedman in his analysis of the behavior of the Mexican currency, refers to a situation where rational agents anticipate the possibility of future changes in the data-generating mechanism of economic variables.

3. Expectation formation using statistical predictors

Chapter three presents expectation formation using statistical predictors. Statistical predictors are used by economic agents to generate forecasts on future values of variable of interest. From this point of view, statistical predictors may be regarded as simple approaches to expectation formation that are more complex than simple adaptive expectations but less demanding than the concept of rational expectations. Basically, statistical predictors are backward-looking functions of past observations that provide estimates of future values. Broadly, for statistical predictor, a distinction can be made between parametric and nonparametric approaches. The parametric approaches including autoregressive integrated moving average (ARIMA) models, state space models, and Kalman filter and nonparametric regressions such as the additive model (AD), multiple adaptive regression splines (MARS), projection-pursuit regression (PPR) are discussed.

3.1 Parametric prediction models

The parametric regression approach is based on the prior knowledge of the functional form relationship. If knowledge is correct, the parametric method can model most data sets well. However, if the wrong functional form is chosen a priori, this will result in larger bias as compared to competitive models. Parametric linear models, as a type of parametric regression, are frequently used to describe the association between the dependent variable and explanatory variables. They require the estimation of a finite number of parameters. We will apply ordinary least square (OLS) and two-stage least square (2SLS) estimators for linear models. Also, parametric linear dynamic models such as autoregressive and moving-average models which are based on a atheoretical or data-driven approach will be employed. Now we review some basic theory for time series and present a brief discussion of state-space model and Kalman filter.

3.1.1 ARIMA modeling

Autoregressive integrated moving average (ARIMA) or (Box-Jenkins) models are the basis of many fundamental ideas in time-series analysis. In order to analyze a time

series, it must be assumed that the structure of the stochastic process which generates the observations is essentially invariant through time. The important assumption is that of stationarity, which requires the process to be in a particular state of 'statistical equilibrium' (Box and Jenkins, 1976). A stochastic process is said to be second-order (or weak) stationary if its first and second moments are finite and do not change through time

$$E[X_t] = \mu$$

$$Var[X_t] = \sigma^2$$

$$Cov[X_t, X_{t+k}] = E[(X_t - \mu)(X_{t+k} - \mu)]$$

$$= \gamma_k$$

Note that γ_0 equals the variance, σ^2. The set of autocovariance coefficients $\{\gamma_k\}$ for k=0,1,2,... constitute the autocovariance function (acv.f.) of the process. The autocorrlation coefficients, $\{\rho_k\}$ are also obtained as

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

The set of autocorrlation coefficients, $\{\rho_k\}$ constitute the autocorrlation function (ac.f.). If the time series X_t is stationary, ρ_k measures the correlation at lag k between X_t and X_{t+k}. Another useful function in model identification is the partial autocorrelation function. It measures the excess correlation at lag k which has not already been accounted for by autocorrelations at lower lags.

The pure random process (ε_t) is a sequence of uncorrelated, identically distributed random variables with zero mean and constant variance. This process is stationary and has the following ac.f.

$$\rho_k = \begin{cases} 1 & k=0 \\ 0 & \text{otherwise} \end{cases}$$

It is also called uncorrelated white noise or innovation process. Using this process, the random walk model is stated as

$$X_t = X_{t-1} + \varepsilon_t$$

Where $\{\varepsilon_t\}$ denotes a pure random process. Since the variance increases through time, the series X_t is not stationary. However, it would be stationary if we take the first differences of the series $(X_t - X_{t-1}) = \varepsilon_t$.

Autoregressive (AR) processes

A process $\{X_t\}$ is said to be an autoregressive process of order p, AR (p), if it s a weighted linear sum of the past p values plus a random shock so that

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \varepsilon_t$$

where α and ϕ_1 to ϕ_p are unknown parameters. The process $\{\varepsilon_t\}$ denotes a white noise with zero mean and variance σ_ε^2. Using the lag operator L with $L^k x_t = x_{t-k}$, the AR (p) model can then be written in a more concise form as

$$\phi(L)X_t = \alpha + \varepsilon_t$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$ is a polynomial in L of order p. The statistical properties of AR process are determined by values of the parameters $\phi_1, ..., \phi_p$. For instance, the condition for stationary can be expressed in terms of the roots of the polynomial $\phi(z)$ by factorizing this polynomial in terms of its p roots $z_i = \dfrac{1}{\alpha_i}$ as

$$\phi(z) = (1 - \alpha_1 z)(1 - \alpha_2 z)...(1 - \alpha_p z).$$

The process is stationary if and only if $|\alpha_k| < 1$ for all k=1,...,p- that is the roots of $\phi(z) = 0$ should lie outside the unite circle.

The simplest type of AR process is AR (1), given by

$$X_t = \phi X_{t-1} + \varepsilon_t$$

Here, for simplicity, we assume that $\alpha = 0$. It is clear that if $\phi = 1$, the model reduces to a random walk, when the model is non-stationary. This process, by recursive substitution of the lagged values of X_t, can be rewritten as

$$X_t = \phi^{t-1} X_1 + \sum_{j=0}^{t-2} \phi^j \varepsilon_{t-j}, \text{ t=2,...,n.}$$

If $|\phi| > 1$, then the impact of the white noise ε grows over time, the series will be explosive and hence non-stationary. However, if $|\phi| < 1$ the impact dies out over time and the process will be stationary.

It can be shown that the variance and the ac.f. of a stationary AR(1) process (with $|\phi| < 1$) are given by

$$\gamma_0 = \frac{\sigma^2}{1 - \phi^2}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k$$

The autocorrelations approach exponentially a value of zero as k→∞. For $\phi = 1$ the series X_t, which is non-stationary, does not have a finite variance and it has a trending behavior. For the AR(p) process, the partial ac.f. is zero at all lags greater than p which implies that we can determine the order of an AR process by looking for the lag value at which the sample ac.f. "cuts off" (not significantly different from zero).

Moving average (MA) processes

A process $\{X_t\}$ is called a moving average process of order q, MA (q), if

$$X_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} \tag{1}$$

where ε_t is white noise. This process may be written in the form

$$X_t = \alpha + \theta(L)\varepsilon_t$$

where $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$ is a polynomial in L of order q.

This process is stationary for all parameter values with the following properties

$$E[X_t] = \alpha$$

$$\gamma_0 = \sigma^2 (1 + \sum_{j=1}^{q} \theta_j^2)$$

$$\gamma_k = \sigma^2 (\theta_k + \sum_{j=k+1}^{q} \theta_j \theta_{j-k}) \text{ for k≤q and } \gamma_k = 0 \text{ for k>q.}$$

In order to ensure that there is a unique MA model, we need to impose some restrictions on the parameters, called invertibility condition, of the model. In fact, if a MA model can be expressed as an autoregressive model, then the MA model is called invertible. In this case the error terms ε_t in (1) are equal to the innovations $\varepsilon_t = X_t - E(X_t \mid I_{t-1})$, where I_{t-1} is the information set available at time t-1, $I_{t-1} = \{X_{t-1}, X_{t-2}, ...\}$, so that

$$E(X_t \mid I_{t-1}) = \alpha + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

The invertibility condition can be expressed in terms of the roots of the polynomial $\theta(z)$ by factorizing the MA polynomial in terms of its q roots as

$$\theta(z) = (1 - \beta_1 z)(1 - \beta_2 z)...(1 - \beta_q z)$$

Invertibility is equivalent to the condition that $|\beta_j| < 1$ for all j=1,...,q (Heij et al, 2004)- that is the roots of $\theta(z) = 0$ should lie outside the unite circle.

The simplest type of the MA (q) model is the first order case, MA (1), given by

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

Here, for simplicity, we assume that $\alpha = 0$. This process is stationary for all values of θ with an ac.f. given by

$$\rho_k = \begin{cases} 1 & k = 0 \\ \theta \big/ (1+\theta^2) & k = 1 \\ 0 & k > 1 \end{cases}$$

Hence the ac.f. 'cuts off' at lag 1. For the MA (1) process to be invertible, ε_t should be expressed in terms of current and past values of the observed process. Therefore

$$\varepsilon_t = X_t - \theta \varepsilon_{t-1} \qquad (2)$$

$$\varepsilon_{t-1} = X_{t-1} - \theta \varepsilon_{t-2} \qquad (3)$$

Plugging (3) into (2) results in

$$\varepsilon_t = X_t - \theta(X_{t-1} - \theta \varepsilon_{t-2}) = X_t - \theta X_{t-1} + \theta^2 \varepsilon_{t-2}$$

By further substitutions we obtain

$$\varepsilon_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - ... + (-\theta)^{t-2} X_2 + (-\theta)^{t-1} \varepsilon_1$$

Invertibility requires that, in the limit, the error term on the right-hand side vanishes. This holds if and only if $|\theta| < 1$.

Autoregressive moving average process: ARMA (p,q)

An autoregressive moving average process: ARMA (p,q) is obtained by combining p autoregressive terms and q moving average terms and can be written as

$$\phi(L)X_t = \alpha + \theta(L)\varepsilon_t$$

with AR polynomial $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$ and

MA polynomial $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$. An ARMA model is stationary provided that the roots of $\phi(L) = 0$ lie outside the unite circle. This process is invertible if the roots of $\theta(L) = 0$ lie outside the unite circle. Low order ARMA models are of much interest since many real data sets are well approximated by them rather than by a pure AR or pure MA model. In general, ARMA models need fewer parameters to describe the process.

In most cases economic time series are non-stationary and therefore we cannot apply ARMA models directly. One possible way to remove the problem is to take difference so as to make them stationary. Non-stationary series often become stationary after taking first difference ($X_t - X_{t-1} = (1 - L)X_t$). If the original time series is differenced d times, then the model is said to be an ARIMA (*p, d, q*) where 'I' stands for integrated and *d* denotes the number of differences taken. Such a model is described by

$$\phi(L)(1 - L)^d X_t = \alpha + \theta(L)\varepsilon_t$$

The combined AR operator is now $\phi(L)(1 - L)^d$. The polynomials $\phi(z)$ and $\theta(z)$ have all their roots outside the unit circle. The model is called integrated of order d and the process is said to have d unit roots.

3.1.2 State-space modeling

State space models originate from control theories (Kalman, 1960) but have been received much attention in the economics literature since 1990s. A state space model includes two equations: measurement (or observation) equation and transition (or state) equation. The measurement equation specifies the relationship between the observed and unobserved (state) variables while transition equation models the dynamics of state variables. For a linear Gaussian state space model, the Kalman filtering approach provides optimal estimates for the state variables based on the information from the transition equation and the observations.

State-space model

Following Harvey (1991, 1993), let y_t be a $N \times 1$ vector of observed variables at time t which is related to an $m \times 1$ state vector, α_t, through a measurement equation

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \qquad t = 1, \ldots, T \tag{1}$$

where Z_t is an $N \times m$ matrix, d_t an $N \times 1$ vector and ε_t an $N \times 1$ vector of serially uncorrelated disturbances with mean zero and covariance matrix H_t. The unknown vector α_t is assumed to follow a first order Markov process,

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t, \qquad t = 1, \ldots, T \tag{2}$$

where T_t is an $m \times m$ matrix, c_t an $m \times 1$ vector, R_t an $m \times g$ matrix, and η_t a $g \times 1$ vector of serially uncorrelated disturbances with mean zero and covariance matrix Q_t. Equation (2) is called the transition equation. The matrices Z_t, d_t, and H_t in the measurement equation and the matrices T_t, c_t, R_t, and Q_t in the transition equation are referred to as the system matrices. The model is said to be time invariant or time homogeneous if the system matrices do not change over time, otherwise, it is time variant. For instance, the AR (1) plus noise model

$$y_t = \mu_t + \varepsilon_t$$

$$\mu_t = \phi \mu_{t-1} + \eta_t$$

is a time invariant state space model with η_t being the state.

Kalman filter

The Kalman filter can be applied to the state-space form equations to estimate time-varying parameters. The estimations can be carried out in three steps: prediction, updating and smoothing. The first step is to calculate the optimal estimator of the state vector given all the currently available information. Reaching the end of series, optimal predictions of state vector for the next period can be made. Updating step is done as new observation becomes available. Using a backward recursion, the

estimators are smoothed based on the full sample in the final step. These steps are presented below in more detail.

Let a_t denote the optimal estimate of the state vector, α_t, based on all observations (t=1,…,t), and P_t the $m \times m$ covariance matrix of the estimate error, that is

$$P_t = E[(\alpha_t - a_t)(\alpha_t - a_t)^{'}]$$

Now assume that we are at time t-1, and that a_{t-1} and P_{t-1} are given. The optimal estimate of α_t is then given by the prediction equations

$$a_{t|t-1} = T_t a_{t-1} + c_t$$

and

$$P_{t|t-1} = T_t P_{t-1} T_t^{'} + R_t Q_t R_t^{'}, \qquad t = 1,...,T$$

While the corresponding estimate of y_t is

$$\tilde{y}_{t|t-1} = Z_t a_{t|t-1} + d_t, \qquad t = 1,...,T$$

Once the new observations of y_t becomes available, the estimator of the state can be updated with updating equations

$$a_t = a_{t|t-1} + P_{t|t-1} Z_t^{'} F_t^{-1} v_t,$$

and

$$P_t = P_{t|t-1} - P_{t|t-1} Z_t^{'} F_t^{-1} Z_t P_{t|t-1}, \qquad t = 1,...,T$$

where $v_t = y_t - Z_t a_{t|t-1} - d_t$ is the prediction error and $F_t = Z_t P_{t|t-1} Z_t^{'} + H_t$ the MSE of the prediction error.

The prediction and updating equations utilize information available at time t in estimating the state vector while the smoothing step is done using the information available after time t. Applying the fixed-interval smoothing algorithm, the last step start with the final quantities, a_T and P_T , and work backwards. The smoothing equations are

$$a_{t|T} = a_t + P_t^{*}(a_{t+1|T} - T_{t+1} a_t - c_{t+1})$$

and

$$P_{t|T} = P_t + P_t^{*}(P_{t+1|T} - P_{t+1|t})P_t^{*'}$$

where

$$P_t^* = P_t T_{t+1}' P_{t+1|t}^{-1}, \qquad t = 1, ..., T$$

with $a_{T|T} = a_T$ and $P_{T|T} = P_T$.

Estimating the state vector

Estimation of parameters in the state equation (vector ψ which is referred to as hyperparameters), assuming initial values of a_t and P_t (a_0 and P_0)[1], can be carried out by the method of maximum likelihood. For a multivariate

$$L(y; \psi) = \prod_{t=1}^{T} p(y_t | Y_{t-1})$$

where $p(y_t | Y_{t-1})$ denotes the distribution of y_t conditional on $Y_{t-1} = \{y_{t-1}, y_{t-2}, ;, ..., y_1\}$. For a Gaussian model, the likelihood function above can be written as

$$\log L(\psi) = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log |F_t| - \frac{1}{2} \sum_{t=1}^{T} v_t' |F_t^{-1}| v_t$$

Time-varying parameter models and state-space form

It is possible to analyze the time-varying parameter model which can be cast in state-space form. Consider a linear model

$$y_t = x_t' \beta + \varepsilon_t, \qquad t = 1, ..., T$$

where x_t is a $k \times 1$ vector of exogenous variables and β the corresponding $k \times 1$ vector of unknown parameters. We can use the state-space model and Kalman filter to estimate the time-varying parameter model. In this case, β is allowed to evolve over time according to various stochastic processes. Now let us examine different

----

1. The initial values for a stationary and time-varying transition equation are given as $a_0 = (I - T)^{-1} c$ and $vec(P_0) = [I - T \otimes T]^{-1} vec(RQR')$. If the transition equation is non-stationary, the initial values must be estimated from the model. To do so, there are two approaches. The first assumes the initial state is fixed with $P_0 = 0$ and is estimated as unknown parameters in the model. The second assumes that the initial state is random and has a diffuse distribution with $P_0 = \kappa I$, where κ goes to ∞.

forms of time-varying parameter model.

Consider first, the *random walk* model. Here the time-varying coefficients follow a random walk. The state space form is as follows

$$y_t = x_t' \beta_t + \varepsilon_t, \qquad t = 1, \ldots, T$$

$$\beta_t = \beta_{t-1} + \eta_t$$

where $\varepsilon_t \sim NID(0, \sigma^2)$, $\eta_t \sim NID(0, \sigma^2 Q)$, and β_t denotes the state vector. The $k \times k$ positive semi-definite matrix Q determines to what extent it may vary. In case $Q = 0$, the model reduces to an ordinary linear regression model because $\beta_t = \beta_{t-1}$. But if Q is positive definite, all coefficients will be time-varying.

The second time-varying form might be referred to as *return to normality* model. In this model, the time-varying coefficients are generated by a stationary vector AR (1) process. The state-space form could be represented as

$$y_t = x_t' \beta_t + \varepsilon_t, \qquad t = 1, \ldots, T$$

$$\beta_t - \overline{\beta} = \phi(\beta_{t-1} - \overline{\beta}) + \eta_t$$

where $\varepsilon_t \sim NID(0, \sigma^2)$, $\eta_t \sim NID(0, \sigma^2 Q)$. The stationary coefficients evolve around a constant mean, $\overline{\beta}$. If the matrix $\phi = 0$, the model is called *random-coefficient* model. In this case, the coefficients have a fixed mean ($\overline{\beta}$) but are allowed to evolve randomly around it.

Applying the Kalman filter and letting $\beta_t^* = \beta_t - \overline{\beta}$, the return to normality model could be rewritten as

$$y_t = (x_t' x_t') \alpha_t + \varepsilon_t, \qquad t = 1, \ldots, T$$

and

$$\alpha_t = \begin{bmatrix} \overline{\beta}_t \\ \beta_t^* \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & \phi \end{bmatrix} = \begin{bmatrix} \overline{\beta}_{t-1} \\ \beta_{t-1}^* \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix}$$

A diffuse prior is used for $\overline{\beta}_t$, meaning that starting values are constructed from the first k observations. The initial values of β_t^* is given by a zero vector and the initial values of covariance matrix is given as

$$vec(P_0) = [I - T \otimes T]^{-1} vec(RQR').$$

3.2 Nonparametric prediction models

Over the last decade, increasing attention has been devoted to nonparametric regression as a new technique for estimation and forecasting in different sciences including economics. This section will examine nonparametric regression.

Why is nonparametric regression important? A comparison between parametric and nonparametric estimation is needed to answer this question. In parametric regression estimation such as linear regression one assumes the regression function is known and depends only on a few parameters, and one uses data to estimate these parameters. As a result, we can easily interpret the coefficients but this method has a limited flexibility and is useful only when the underlying relationship is close to the pre-specified estimation function in the model. In fact if the true underlying regression function is not linear then a linear regression estimate will produce a large error for every sample size.

Nonparametric regression analysis relaxes the assumption of linearity in regression analysis and allows to explore data more flexibly. However, in high dimensions the variance of the estimates increases rapidly, known as the "curse of dimensionality", due to the sparseness of data. To overcome this problem, some nonparametric methods have been proposed such as the additive model (AD), multiple adaptive regression splines (MARS), projection-pursuit regression (PPR).

3.2.1 Nonparametric Smoothers

The general nonparametric regression model (Fox, 2000, 2005) is as follows:

$$y_i = f(X_i^{'}) + \varepsilon_i$$

$$= f(x_{i1}, x_{i2}, ..., x_{ik}) + \varepsilon_i \qquad \varepsilon_i \sim NID(0, \sigma^2)$$

The regression function is $f(.)$ unspecified in advance and is estimated directly. In fact, there is no parameter to estimate. It is implicitly assumed that $f(.)$ is a smooth, continuous function. If there is only one predictor $y_i = f(x_i) + \varepsilon_i$ then it is called 'scatter plot smoothing' because it traces a smooth curve through a scatter plot of y against x.

