Applications of Evolutionary Learning in Macroeconomic Models

by

Michael Kurtis Maschek

B.A.(Hon), Simon Fraser University, 1999.

M.A., University of British Columbia, 2000.

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department
of
Economics

SIMON FRASER UNIVERSITY
Spring 2005

APPROVAL

Name: **Michael Maschek**

Degree: **PhD (Economics)**

Title of Thesis **Applications of Evolutionary Learning in Macroeconomic Models**

Examining Committee:

Chair: Gordon Myers

Jasmina Arifovic
Senior Supervisor

Ken Kasa
Supervisor

Alex Karaivanov
Supervisor

Jenny Xu
Internal Examiner

Dr. James Bullard

External Examiner

Date Approved: Friday, April 8, 2005

Abstract

Genetic Algorithms are the best known representation of a class of direct random search methods called evolutionary algorithms which are widely used to solve complex optimization and adaptation problems. They have grown in popularity within economics due to their ability to represent the adaptation of individuals to the underlying parameters of their economic system. This work examines three applications of genetic algorithm adaptation in macroeconomic environments.

In the first of these applications, the Arifovic and Masson (2003) model of currency crisis is simulated in controlled laboratory experiments with human subjects. An extended model of agents expectations is considered in which each investor has multiple rules, choosing one of them probabilistically in each period. The properties of time series generated by computer simulations are compared to those of human data. In each framework the time series of returns on emerging market debt is characterized by fat tails which matches features of empirical data. Additionally, the extended model of expectations better matches the duration statistics found in the experimental setting.

The second application investigates the sufficiency of learning-by-doing for explaining negative macroeconomic output shocks in an evolutionary model of technological transition. The model allows firms to divide labour between two distinct technologies in a continuous manner. The ability of each firm to innovate within each technology is dependent on this choice for the division of labour. Contrary to previous literature, innovations are not transferable between technologies. It is argued that in such a framework learning-by-doing remains sufficient for periodic observations of negative macroeconomic growth.

The final examination represents the first application of *two-level learning* in an economic environment in which the performance of potential rules is complementary across individuals. *Two-level learning*, or *self-adaptation*, incorporates certain strategy parameters into the representation of each individual. In this work, these strategy parameters determine the level of heterogeneity introduced into the environment. They evolve by means of mutation and recombination, just as the object variables do. It is argued that self-adaptation over these parameters can replace the election operator proposed by Arifovic (1994) in order to attain convergence to a rational expectations equilibrium.

Dedication

To my parents, Mike and Thyiela, and fiancé, Hayley.

Acknowledgements

Without the encouragement and insight inherent in the supervision by Dr. Jasmina Arifovic, this work would not have been possible. In addition to frequently displaying generous confidence in my ability, her willingness to provide direction and intuition helped foster an environment that culminated in a work I am truly proud of.

I would like to thank Dr. Ken Kasa, Dr. Alex Karaivanov, Dr. Jenny Xu, and Dr. James Bullard for their perspicacious contributions to the examining committee.

In addition to thanking all members of the Department of Economics, I would be remiss if I did not specifically thank Dr. Clyde Reed, Dr. Steeve Mongrain, Dr. Philip Curry, Dr. Gord Myers, and Dr. Ken Kasa for their encouragement and advice throughout the duration of my work at Simon Fraser University. I am indebted to Dr. David Andolfatto and Dr. Lawrence Bolland for their unique offering of penetrating criticism coupled with support, advice and insight.

I would like to thank the entire support staff within the department, especially Tracey Sherwood, Barb Clark and Sherrill King. On countless occasions, you went well beyond what was required in an effort to make my life a little easier. I am incredibly thankful.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> We use economic theory to calculate how certain variations in the situation are predicted to affect behavior, but these calculations obviously do not reflect or usefully model the adaptive process by which subjects have themselves arrived at the decision rules they use. Technically, I think of economics as studying decision rules that are steady states of some adaptive process, decision rules that are found to work over a range of situations and hence are no longer revised appreciably as more experience accumulates. (Lucas, 1986)

The rational expectations hypothesis asserts that economic outcomes do not differ systematically from what economic agents expect them to be. Rational agents form these expectations by optimally using all information pertaining to the expected outcome available. The rational expectations assumption has been used with great success in an enormous number of economic problems. It has proved exceptionally powerful in solving economic models and developing comparative statics pertaining to important economic relationships. However, its use encompasses assumptions that prove very demanding and unrealistic in many applications. Alternatively, the assumption of bounded rationality maintains that economic agents form expectations in a manner that is only as optimal as information, resource and cognitive constraints will allow. It is argued that bounded rationality better describes the actual behavior of economic agents. In reality, individuals do not always behave in a truly optimal manner. Furthermore, individuals must discover the manner in which information is represented and the strategies for forming expectations using this information (Simon, 1957).

Though work utilizing the rational expectations assumption typically outlines the comparative statics of the economic system in question, it devotes rather little effort to explaining the process of moving from one equilibrium to another. Utilizing the less demanding assumption of bounded rationality, many economists have turned their attention to the behavior of agents when the system in question is out of equilibrium. Models that result from this approach work to explain how equilibrium behavior can emerge as the limit of an adaptation or learning process of boundedly rational agents (Dawid (1996), Lucas (1986)). If equilibrium is to result from the limiting nature of an adaptation process, this work must begin, of course, with a model of this adaptation.

Some of these models assume agents behave as if econometricians, forming expectations according to econometric methods and based on historical observation. Examples of such work include those with expectation updating based on a simple moving average (Lucas (1986)), least squares estimation (Sims (1980), Marcet and Sargent (1989)), stochastic approximation (Robbins and Monro (1951), Woodford (1990)), or Bayesian learning (Blume and Easly (1982), Turnovsky (1969)).

This work examines the application of a particular set of adaption models. Originally developed by Holland (1975), *genetic algorithms* are the best known representation of a class of direct random search methods called *evolutionary algorithms* which are widely used to solve complex optimization and adaptation problems. Their use within economics is grounded on their ability to represent the adaptation of individuals to the underlying parameters of their economic environment.

Based on the principles and mechanics of natural evolution, genetic algorithms are a set of search algorithms in which a population of potential solutions is encoded and subject to three basic genetic operators - *selection*, *crossover*, and *mutation*. These potential solutions may be candidates for an optimization problem or - as in economic systems - the belief or decision rules of an agent regarding an economic problem. The algorithms combine the principle of "survival of the fittest" with a "structured yet randomized information exchange to form a search algorithm with some of the innovative flair of human search." (Goldberg, 1989). The initial population of rules is often randomly generated. Referred to as *generations*, subsequent populations are created by taking the best performing candidates from the old generation (*selection*) and exploiting these rules to introduce new solution candidates (*crossover* and *mutation*).

The *selection* operator determines which candidates from the population of rules will be used to create the subsequent population. It is through application of this rule that the idea of "survival of the fittest" takes form. The performance of each specific rule is determined according to a *fitness function*. *Fitness* measures the performance of each rule with respect to the surrounding system and facilitates comparison between competing rules within the population. The standard selection operator, referred to as *proportional selection*, creates a new generation of potential rules by randomly drawing (probabilistically) from the old population. The likelihood of each rule being selected for use in the subsequent population is proportional to its level of fitness. Once a rule is selected, it is replaced in the population of potential rules and has the opportunity of being randomly selected again. Therein, a rule that has twice the fitness of another candidate has twice the likelihood of being selected for the subsequent population. Over repeated draws on the current population it is expected to generate twice as many copies in the new generation.

The replication inherent in the selection process represents "imitation of the successful" (Dawid, 1996). It is analogous to the imitation effect within a population. In a population where the payoffs of the individual actions are known to other members of the population, it is very plausible to assume that low payoff individuals will imitate the actions of those individuals employing highly profitable rules.

The *crossover* and *mutation* operators generate new rules in the population; they introduce and maintain diversity over this population. Following application of the selection process, *crossover* works by randomly assigning each rule to a pair. The crossover operator is applied to each pair according to a given probability. If crossover occurs, a portion of the genetic material encoding each rule in a pair is swapped, yielding two new rules distinct from the pair that created them. This operator is often interpreted as analogous to some process of communication in which information is exchanged. Members of the population exchange information regarding their planned action. Some agents utilize this information, adopting part of the strategy of the other agent in combination with their own plan. The *mutation* operator is utilized following crossover. The genetic material encoding each rule is randomly altered from its current state according to a given probability. This operator incorporates the effect of purposeful innovations, or trembling hand mistakes.