There are several smoothers such as local averaging, kernel smoother, Weighted Scatterplot Smoothing (lowess) and spline Smoother that fit a linear or polynomial

regression to the data points in the vicinity of x and then use the smoothed value as the predicted value at x.

3.2.1.1 Local Averaging

In local averaging procedures, we move a window continuously over the data, averaging the observations that fall in the window. The estimated values $\hat{f}(x)$ at a number of focal values of x are calculated and connected. It is possible to use a window of fixed width or to adjust the width of window to include a constant number of observations. Local averages are usually subject to boundary bias, roughness and distortion (when outliers fall in the window).

3.2.1.2 Kernel Smoother

A Kernel smoother is an extension of local averaging and usually produces a smoother result. At the focal value x_0, it is of the form

$$\hat{f}(x_0) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{x_i - x_0}{b}\right)}{\sum_{i=1}^{n} K\left(\frac{x_i - x_0}{b}\right)}$$

where b is a bandwidth parameter, and K a kernel function. The Gaussian kernel $(K_N(z))$ and the tricube kernel $(K_T(z))$ are popular choices of kernel functions.

$$K_N(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$K_T(z) = \begin{cases} (1-|z|^3)^3 & \text{for} \quad |z|<1 \\ 0 & \text{for} \quad |z|\geq 1 \end{cases}$$

For the Gaussian kernel the bandwidth b is the standard deviation of a normal distribution and for the tricube kernel b is the half-width of a window enclosing the observations for the local regression. Although the kernel smoother has a better performance as compared to the local average regression, it is still subject to boundary bias.

It is implicitly assumed that the bandwidth b is fixed, but it is possible for kernel smoothers to be adapted to nearest-neighbor bandwidths. We can adjust b (x) so that a constant number of observations m are included in the window. The fraction m/n is called the span of the kernel smoother and is chosen based on a cross-validation

approach. The kernel estimator can produce smoother results using larger bandwidths. In fact, there is a direct relationship between the span and smoothing degree: the larger the span, the smoother the result.

3.2.1.3 Lowess Smoother

As mentioned above, the kernel estimation has some problems. Local polynomial regression tries to overcome these difficulties and provides a generally adequate method of nonparametric regression which extends to additive regression (Fox, 2005). An implementation of local polynomial regression is lowess (Cleveland, 1979). The algorithm used by lowess smoothing applies robust locally linear fits. It is similar to local averaging but the data points that lie in the window are weighted so that nearby points get the most weight and a robust weighted regression is used.

We can examine local polynomial regression in two cases: simple regression and multiple regression.

Simple Regression: suppose we want to estimate the simple regression $y_i = f(x_i) + \varepsilon_i$ at a particular x-value, for example x_0. Local polynomial regression extends kernel estimation to a polynomial fit at x_0, using local kernel weights, $w_i = K[(x_i - x_0)/b]$. We implement a pth-order weighted-least-squares polynomial regression of y on x,

$$y_i = \alpha + \beta_1(x_i - x_0) + \beta_2(x_i - x_0)^2 + ... + \beta_p(x_i - x_0)^p + e_i$$

to minimize the weighted residual sum of squares, $\sum_{i=1}^{n} w_i e_i^2$. This procedure is repeated for representative values of x. As in kernel regression, the bandwidth b can either be fixed or variable, b(x), and the span of the local-regression smoother is selected based on a cross-validation approach.

Multiple Regression: in this case, $y_i = f(X_i') + \varepsilon_i$, we need to define a

a multivariate neighborhood around a focal point $x_0' = (x_{01}, x_{02}, ..., x_{0k})$. Furthermore, Euclidean distance is employed in the lowess function as:

$$D(x_i, x_0) = \sqrt{\sum_{j-=1}^{k} (z_{ij} - z_{0j})^2}$$ where the z_{ij} are the standardized predictors,

$z_{ij} = x_{ij} - \overline{x}_j / s_j$, $\overline{x}_j$ is the mean of the jth predictor and s_j is its standard deviation. Calculating weights are based on the scaled distances:

$$w_i = W\left[\frac{D(x_i, x_0)}{b}\right]$$

Where w (.) is a weight function. In some cases, b needs to be adjusted to define a neighborhood including the [ns] nearest neighbors of x_0 (where the square brackets denote rounding to the nearest integer).

As a simple example, a local linear fit takes the form:

$$y_i = \alpha + \beta_1(x_{i1} - x_{01}) + \beta_2(x_{i2} - x_{02})^2 + \ldots + \beta_k(x_{ik} - x_{0k}) + e_i$$

The combinations of predictor values are used repeatedly to create the regression surface.

3.2.1.4 Spline Smoother

Suppose we have n pairs (x_i, y_i). A smoothing spline equation is considered as

$$ss(h) = \sum_{i=1}^{n}\left[y_i - f(x_i)\right]^2 + h\int_{x_{\min}}^{x_{\max}}\left[f''(x)\right]^2 dx$$

The equation consists of two terms. The first term is the residual sum of squares and the second term is a roughness penalty. The object is to find the function $\hat{f}(x)$ with two continuous derivatives that minimized the penalized sum of squares. Here h is a smoothing parameter. For h=0, $\hat{f}(x)$ will interpolate the data if the x_i are distinct; this is similar to a local-regression estimate with span=1/n. If h is very large, then $\hat{f}$ will be selected so that $\hat{f}''(x)$ is everywhere 0, which implies globally linear least-squares fit to the data. This is again similar to local regression with infinite neighborhoods.

The Spline Smoother is more attractive than local regression because there is an explicit objective-function to optimize. But it is not easy to generalize splines to multiple regression. Generally, the smoothing parameter h is selected indirectly by setting the equivalent number of parameters for the smoother .Both smoothing-spline and local-regression fits with the same degree of freedom are usually very similar.

3.2.2 Nonparametric Models

3.2.2.1 Additive model (AD)

Nonparametric regression based on kernel and smoothing spline estimates in high dimensions faces two problems, that is, the curse of dimensionality and interpretability. Stone (1985) proposed the additive model to overcome these problems. In this model, since each of the individual additive terms is estimated using a univariate smoother, the curse of dimensionality is avoided. Furthermore, while the nonparametric form makes the model more flexible, the additivity allows us to interpret the estimates of the individual terms. Hastie and Tibshirani (1990) proposed generalized additive models for a wide range of distribution families. These models allow the response variable distribution to be any member of the exponential family of distributions. We can apply additive models to Gaussian response data, logistic regression models for binary data, and loglinear or log-additive models for Poisson count data.

A generalized additive model has the form

$$Y = \alpha + f_1(X_1) + f_2(X_2) + ... + f_p(X_p) + \varepsilon$$

where $f_j(.)$ are unspecified smooth (partial-regression)functions. We fit each function using a scatterplot smoother and provide an algorithm for simultaneously estimating all j functions. Here an additive model is applied to a logistic regression model as a generalized additive model. Consider a logistic regression model for binary data. The mean of the binary response $\mu(X) = \Pr(Y = 1 | X)$ is related to the explanatory variables via a linear regression model and the logit link functions:

$$\log\left(\frac{\mu(X)}{1 - \mu(X)}\right) = \alpha + \beta_1 X_1 + ... + \beta_j X_j$$

The additive logistic model replaces each linear term by a more general functional form

$$\log\left(\frac{\mu(X)}{1 - \mu(X)}\right) = \alpha + f_1(X_1) + ... + f_j(X_j)$$

In general, the conditional mean $\mu(X)$ of a response Y is related to an additive function of the explanatory variables via a link function g:

$$g[\mu(X)] = \alpha + f_1(X_1) + \ldots + f_j(X_j)$$

The functions f_j are estimated in a flexible way using the backfitting algorithm. This algorithm fits an additive model using regression-type fitting mechanisms.

Consider the jth set of partial residuals

$$\varepsilon_j = Y - (\alpha + \sum_{k \neq j} f_k(X_k))$$

Then $E(\varepsilon_j | X_j) = f_j(X_j)$. This observation provides a way for estimating each $f_j(.)$ given estimates $\{\hat{f}_j(.), i \neq j\}$ for all the others. The iterative process is called the backfitting algorithm (Friedman and Stuetzle, 1981).

3.2.2.2 Multiple Adaptive Regression Splines (MARS)

This approach (Friedman (1991), Hastie et al (2001)) fits a weighted sum of multivariate spline basis functions and is well suited for high-dimensional problems, where the curse of dimensionality would likely create problems for other methods. The MARS uses the basis functions $(x-t)_+$ and $(t-x)_+$ in the following way

$$(x-t)_+ = \begin{cases} \text{x-t} & \text{if} \quad \text{x>t} \\ 0 & \text{otherwise} \end{cases}$$

$$(t-x)_+ = \begin{cases} \text{t-x} & \text{if} \quad \text{x<t} \\ 0 & \text{otherwise} \end{cases}$$

The "+" denotes positive part. Each function is piecewise linear or linear spline, with a knot at value t. These functions are called a reflected pair for each input X_j with knots at each observed value x_{ij} of that input, and then the set of basis functions is defined as

$$C = \{(X_j - t)_+, (t - X_j)_+\}$$

The strategy for model-building is a forward stepwise linear regression using functions from the set C and their products. Thus the MARS model has the form

$$f(X) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(X)$$

where the coefficients β_m are estimated by minimizing the residual sum-of-squares and each $h_m(X)$ is a function in C. By setting $h_0(X) = 1$ (constant function), the other multivariate splines are products of univariate spline basis functions:

$$h_m(X) = \prod_{s=1}^{k_m} h(x_{i(s,m)} | t_{s,m}) \qquad 1 \le m \le k$$

where the subscript $i(s,m)$ means a particular explanatory variable, and the basis spline in that variable has a knot at $t_{s,m}$. k_m is the level of interactions between $i(s,m)$ variables and the values of m, $k_1, k_2, ..., k_m$, are the knot sets. Explanatory variables in the model can be linearly or non-linearly and are chosen for inclusion adaptively from the data. The model will be additive if the order of interactions equals one ($k = 1$).

A backward deletion procedure is used in the MARS model to prevent overfitting. The basis functions which have little contributions to the accuracy of fit are deleted from the model at each stage, producing an estimated best model $\hat{f}(\lambda)$ of each size λ. We can apply a generalized cross-validation criterion to estimate the optimal value of λ in the following way

$$GCV(\lambda) = \frac{\sum_{i=1}^{N}(y_i - \hat{f}_\lambda(x_i))^2}{(1 - M(\lambda)/N)^2}$$

The value of M (λ) includes the number of basis functions and the number of parameters used in selecting the optimal positions of the knots.

3.2.2.3 Projection-Pursuit Regression (PPR)

If the explanatory vector X is of high dimension, the additive model does not cover the effect of interactions between the independent variables. Projection-Pursuit Regression (Friedman and Stuetzle, 1981) applies an additive model to projected variables, projecting predictor variables X in M, as follows

$$Y = \sum_{m=1}^{M} g_m(w_m^T X) + \varepsilon \qquad E(\varepsilon) = 0, \text{var}(\varepsilon) = \sigma^2$$

where w_m are unit p-vectors of unknown parameters. The functions g_m are unspecified and estimated along with the direction w_m using some flexible smoothing

method. The PPR model employs the backfitting algorithm and Gauss-Newton search to fit Y.

The functions $g_m(w_m^T X)$ are called the ridge functions because they are constant in all but one direction. They vary only in the direction defined by the vector w_m. The scalar variable $V_m = (w_m^T X)$ is the projection X onto the unit vector w_m. The aim is to find w_m to yield the best fit to the data. If M is chosen large enough then the PPR model can approximate arbitrary continuous function of X (Diaconis and Shahshahani, 1984). However, in this case there is a problem of interpretation of the fitted model since each input enters into the model in a complex and multifaceted way (Hastie et al, 2001). As a result, the PPR model is a good option only for forecasting.

To fit a PPR model, we need to minimize the error function

$$E = \sum_{i=1}^{N} [y_i - \sum_{m=1}^{M} g_m (w_m^T x_i)]^2$$

over functions g_m and direction vectors w_m. The g and w are estimated by iteration. Imposing complexity constraints on the g_m is needed to avoid overfitting. There are two stages to estimate g and w. First, to obtain an estimate of g, suppose there is one term (M=1). We can form the derived variables $v_i = w^T x_i$ for any value of w. This implies a one-dimensional smoothing problem and any scatterplot smoother such as smoothing spline can be used to estimate g. Second, we minimize E over w for any value of g. These two steps are iterated until convergence. If there is more than one term in the PPR model then the model is built in a forward stage-wise manner that at each stage a pair (w_m, g_m) is added.

4. Neural Networks

In chapter four, first the basics of neural networks are presented. Then the process of learning in these models using backpropagation algorithm is discussed in details. The convergence of learning to the rational expectations using different approaches was presented before. In this chapter, an innovation based on computational intelligence is used to describe learning procedure. The convergence of learning to the rational expectations equilibrium using neural networks is examined. In fact, we are interested in knowing whether the private agents are able to learn to form rational expectations with help of neural networks.

4.1 Basics of neural networks

Many recent methods to developing data-driven models have been inspired by the learning abilities of biological systems. For instance, most adults drive a car without knowledge of the underlying laws of physics and humans as well as animals can recognize patterns for the tasks such as face, voice or smell recognition. They learn them only through data-driven interaction with the environment. The field of pattern recognition considers such abilities and tries to build artificial pattern recognition systems that can imitate human brain. The interest to such systems led to extensive studies about neural networks in the mid-1980s (Cherkassky and Mulier, 2007).

Why use Neural Networks? Neural network modeling has seen an explosion of interest as a new technique for estimation and forecasting in economics over the last decades. They are able to learn from experience in order to improve their performance and to adapt themselves to changes in the environment. In fact, they can derive trends and detect patterns from complicated or imprecise data, and then model complex relationships between explanatory variables (inputs) and dependent variables (outputs). They are resistance to noisy data due to a massively parallel distributed processing.

The basics of neural networks from a biological point of view are now considered (Gleitman, 1991). The neuron is the basic functional element of the brain. An individual neuron has three principal components: a cell body, dendrites, and an axon. The dendrites are tree-like respective networks of nerve fibers that carry electrical signals into the cell body. The cell body sums and thresholds these

incoming signals. The axon is a single long fiber that carries the signal from the cell body out to other neurons. The connection between dendrites of two neurons is called a synapse.

Each individual neuron receives electrical stimuli from other neurons through the dendrites, which is then amplified or de-amplified by the synapse and summated. If the sum of all stimuli exceeds the neuron's resistance threshold, then the neuron fires, producing a stimulus that passes through the axon to another neuron. Figure 4.1 shows a schematic diagram of two biological neurons.
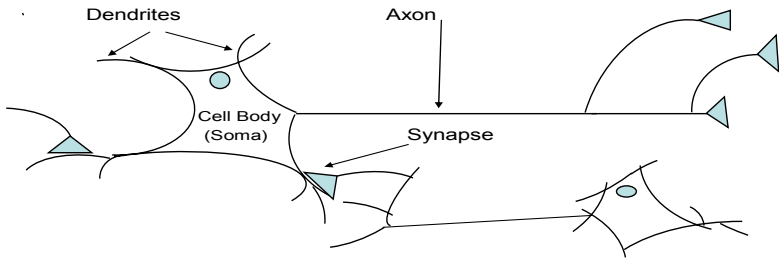
**Figure 4.1: Schematic Diagram of Biological Neurons**

In a typical network, we have a set of inputs x_i, a set of weights w_i, a threshold, u, a transfer function, f, and a signal neuron output, y, where i is the degree (number of inputs) of the neuron. The weights represent the amplification or de-amplification of the process. The sign of this weight is positive if the effect is excitatory, and negative if it is inhibitory; the magnitude of the weight represents the strength of the interaction.

Consider a single neuron with a set of weights w_i. The neuron produces an output which is a function of the weighted sum of the inputs from the incoming neurons.

$$y = f\left(w_1 x_1 + w_2 x_2 + \ldots + w_n x_n\right)$$

Here y is the output of neuron. The inputs x_i to the neuron could be outputs of the neurons feeding into this neuron or could come from sensory cells. The weights

w_i are the interaction strengths. The output of neuron is a function of the weighted sum of its inputs (see Figure 4.2):

$$wx = w_1x_1 + w_2x_2 + \ldots + w_nx_n$$
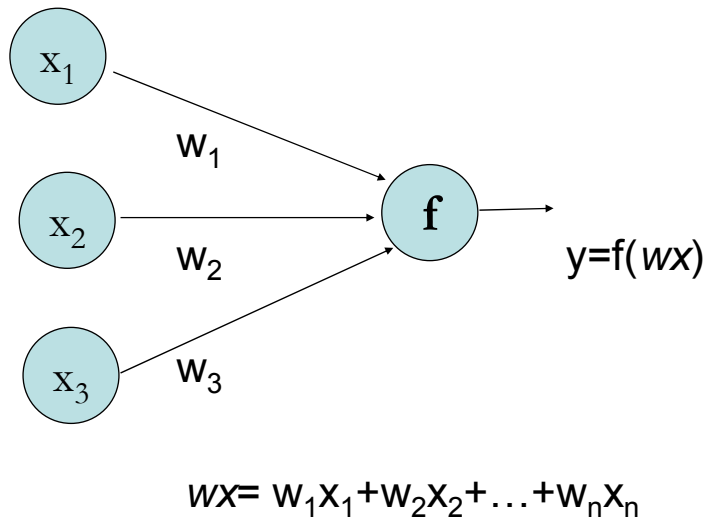
Figure 4.2: The Model of a neuron

The choice of the transfer function $f(.)$ varies in different models. For the binary representation, a common choice is to state the output is 1 if the weighted sum of input exceeds some threshold (u), and is 0 otherwise,

$$y = \begin{cases} 1 & \text{if } w_1x_1 + w_2x_2 + \ldots + w_nx_n > u \\ 0 & \text{otherwise} \end{cases}$$

This transfer function is called hard-limiting function. An alternative transfer function for the neurons in a neural network is log-sigmoid function. This squashes the linear combinations of inputs within the interval [0, 1]. The log-sigmoid equation is as follows:

$$F(x) = 1/(1 + e^{-x})$$

This nonlinear function is often used to construct the neural networks. It is mathematically well behaved, differentiable and strictly increasing function (Zilouchian, 2001). We could shift the threshold to the other side and could write the output function as

$$y = \begin{cases} 1 & \text{if } w_1x_1 + w_2x_2 + \ldots + w_nx_n + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

Here $b = -u$ and it is called a bias. It can be treated as another weight to an input which always has the value one. Figure 4.3 shows the behavior of the log-sigmoid and hard-limiting transfer functions.
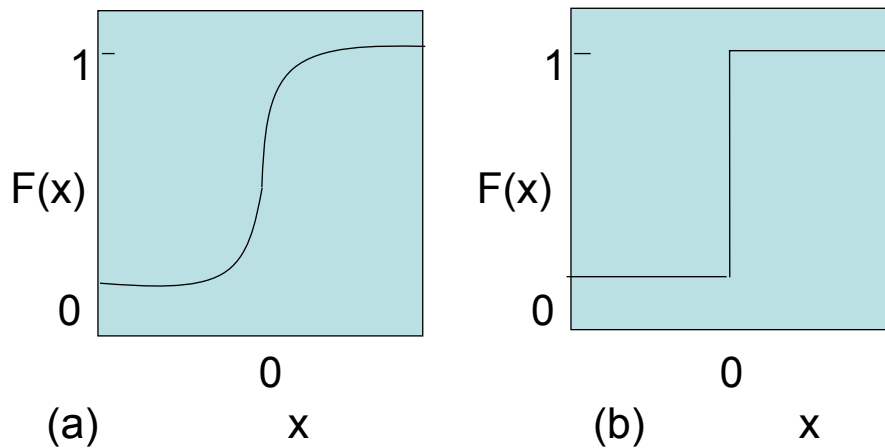
Figure 4.3: A log-sigmoid function (a) and a hard-limiting function (b)

Also, figure 4.4 illustrates a neuron model with bias. This model is of the usual *feed-forward* type since there are no feedback loops (Hagan et al, 1996). The information moves in only one direction, forward, from the input neurons, through the hidden neurons and to the output neurons.

$$y = \begin{cases} 1 & \text{if } w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + b > 0 \\ 0 & \text{otherwise} \end{cases}$$
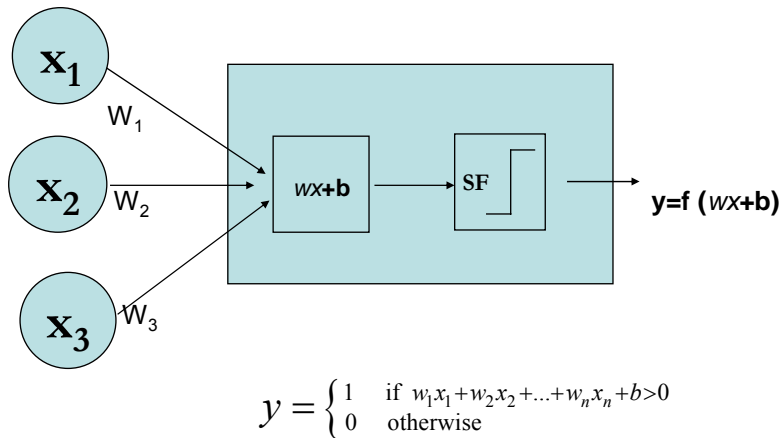
Figure 4.4: The model of a neuron with bias

Now we are interested in examining the learning process of neural networks.