Though the selection operator reduces the level of diversity over the population of rules comprising a generation, the crossover and mutation operators introduce diversity following this replication process. Of course, only newly created rules that outperform relative to the rest of the population will be replicated under the selection process in subsequent generations.

Arifovic (2000) notes that these models of adaptation hold several advantages over competing approaches. First, they allow for a population of heterogeneous agents' beliefs. Each belief's propagation is dependent on performance as measured by the payoff an agent receives in holding that belief. Modelling economic agents utilizing these algorithms imposes low requirements pertaining to their computational ability. In terms of their ability to explain the observations of actual empirical outcomes and experimental economics, they perform better than models with rational agents or alternative models of adaptive behavior.

Arifovic classifies the research questions addressed by the study of these algorithms into four different categories. The first category contains research related to the convergence and stability of equilibria in models with unique rational expectations equilibria. Many examples of this first category consider the application of genetic adaptation to the cobweb model. The original rational expectations consideration of this model is attributed to Muth (1961). Arifovic (1994) considers the learnability and stability of this equilibrium in a model of genetic algorithm adaptation. She compares these results to those of other learning algorithms and experimental evidence. This work demonstrates that genetic learning provides a better approximation of experimental results than other models of adaptation.[1]

[1] Arifovic (1994) compares the results of genetic learning to those of three other commonly used learning algorithms: cobweb expectations, sample average of past prices, and least squares learning.

The second category contains research where genetic algorithms are used as an equilibrium selection device in models with multiple equilibria. Examples of such work include the examinations of an overlapping generations framework with fiat money (Arifovic (1995), Bullard and Duffy (1998a)). Here, the government finances a constant deficit through seignorage. The model has two stationary equilibria, one associated with a lower level of inflation than the other. An important characteristic of these frameworks is their convergence to the low stationary inflation equilibrium for deficit values and initial conditions under which least squares learning exhibited divergent behavior.[2]

The third category includes work that examines transitional dynamics that accompanies the equilibria selection process. For example, Bullard and Duffy (1998c) examine the application of genetic algorithm learning in the Grandmont (1985) environment. They consider a two period overlapping generations model with preferences dependent on the coefficient of relative risk aversion. In addition to two steady states, certain parameter choices are associated with periodic and chaotic equilibrium trajectories. The work demonstrates qualitatively complicated dynamics for long periods of time prior to convergence.

The final category contains examinations of learning dynamics that are intrinsically different from the dynamics of rational expectation considerations. An example of such work is found in the modelling of recurrent currency crises. Rational expectations solutions to the problem are often dependent on the existence of sunspot equilibria (Cole and Kehoe (1996), Jeanne and Masson (2000)). These solutions, however, do not explain how investors coordinate on a currency crisis path. Arifovic and Masson (2003) describe an evolutionary model that results in recurrent episodes of currency crisis that are driven solely by changes in investors beliefs; periods of excessive optimism are followed by periods of excessive pessimism. Currency crises characterized by recurrent periods of devaluations are purely expectationally driven. The model yields predictions regarding the behavior of the distribution of beliefs that are linked to recurrent devaluations.

This dissertation contains three distinct applications of genetic algorithm learning. Each work contributes to the literature within one of the categories described above. In the chapters that follow this introduction, the second and third examine learning dynamics that are intrinsically different from the dynamics of the rational expectations versions of the models.

In Chapter 2, the Arifovic and Masson (2003) model of currency crisis is simulated in laboratory experiments with human subjects. The framework modelling agents' expectations utilized by Arifovic and Masson is extended. In this model, each investor has multiple rules and in each period chooses one of them probabilistically. The properties of time series generated by computer simulations are compared to those of the experimental data and the original results of Arifovic and Masson. Both the simulations and experiments generate times series of returns on emerging debt whose distributions

[2] In an extension of this work, Bullard and Duffy (1998b) model agents which live for more than two periods. They show that for larger parameter choices for the agents' length of life, convergence to the low inflation stationary equilibrium becomes less likely. In this extension, a similar result holds for larger specifications of the government deficit.

are characterized by fat tails. This feature matches the empirical data.