4.2 Learning in neural network model

Stochastic approximation (or gradient descent) is one of the basic nonlinear optimization strategies commonly used in statistical and neural network methods (Cherkassky and Mulier, 2007). The gradient-descent methods are based on the first –order Taylor expansion of a risk functional

$$R(w) = \int L(y, f(x, w)) p(x, y) dx dy \tag{1}$$

where $R(w)$ is the risk functional, $L(y, f(x, w))$ the loss function and $p(x, y)$ the joint probability density function. For regression, a common loss function is the squared error

$$L(y, f(x, w)) = (y - f(x, w))^2 \tag{2}$$

Learning is then defined as the process of estimating the function $f(x, w_0)$ that minimizes the risk functional

$$R(w) = \int (y - f(x, w))^2 p(x, y) dx dy$$

using only the training data. Although the gradient-descent methods are computationally rather slow, their simplicity has made them popular in neural networks. We will examine two cases to describe such methods: linear parameter estimation and nonlinear parameter estimation.

4.2.1 Linear Parameter Estimation

Consider a linear (in parameters) approximating function and the loss function specified above. For the task of regression, it can be shown that the empirical risk is as follows

$$R_{emp}(w) = \frac{1}{n}\sum_{i=1}^{n} L(x_i, y_i, w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i, w))^2 \tag{3}$$

This function is to be minimized with respect to the vector of parameters w. Here the approximating function is a linear combination of fixed basis functions

$$\hat{y} = f(x, w) = \sum_{j=1}^{m} w_j g_j(x) \tag{4}$$

For some (fixed) m. The updating equation for minimizing $R_{emp}(w)$ with respect to w is

$$w(k+1) = w(k) - \gamma_k \frac{\partial}{\partial w} L(x(k), y(k), w) \tag{5}$$

where $x(k)$ and $y(k)$ are the sequences of input and output data samples presented at iteration step k. The gradient above can be written as

$$\frac{\partial}{\partial w_j} L(x, y, w) = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_j} = 2(\hat{y} - y)g_j(x) \tag{6}$$

Now the local minimum of the empirical risk can be computed using the gradient (6). Let us start with some initial values $w(0)$. The stochastic approximation method for parameter updating during each presentation of kth training sample is as:

- Step 1: Forward pass computations.

$$z_j(k) = g_j(x(k)), \quad j=1,..,m \tag{7}$$

$$\hat{y}(k) = \sum_{j=1}^{m} w_j(k)z_j(k). \tag{8}$$

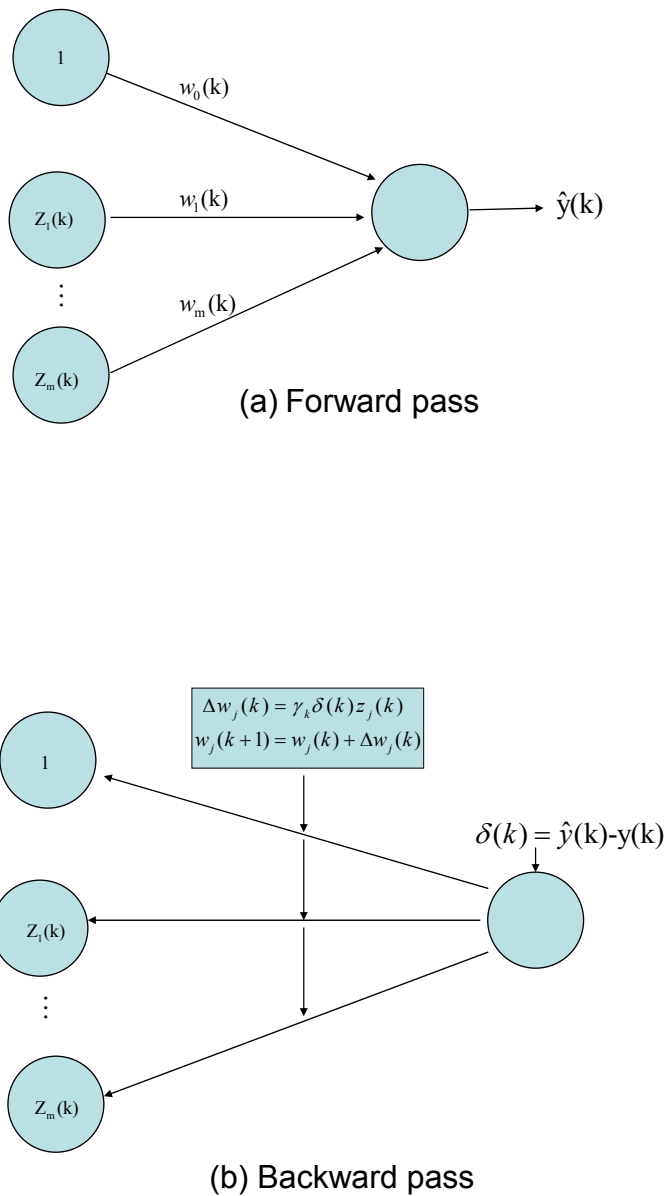- Step 2: Backward pass computations.

$$\delta(k) = \hat{y}(k) - y(k) \qquad\qquad (9)$$

$$w_j(k+1) = w_j(k) - \gamma_k \delta(k) z_j(k) \quad \text{j=1,..,m} \qquad (10)$$

where the learning rate γ_k is a small positive number decreasing with k. In the forward pass, the output of the approximating function is computed whereas in the backward pass, the error term (9), which is called "delta" in neural network literature, for the presented sample is calculated and utilized to modify the parameters. The parameter updating equation (10), known as delta rule, updates parameters with every training sample.

Figure 4.5 demonstrates the forward and backward passes of the neural network. Based on the delta rule (equation 10), the change in connection strength is proportional to the error and the activation of the input layer.

(a) Forward pass

$$\Delta w_j(k) = \gamma_k \delta(k) z_j(k)$$
$$w_j(k+1) = w_j(k) + \Delta w_j(k)$$

$$\delta(k) = \hat{y}(k) - y(k)$$

(b) Backward pass

Figure 4.5: Neural network interpretation of the delta rule

4.2.2 Nonlinear Parameter Estimation

The standard method used in the neural network literature is the *backpropagation* algorithm which is an example of stochastic approximation strategy for nonlinear approximating functions. As it was considered already, the mapping from inputs to output given by a single layer of hidden units is as follows

$$f(x,w,V) = w_0 + \sum_{j=1}^{n} w_j g\left(v_{0j} + \sum_{i=1}^{d} x_i v_{ij}\right)$$ (11)

In contrast to (4), the set of functions is nonlinear in the parameters V. We seek values for the unknown parameters (weights) V and w that make the model fit the training data well. To do so, the sum of squared errors as a measure of fit must be minimized:

$$R_{emp} = \sum_{i=1}^{n} (f(x_i, w, V) - y_i)^2$$ (12)

The stochastic approximation procedure for minimizing R_{emp} with respect to the parameters V and w is

$$V(k+1) = V(k) - \gamma_k \nabla_V L(x(k), y(k), V(k), w(k)),$$ (13)

$$w(k+1) = w(k) - \gamma_k \nabla_w L(x(k), y(k), V(k), w(k)), k = 1,...,n,$$ (14)

where $x(k)$ and $y(k)$ are the kth training samples, presented at iteration step k. The loss function L is

$$L(x(k), y(k), V(k), w(k)) = \frac{1}{2}(f(x, w, V) - y)^2$$ (15)

where the factor ½ is included only for simplifying gradient calculations in the learning algorithm. We need to decompose the approximation function (11) for computations of the gradient of loss function (15) as follows

$$a_j = \sum_{i=0}^{d} x_i v_{ij}, \quad \text{j=1,...,m}$$ (16)

$$z_j = g(a_j), \quad \text{j=1,...,m}$$ (17)

$$z_0 = 1,$$

$$\hat{y} = \sum_{j=0}^{m} w_j z_j$$ (18)

For simplicity, we drop the iteration step k, consider calculation/parameter update for one sample at a time and incorporate the terms w_0 and v_{0j} into the summations ($x_0 \equiv 1$). The relevant gradients, based on the chain rule of derivatives, are

$$\frac{\partial R}{\partial v_{ij}} = \frac{\partial R}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial a_j} \frac{\partial a_j}{\partial v_{ij}},$$ (19)

$$\frac{\partial R}{\partial w_j} = \frac{\partial R}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_j}. \tag{20}$$

In order to calculate each of the partial derivatives, we need equations (15) to (18). Therefore,

$$\frac{\partial R}{\partial \hat{y}} = \hat{y} - y \tag{21}$$

$$\frac{\partial \hat{y}}{\partial a_j} = g'(a_j)w_j \tag{22}$$

$$\frac{\partial a_j}{\partial v_{ij}} = x_i \tag{23}$$

$$\frac{\partial \hat{y}}{\partial w_j} = z_j \tag{24}$$

If we plug the partial derivatives (21)-(24) into (19) and (20), the gradient equations are

$$\frac{\partial R}{\partial v_{ij}} = (\hat{y} - y)g'(a_j)w_j x_i \tag{25}$$

$$\frac{\partial R}{\partial w_j} = (\hat{y} - y)z_j \tag{26}$$

Using these gradients and the updating equations, we can construct a computational method to minimize the empirical risk. Starting with some initial values w (0) and V (0), the stochastic approximation method updates weights upon presentation of a sample (x (k), y (k)) at iteration step k with learning rate γ_k as

- Step 1: Forward pass computations.
  "Hidden layer"

$$a_j(k) = \sum_{i=0}^{d} x_i(k)v_{ij}(k), \quad j=1,..,m \tag{27}$$

$$z_j(k) = g(a_j(k)), \quad j=1,..,m \tag{28}$$

$$z_0(k) = 1$$

  "Output layer"

$$\hat{y}(k) = \sum_{j=0}^{m} w_j(k)z_j(k). \tag{29}$$

- Step 2: Backward pass computations.

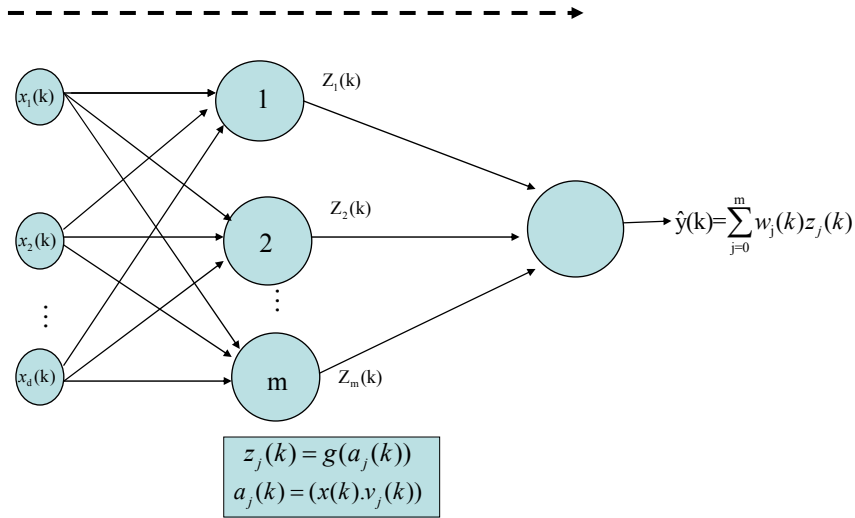 "Output layer"

$$\delta_0(k) = \hat{y}(k) - y(k) \tag{30}$$

$$w_j(k+1) = w_j(k) - \gamma_k \delta_0(k) z_j(k) \quad \text{j=0,..,m} \tag{31}$$
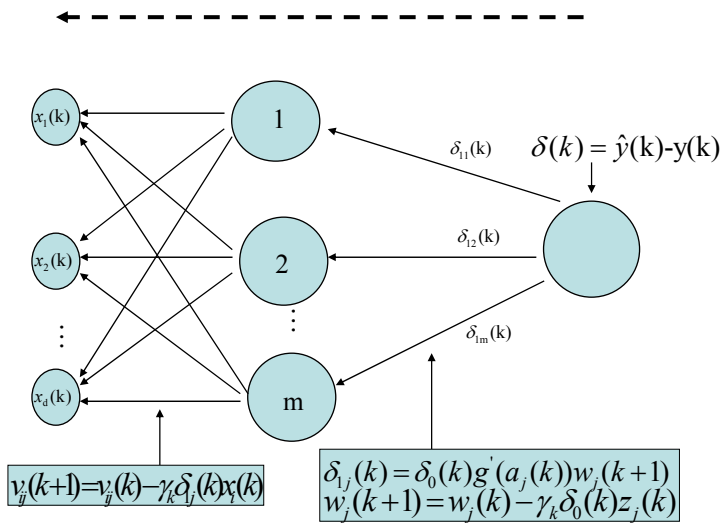
  "Hidden layer"

$$\delta_{1j}(k) = \delta_0(k) g^{'}(a_j(k)) w_j(k+1), \quad \text{j=0,..,m} \tag{32}$$

$$v_{ij}(k+1) = v_{ij}(k) - \gamma_k \delta_{1j}(k) x_i(k), \text{ i=0,…,d, } \text{j=0,..,m} \tag{33}$$

In the forward pass, the output of the approximating function is computed whereas in the backward pass, the error term for the presented sample is calculated and utilized to modify the parameters in the output layer. Since it is possible to propagate the error at the output back to an error at each of the internal nodes a_j through the chain rule of derivatives, the procedure is called *error backpropagation*. In fact it is a propagation of the error signals from the output layer to the input layer. Figure 4.6 below demonstrates the forward and backward passes of the backpropagation training.

$$z_j(k) = g(a_j(k))$$
$$a_j(k) = (x(k).v_j(k))$$

(a) Forward pass

$$\hat{y}(k)=\sum_{j=0}^{m} w_j(k)z_j(k)$$

$$\delta(k) = \hat{y}(k)-y(k)$$

$$v_{ij}(k+1)=v_{ij}(k)-\gamma_k\delta_{lj}(k)x_i(k)$$

$$\delta_{1j}(k) = \delta_0(k)g'(a_j(k))w_j(k+1)$$
$$w_j(k+1) = w_j(k) - \gamma_k\delta_0(k)z_j(k)$$

(a) Backward pass

Figure 4.6: Backpropagation training

The updating steps for output layer are similar to those for the linear case. Besides, the updating rule for the hidden layer is the same as the linear one but for the delta term (32). For this reason, backpropagation update rules (32) and (33) are usually called the "generalized delta rule". The parameter updating algorithm holds if the sample size is large (infinite). However, if the number of training samples are finite, the asymptotic conditions of stochastic approximation are (approximately) satisfied by the repeated presentation of the finite training sample to the training algorithm. This is

called recycling and the number of such repeated training samples is called the number of cycles (or epochs).

It is possible to use the backpropagation algorithm for networks with several output layers and networks with several hidden layers. For instance, if additional layers are added to the approximation function, then errors are 'propagated' from layer to layer by repeated application of generalized delta rule.

It should be noted that a neural network model can be identified as a pursuit projection regression (PPR) model (Hastie et al, 2001). In fact, the neural network with one hidden layer has the exactly the same form as the PPR model. The only difference is that the PPR model uses nonparametric functions ($g_m(v)$) while the neural network employs a simpler function which is based on a sigmoid transfer function.

4.3 Learning of rational expectations using a neural network

The approaches to learning in macroeconomics including eductive learning, adaptive learning, and rational learning were discussed in chapter two. An innovation based on computational intelligence has been the use of neural networks as a semi parametric approach to describe learning procedure (Salmon, 1995; Packalén 1998; Barucci and Landi, 1998; Heinemann, 2000). What we are interested in examining is whether rational expectations are learnable with help of neural networks.

Assume the reduced form for prices is as follows

$$p_t = \alpha p_t^e + h(x_t) + \varepsilon_t \tag{1}$$

Here p_t and p_t^e are as before, x_t denotes a vector of exogenous variables which is assumed to have independent and identical distribution (*i.i.d*) and to be bounded for all t, i.e., x_t takes only values in a set $\Omega_x \subset R^k$. The unobservable error, ε_t, is also *i.i.d* random variable which satisfies $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, $E(\varepsilon_t | x_t) = 0$ and is bounded for all t. Last, $h(x)$ is a continuous function for all $x \in \Omega_x$.

Taking expectation of both sides of the reduced form gives rational expectation

$$E(p_t | x_t) = E\left[\frac{h(x_t) + \varepsilon_t}{1 - \alpha}\right] = \frac{h(x_t)}{1 - \alpha} = \varphi(x_t) \tag{2}$$

If $\alpha \neq 1$, there exists a unique rational expectation of p_t which is given by the rational expectation function $\varphi(x_t)$. If agents do not know the reduced form of the model and the form of $h(x)$, rational expectation may not be reachable. However, they may learn to form RE using the past values of p_t and x_t. In other words, it is assumed that agents have an auxiliary model showing the relationship between the exogenous variables (x_t) and the endogenous variable (p_t).

If $h(x)$ is linear in x_t, the reduced form (1) becomes the linear model $p_t = \alpha p_t^e + \beta x_t + \varepsilon_t$, where β is a vector of parameters. If it is assumed agents use the auxiliary model $p = \delta' x$ where δ are estimated using recursive least squares, the following results hold (Bray and Savin, 1986; Marcet and Sargent, 1989)

(a) If the estimator $\hat{\delta}$ for δ converges, this results in rational expectations, i.e.

$$\hat{\delta} = \frac{\beta'}{1-\alpha}.$$

(b) The estimator for δ will converge towards $\dfrac{\beta'}{1-\alpha}$ if and only if $\alpha < 1$.

If the function $h(x)$ is not linear, $\varphi(x_t)$ is not linear too. In such cases, agents, having no prior knowledge about the functional form of $\varphi(x_t)$, may use an auxiliary model such as neural networks which is flexible enough to approximate the rational expectation function $\varphi(x_t)$.

The following equations describe the neural network by mapping inputs x_j to the output y as

$$n_i = w_{i,0} + \sum_{j=1}^{k} w_{i,j} x_j$$

$$S_i = L(n_i) = \frac{1}{1 + e^{-n_i}}$$

$$y = q_0 + \sum_{i=1}^{m} q_i S_i$$

$$= f(x, \theta), \tag{3}$$

where $x^{'} = (x_1,...,x_k)$, $\theta^{'} = (q_0, q_1, w_{1,0},..., w_{1,k}, q_2,..., w_{m,k})$ and $L(n_i)$ shows the log-sigmoid transfer function. A linear combination of the input variables x_j, with the coefficient vectors $w_{i,j}$, as well as the constant term, $w_{i,0}$, form the variable n_i. This variable is squashed by log-sigmoid function, and becomes a neuron S_i. The set of m neurons are combined in a linear way with the coefficient vector q_i, and taken with a constant term q_0 to forecast y.