Chapter 3 develops an evolutionary model of technological transition in order to investigate the sufficiency of learning-by-doing for explaining negative macroeconomic output shocks. Contrary to previous literature, innovations are not transferable between technologies. The assumption maintained within this work is that the only manner in which a firm may learn about a newer grade of technology is to devote resources toward production within it. However, firms are not required to fully commit to a single technology at any given point in time. Firms may divide labour between two technological paradigms in a continuous manner. The ability of each firm to innovate within each paradigm is dependent on it's choice for the division of labour. Productivity gaps between old and new technologies result from a lack of accrued incremental innovations in the newer technologies. In all simulations, these gaps result in periods of negative economic growth in real income per capita despite the ability of firms to adopt these new technologies in a non-discrete manner. For selective parameterizations of the model, although average quarterly growth statistics match the data well, periods of negative economic growth occur with less frequency than observed in reality. It is shown that in such a framework learning-by-doing remains sufficient for periodic observations of negative macroeconomic growth, though not with enough frequency to match actual data.

The fourth and final chapter of this work contains research related to the convergence and stability of equilibria in the models with unique rational expectations equilibria. Here we extend the work of Arifovic (1994) described in the exposition of this category.

The work contained in Chapter 4 represents the first application of *two-level learning* in genetic algorithms in an economic environment in which the fitness value of potential rules are complementary across individuals. *Two-level learning*, or *self-adaptation*, incorporates certain strategy parameters into the representation of each individual. In this work, these strategy parameters provide the likelihood of mutation for the individual. These strategy parameters evolve by means of mutation and recombination, just as the object variables do. It is argued that self-adaptation over the parameter governing mutation can replace the election operator proposed by Arifovic (1994) in order to attain convergence to a rational expectations equilibrium. While both adaptive mutation and the election operator are sufficient for convergence, self-adaptation may be more appropriate in non-static environments. This convergence, however, will require a strong selective pressure only attained through a transformation of the baseline fitness function.

References

Arifovic, J. (1994), "Genetic Algorithm Learning and the Cobweb Model," *Journal of Economic Dynamics and Control*, 18, 3-28.

Arifovic, J. (1995), "Genetic Algorithms and Inflationary Economies," *Journal of Monetary Economics*, 36, 219-243.

Arifovic, J. (2000), "Evolutionary Algorithms in Macroeconomic Models," *Macroeconomic Dynamics*, 4, 373-414.

Arifovic, J. and Masson, P. (2003). "Heterogeneity and Evolution of Expectations in a Model of Currency Crisis," *Nonlinear Dynamics, Psychology and Life Sciences*, forthcoming.

Blume, L.E. and Easley, D. (1982), "Learning to Be Rational," *Journal of Economic Theory*, 26, 340-351.

Bullard, J. and Duffy, J. (1998a), "A Model of Learning and Emulation with Artificial Adaptive Agents," *Journal of Economic Dynamics and Control*, 22, 179-207.

Bullard, J. and Duffy, J. (1998b), "Using Genetic Algorithms to Model the Evolution of Heterogenous Beliefs," *Computational Economics*, 1-20.

Bullard, J. and Duffy, J. (1998c), "Learning and the Stability of Cycles," *Macroeconomic Dynamics*, 2, 22-48.

Cole, H. and Kehoe, T. (1996), "A Self-Fulfilling Model of Mexico's 1994-95 Debt Crisis," *Journal of International Economics*, 41, 309-330.

Dawid, H. (1999), **Adaptive Learning by Genetic Algorithms, Analytical Results and Applications to Economic Models**, 2nd revised and extended edition, Springer-Verlag, Berlin.

Goldberg, D. (1989), **Genetic Algorithms in Search, Optimization and Machine Learning**, Addison-Wesley Publishing Company, New York.

Grandmont, J-M. (1985). "On Endogenous Competitive Business Cycles," *Econometrica*, 53, 995-1045.

Holland, J.H. (1975), **Adaptation in Natural and Artificial Systems**, The University of Michigan Press, Ann Arbor, MI.

Jeanne, O., and P. Masson (2000). "Currency Crisis, Sunspots, and Markov-Switching Regimes," *Journal of International Economics* 50:327-350.

Lucas, R.E. (1986), "Adaptive Behavior and Economic Theory," *Journal of Business*, 59, 401-426.

Marcet, A. and Sargent, T.J. (1989), "Convergence of Least Squares Learning Mechanisms in Self Referential Linear Stochastic Models," *Journal of Economic Theory*, 48, 337-368.

Muth, J. (1961), "Rational Expectations and the Theory of Price Movements," *Econometrica*, 29, 315-335.

Robbins, J. and Monro, S. (1951), "A Stochastic Approximation Method," *Annals of Mathematical Statistics*, 22, 400-407.

Simon, H. (1957), **Models of Man**, Wiley, New York.

Sims, C.A. (1980), "Macroeconomics and Reality," *Econometrica*, 69, 1-48.

Turnovsky, S.J. (1969), "A Bayesian Approach to the Theory of Expectations," *Journal of Economic Theory*, 1, 220-227.

Woodford, M. (1990), "Learning to Believe in Sunspots," *Econometrica*, 58, 277-307.

Chapter 2

Currency Crises

Evolution of Beliefs, Experiments with Human Subjects and Real World Data[1]

2.1 Introduction

The role of investors' expectations has always been emphasized as a very important factor affecting the behavior observed in the financial markets. In particular, conventional accounts of the episodes of currency crisis focus on changes and shifts in investors' beliefs. However, modeling the changes in investors' expectations that might trigger currency crisis, without any apparent change in economic fundamentals, has not been given much attention in the existing literature.

The traditional rational expectations approach leaves little room for modeling endogenous changes in investors' expectations that would trigger recurrent speculative attacks on currency.[2] The exception are the models that due to the features of the underlying fundamentals exhibit multiple (static) equilibria where it is usually possible to add an exogenous sunspot process that governs switches between the neighborhoods of these equilibria. As a result, sunspot models generate dynamics of the recurrent currency crises.[3] However, they require coordination of investors' beliefs on a particular sunspot process falling short of explanation of why and how this coordination might take place.

Over the last few years, advances have been made with the models that depart from the rational expectations hypothesis, and instead assume that investors are boundedly rational agents who have

[1]This chapter is based on a work co-written with Dr. Jasmina Arifovic, Simon Fraser University. Jasmina Arifovic acknowledges the support of the SSHRC Standard Research Grant Program. We would like to thank Dan Friedman, Rob Oxoby, as well as the participants at the CEF Meetings in Seattle (2003), ESA Meetings in Tucson (2003), and the Experimental Workshop held at the University of Calgary (October 2004) for helpful comments.

[2]Models that incorporate imperfect and asymmetric information can give rise to one-time speculative attacks, but cannot generate recurrent currency crises.

[3]See, for example, Cole and Kehoe (1996, 2000), Jeanne and Masson (2000).

to learn and adapt over time. Kasa (2001) introduces adaptive learning into Obstfeld's (1997) 'escape clause' model and shows that learning dynamics, rather than sunspots, can generate switches between multiple steady states. Cho and Kasa (2003) introduce learning into a model of Aghion, Bacchetta and Banerjee (2001). Even when equilibrium is unique in this model, they show that the 'escape dynamics' of the learning algorithm produce the kind of Markov-Switching exchange rate behavior that is typically attributed to sunspots. Both of these studies assume homogeneity of investors' beliefs.

Arifovic and Masson (2003) take a different approach and study a dynamic model of currency crisis in which heterogenous expectations of boundedly rational agents evolve through a very simple algorithm that involves imitation and experimentation. Their model generates recurrent crises that result from investors' change in expectations; periods of excessive optimism are followed by periods of excessive pessimism. Currency crises characterized by recurrent periods of devaluations are purely expectationally driven. The model also yields some predictions about the behavior of distributions of beliefs over time (that in fact are linked to recurrent devaluations). Direct empirical tests of these predictions cannot be done as we do not have any data concerning the behavior of investors' beliefs in real markets.