The model with one layer of hidden units and log-sigmoid transfer function is able to approximate any continuous function if a sufficient number of hidden units are used (Hornik, 1989). The interesting feature of neural networks is their ability to learn. Therefore, there exists a neural network and a vector of parameters θ^* such that $\varphi(x_t) = f(x, \theta^*)$. However, since the exact number of hidden units required to obtain a perfect approximation is not known with certainty, a perfect approximation of rational expectation function $\varphi(x_t)$ can not be guaranteed.

Objectives of learning

Assume agents use the neural network of the form (3) as an auxiliary model. If the expectation of p which is given by $p^e = f(x, \theta)$ is found to be incorrect, agents will improve the predictive power of their model by changing the values of parameters. This process, in fact, is called learning.

The mean squared error (MSE) of expectations is a measure for success of learning. It is defined as the expected value of the squared deviation of the agents' expectation $p^e = f(x, \theta)$ from its actual value $p = \alpha f(x, \theta) + g(x) + \varepsilon$. Denoting this MSE as λ_θ we obtain

$$\lambda_\theta = E\left[\alpha f(x,\theta) + g(x) + \varepsilon - f(x,\theta)\right]^2 = (1-\alpha)^2 E\left[\frac{g(x)+\varepsilon}{1-\alpha} - f(x,\theta)\right]^2$$

$$= (1-\alpha)^2 E\left[\varphi(x) + \frac{\varepsilon}{1-\alpha} - f(x,\theta)\right]^2 \tag{4}$$

The optimal vector of parameters θ^* is achieved by minimizing λ_θ with respect to θ

$$\theta^* = \arg\min \lambda(\theta) \tag{5}$$

Using ∇_θ for the gradient vector of $\lambda(\theta)$, the necessary condition for this problem can be written as

$$\nabla_\theta \lambda(\theta) = -2(1-\alpha)^2 E\left\{\nabla_\theta f(x,\theta)\left[\varphi(x) + \frac{\varepsilon}{1-\alpha} - f(x,\theta)\right]\right\} = 0 \qquad (6)$$

It is clear that equation (6) may have a multiple solutions. In case of existence of a solution θ satisfying the necessary condition, a (local) minimum of MSE is obtained if the Jacobian matrix $J_\lambda(\theta)$ is positive semidefinite.

$$J_\lambda(\theta) = \nabla_\theta^2 \lambda(\theta) = -E\left\{\nabla_\theta^2 f(x,\theta)\left[\varphi(x)\frac{\varepsilon}{1-\alpha} - f(x,\theta)\right]\right\} + E\left\{\nabla_\theta f(x,\theta)\nabla_\theta f(x,\theta)'\right\} \quad (7)$$

A (local) minimum at θ^* is (locally) identified if $J_\lambda(\theta^*)$ is positive definite. Otherwise, at least one eigenvalue of $J_\lambda(\theta^*)$ is equal to zero, such that the minimum is not (locally) identified.

Now consider the set Θ^L which includes all vectors of parameters for neural network implying a (local) minimum of MSE

$$\Theta^L = \{\theta \in R^q | \nabla_\theta \lambda(\theta) = 0,\ J_\lambda(\theta) \text{ is positive semidefinte}\}$$

If a neural network can perfectly approximate the unknown rational expectation function $\varphi(x)$, there exist vectors of parameters implying $\lambda(\theta) = \sigma_\varepsilon^2$. Since all hidden units in the neural network stated here employ identical activation functions, there will be no unique vector having this property. To remove this problem, let

$$\Theta^G = \{\theta \in R^q | \lambda(\theta) = \sigma_\varepsilon^2\}$$ denote the set of all these vectors of parameters. Any $\theta \in \Theta^G$ implies that the expectations formation using the neural network model and rational expectation function $\varphi(x)$ are identical. This is not true for the remaining $\theta \in \Theta^L \setminus \Theta^G$: All θ result in (local) minima of the $\lambda(\theta)$, but they do not imply $\varphi(x) = f(\theta,x)$. These vectors of parameters result in approximate unknown rational expectation functions and the resulting equilbria are called rational expectation equilibria. (Sargent 1993).

Learnability of the rational expectations

Learning implies that agents estimate the parameters of the neural network model using exogenous and endogenous variables. Here the question may arise whether

the agents can learn to form rational expectations or equivalently whether there will result asymptotically correct parameters values. Do the estimated parameter vectors converge to a $\theta \in \Theta^G$ or at least to a $\theta \in \Theta^L$?

Substituting the expectation function $p^e = f(x_t, \theta_t)$ into the reduced form (1), we get the actual value of endogenous variable

$$p_t = \alpha f(x_t, \theta_t) + h(x_t) + \varepsilon_t.$$

If $f(x_t, \theta_t) \neq \varphi(x)$, the agents' expectation turns out to be incorrect and p_t diverges from the rational expectation equilibrium. Assume the learning algorithm used by agents is the 'backpropagation' algorithm. It changes the vector of parameters θ_t according to the product of the actual expectation error $p_t - p^e = p_t - f(x_t, \theta_t)$ and the gradient of the neural network with respect to θ_t:

$$\theta_{t+1} = \theta_t + \gamma_{t+1} \left[\nabla_\theta f(x_t, \theta_t)(p_t - f(x_t, \theta_t)) \right] \tag{8}$$

Here γ_t is a declining learning rule that satisfies $\gamma_t = t^{-k}$, $0 < k \leq 1$. It implies that changes of θ_t becomes smaller over time and this helps us to answer the question whether agents will asymptotically learn to form (approximate) rational expectations or equivalently whether θ_t converge to a $\theta \in \Theta^L$. Since the analysis of the stochastic difference equation (8) is difficult, we follow Ljung (1977) in approximating θ_t using the differential equation

$$\dot{\theta}(\tau) = \bar{Q}(\theta(\tau)), \tag{9}$$

where

$$\bar{Q}(\theta) = E\left\{ \nabla_\theta f(x, \theta) \left[p - f(x, \theta) \right] \right\}$$

$$= E\left\{ \nabla_\theta f(x, \theta) \left[g(x) + \varepsilon - (1 - \alpha) f(x, \theta) \right] \right\}$$

As equation (9) is a deterministic differential equation, all conclusions resulting from (9) about the stochastic difference equation (8) are valid in a probabilistic sense. In other words, the time path of θ_t according to (8) is asymptotically equivalent to the trajectories of θ resulting from (9). This means that for t→∞, θ_t from (8) will- if ever- converge only to stationary points of (9) which are (locally) stable. It will not converge to stationary points that are unstable.

Analyzing the asymptotic properties of the learning algorithm (8) requires to examine the stationary points of the differential equation (9). Since α is constant, $\bar{Q}(\theta)$ can be written as

$$\bar{Q}(\theta) = E\left\{\nabla_\theta f(x,\theta)\left[g(x) + \varepsilon - (1-\alpha)f(x,\theta)\right]\right\}$$

$$= (1-\alpha)E\left\{\nabla_\theta f(x,\theta)\left[\frac{g(x)+\varepsilon}{1-\alpha} - f(x,\theta)\right]\right\}$$

$$= (1-\alpha)E\left\{\nabla_\theta f(x,\theta)\left[\varphi(x) + \frac{\varepsilon}{1-\alpha} - f(x,\theta)\right]\right\}$$

$$= -\frac{1}{2(1-\alpha)}\nabla_\theta \lambda(\theta) \tag{10}$$

According to equation (10), differential equation (9) is a gradient system[1], the potential of which is proportional to $\lambda(\theta)$ from (4). Therefore:

Proposition 1: Any θ implying the mean squared error $\lambda(\theta)$ from (4) takes an extreme value is a stationary point of the differential equation (10).

We can state the conditions for (local) stability of a fixed point using the Jacobian matrix of $\bar{Q}(\theta)$. Hence, according to (8), we obtain

Proposition 2: let θ^* be a stationary point of differential equation (9). The probability that for $t\to\infty$, θ_t according to (8), will converge to θ^* is positive only if the real parts of all eigenvalues of the following Jacobian matrix are nonpositive

$$J(\theta^*) = \frac{\partial \bar{Q}(\theta)}{\partial \theta'}\big|_{\theta^*}$$

Since the equation (9) is a gradient system, we obtain together with (7)

$$J(\theta) = (\alpha - 1)J_\lambda(\theta) \tag{11}$$

From equation (11) we get

----

1. A gradient system in $\mathbb{R}^n$ is an autonomous ordinary differential equation $\dot{x} = -gradF(x)$ where F: $\mathbb{R}^n \to$ R. (Hirsch, Smale and Devaney, 2004) For the dynamic system $\dot{\theta} = \bar{Q}(\theta)$ we have $-\nabla_\theta F(\theta) = \bar{Q}(\theta)$ where $F(\theta) = \left[2(1-\alpha)\right]^{-1}\lambda(\theta)$

Proposition 3: let θ^* be any element of the set Θ^L , i.e. θ^* implies a local minimum of the mean-squared error $\lambda(\theta)$ .The probability that θ_t from (8) converges to θ^* asymptotically is positive only if $(\alpha - 1) < 0$.

The set Θ^L includes the rational expectation equilibrium if the neural network can perfectly approximate the unknown rational expectation function. As a result, according to Proposition 3, this rational expectation function will be learnable if $(\alpha - 1) < 0$. This result is similar to that of linear models.

Now consider the learnability of the correct rational expectations graphically. To do so, we need to examine the stability condition $\alpha < 1$. In case the expectation of the endogenous variable $p^e = f(x_t, \theta_t)$ (under) overestimates the actual value p, the learning algorithm (8) changes θ_t in a way that given x_t, there results a lower (higher) expectation. Convergence to the correct rational expectation depends on the value of α. Figure 4.7(a) shows that the expectation error $p_t - p_t^e$ becomes smaller if $\alpha < 1$. In this case, the learning algorithm may converge. With $\alpha > 1$, figure 4.7(b), this error becomes larger and as a result such an algorithm would never converge. In this case the learning process directs towards (local) maxima of the mean squared error $\lambda(\theta)$. But there exists no θ^* satisfying the sufficient conditions for a maximum.
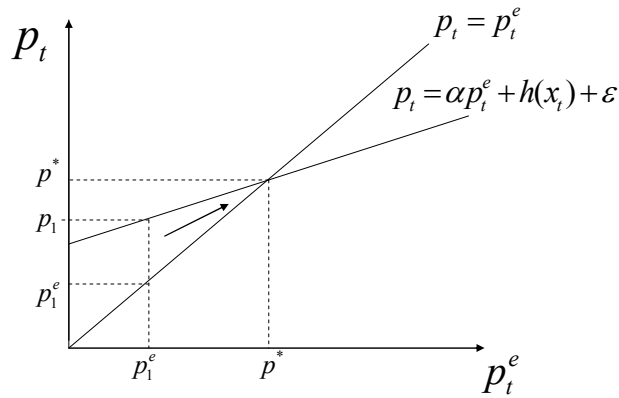
a) Stability: $\alpha < 1$
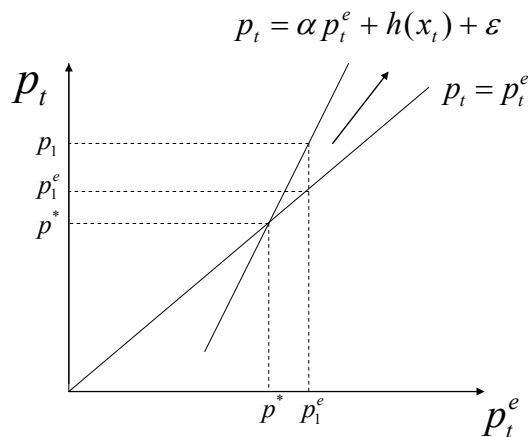
b) Instability: $\alpha > 1$

Figure 4.7: Learnability of correct expectations

Propositions 2 and 3 provide necessary and sufficient conditions for a parameter vector $\theta^* \in \Theta^L$ to be a locally stable fixed point of differential equation (9). They are conditions for the probability that θ_t converges to an element of Θ^L to be nonzero. However, this does not mean that θ_t will converge almost surely to an element of Θ^L. Thus, we need an additional case guaranteeing convergence. This can be done by

augmenting algorithm (8) with a projection facility. But formulating a projection facility in nonlinear models is quite complex task.

5. Empirical Results

Much research has been done in the field of expectations in economics. We will examine different approaches to the formation of expectations. This study has tried to extract expectations from past data, on the assumption that people look to past experience as a guide to the future.

Two approaches will be analyzed in this section: simple forecast and a multi-equation model. In the first case, we apply parametric and nonparametric methods and then evaluate whether nonparametric models yield better estimates of inflationary expectations than do parametric alternatives. In fact, out-of-sample estimates of inflation generated by the parametric and nonparametric models will be compared. In the case of a multi-equation model, expected inflation will be considered in the augmented Philips curve. The expectation hypothesis will be tested and our main concern is whether inflation expectations play a main role in the determination of the wages. Finally the best model will be selected based on the two criteria, i.e. the standardized expected inflation coefficient and adjusted R-squared.

Background

Iran has an area of 1,648,000 km². According to the Central Bank, Iran's population was 33.5 million people in 1976 and increased radically in 1980s so that it reached at 49.4 million in 1986. The population was estimated at 66.4 millions in 2003 which 43.9 million people (66 percent of total population) live in urban areas. More than two-third of the population is under the age of 30, and the literacy rate is 82%.

The Iranian economy is oil-reliant so that any change in oil price can directly affect all economic sectors. It should be noted that Iran ranks second in the world in natural gas reserves and third in oil reserves. It is also OPEC's second largest oil exporter. The economic sectors include the services sector, industry (oil, mining and manufacturing) and the agricultural sector. During the recent decades, the services sector has contributed the largest percentage of the GNP, followed by industry and agricultural sectors. The share of the services sector was 51 percent of GNP in 2003 while those of the industry and agricultural sectors were 35.1 and 13.9 percent of GNP respectively.

The Iranian economy has been subject to a number of critical events over the past five decades including the 1979 revolution, the eight-year war with Iraq (1980-88), volatility in global oil prices, and the 1993 balance of payment crisis. These events plus government controls of the major parts of the economy have substantially changed the behavior of the macroeconomic variables.

Over the period 1959-2003[1], the economy has experienced a relatively high inflation averaging about 15 percent per year. The inflation rate has even been more than 21 percent on average after the 1973 oil crisis. Furthermore, there is a general agreement over the underestimating of the measured inflation due to price controls and government subsidies. Although there are some differences between official data and private estimates, the official figures seem to show reasonably economic trends. Another major problem is high unemployment in Iran. The unemployment rate increased from 2.8 percent in 1959 to 8.7 percent in 1974 and this increase has continued so that it peaked at 14.7 percent in 2001. Unemployment has been on average more than 12 percent for the period 1974-2003 (see Figure 5.1). High inflation along with high unemployment, which is referred to as stagflation, have been the major concerns in Iran's economy. Although the oil shock is the source of stagflation in developed countries, the massive currency depreciation is the main factor for the case of Iran (Bahmani-Oskooee, 1996). In order to remove these
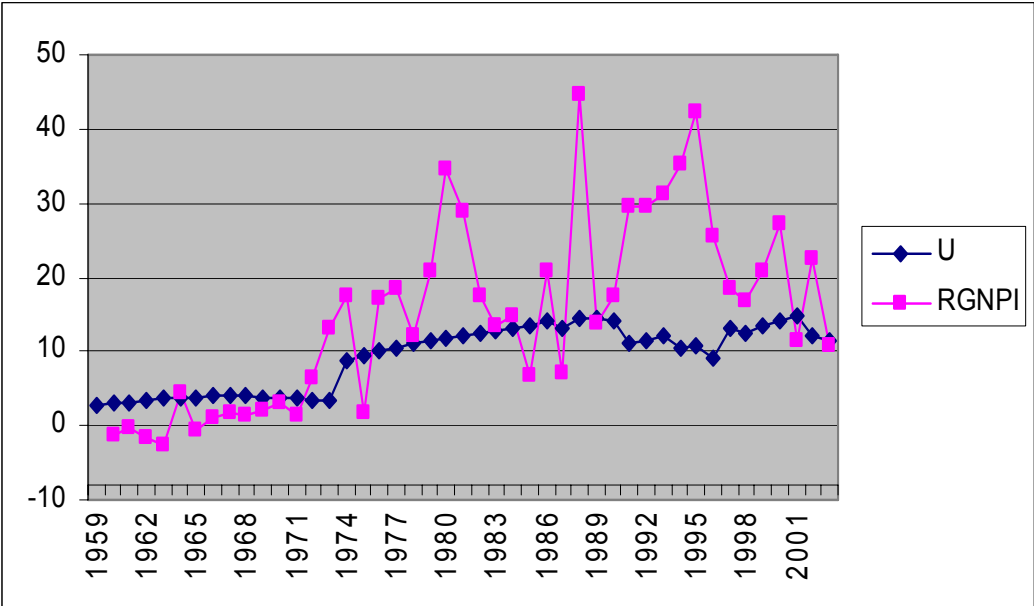
Figure 5.1: Unemployment rate (U) and Inflation rate (RGNPI)

----

1. It should be noted that the Iranian calendar starts on 21st of March. However, for convenience, the "1959-2003" notation, instead of "1959/1960-2003/2004", is used.

problems, the government has taken some actions including trade liberalization, tax reform, exchange rate unification and financial sector reform. However, little progress has been made in the areas of privatization and subsidy reform.

Oil revenues have been one of the main sources of money creation, fueled by government spending. Another important factor for increasing liquidity (M2) is subsidies on energy, food, bank credit and the large number of government-controlled enterprises which increase the budget deficit through borrowing from the Central Bank, and thus increase the monetary base. Money supply growth has been 24.48 percent on average for the period 1959-2003, whereas real GNP growth recorded only on average 6.12 percent during the same period. Furthermore, money supply has become 10127 fold while real GNP recorded only a 10 fold increase during the same period (see Figure 5.2). As the same time population has been growing, resulting in an increasing demand which adds to inflation pressure.
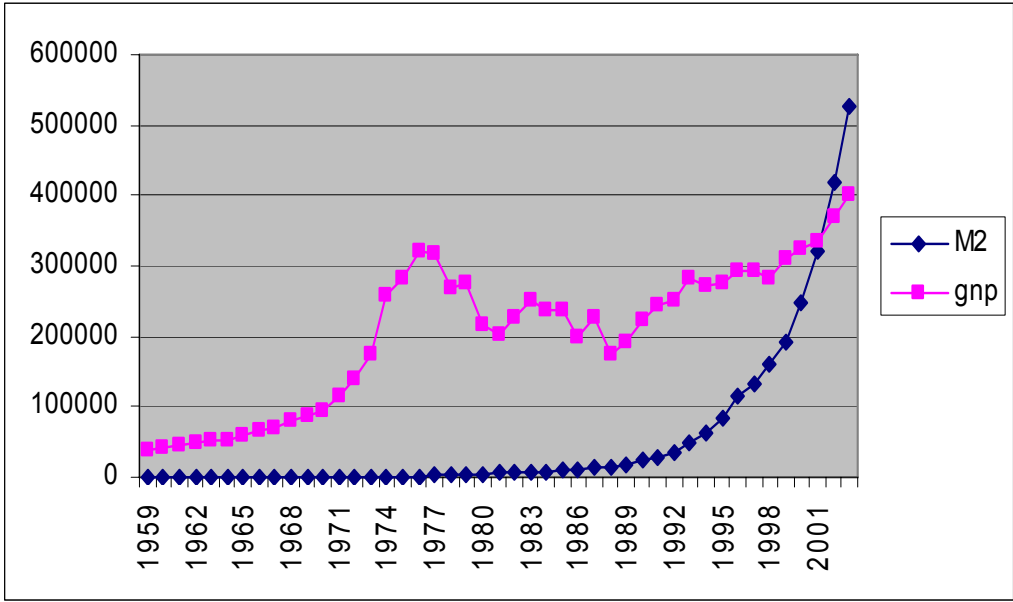
Figure 5.2: Liquidity (M2) and real Gross National Product (gnp)

The empirical evidence implies that inflation is persistent in Iran. In other words, the effects of a shock to inflation results in a changed level of inflation for an extended period. To see this, the inflation rate is regressed on its own lags.

$$\pi_t = 0.44\pi_{t-1} + 0.49\pi_{t-2}$$

(t-value)   (3.15)   (3.42)

As the sum of coefficients on lagged inflation (0.93) is close to one, shocks to inflation have long-lasting effects on inflation.