Arifovic's and Masson's model is based on the idea of *social learning* where a population of beliefs of a large number of agents evolves together over time. This concept captures well the fact that a large number of investors participate in trading in real markets. Investors in real markets can also observe the behavior of some of the other investors (captured well by imitation).

We extend Arifovic's and Masson's original framework by using a model (see Arifovic and Ledyard, 2003) where each investor has a collection of alternative beliefs and chooses one of them probabilistically. (The evolution of beliefs takes place at the level of an individual.) In addition to being interested in the robustness of the dynamics with respect to two different learning paradigms, we employ a model of individual learning as it is better suited for direct mapping into the design of the experiments with human subjects.

We simulate both models of social and individual learning for a large number of different parameter values, and examine the observed dynamics. It is noteworthy that the model of individual learning is also characterized by recurrent currency crises. Other features such as duration of periods of devaluation and no-devaluation and the characteristics of the times series of the models' variables that are generated vary across different types of simulations.

As the appropriate data regarding investors' beliefs is not available, the approach we take in this paper is to test the model's predictions in simulations with the data collected in the experiments with human subjects. This way we can directly observe the evolution of investors' beliefs over time and compare the properties of the distributions generated in a model and those that result from the experiments with human subjects. The observed experimental behavior matches well the behavior of the boundedly rational, artificial agents along many dimensions. Most importantly, experiments do result in recurrent instances of currency crises. We also examine the time series properties of the returns, both those generated by our model and those collected in the experiments. Both time series are characterized by 'fat tails' which is the feature observed in the real data on returns from the emerging markets (see Masson, 2003).

In section 2.2, we first describe a simple balance of payments model with a representative agent and characterize its rational expectations equilibrium. This description is followed by an introduction of a model in which agents have heterogenous beliefs. We present our two models of learning, social and individual, in section 2.3. We describe our simulation and experimental design in section 2.4. The results of simulations are presented in section 2.5. The analysis of the results of the experiments with human subjects and the features of the dynamics of the changes in expectations are discussed in section 2.6. Finally, concluding remarks are given in section 2.7.

2.2 A Model of Currency Crises

2.2.1 Representative agent model

We follow Arifovic and Masson (2003) in describing a simple model of a portfolio allocation between mature and emerging markets in which risk neutral investors decide to put their wealth either in an emerging market country or the United States. An emerging market central bank defends a currency peg using its foreign exchange reserves until those reserves reach some minimum value.

The U.S. asset is riskless, and pays a known rate r^*, while the emerging market asset's return, r_t, is subject to devaluation (or default) risk as well as potentially decreasing returns to the amount invested. The agent puts a fraction λ_t of her fixed wealth $\overline{W}$ in emerging market assets, such that expected returns on the two assets are equalized.

Making explicit the dependence of r_t on λ_t, letting π_t be the probability of a devaluation and δ_t the size of devaluation, the condition for portfolio equilibrium is[4]

$$r^* + \pi_t \delta^e = r_t = r(\lambda_t) \tag{2.1}$$

Inverting (2.1), we can write this dependence as

$$\lambda_t = \lambda(\pi_t) \tag{2.2}$$

As in the canonical currency crisis model (Krugman, 1979), devaluations are triggered by the decline of reserves to some threshold level, which we assume to be zero. The change in reserves is equal to the capital inflow plus the trade balance, minus the interest payments on outstanding debt:

[4] For convenience, cross product terms are ignored here.

$$R_t = R_{t-1} + T_t + D_t - D_{t-1} - r_{t-1}D_{t-1} \tag{2.3}$$

where $D_t = \lambda_t \overline{W}$. The trade balance T_t is stochastic and is assumed to follow a Markov process; that is, it depends only on its lagged value.

A rational expectation for the devaluation probability will satisfy

$$\pi_t = \Pr_t(R_{t+1} < 0 | no\ devaluation) \tag{2.4}$$

This probability can be rewritten

$$\pi_t = \Pr_t(R_t + T_{t+1} + \lambda(\pi_{t+1})\overline{W} - (1 + r^* + \pi_t\delta_t^e)\lambda(\pi_t)\overline{W} < 0) \tag{2.5}$$

Assuming that the reserve level R_t is part of the representative agent's information set, and using the notation in Jeanne and Masson (2000), we can write this as

$$\pi_t = \Pr_t(B(T_{t+1}, \pi_{t+1}, \pi_t) < 0 | T_t, R_t) \tag{2.6}$$

This latter equation determines the rational expectation for the devaluation probability, given the stochastic process for T_t.

The dynamics of (2.6) are difficult to characterize. However, it is shown in Jeanne and Masson that a simplified version of equation (2.6) can have multiple solutions. In particular, in the simplified case where B does not depend on π_t (just on π_{t+1})[5], nor on R_t, and if transitions between equilibria are described by a Markov transition matrix, then there is an unlimited number of rational expectations solutions. In particular, for any set of n equilibria, another rational expectations equilibrium can also be constructed.

2.2.2 A Simplified Model

Arifovic and Masson (2003) have shown that the model of social learning in which heterogenous beliefs about π_t, and δ_t that evolve over time results in recurrent currency crisis. In order to test the robustness of their model, they also examined the behavior of a simplified model in which only beliefs about π_t evolved, and the belief about δ_t was kept at the constant level. This model resulted in the same type of dynamics. Finally, a further simplification in which there is no stochastic element of the trade balance (resulting in $T_t = T_{t+1}$ for all t) did not affect the qualitative features of the dynamics. As the main objective of this paper is to compare the results of the simulations with

[5]The case where B depends on both π_{t+1} and π_t can generate chaotic dynamics, as shown in Jeanne and Masson (2000).

the experimental data, we will work with this simplified model because it lends itself better to the experimental implementation.

Thus, we abstract from an evolving trade balance to one in which T_t equals zero for all periods. In addition, we assume that all individuals share the same expectation regarding the size of devaluation. Specifically, $\delta^e_{i,t} = \delta^e = 1$ for all i and over all periods t. In this simplified model, equilibrium is no longer characterized by an infinite number of solutions. (The inclusion of a non-stochastic trade balance will instead decrease the number of rational expectation solutions to just two.)

Reserve levels are determined identically to the specification in equation (2.3), setting T_t equal to zero for all t.[6] The rational expectation solution for an individual's probability assessment is therefore still characterized by equation (2.5). We make the following assumption for the function $\lambda_t = \lambda(\pi_t)$

$$\lambda'(\pi_t) < 0 \tag{2.7}$$

ensuring that as individuals become more pessimistic, their investment in the emerging market decreases (ceteris paribus). We also assume $\lambda(0) = 1$ and $\lambda(\pi_{max}) = 0$. Under these simplifying assumptions, the rational expectations solution for π_t (equation (2.5)) therefore becomes

$$\pi_t = Pr_t(R_t + \lambda(\pi_t)\overline{W} - (1 + r^* + \pi_{t-1})\lambda(\pi_{t-1})\overline{W} < 0 | no \quad devaluation) \tag{2.8}$$

In any situation in which $R_t > (1+r^*+\pi_{t-1})\lambda(\pi_{t-1})\overline{W}$ holds, the solution to this assessment has a unique solution. Specifically, $\pi_t = 0$. Here, even as no funds are invested in the emerging market, it is impossible for a devaluation to occur. The reserve level of the emerging market's central bank is sufficient to cover *all* of its economy's current debt.

A unique solution also results in any situation in which it is impossible to meet a shortfall in reserves with incoming emerging market investment. That is, when $(1 + r^* + \pi_{t-1})\lambda(\pi_{t-1})\overline{W} - R_t > \overline{W} > 0$ holds, a devaluation is certain, and $\pi_t = \pi_{max}$ is the unique solution.