Since any decision or news announced by the government or the Central Bank could severely change the distribution of resources in the economy, it matters for the Central Bank to know how private agents form their expectations. Moreover, optimal monetary policy depends considerably on the assumed nature of expectations formation process.

5.1 Simple forecast

It is assumed agents use the lagged values of inflation and real GNP growth to forecast inflation. Figures 5.3.a and Figures 5.3.b demonstrate a local linear regression fit of inflation rate (rgnpi), defined as the rate of change of GNP deflato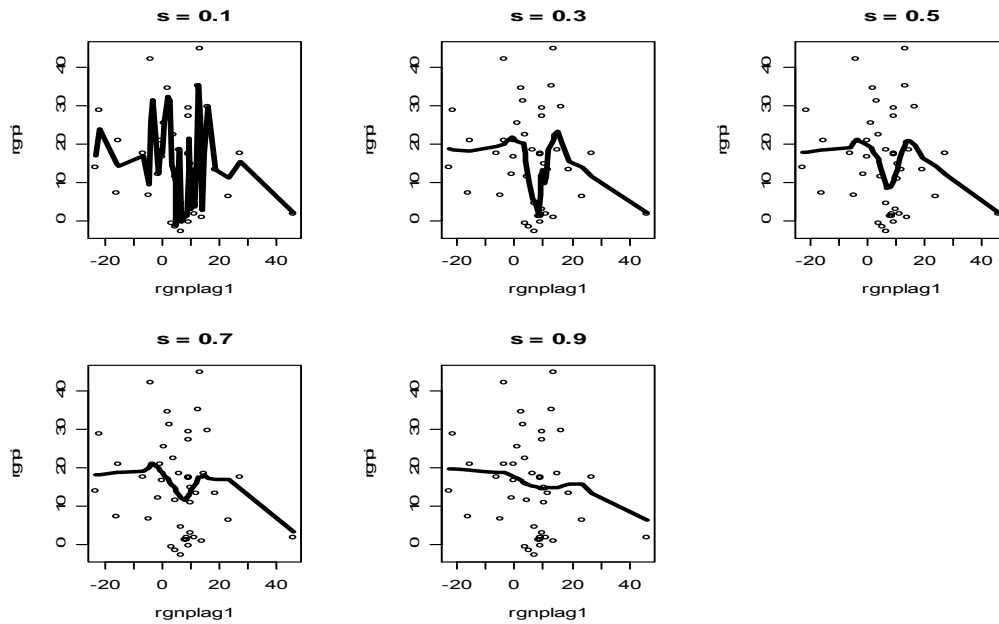r, on the lagged inflation rate (rgnpilag1) and lagged real GNP growth rate (rgnplag1) using the Lowess function for a variety of spans. If the fitted regression looks too rough, then we try to increase the span but if it looks smooth, then we will examine whether the span can be decreased without making the fit too rough. The objective is to find the smallest value of span (s) that provides a smooth fit. A trial and error procedure suggests that the span s=0.5 is suitable and it seems to provide a reasonable compromise between smoothness and fidelity to the data.
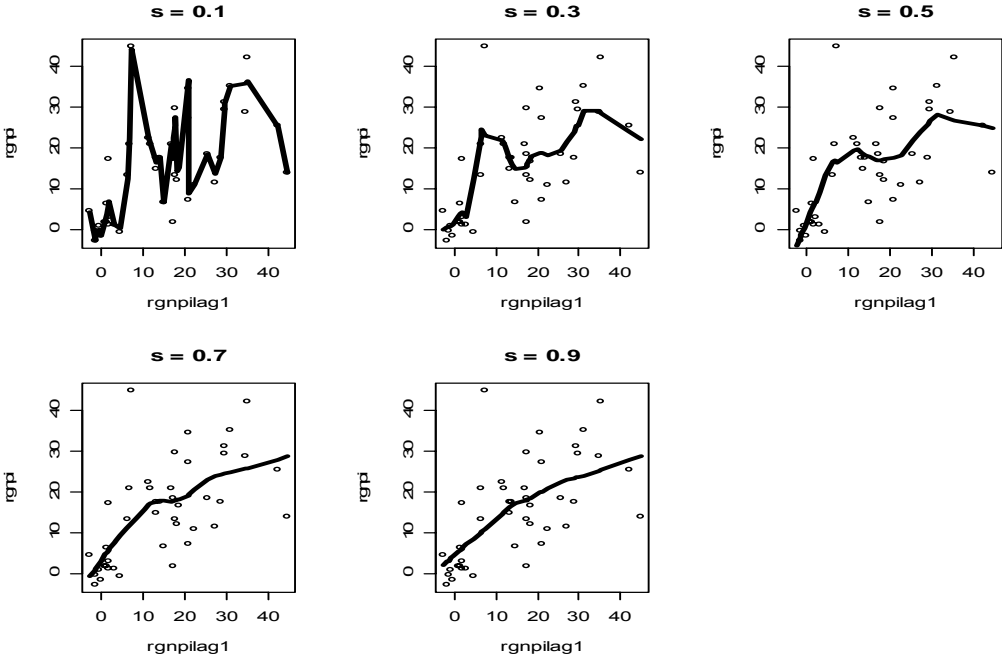
Figure 5.3.a: Local linear regression fit of inflation rate (rgnpi) on the lagged inflation rate (rgnpilag1) using Lowess function for a variety of spans

Figure 5.3.b: Local linear regression fit of inflation rate (rgnpi) on the lagged real GNP growth rate (rgnplag1) using Lowess function for a variety of spans

A test of nonlinearity is performed by contrasting the nonparametric regression model with the linear simple-regression model. We regress inflation on rgnpilag1 (Case 1) and rgnplag1 (Case2) separately. As a linear model is a special case of a nonlinear model, two models are nested. An F-test is formulated by comparing alternative nested models. The results is as follows

Linear model vs Nonparametric regression (Case1): F=8.78(p-value=0.008)
Linear model vs Nonparametric regression (Case2): F=6.48(p-value=0.04)

It is obvious that the relationship between the dependent variable and explanatory variables are significantly nonlinear. It should be noted that the variable rgnplags1 will not be significant if a linear regression is considered. It is generally not easy to discover nonlinearity in multiple regressions because the explanatory variables are usually correlated. In this case, partial-residual plots or component+residual plots can help to detect nonlinearity. These plots are given in figure 5.4.a and figure 5.4.b, suggesting a nonlinear relationship between inflation and the explanatory variables.
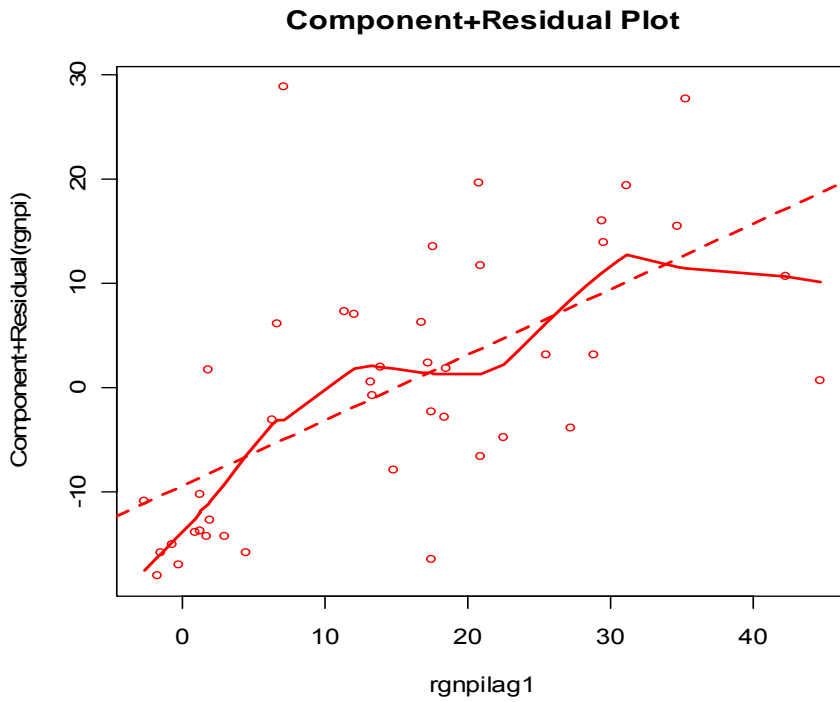
Figure 5.4.a: Partial residual plot for the lagged inflation rate (rgnpilag1) from the fit to the multiple regression of the inflation rate (rgnpi) on rgnpilag1 and rgnplag1

Figure 5.4.b: Partial residual plot for the lagged real GNP growth rate (rgnplag1) from the fit to the multiple regression of the inflation rate (rgnpi) on rgnpilag1 and rgnplag1

Since nonparametric regression based on smoothing functions faces the curse of dimensionality, the additive model has been proposed.

The result of fitting an additive model using Lowess smoother can be written as

$$rgnpi = S(\text{rgnpilag1}) + S(\text{rgnplag1})$$

F	(4.13)	(4.43)	
p-value	(0.01)	(0.03)	

where S denotes the Lowess smoother function. It is obvious that both smoothers are significantly meaningful. Furthermore, the linear model is nested by the additive model with p-value being equal to 0.01. Figure 5.5 illustrates plots of the estimated partial-regression functions for the additive regression model. The points in each graph are partial residuals for the corresponding explanatory variable, removing the effect of the other explanatory variable. The broken lines demonstrate pointwise 95-percent confidence envelopes for the partial fits.
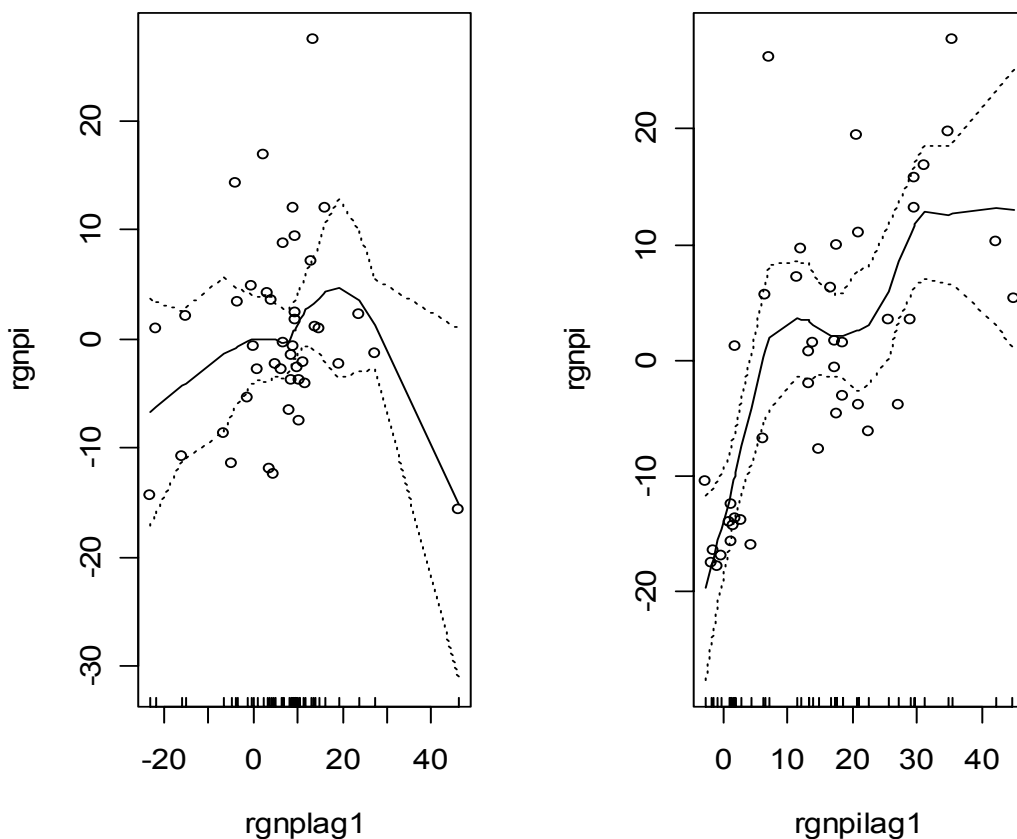
Figure 5.5: Plots of the estimated partial-regression functions for the additive regression of the inflation rate (rgnpi) on the lagged real GNP growth rate (rgnplag1) and the lagged inflation rate (rgnpilag1)

84

We use MARS model to fit a piecewise linear model with additive terms to the data. The results indicate that pairwise interaction terms (by degree=2 and degree=3) make little difference to the effectiveness of explanatory variables. Finally we computed the residuals of this model to compare to the alternative models.

The additive model seems to be too flexible and it is not able to cover the effect of interactions between explanatory variables. To remove this problem, the Projection-Pursuit Regression model has been proposed. The PPR model applies an additive model to projected variables. Figure 5.6 shows plots of the ridge functions for the three two-term projection pursuit regressions fitted to the data. As MARS model, residuals of PPR model have been computed.
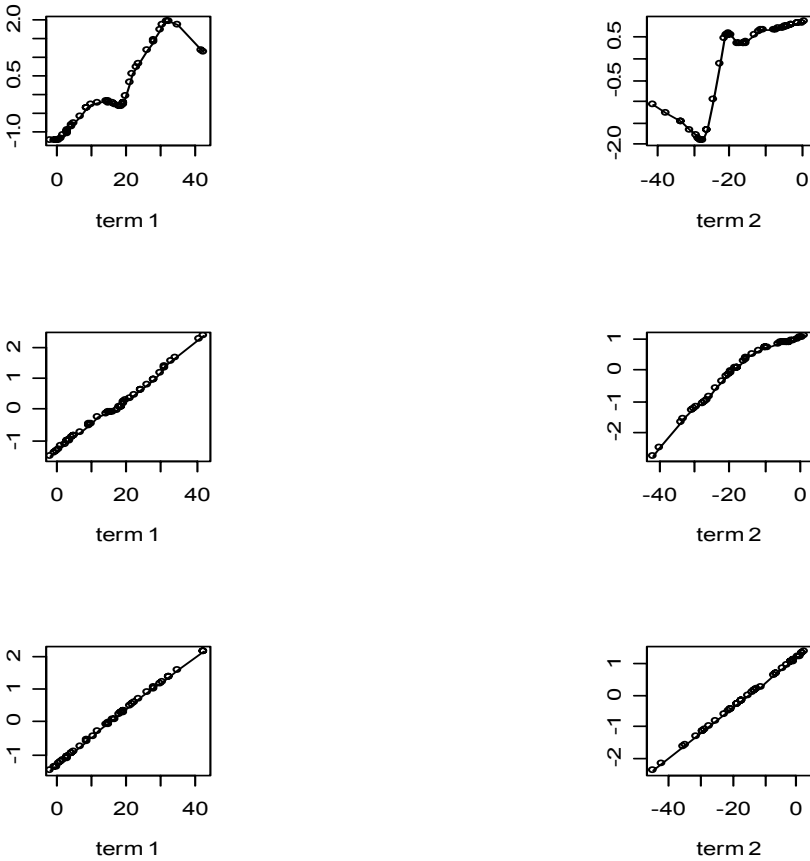
Figure 5.6: Plots of the ridge regression for three two-term projection pursuit regressions fitted to the data.

Although MARS model is an accurate method, it is sensitive to concurvity. Neural networks do not share this problem and are better able to predict in this situation. In fact, as neural networks are nonlinear projection methods and tend to overparameterize, they are not subject to concurvity. We examined several neural

network models and the results indicate that a 2-3-1 network has a better performance.

The Wilcoxon test has been used to compare the squared error of a neural network model and a rival model. The performance of PPR and AD models appears to differ from the neural network model, implying that the NN model can significantly outperform the PPR model and it has a better performance than the AD model, but not by much. Furthermore, the NN model is significantly better than the linear model (LM). However, there is no possibility that the NN model can outperform the MARS model. Table 5.1 presents the result of model comparison based on Wilcoxon test.

Table 5.1: Model Comparison based on Wilcox test

	p-value
PPR vs. NN	0.01
LM vs. NN	0.00
MARS vs. NN	1
AD vs. NN	0.38

Now we compare the NN model to the parametric autoregressive moving average (ARMA) model for inflation. Riddell and Smith (1982) used an "economically rational" expectations approach, proposed by Feige and Pearce (1976), by applying the Box-Jenkins (1970) model to the inflation series and then computing the predicted values as the expected inflation.

A collection of ARMA (p, q) models, for different orders of p and q, have been estimated and then the best model was selected according to the Akaike information criterion (AIC) and the Schwarz information criterion (SIC). Examining the ARMA models for the inflation series indicates that ARMA (1, 1) is the best-fitting model (see Table 5.2).

Table 5.2: Model Selection based on AIC and SIC

ARMA(p, q)	AIC	SIC
ARMA(3, 3)	7.30	7.60
ARMA(3, 2)	7.37	7.63
ARMA(2, 3)	7.34	7.60
ARMA(2, 2)	7.30	7.52
ARMA(2, 1)	7.26	7.43
ARMA(1, 2)	7.24	7.41
ARMA(1, 1)	7.20	7.33
ARMA(0, 1)	7.50	7.59
ARMA(1, 0)	7.76	7.85

Diagnostic checking, the correlogram (autocorrelations) of inflation from the regression tests was examined and confirmed the results. The last 5 observations are used for comparing the ex post forecasts generated by the two models. Furthermore, the Root Mean Square Error (RMSE) is used to evaluate ex post forecasts. We apply the feed-forward backpropagation as learning algorithm and a 1-2-1 network, where only lagged inflation is used as input. The results imply that the forecasting performance of the NN model (RMSE=0.05) is significantly better than that of the ARMA model (RMSE=11.73). It should be noted that the results from the inflation lags exceeding one and more number of hidden layers are almost the same. Therefore, the NN model outperforms the parametric ARMA model.

5.2 A multi-equation model

How do private agents form their inflation expectations? A variety of expectation formation schemes in the context of a multi-equation model will be considered to answer this question.

The multi-equation model to be estimated consists of three equations. The first equation is the wage equation:

$$W = \alpha_0 + \alpha_1 U + \alpha_2 OG + \alpha_3 \pi + \alpha_4 \pi^e \tag{1}$$

$$\alpha_1 < 0, \alpha_2, \alpha_3, \alpha_4 > 0$$

where W is nominal wage growth rate, U unemployment rate, OG output gap, π inflation rate, and π^e expected inflation rate. Equation (1) is the expectations-augmented Phillips curve. The output gap is defined as the percentage deviation of real GNP from its long-term trend (derived by the Hodrick-Prescott filter[1])

$$OG = \frac{y - y^*}{y^*}$$

where y is real GNP and y^* is the potential output of the economy. The unemployment rate (U) is a proxy for excess supply of labor. An increase in U will result in a decrease in W so that one expects $\alpha_1 < 0$. Since y is a proxy for the demand for labor and there is a positive relationship between y and OG, one would expect $\alpha_2 > 0$. The increases in wages caused by increases in prices are given by the equation $W = MP_L.P$, where MP_L is the marginal product of labor. The role of price inflation and expected price inflation in the determination of wages has been emphasized by many researchers (Gordon, 1971; Lahiri, 1981; Chen and

----

1. The Hodrick-Prescott (HP) filter is a two-sided linear filter which computes the potential output y_t^* of actual y_t by minimizing the variance of y_t around y_t^*, subject to a penalty that constrains the second difference of y_t^*. The HP filter selects y_t^* to minimize:

$$\sum_{t=1}^{T}(y_t - y_t^*)^2 + \lambda \sum_{t=2}^{T-1}((y_{t+1}^* - y_t^*) - (y_t^* - y_{t-1}^*))^2$$

The parameter, λ , controls the smoothness of the y_t^* series. The larger the parameter, the smoother the series.

Flaschel, 2006).

In a competitive economy, the expected inflation should have a coefficient equal to unity while in noncompetitive situations it lies between zero and unity depending on the strength of unions and other noncompetitive elements in the bargaining process. (Turnovsky and Wachter, 1972). Furthermore, the expected inflation coefficient will be unity only if workers can fully account for the amount of expected inflation in their current wage settlements (Turnovsky, 1972). According to the natural rate hypothesis, there is no way for the government to keep the unemployment rate permanently below the natural rate. Therefore, there is no long-run trade-off between inflation and unemployment, which implies that the coefficient of expected inflation will be unity.

Equation (2) is the aggregate demand function

$$y = \beta_0 + \beta_1(M_2/P) + \beta_2 g + \beta_3 c_{-1} + \beta_4 D \qquad (2)$$

$$\beta_1, \beta_2, \beta_3 > 0, \beta_4 < 0$$

where M_2/P is real money (M_2 broad money and P the implicit price deflator), g real government expenditure, c real consumption, and D dummy variable. This function can be derived as a solution of IS-LM relationships, with the lagged value of c introduced to make the implied consumption function of permanent-income hypothesis and the dummy variable included to capture the effect of the 1979 revolution on production. All variables in equation (2) except the dummy variable will be employed in the growth rates.

Equation (3) determines price level changes as

$$\pi = \gamma_0 + \gamma_1 M_2 + \gamma_2 X + \gamma_3 \pi^m + \gamma_4 T \qquad (3)$$

$$\gamma_1, \gamma_3, \gamma_4 > 0, \gamma_2 < 0$$

where X is the labor productivity measured as the ratio of real GNP to employment, π^m import inflation, and T is the trend variable. The variables M_2 and X will be used in the growth rates. Bahmani-Oskooee (1995), Liu and Adedeji (2000), Valadkhani (2006) and Bonato (2007) have supported money supply growth as one of the main determinants of inflation in Iran. It should be noted that wages are not the

cause of inflation in Iran because labor unions have not enough power and play no determining role in the economy.

Since all equations are over-identified, applying the ordinary least square (OLS) method to the equations will be inappropriate. One solution to this problem is to use two-stage least square procedure (2SLS).

Table 5.3 displays summary statistics of the data including the mean, the maximum, the minimum, and the standard deviation for the period 1959-2003 (see appendix I for the data source and definitions).

Table 5.3: Descriptive Statistics of the data (1959-2003)

Variable		Mean	Maximum	Minimum	Std. dev.
Wage growth arte (%)	(W)	16.59	47.47	-6.24	11.76
Unemployment rate (%)	(U)	9.28	14.70	2.80	4.27
Output Gap (%)	(OG)	0.28	38.80	-28.44	15.52
Inflation rate (%)	(π)	14.96	44.75	-2.62	12.49
Real GNP growth rate (%)	(y)	6.12	46.12	-22.93	12.06
Nominal growth rate of M2 (%)	(M_2)	24.48	57.06	6.02	9.47
Real growth rate of M2 (%)	(M_2/P)	9.20	38.85	-14.45	11.72
Real growth rate of government expenditure (%) (g)		6.45	61.76	-20.57	13.89
Real growth rate of consumption (%) (c)		4.97	31.47	-10.13	7.18
Import price growth rate (%)	(π^m)	13.59	71.89	-2.06	14.82
Productivity growth rate (%)	(X)	3.54	43.05	-24.70	11.69

Before running the regressions, we examined whether the data are stationary. The empirical results of the Augmented Dicky-Fuller test indicate that all the variables employed are stationary, and thus this issue helps us to avoid the problem of the spurious relationships (see Table 5.4).

Table 5.4: Results of Augmented Dicky-Fuller test

	Augmented ADF test statistic			
	levels	p-value	growth rates	p-value
W				
Case 1	21.18	1	- 3.35	0.01
Case 2	20.67	1	-3.61	0.04
Case 3	21.44	1	-0.85	0.34
u				
Case 1	-1.54	0.50	-6.62	0.00
Case 2	-1.43	0.84	-6.65	0.00
Case 3	0.26	0.76	-6.45	0.00
OG				
Case 1	-3.09	0.03	-5.83	0.00
Case 2	-4.16	0.01	-5.76	0.00
Case 3	-3.13	0.00	-5.90	0.00
P				
Case 1	2.58	1	-2.10	0.25
Case 2	2.73	1	-4.54	0.00
Case 3	2.80	0.99	-0.90	0.32
gnp				
Case 1	-0.49	0.88	-4.69	0.00
Case 2	-1.47	0.83	-4.80	0.00
Case 3	1.68	0.98	-2.30	0.02
M_2				
Case 1	7.29	1	-3.23	0.02
Case 2	7.06	1	-3.22	0.09
Case 3	7.41	1	-0.96	0.30
g				
Case 1	-2.09	0.25	-2.40	0.15
Case 2	-2.13	0.51	-4.47	0.00
Case 3	0.14	0.72	-2.07	0.03
c				
Case 1	0.72	0.99	-4.14	0.00
Case 2	-2.16	0.50	-4.13	0.01
Case 3	2.42	0.99	-3.13	0.00
X				
Case 1	-2.50	0.12	-4.52	0.00
Case 2	-2.46	0.35	-4.74	0.00
Case 3	0.19	0.74	-.2.49	0.01
p^m				
Case 1	2.54	0.11	-2.87	0.05
Case 2	-2.99	0.15	-3.09	0.12
Case 3	-2.44	0.01	-1.99	0.04

Case 1: constant & no trend model
Case 2: constant & trend model
Case 3: no constant & no trend model

5.2.1 Rational expectations

According to the rational expectation hypothesis (Muth, 1961), people expect inflation based on the information available and their expectation is the same as the prediction of the relevant economic theory. Following McCallum (1976), we apply the rational expectations hypothesis to the model. The rationality assumption could be written as

$$\pi_{t+1}^{e} = E(\pi_{t+1}|\Omega_t) = \pi_{t+1} - \eta_t$$

$$E(\eta_t) = 0$$

where Ω_t is the information set as of time t including the predetermined and lagged variables of the system. Muth assumes that the error term η_t is uncorrelated with each variable that appears in the information set.

A variety of estimates of the wage equation are obtained using different information sets for the expected inflation. We apply six different information sets which are known to market participants to see whether estimates of the wage equation, and especially the coefficient of π^e, are sensitive to any restrictions on the information sets. The private agents may not have access to some information when forming expectations. In case I, lagged values of the exogenous variables of the system have been used. Case II is the same as Case 1 but in addition including dummy and trend variables. Contemporaneous values of exogenous variables are assumed to be known in Case III and in Case IV lagged exogenous variables are added to the instrument set in Case III. Expected inflation is generated under the assumption that private agents are "partly rational" in forming expectations in Case V. We assume that π_{t-1} and π_{t-2} are used to create expected inflation. Finally, in case VI only π_{t+1} is included in the instrument set and thus the wage equation will be estimated by ordinary least squares (OLS).

Consider the estimated multi-equation model for rational expectations which has been presented in Table 5.5. Since there was evidence of autocorrelation in the wage equation, the estimated function was corrected for this problem using the Eviews procedure. In the income equation, lagged real consumption was excluded from specification because this variable was not significant.

Table 5.5: Multi-Equation model for Rational Expectations

	Case I	Case II	Case III	Case IV	Case V	Case VI
Wage equation						
Const. (t-value)	19.92 (2.59)	19.54 (2.57)	13.13 (1.63)	14.48 (2.56)	20.07 (3.03)	16.34 (1.74)
U (t-value)	-0.70 (-1.00)	-0.69 (-1.00)	-0.24 (-0.33)	-0.72 (-1.04)	-0.67 (-1.06)	-0.33 (-0.40)
OG (t-value)	0.32 (2.43)	0.33 (2.42)	0.25 (1.69)	0.34 (2.52)	0.36 (2.75)	0.24 (1.69)
π (t-value)	0.30 (1.22)	0.34 (1.37)	0.53 (1.66)	0.48 (1.30)	0.31 (1.46)	0.25 (1.85)
π^e (t-value)	0.05 (0.33)	0.02 (0.18)	-0.12 (-0.69)	-0.04 (-0.18)	0.03 (0.24)	0.02 (0.18)
$\overline{R}^2$	0.59	0.59	0.59	0.56	0.56	0.56
DW	2.07	2.06	1.75	2.06	2.02	1.73
Income equation						
Const. (t-value)	0.224 (0.139)	0.223 (0.138)	0.143 (0.090)	0.278 (0.173)	0.156 (0.095)	0.680 (0.447)
M_2/P (t-value)	0.441 (2.897)	0.441 (2.899)	0.445 (3.029)	0.430 (2.844)	0.478 (2.909)	0.358 (2.780)
g (t-value)	0.359 (3.009)	0.359 (3.008)	0.357 (3.065)	0.365 (3.063)	0.335 (2.629)	0.401 (3.686)
D57 (t-value)	-20.455 (-2.493)	-20.455 (-2.493)	-20.412 (-2.518)	-20.418 (-2.491)	-20.708 (-2.481)	-20.182 (-2.537)
$\overline{R}^2$	0.57	0.57	0.57	0.57	0.56	0.57
DW	1.96	1.96	1.95	1.97	1.93	2.05
Price equation						
Const. (t-value)	-5.428 (-1.733)	-5.428 (-1.733)	-5.612 (-1.863)	-5.428 (-1.733)	-5.070 (-1.557)	-4.943 (-1.630)
M_2 (t-value)	0.362 (3.088)	0.362 (3.088)	0.363 (3.135)	0.362 (3.088)	0.362 (3.051)	0.357 (3.034)
X (t-value)	-0.366 (-3.853)	-0.366 (-3.853)	-0.366 (-3.897)	-0.366 (-3.853)	-0.368 (-3.833)	-0.371 (-3.888)
π^m (t-value)	0.309 (3.828)	0.309 (3.828)	0.308 (3.872)	0.309 (3.828)	0.310 (3.800)	0.339 (4.326)
T (t-value)	0.377 (3.941)	0.377 (3.941)	0.383 (4.155)	0.377 (3.941)	0.365 (3.658)	0.333 (3.790)
$\overline{R}^2$	0.75	0.75	0.76	0.75	0.74	0.74
DW	2.00	2.00	2.00	2.00	2.01	1.98

All coefficients of the explanatory variables in the wage equation have the expected signs but they are not statistically significant. The size of the coefficient of expected inflation (π^e) is very small and even negative in cases III and IV. This implies that the expectation hypothesis is rejected for the rational expectations model. It should be noted that all coefficients of the explanatory variables in the income and price equations are of the appropriate sign and are statistically significant.

Although the idea of rational expectations is attractive, it does not hold in the case of Iran. Apparently access to the information is not symmetric and inflation expectations cannot be formed in a rational manner. The structure of the economy is unstable in a way that is imperfectly understood by both the public and policymakers and the policymakers' objective function seems to be not completely known by private agents.

Now we examine whether the Hodrick-Prescott (HP) filter might stand as a proxy for a rational expectations series. The reasoning behind this is that both the HP filter and rational expectations use all available information. This method has been supported by some studies (Orr et al, 1995; Martins and Scarpetta, 1999; Ash et al, 2000). Some rationality tests including unbiasedness and efficiency are applied to the HP-filtered series. Assume that π_t^f denotes a proxy for a rational forecast of π_t, where π_t is inflation rate series and π_t^f the corresponding HP filtered series. To test for unbiasedness, we first run the following regression

$$\pi_t = a_0 + a_1 \pi_t^f + \varepsilon_t$$

Then we examine the necessary condition of unbiasedness by testing the joint hypothesis $a_0 = 0$ and $a_1 = 1$. The sufficient condition is given by

$$\pi_t - \pi_t^e = E_t = \mu_t + \varepsilon_t$$

The hypothesis to test is µ = 0. The estimated regression of π_t on π_t^f is as follows

$$\pi_t = -0.69 + 1.04 \pi_t^f$$

(t)   (-0.34) (9.09)

The joint test of the hypothesis $a_0 = 0$ and $a_1 = 1$ can not be rejected with $\chi^2 = 0.163$ (p-value=0.92). Furthermore, calculating the mean forecast error

(E=3.25e-14), the sufficient condition holds with $t = 2.97e - 14$ (p-value=1). As a result, the HP series may be assumed to be unbiased.

To test for efficiency, a test of the joint hypothesis $b_1 = b_2 = 0$ is conducted based on the following regression

$$\pi_t - \pi_t^e = b_1 (\pi_{t-1} - \pi_{t-1}^e) + b_2 (\pi_{t-2} - \pi_{t-2}^e) + \varepsilon_t$$

The estimated coefficients ($\hat{b}_1$ and $\hat{b}_2$) are -0.040 and 0.045 respectively. The joint hypothesis $b_1 = b_2 = 0$ can not be rejected with chi-square (χ^2) being equal to 0.167 (p-value=0.91). Therefore, the filtered series π_t^f may be regarded to be efficient. Since our results indicate that the series π_t^f is unbiased and efficient, we may conclude that π_t^f is rational in the sense of Muth (1960) and it can be used as a proxy for rational expectations in the case of Iran. Ash et al. (2000) applied the rationality tests to the US data and came to the conclusion that the HP series are 'weakly rational', i.e. the series is unbiased but inefficient.

5.2.2 Backward-looking expectations

As backward-looking models can help to explain inflation inertia, many studies have applied these models for inflation expectations (Ball, 1991; Roberts, 1997, 1998; Rudebusch and Svensson, 1999). On the other hand, since past inflation data is a cheap and relatively informative signal about the central bank policies, the agents can easily use these models by extrapolating from observed past inflation. However, it is asserted that they are subject to the Lucas critique. Rudebusch and Svensson (1999) concluded that this critique seems to be irrelevant since empirically the estimated parameters do not show significant instability. Stanley (2000) also concluded that there is little evidence supporting the empirical relevance of the Lucas critique. Linde (2001a) arrives at the conclusion that instability tests cannot detect the relevance of the Lucas critique in small samples. On the other hand, the results of some studies indicate that forward-looking models are also to be subject to this critique (Linde 2001b, Rudd and Whelan 2007). Therefore, there is no general agreement to choose backward-looking or forward-looking models if instability tests are considered as a criterion.

5.2.2.1 Adaptive expectations

According to adaptive expectations, expectations are revised upward or downward based on the most recent error (Cagan, 1956). As it was mentioned in chapter two, adaptive expectations is defined as

$$\pi_t^e = \pi_{t-1}^e + \lambda(\pi_{t-1} - \pi_{t-1}^e)$$

Setting $\lambda = 1$, adaptive expectations will result in a special type of expectation scheme called static expectations ($\pi_t^e = \pi_{t-1}$). Private agents most likely consider the lagged inflation rates in forecasting inflation (McCallum, 1976).

Since there is some inertia in the inflation process, some researchers have proposed the use of a less-than-fully rational model of expectations called the optimal univariate model (Riddell and Smith, 1982; Staiger et al, 1997; Ball, 2000). It is assumed that agents use only the past values of inflation but in a way different from the backward-looking models. Applying the Box-Jenkins approach to select an autoregressive moving average (ARMA) model for inflation, agents make optimal univariate forecasts. Although they ignore other relevant variables, they use inflation data as best they can. This model is not subject to the Lucas critique because the univariate process for inflation can be different as the monetary regime changes. Since such model use only the lagged values of inflation, we compare it to the static and adaptive alternatives here and then to all expectation models in the last part of this chapter.

Table 5.6 shows the estimated multi equation model for static, adaptive and univariate expectations. Using three lags of inflation, adaptive expectations are generated for the case in which $\lambda = 0.3$, 0.5 and 0.7. Applying the Box-Jenkins approach, the expected inflation is given by ARMA (1, 1) process without a constant.

The results imply that all coefficients in the wage equation are statistically significant with the expected signs for all cases in which the expectations are generated by the static, adaptive and optimal univariate models (except the coefficient of π_t^e when adaptive expectations ($\lambda = 0.7$) is considered (This model will be excluded when various expectation formation schemes are compared to select the best model.). As the coefficient of expected inflation in most cases is significant, we can conclude that the expectation hypothesis is supported by each of these near rational expectations models. Therefore, inflation expectations play a major role in the determination of wages.

Table 5.6: Multi-Equation Model for Static, Univariate and Adaptive Expectations

	Static expectations	Univariate expectations	Adaptive expectations		
	.		$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 0.7$
Wage equation					
Const. (t-value)	19.66 (2.52)	20.73 (2.50)	21.90 (2.98)	21.87 (2.97)	21.97 (3.09)
U (t-value)	-1.59 (-2.02)	-1.90 (-2.14)	-1.79 (-2.25)	-1.77 (-2.19)	-1.63 (-2.07)
OG (t-value)	0.38 (2.72)	0.31 (2.29)	0.38 (2.79)	0.36 (2.64)	0.35 (2.60)
π (t-value)	0.62 (3.58)	0.48 (3.08)	0.54 (3.25)	0.50 (3.10)	0.46 (2.91)
π^e (t-value)	0.36 (2.68)	0.69 (2.14)	0.46 (2.24)	0.56 (2.06)	0.68 (1.78)*
$\bar{R}^2$	0.57	0.58	0.51	0.52	0.52
DW	1.86	1.84	1.80	1.81	1.82
Income equation					
Const. (t-value)	-0.285 (-.172)	-0.247 (-.149)	-0.393 (-0.228)	-0.365 (-0.213)	-0.337 (-0.197)
M_2/P (t-value)	0.540 (3.244)	0.533 (3.230)	0.645 (3.268)	0.639 (3.256)	0.633 (3.242)
g (t-value)	0.307 (2.474)	0.310 (2.516)	0.240 (1.691)*	0.244 (1.722)*	0.247 (1.753)*
D57 (t-value)	-20.822 (-2.526)	-20.799 (-2.527)	-21.626 (-2.526)	-21.599 (-2.527)	-21.572 (-2.528)
$\bar{R}^2$	0.55	0.55	0.54	0.54	0.55
DW	1.86	1.87	1.84	1.84	1.85
Price equation					
Const. (t-value)	-4.713 (-1.497)	-4.713 (-1.497)	-3.698 (-1.104)	-3.698 (-1.104)	-3.698 (-1.104)
M_2 (t-value)	0.356 (2.989)	0.356 (2.989)	0.358 (2.966)	0.358 (2.966)	0.358 (2.966)
X (t-value)	-0.371 (-3.848)	-0.371 (-3.848)	-0.379 (-3.866)	-0.379 (-3.866)	-0.379 (-3.866)
π^m (t-value)	0.340 (4.278)	0.340 (4.278)	0.339 (4.219)	0.339 (4.219)	0.339 (4.219)
T (t-value)	0.326 (3.581)	0.326 (3.581)	0.293 (2.989)	0.293 (2.989)	0.293 (2.989)
$\bar{R}^2$	0.73	0.73	0.71	0.71	0.71
DW	1.98	1.98	2.03	2.03	2.03

* Not significant at 5 percent level.

Although the magnitude of the coefficient of π^e for static expectations is less than that of adaptive alternatives, its adjusted R-squared is larger than for the adaptive ones. Comparing univariate type of formation expectation to the other alternatives, we find that $\overline{R}^2$ for the univariate model is marginally larger than for the static and the adaptive schemes. Furthermore, the coefficient of expected inflation in the univariate model is much larger than the static case and even lager than for the adaptive expectation models.

Instability tests including the Chow breakpoint test and the Chow forecast test for some specific years have been conducted and the results imply that there are no structural breaks and thus the Lucas critique is irrelevant.

5.2.2.2 Forming expectations using a mix of extrapolative and regressive expectations

Some studies have utilized the hybrid model of expectations (Modigliani and Sutch, 1966; Hara and Kamada, 1999; Westerhoff, 2006). The basic idea is that the agents' final expectations may combine extrapolative and regressive elements. It is assumed that agents use a weighted average of extrapolative and regressive expectations to forecast inflation as

$$\pi_t^e = W_t \pi_t^{ex} + (1 - W_t)\pi_t^{re}$$

$$\pi_t^{ex} = \pi_{t-1} + \tau(\pi_{t-1} - \pi_n)$$

$$\pi_t^{re} = \pi_{t-1} + \kappa(\pi_n - \pi_{t-1})$$

$$W_t = \frac{1}{1 + (\pi_{t-1} - \pi_n)^2}$$

where π_t^{ex} and π_t^{re} are extrapolative and regressive expectations respectively. If inflation rises from its "normal" level (π_n), then this increase is extrapolated and expected inflation increases. But there is a possibility that some agents expect the inflation to regress to its previous level. As it is obvious, the relative impact of extrapolative and regressive expectations (W) is time-varying and thus agents' expectations will be nonlinear then. The more lagged inflation deviates from its "normal" level, the less weight the agents put on extrapolative expectations and the more weight on regressive expectations.

Table 5.7 reports the estimated multi equation model for a mix of extrapolative and regressive expectations for the cases which τ and κ are set to be 0.3, 0.5, and 0.7. As for other backward-looking models, all the parameter estimates are of the appropriate sign with the associated t-ratios in excess of 2.0.

Table 5.7: Multi-Equation Model for a mix of extrapolative and regressive expectations with time-varying weights

	$\tau = \kappa = 0.3$	$\tau = \kappa = 0.5$	$\tau = \kappa = 0.7$
Wage equation			
Const.	21.16	21.51	21.82
(t-value)	(2.87)	(2.88)	(2.94)
U	-1.72	-1.78	-1.76
(t-value)	(-2.20)	(-2.20)	(-2.13)
OG	0.39	0.38	0.38
(t-value)	(2.84)	(2.80)	(2.78)
π	0.58	0.56	0.53
(t-value)	(3.48)	(3.40)	(3.26)
π^e	0.41	0.44	0.44
(t-value)	(2.47)	(2.36)	(2.12)
$\overline{R}^2$	0.54	0.54	0.54
DW	1.83	1.84	1.86
Income equation			
Const.	-0.314	-0.307	-0.297
(t-value)	(-0.186)	(-0.182)	(-0.176)
M_2 / P	0.578	0.576	0.574
(t-value)	(3.206)	(3.203)	(3.198)
g	0.281	0.281	0.283
(t-value)	(2.109)	(2.218)	(2.128)
D57	-21.118	-21.112	-21.104
(t-value)	(-2.514)	(-2.514)	(-2.515)
$\overline{R}^2$	0.55	0.55	0.55
DW	1.84	1.84	1.85
Price equation			
Const.	-4.305	-4.305	-4.305
(t-value)	(-1.319)	(-1.319)	(-1.319)
M_2	0.356	0.356	0.356
(t-value)	(2.957)	(2.957)	(2.957)
X	-0.374	-0.374	-0.374
(t-value)	(-3.837)	(-3.837)	(-3.837)
π^m	0.340	0.340	0.340
(t-value)	(4.246)	(4.246)	(4.246)
T	0.313	0.313	0.313
(t-value)	(3.305)	(3.305)	(3.305)
$\overline{R}^2$	0.72	0.72	0.72
DW	2.00	2.00	2.00

Now consider the case that agents apply another procedure to update the weights. They use a weighted linear combination of extrapolative and regressive expectations to forecast inflation as

$$\pi_t^e = W_t^{ex}\pi_t^{ex} + W_t^{re}\pi_t^{re}$$

where W_t^{ex} and W_t^{re} are the relative weights of extrapolative and regressive expectations respectively. It is assumed that these weights are updated via a discrete-choice model as

$$W_t^{ex} = \frac{\exp(\delta a_t^{ex})}{\exp(\delta a_t^{ex}) + \exp(\delta a_t^{re})}$$

$$W_t^{re} = \frac{\exp(\delta a_t^{re})}{\exp(\delta a_t^{ex}) + \exp(\delta a_t^{re})}$$

where a_t^{ex} and a_t^{re} are the attractiveness of extrapolative and regressive expectations defined as

$$a_t^{ex} = -(\pi_{t-1}^{ex} - \pi_{t-1})^2$$

$$a_t^{re} = -(\pi_{t-1}^{re} - \pi_{t-1})^2$$

The parameter $\delta \geq 0$ measures degree of agents' sensitivity to choosing the most attractive predictor. In case $\delta = 0$, agents cannot distinguish between extrapolative and regressive expectations so that $W_t^{ex} = W_t^{re} = 0.5$. Therefore, we may interpret an increase in δ as an increase in the rationality of the agents.

Estimating the multi equation model in the case mentioned above, it is concluded that if δ is selected to be equal to zero, then the coefficient of expected inflation is significant. Moreover, market participants seem to be able to distinguish between the two predictors and there is possibility to increase rationality of agents as we increase δ. Table 5.8 shows the results for the cases $\delta = 0$, $\delta = 1$, and $\delta = 5$. The results of comparing the discrete-choice rule $(\delta = 0)$ to time-varying weights counterparts imply that the adjusted coefficient of determination ($\bar{R}^2$) in the wage equation for the discrete-choice rule (0.55) exceeds that of time-varying peers but its coefficients of π^e (0.28) is less than that of time-varying weighting rules.

Table 5.8: Multi-Equation Model for a mix of extrapolative and regressive expectations with discrete-choice updating weights

	$\delta = 0$	$\delta = 1$	$\delta = 5$
Wage equation			
Const.	20.43	21.39	21.50
(t-value)	(2.91)	(3.15)	(3.11)
U	-1.46	-1.55	-1.60
(t-value)	(-2.03)	(-2.16)	(-2.18)
OG	0.39	0.41	0.41
(t-value)	(2.91)	(3.09)	(3.06)
π	0.58	0.58	0.59
(t-value)	(3.50)	(3.45)	(3.43)
π^e	0.28	0.28	0.29
(t-value)	(2.50)	(2.31)	(2.29)
$\bar{R}^2$	0.55	0.50	0.50
DW	1.83	1.80	1.79
Income equation			
Const.	-0.325	-0.422	-0.422
(t-value)	(-0.192)	(-0.245)	(-0.245)
M_2/P	0.580	0.652	0.652
(t-value)	(3.210)	(3.278)	(3.277)
g	0.279	0.237	0.237
(t-value)	(2.097)	(1.658)	(1.658)
D57	-21.126	-21.654	-21.654
(t-value)	(-2.514)	(-2.525)	(-2.525)
$\bar{R}^2$	0.55	0.54	0.54
DW	1.84	1.83	1.83
Price equation			
Const.	-4.305	-3.698	-3.698
(t-value)	(-1.319)	(-1.104)	(-1.104)
M_2	0.356	0.358	0.358
(t-value)	(2.957)	(2.966)	(2.966)*
X	-0.374	-0.379	-0.379
(t-value)	(-3.837)	(-3.866)	(-3.866)
π^m	0.340	0.339	0.339
(t-value)	(4.246)	(4.219)	(4.219)
T	0.313	0.293	0.293
(t-value)	(3.305)	(2.989)	(2.989)
$\bar{R}^2$	0.72	0.71	0.71
DW	2.00	2.03	2.03

5.2.3 Forming expectations using a learning model

Since the optimal monetary policy depends considerably on the assumed nature of the expectations formation process, some researchers have considered more realistic ways of modeling expectation formation, i.e. the learning approach (Sargent, 1999; Evans and Honkapohja, 2001; Orphanides and Williams, 2004; Basdevant, 2005; Evans and McGough, 2006; Waters, 2007). They believe that neither backward-looking models and nor models with rational expectations are reasonable and realistic because the former assumes that historical econometric relationships are invariant to changes in the economic policy and that agents do not react to systematic mistakes they made while the latter assumes that agents fully know the structure of the economy and form their expectations accordingly.

Since private agents are not endowed with a priori knowledge of inflation behavior, they must learn about it over time as new data becomes available (Sargent, 1999). Although much research has been done on adaptive learning models in recent years, they are largely theoretical. In this research, expected inflation is generated through a least square learning rule and then expected inflation is being analyzed in the augmented Phillips curve equation as for previous models. An econometric tool to study learning is the Kalman filter which can be used to estimate time-varying economic relationships.

Some economists, especially in 1990s, have presumed that the Phillips curve is dead since inflation and unemployment fell. During that period, some favorable supply shocks happened such as a reduction in oil prices, labor-market changes which resulted in reducing the natural rate of unemployment, and improvements in production technology. However, other scientists maintain that the Phillips curve is still relevant (Mankiw, 2001; Eller and Gordon, 2003; Fischer, 2007).

The traditional Phillips curve focuses mainly on backward-looking behavior, while the New Keynesian Phillips curve considers forward-looking behavior. In fact, the main difference between these two is in the way expectations are estimated. The Phillips curve equation has not changed, only the expected inflation term is estimated in a different way (Fischer, 2007).

The learning approach assumes that the agents' expectations of inflation are on average correct but a limited set of information is used. Different information sets are used to test whether estimates of the wage equation are sensitive to any change in

the information sets. First it is assumed that agents forecast inflation using π_{t-1} and y_{t-1} through recursive least squares (RLS)

$$\pi_{t|t+1} = b_{1,t} + b_{2,t}\pi_{t-1} + b_{2,t}y_{t-1} + \varepsilon_t$$

where $\pi_{t|t+1}$ is the inflation rate in the next period expected by the agents at time t. Agents forecast inflation in the next period by updating the parameters period by period. The process of updating is based on RLS as follows (Bullard 1992, Sargent 1999, Evans and Honkapohja 2001)

$$B_t = B_{t-1} + t^{-1}R_t^{-1}X_t(\pi_t - X_t'B_{t-1})$$

$$R_t = R_{t-1} + t^{-1}(X_t X_t' - R_{t-1})$$

where $B_t = (b_{1,t}, b_{2,t}, b_{3,t})'$ and $X_t = (1, \pi_{t-1}, y_{t-1})$. The equations above correspond to the following state space model:

$$\pi_{t|t+1} = b_{1,t} + b_{2,t}\pi_{t-1} + b_{2,t}y_{t-1} + \varepsilon_t$$

$$b_{i,t} = b_{i,t-1} + v_{i,t}$$

The expected inflation is computed as the predicted value for $\pi_{t|t+1}$.

Table 5.9 presents the results of different information sets used for learning. In Case 1, it is assumed that agents use only lagged inflation π_{t-1} to create expected inflation. Case2, which has been explained above, includes π_{t-1} and y_{t-1}. Finally in Case 3, market participants apply π_{t-1} and $M_{2,(t-1)}$. There is little difference in the results as the information sets change.

All coefficients of the explanatory variables in the multi equation model have the expected signs and are statistically significant. The size of the coefficient of expected inflation (π^e) in the wage equation is relatively large. Moreover, the values of adjusted R-squared for the wage equation for learning models are larger than other alternative models, implying that learning models seem to be better suited modeling expectation formation than the traditional approaches.

Table 5.9: Multi-Equation Model for learning

	Case I	Case II	Case III
Wage equation			
Const. (t-value)	15.79 (1.97)	15.19 (1.90)	15.17 (1.90)
U (t-value)	-1.59 (-2.02)	-1.57 (-2.05)	-1.56 (-2.03)
OG (t-value)	0.39 (2.72)	0.36 (2.63)	0.36 (2.62)
π (t-value)	0.62 (3.58)	0.60 (3.61)	0.60 (3.60)
π^e (t-value)	0.61 (2.68)	0.66 (2.99)	0.65 (2.92)
$\overline{R}^2$	0.57	0.60	0.59
DW	1.86	1.80	1.82
Income equation			
Const. (t-value)	-0.285 (-0.172)	-0.287 (-0.173)	-0.285 (-0.172)
M_2/P (t-value)	0.540 (3.224)	0.540 (3.224)	0.540 (3.224)
g (t-value)	0.307 (2.474)	0.307 (2.471)	0.307 (2.474)
D57 (t-value)	-20.822 (-2.526)	-20.824 (-2.526)	-20.822 (-2.526)
$\overline{R}^2$	0.56	0.56	0.56
DW	1.87	1.87	1.87
Price equation			
Const. (t-value)	-4.713 (-1.497)	-4.713 (-1.497)	-4.713 (-1.497)
M_2 (t-value)	0.356 (2.989)	0.356 (2.989)	0.356 (2.989)
X (t-value)	-0.371 (-3.848)	-0.371 (-3.848)	-0.371 (-3.848)
π^m (t-value)	0.340 (4.278)	0.340 (4.278)	0.340 (4.278)
T (t-value)	0.326 (3.581)	0.326 (3.581)	0.326 (3.581)
$\overline{R}^2$	0.74	0.74	0.76
DW	1.99	1.99	1.99

5.2.4 Forward-looking expectations

The basic New Keynesian Phillips curve (NKPC) can be represented as (Gali and Gertler, 1999; Galí, Gertler and López-Salido, 2001)

$$\pi_t = \beta E_t \pi_{t+1} + \lambda mc_t$$

where $E_t \pi_{t+1}$ is the expected rate of inflation at time t+1 based upon information available at time t and mc_t is real marginal cost. This model can be derived from the Calvo price-setting mechanism (Calvo, 1983). Monopolistically competitive firms would set prices as a fixed markup over marginal cost. Since marginal cost can be related to the output gap, the NKPC can be specified as

$$\pi_t = \beta E_t \pi_{t+1} + \lambda y_t$$

where y_t is the output gap. Empirical evidence demonstrates that there are three problems regarding the NKPC (Mankiw, 2001): (1) It results in "disinflationary booms", (2) Inflation inertia can not be explained, and (3) It is not able to give a proper description of the impulse response functions to monetary policy shocks. In order to remove these problems, the hybrid NKPC, which includes an additional lagged inflation term, has been suggested.

In this section, the hybrid New Keynesian Phillips curve will be analyzed[1] (Fuhrer and Moore, 1995)

$$\pi_t = \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + ky_t$$

In the backward-looking Phillips curve, agents form their expectations using lagged inflation rates. These are included into current wage and price contracts. In the hybrid NKPC, past inflation matters just due to its correlation with $E_t \pi_{t+1}$. This term ($E_t \pi_{t+1}$) can be proxied by the fitted values from a regression of π_{t+1} on the information set including π_{t-1} and y_t. Although the backward-looking Phillips curve and the NKPC are apparently similar, policy implications will be different under each of these views. The estimated equation is as follows

----

1. The New Keynesian models usually include the NKPC, the IS curve equation and the Taylor-type interest rate rule (see Clarida et al., 1999 or Gali, 2000) However, since there is no data on interest rate in Iran after the 1979 Islamic revolution, we dropped the IS curve equation and the Taylor rule from our specification. It should be noted that after 1979 profit rates, instead of interest rate, have been introduced but such rates are not compatible with pre-revolution's interest rate data.

$$\pi_t = 2.52 + 0.61 E_t \pi_{t+1} + 0.23 \pi_{t-1} + 0.02 OG$$

(t-value) (1.68)   (8.63)   (2.74)   (0.28)

$$\bar{R}^2 = 0.77, \; D.W = 1.54$$

In this model, inflation outcome is related to both forward-looking and backward-looking terms. The large estimate of the forward-looking coefficient and the small estimate of the backward-looking coefficient should not be interpreted in favor of the forward-looking behavior. Such estimates can be obtained even if the true model is purely backward-looking (Rudd and Whelan, 2005). This situation may occur because of model misspecification and especially due to omitted variable bias. If an omitted variable z which is one of the determinants of inflation is correlated with π_{t+1} and the variables employed to instrument for it, the estimates of the forward-looking coefficient will be biased upwards (see appendix II for further detail of the effects of omitted variable bias).

Empirical evidence about the hybrid NKPC gives the contrasting results. For instance, Gali and Gertler (1999) using marginal cost find that forward-looking behavior is dominant while Fuhrer (1997) and Roberts (2001) using output gap as a proxy for the marginal cost conclude that forward-looking behavior is unimportant. Therefore, in case one uses marginal cost then the forward-looking term will be dominant while models based on output gap tend to reject the forward-looking behaviors.

Selecting the best model

Two criteria are used to compare the models: the standardized expected inflation coefficient, i.e. the expected inflation coefficient (α_4) divided by the standard deviation of expected inflation, and adjusted R-squared ($\bar{R}^2$). Table 5.10 reports the results.

Table 5.10: Comparing inflation expectations schemes in the wage equation

Expectations schemes	The coefficient of π^e (α_4)	$\alpha_4 / S(\pi^e)$	$\bar{R}^2$
Static Exp.	0.3682	0.0292	0.572
Univariate Exp.	0.6953	0.0658	0.587
Adaptive Exp. ($\lambda = 0.3$)	0.4660	0.0421	0.518
Adaptive Exp. ($\lambda = 0.5$)	0.5606	0.0580	0.523
Mix Exp. ($\tau = \kappa = 0.3$)	0.4110	0.0344	0.548
Mix Exp. ($\tau = \kappa = 0.5$)	0.4477	0.0383	0.547
Mix Exp. ($\tau = \kappa = 0.7$)	0.4469	0.0389	0.545
Mix Exp (discrete-choice rule, $\delta = 0$)	0.2869	0.0218	0.550
Mix Exp (discrete-choice rule, $\delta = 1$)	0.2803	0.0215	0.509
Mix Exp (discrete-choice rule, $\delta = 5$)	0.2963	0.0230	0.504
Learning (Case 1)	0.6137	0.0810	0.572
Learning (Case 2)	0.6590	**0.0865**	**0.598**
Learning (Case 3)	0.6498	0.0856	0.593

The learning approach is better suited for modeling inflation expectations than other alternative models if the two criteria mentioned above are considered. The learning approach (Case 2) has the maximum adjusted R-squared and the standardized expected inflation coefficient among other models.

The message of the learning models is: "being more aggressive to inflation". According to Orphanides and Williams (2002), the optimal monetary policy under a learning process should be more aggressive and narrowed to inflation stability. In case a learning model is considered, any inflation shock can feed into the future which contradicts with stabilization polices. In such conditions, a tight monetary policy geared at solidly anchored inflation expectations is recommended.

6. Summary and Conclusions

The role of expectations in the inflation process has been hotly debated over the years. Although economists agree that inflation expectations matter, there is not yet consensus about which inflation expectations matter (Mankiw, 2007). Does current inflation depend on the *current* expectation of *future* inflation (forward-looking new Keynesian models) or on the *past* expectations of *current* inflation (backward-looking models)? Failure to investigate this issue fully could lead to flawed economic policy.

The Iranian economy has experienced a relatively high inflation with an average inflation rate of about 15 percent over the period 1959-2003. It has even been more than 21 percent on average after the 1973 oil crisis. There is also a general agreement over the underestimating of the measured inflation due to price controls and government subsidies. Since the economy depends largely on oil revenues, any change in oil prices can directly affect all economic sectors.

The purpose of this study was to examine how market participants form their inflation expectations in the Iranian economy over the period 1959-2003. Inflation expectations are very unstable in Iran's economy because the Central Bank is unable to adhere to an inflation target in practice. Thus, inflation expectations are not well-anchored and any oil price increase, which seems apparently to be a favorable shock, results in money creation, fueled by government spending out of oil revenues, and inflation and causes private agents to raise inflation expectations. This in turn will increase inflation. As a result, poor anchored inflation expectations make price stability much more difficult to achieve in the long run and decrease the Central Bank's ability to stabilize output and employment in the short run. Furthermore, subsidies on energy, food, bank credit and the large number of government-controlled enterprises, which increase the budget deficit through borrowing from the Central Bank, have increased the monetary base. Money supply has become 10127 fold over the period 1959-2003 while real GNP recorded only a 10 fold increase during the same period. With such very high liquidity, any decision or news announced by the government or the Central Bank could severely change distribution of resources in the economy. In such circumstances, it matters for the Central Bank to know how private agents form their expectations. Moreover, optimal monetary policy depends considerably on the assumed nature of expectations formation process.

Empirical analyses on the formation of expectations can be divided into two categories: first, those studies that have been done by asking people about the future values of inflation (survey studies). Second, those studies that have tried to extract expectations from past data, on the assumption that people look to past experience as a guide to the future. This study followed the latter way.

The study found that the expectation hypothesis is accepted for the models under backward-looking expectations and learning approach. In other words, the expected inflation generated by the backward-looking expectations and learning approach are significant in the augmented Philips curve equation, and thus inflation expectations play a major role in the determination of the wages. It should be noted that the expectation hypothesis was rejected for rational expectations model.

Although the idea of rational expectations is attractive, it does not hold in the case of Iran. Since having access to the information is not apparently symmetric, inflation expectations can not be formed in a rational manner. The structure of the economy is unstable in ways that are imperfectly understood by both the public and policymakers and the policymakers' objective function is not completely known by private agents.

One interesting result was that the Hodrick-Prescott (HP) filtered series can be used as a proxy for rational expectations. Applying some rationality tests regarding unbiasedness and efficiency to the HP-filtered series, the results indicated that the filtered series is unbiased and efficient. Therefore, the filtered series is rational in the sense of Muth (1960).

This study compared two approaches to modeling inflation expectations: simple forecast and a multi-equation model. The results of simple statistical predictors revealed that the Neural Network model yields better estimates of inflationary expectations than do parametric autoregressive moving average (ARMA) and linear models. The agents were assumed to use a parametric autoregressive moving average (ARMA) model, proposed by Feige and Pearce (1976), or nonparametric models to form their expectations. Comparing to the nonparametric alternatives, the results of Wilcoxon tests demonstrated that the forecasting performance of Projection-Pursuit Regression and Additive models appeared to differ from the Neural Network model, implying that the Neural Network model can significantly outperform Projection-Pursuit Regression model and it has a better performance than Additive model, but not by much. However, there was no possibility that the Neural Network model can outperform the Multiple Adaptive Regression Splines model.

The results of estimated multi equation model indicated that the static expectations, adaptive expectations, optimal univariate expectations, a mix of extrapolative and regressive expectations with time-varying weights (and with discrete-choice updating weights) and learning approach are acceptable. Among near rational expectation schemes and the learning approach, the learning model was better suited for modeling inflation expectations than other alternative methods if the criteria, adjusted R-squared and the standardized expected inflation coefficient, are considered.

The hybrid New Keynesian Phillips Curve (NKPC), as an alternative to the augmented Phillips curve, was also considered. The results of estimated model indicated that the forward-looking term is dominant. However, this result should not be interpreted in favor of the forward-looking behavior. Such estimates can be obtained even if the true model is purely backward-looking (Rudd and Whelan, 2005). This situation may occur because of model misspecification and especially due to omitted variable bias. It should be noted that the New Keynesian models usually include the NKPC, the IS curve equation and the Taylor-type interest rate rule. However, since there is no data on interest rate in Iran after the 1979 Islamic revolution, we dropped the IS curve equation and the Taylor rule from our specification. After the 1979 profit rates, instead of interest rate, have been introduced but such rates are not compatible with pre-revolution's interest rate data.

Since the learning approach was better suited for modeling inflation expectations than other alternative methods, the Central Bank should be more aggressive towards inflation. The optimal monetary policy under a learning process should be more aggressive and narrowed to inflation stability (Orphanides and Williams, 2002). Furthermore, as any decrease in inflation is highly desirable and is one of the main macroeconomic goals, solidly anchored inflation expectations are suggested. To do so we need to keep monetary policy tight for a considerable period. However, it should be noted that conducting such a policy will also decrease output and employment.

The Central Bank should be independent so that it is able to adhere to an inflation target in practice. In this case, the monetary policy will be more credible so that it makes the private agents' expectations more responsive to signals from the Central Bank and the agents know what to expect following a set of published inflation targets.

According to the findings mentioned above, further research to design optimal monetary policy under adaptive learning is necessary. Furthermore, since the Neural Network model outperformed the linear, autoregressive moving average (ARMA), and nonparametric models (except MARS), there is a need for an empirical investigation on adaptive learning of rational expectations using Neural Networks. In this case, the question may arise whether the agents' expectations can converge to rational expectations with the help of Neural Networks.

Lists of Tables, Figures and Acronyms

Lists of Tables

Lists of Figures

List of Acronyms

AEH   **Adaptive Expectations Hypothesis**

RE   **Rational Expectations**

REE   **Rational Expectations Equilibrium**

REH   **Rational Expectations Hypothesis**

NKPC   **New Keynesian Phillips Curve**

OLS   **Ordinary Least Square**

2SLS   **Two-Stage Least Square**

ARIMA   **Autoregressive integrated moving average**

AR   **Autoregressive**

MA   **Moving average**

AD   **Additive Model**

PPR   **Projection-Pursuit Regression**

MARS   **Multiple Adaptive Regression Splines**

NN   **Neural Networks**

BP   **Backpropagation**

References

Ash JCK, .Easaw JZ, Heravi SM, Smyth DJ (2000), Are Hodrick-Prescott 'Forecasts' Rational?,

Attfield C.L.F, Demery D, and Duck N W (1991), Rational Expectations in Macroeconomics, 2nd edition, Blackwell Publishers

Bahmani-Oskooee, M (1995), Source of Inflation in Post-Revolutionary Iran, International Economic Journal, Vol. 9, No. 2, pp. 61-72.

Ball L (1991), The Genesis of Inflation and the Costs of Disinflation, Journal of Money, Credit, and Banking, Vol. 23, pp. 439-452

Banks, DL, Olszewski, RT, Maxion, RA (1999), Comparing Methods for Multivariate Nonparametric Regression, School of Computer Science, Carnegi Mellon University.

Barucci E and Landi L(1998), Nonlinear Versus Linear Devices: A Procedural Perspective, Computational Economics, Vol. 12, pp. 171-191

Bernanke B S (2003), A Perspective on Inflation Targeting, speech delivered at the Annual Washington Policy Conference of the National Association of Business Economists, Washington, D.C., March 25.

Bernanke B S (2003), Panel Discussion at the 28th Annual Policy Conference: Inflation Targeting: Prospects and Problems, Federal Reserve Bank of St. Louis, St. Louis, Missouri, October 17

Bonato L (2007), Money and Inflation in the Islamic Republic of Iran, IMF Working Paper WP/07/119 (Washington: International Monetary Fund).

Branson WH (1989), Macroeconomic Theory and Policy, 3rd edition, New York: Harper and Row Publishers

Bullard J (1991), Learning, rational expectations and policy: a summary of recent research, Federal Reserve Bank of St. Louis, pp. 50-60

Bullard, J (1992), Time-varying parameters and non-convergence to rational expectations under least square learning, Economic Letters vol. 40, issue 2, pp.159-166.

Caplan B (2000), Rational Expectations, Department of Economics, George Mason University

Chatfield C (2000), Time-Series Forecasting, Chapman & Hall/CRC

Chen P and Flaschel P (2006), Measuring the Interaction of Wage and Price Phillips Curves for the U.S. Economy, Studies in Nonlinear Dynamics & Econometrics, Vol. 10, Issue 4

Clarida, R, J Gali and M Gertler (1999), The science of monetary policy: a new Keynesian perspective, Journal of Economic Literature Vol.37, No.4, pp.1661-1707.

DeCanio, SJ (1979), Rational expectations and learning from experience, The Quarterly Journal of Economics, Vol. 93, No. 1, pp. 47-57

Diaconis P and Shahshahani M (1984), On nonlinear functions of linear combinations, SIAM Journal of Scientific and Statistical Computing, Vol. 5, No. 1, pp. 175-191

Evans, G and S Honkapohja (1999) Learning Dynamics in J B Taylor & M Woodford eds. Handbook of Macroeconomics, Elsevier

Evans, G and S Honkapohja (2001), Learning and expectations in macroeconomics, Princeton University Press, Princeton

Evans G and McGough B (2006); Implementing Optimal Monetary Policy in New-Keynesian Models with Inertia, University of Oregon, Economics Department, working papers, No. 2006-5.

Fan, J, Yao, Q (2003), Nonlinear Time Series: Nonparametric and Parametric Methods, Springer-Verlag: Berlin Heidelberg and New York.

Feige EL, Pearce DK (1976), Economically Rational Expectations: Are Innovations in the Rate of Inflation Independent of Innovations in Measures of Monetary and Fiscal Policy? Journal of Political Economy Vol. 84, No. 3, pp 499-522

Fischer S (2007), The Econometrics of Price Determination, Thirty-Five-Years Later, Journal of Money, Credit, and Banking, Supplement to Vol. 39, No. 1, pp. 171-179

Fox, J (2000), Multivariate Generalized Nonparametric Regression, Sage Publications.

Fox, J (2005), Introduction to Nonparametric Regression, McMaster University, Canada.

Friedman, JH, Stuetzle, W (1981), Projection Pursuit Regression, Journal of the American Statistical Association, Vol. 76, No. 376, pp. 817-823.

Friedman, B (1979), Optimal expectations and the extreme information assumptions of 'rational expectations' macromodels, Journal of Monetary Economics, Vol. 5, Issue 1, pp. 23-41

Friedman, HF (1991), Multivariate Adaptive Regression Splines, The Annals of Statistics, Vol. 19, No. 1, pp. 1-67.

Fuhrer, J C (1997), The (un)importance of forward looking behavior in price specifications, Journal of Credit Money and Banking, Vol. 29, No. 3, pp. 338-350 29.

Gali, J, and M Gertler (1999), Inflation Dynamics: A Structural Econometric Approach, Journal of Monetary Economics Vol. 44, No. 2, pp.195-222.

Galí J, Gertler M and D López-Salido (2001), European inflation dynamics, European Economic Review, Vol. 45, pp. 1237-1270.

Gleitman, H (1991), Psychology, W.W. Norton & Co., New York

Györfi, L, Kohler, M, Krzyzak, A, Walk, H (2002), A Distribution- Free Theory of Nonparametric Regression, Springer-verlag: Berlin and New York.

Hagan M T, Demuth H B, Beale M (1996), Neural Network Design, PWS Publishing Company

Hara N and Kamada K (1999), Yen/Dollar Exchange Rate Expectations in 1980-90's, Working Paper 99-1, Research and Statistics Department Bank of Japan

Harvey, AC (1991), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press.

Harvey, AC (1993), Time Series Models, 2nd edition, London, Harvester Wheatsheaf Publication.

Hastie, TJ, Tibshirani, RJ (1995), Generalized Additive Models, Chapman & Hall.

Hastie, TJ, Tibshirani, R, Friedman, J (2001), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer-Verlag: Berlin Heidelberg and New York.

Heij C, de Boer P, Franses P H, Kloek T, van Dijk H K (2004), Econometric Methods with Applications in Business and Economics, Oxford University Press

Heinemann M (2000), Adaptive Learning of Rational Expectations using Neural Networks, Journal of Economic Dynamic & Control, Vol. 24, pp 1007-1026

Lewiss K K (1988), The persistence of the 'Peso Problem' when policy is noisy, Journal of International Money and Finance, Vol. 7, pp. 5-21

Lewis PAW, Ray BK, and Stevens, JG (1994), Modeling Time Series by Using Multivariate Adaptive Regression Splines (MARS), Time Series Prediction: Forecasting the Future and Understanding(ed Neil Gershenfeld, NA, and Weigend, AS), pp. 297-318. Perseus Book Publication, L.L.C.

Lindé, J (2001a), Testing for the Lucas critique: a quantitative investigation, The American Economic Review, Vol. 91, No. 4, pp. 986-1005.

Lindé, J (2001b), The Empirical Relevance of Simple Forward- and Backward-looking Models: A View from a Dynamic General Equilibrium Model, Sveriges Riksbank Working Paper Series No. 130.

Liu O and Adedeji O. (2000), Determinants of Inflation in the Islamic Republic of Iran: A Macroeconomic Analysis, IMF Working Paper WP/00/127 (Washington: International Monetary Fund).

Mankiw N G (2001), The inexorable and mysterious tradeoff between inflation and unemployment, Economic Journal, Vol. 111, pp. 45-61

Mankiw N G (2007), Comments Presented at Federal Reserve Conference Price Dynamics: Three Open Questions, Journal of Money, Credit and Banking, Supplement to Vol. 39, No. 1, pp. 187-192

Martin J and Scarpetta S (1999), The levels and Cyclical Behavior of Mark-Ups Across Coutries and Market Structures, OECD Economics Department working papers, N0. 213

McCallum, B T (1976), Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates, Econometrica, Vol 44, No. 1, pp.43-52.

Minford P and Peel D (2002), Advanced Macroeconomics A Primer, Edward Elgar Publishing

Modigliani F and Sutch R (1966), Innovations in Interest Rate Policy, The American Economic Review, Vol. 56, No. 1/2, pp. 178-197

Muth J F (1961), Rational Expectations and the Theory of Price Movements, Econometrica, Vol. 29, Issue 3, pp. 315-335

Packalén M (1998), Adaptive Learning of Rational Expectations: A Neural Network Approach, University of Helsinki, Discussion Paper

Pesaran M H (1987), The Limits to Rational Expectations, Basil Blackwell

Orphanides A and Williams J C (2002), Imperfect knowledge, inflation expectations and monetary policy, Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 2002-27

Orr A, Edey M, and Kennedy M (1995), The Determination of Real Long Term Interest Rates: 17 Country Pooled-Time-Series Evodence. OECD Economics Department working papers, N0 155

Riddell WC, Smith PM (1982), Expected inflation and wage changes in Canada, Canadian Journal of Economics Vol. 15, No. 3, pp 377-394

Roberts, J M (1997), Is inflation sticky?, Journal of Monetary Economics Vol. 39, pp. 173-196

Roberts, J M (1998), Inflation expectations and the transmission of monetary policy, Board of Governors of the Federal Reserve, Finance and Economic Discussion Series, No. 1998-43

Roberts, J M (2001), How well does the new Keynesian sticky price model fit the data? Federal Reserve Board, FRB FEDS Discussion Paper No 2001-13.

Rudd J and Whelan K (2005), New Tests of the New-Keynesian Phillips Curve, Journal of Monetary Economics, Vol. 52, pp. 1167-1181.

Rudd J and Whelan K (2007), Modeling Inflation Dynamics: A Critical Review of Recent Research, Journal of Money, Credit, and Banking, Supplement to Vol. 39, No. 1, pp. 155-170

Rudebusch,, G D and L Svensson (1999), Policy Rules for Inflation Targeting, in John B. Taylor (ed.) Monetary Policy Rules, University of Chicago Press.

Salmon M (1995), Bounded rationality and learning: procedural learning, in Kirman A, Salmon M (eds.), Learning and rationality in Economics. Basil Blackwell, Oxford, pp. 236-275.

Sargent, T J. (1993), Bounded Rationality in Macroeconomics: The Arne Ryde Memorial Lectures, Oxford University Press

Sargent, T J. (1999), The Conquest of American Inflation, Princeton University Press.

Sent, E M (1998), The Evolving Rationality of Rational Expectations: An Assessment of Thomas Sargent's Achievements, Cambridge University Press

Staiger, D, James H S, and Mark W W (1997), How Precise Are Estimates of the Natural Rate of Unemployment? in Christina Romer and David Romer (eds.), Reducing Inflation: Motivation and Strategy, University of Chicago Press.

Stanley T D (2000), An Empirical Critique of the Lucas Critique, Journal of Socio-Economics, Vol. 29, pp. 91-107

Tesfatsion L (2004), Introduction to Rational Expectations, Iowa State University

Veaux, RD, Ungar, LH (1994), Multicollinearity: A tale of two nonparametric regressions, In Selecting Models from Data: AI and Statistics IV, (ed P.Cheeseman and R.W. Oldford), pp.293-302, Springer-Verlag.

Venables W N and Ripley B D (2002), Modern Applied Statistics with S, Springer-verlag: Berlin and New York.

Waters G (2007), Regime changes, learning and monetary policy, Journal of Macroeconomics, Vol. 29, pp.255-282

Westerhoff F H (2006), Samuelson's multiplier–accelerator model revisited, Vol. 13, pp. 89–92

Zilouchian A (2001), Fundamentals of Neural Networks, in Zilouchian A and Jamshidi M (ed.) Intelligent Control Systems Using Soft Computing Methodologies, CRC Press LLC

Appendices

Appendix I: Data source and definitions

The data are annually for the period 1959-2003 and are collected from the Central Bank of Iran.

W = index of the wage of construction workers (1997=100)

U = unemployment rate

y = real GNP (at the constant 1997 prices)

P = GNP deflator (1997=100)

M_2 = M_1 (currency +demand deposit) + quasi money

g = real government consumption expenditure (at the constant 1997 prices)

c = real private consumption expenditure (at the constant 1997 prices)

P^m = import price index (1997=100)

X = labor productivity ($\frac{\text{real GNP}}{\text{total empolyment}}$)

Appendix II: Effects of omitted variable bias

Assume that the true model is a back-ward looking Phillips curve of the form

$$\pi_t = \beta \pi_{t-1} + \lambda x_t + \mu z_t + u_t \tag{1}$$

where x_t is output gap and z_t denotes a vector of additional determinants of inflation. Suppose λ and μ are positive. Now we fit the following equation using GMM and the instruments which includes z_t

$$\pi_t = w^f E_t \pi_{t+1} + w^b \pi_{t-1} + \gamma x_t \tag{2}$$

GMM and two-stage least squares are equivalent in a linear model. Therefore, in the first-stage regression we obtain the fitted values of π_{t+1} on π_{t-1}, x_t and z_t as

$$\hat{\pi}_{t+1} = \hat{\delta}_1 \pi_{t-1} + \hat{\delta}_2 x_t + \hat{\delta}_2 z_t \tag{3}$$

which $\hat{\pi}_{t+1}$ is a proxy for $E_t \pi_{t+1}$. Then $\hat{\pi}_{t+1}$ is used in a second-stage regression as

$$\pi_t = \hat{w}^f \hat{\pi}_{t+1} + \hat{w}^b \pi_{t-1} + \hat{\gamma} x_t + \varepsilon_t \tag{4}$$

Plugging equation (3) into (4) and rewriting

$$\pi_t = (\hat{w}^f \hat{\delta}_1 + \hat{w}^b) \pi_{t-1} + (\hat{w}^f \hat{\delta}_2 + \hat{\gamma}) x_t + \hat{w}^f \hat{\delta}_3 z_t + \varepsilon_t \tag{5}$$

Comparing equation (5) with the true model (1), we can obtain the following asymptotic properties

$$p\lim(\hat{w}^f \hat{\delta}_1 + \hat{w}^b) = \beta, \tag{6}$$

$$p\lim(\hat{w}^f \hat{\delta}_2 + \hat{\gamma}) = \lambda, \tag{7}$$

$$p\lim \hat{w}^f \hat{\delta}_3 = \mu. \tag{8}$$

Since inflation is highly autocorrelated, it is likely the coefficients $\hat{\delta}_1$, $\hat{\delta}_2$, and $\hat{\delta}_3$ from equation (3) will typically have the same sign as their respective coefficients β, λ, and μ from the true model (1). As a result, it is clear that the estimated value of $\hat{w}^f$ will be positive, even if the true model does not include a forward-looking term. Furthermore, the estimated coefficients $\hat{w}^b$ and $\hat{\gamma}$ will be biased upward, as compared to the true coefficients λ and μ, because the effect of π_{t-1} and x_t on π_t is already partly captured by $\hat{\pi}_{t+1}$. In case this term receives a positive sign, then π_{t-1} and x_t will be crowded out of the second-stage regression.