Multiple solutions exist for situations that fall between these two extremes. That is, when incoming emerging market investment can meet reserve shortfalls, or when $\overline{W} > (1+r^*+\pi_{t-1})\lambda(\pi_{t-1})\overline{W} - R_t > 0$ holds, there are two possible solutions for π_t: $\pi_t = 0$ and $\pi_t = \pi_{max}$. It is impossible, without further specification, to select one of these solutions over the other. When π_t takes the value of π_{max}, a self-fulfilling devaluation takes place in which $\lambda(\pi_t = \pi_{max}) = 0$ and through a devaluation of currency, $R_{t+1} = 0$.[7]

[6] Setting T_t equal to $\overline{T}$ rather than zero does not change the solutions' characterization in any significant manner

[7] This result is in essence a stag-hunt game with a payoff dominated equilibrium. In a model incorporating Bayesian learning, Chamely (2003) considers speculative attacks in a similar spirit. Agents update their expectations regarding the number of other agents that believe the current fundamentals are sufficient for a successful attack. Essentially, there are two states of the economy, one in which there is sufficient speculators for devaluation, and one in which there is not. The mass of these speculators is an uncertain parameter of this economy. While both models are essentially a

In the period following this devaluation, the above problem simplifies to the following

$$\pi_t = Pr_t(R_{t+1} < 0) = Pr_t(\lambda(\pi_t)\overline{W} < 0) \tag{2.9}$$

As is the nature of self-fulfilling phenomena, when investors do not expect a devaluation, that is, when π_t takes the value of 0, a devaluation does not take place. Importantly, this cannot occur indefinitely, as interest payments on emerging market debt will slowly diminish the level of reserves available. Eventually, the economy will find itself with too few reserves to cover its interest outflow and a devaluation occurs.

All of the above analysis is based on a framework where a one-period model (stage game) is repeated over time. In this respect, agents really have expectations of probability of devaluation in the following period. However, if we assumed investors were forward looking, then their rationality will imply the logic of backward induction, i.e. in case that devaluation can occur in some period t, no investment in the emerging market will ever occur.

2.2.3 Heterogeneous agents

We now turn to the model with heterogeneous agents. There are n investors, each with constant wealth $\overline{W}$, who form expectations of the devaluation probability, π_t^i.[8] Since investors are risk neutral, they will be indifferent between investing in the two assets when their *ex ante* returns are equal, and choose between putting all their beginning-of-period wealth into the safe foreign asset, at rate r^*, or into emerging market claims, at rate r_t, depending on which expected return is greater.

We assume that each investor is a price taker, and does not influence the marginal product of capital in the emerging market economy. Short selling of either asset is ruled out; neither portfolio proportion can be negative.[9] If λ_t^i is the share of i's wealth in emerging market debt, then $\lambda_t^i = 0$ or 1 as $(1 + r^*) >$ or $< (1 + r_t)/(1 + \pi_t^i)$.[10] Thus, at any period t, the amount of emerging market deposits held by all foreign investors is

$$D_t = \sum_{i=1}^{n} \lambda_t^i \overline{W}. \tag{2.10}$$

Emerging market banks set the interest rate on bank deposits to reflect market expectations of the return on emerging market debt. We assume that banks do not form expectations of devaluation

game of timing, in Chamely's work multiple periods are necessary for the existence of speculative attacks and these attacks are not recurrent. However, the emphasis of Chamley's work is examining policies' ability to defend the currency peg, not in explaining recurrence.

[8] We continue to assume that each investor has an identical expectation regarding the devaluation size and that this expectation does not change over time, $\delta_t^{e,i} = \delta^e = 1$.

[9] Similar qualitative results can be obtained if borrowing is allowed, but there are limits on leverage (such as a minimum capital requirement).

[10] If the US rate were equal to the gross expected emerging market return discounted by the expected devaluation, λ_t^i would be indeterminate.

themselves; they just use the average of all investors' expectations as a measure of the expected value of devaluation. Thus, the interest rate on emerging market deposits r_t is set equal to the U.S. rate plus a weighted average of the expected rate of devaluation. This equation, which is analogous to an interest parity (no arbitrage) condition, can be written

$$r_t = (1 + r^*) \prod_{i=1}^{n} (1 + \pi_t^i)^{1/n} - 1. \tag{2.11}$$

With different expectations, expected returns will be equalized only for the marginal investor whose expectation equals the average expectation. Each individual investor will make her investment choice on the basis of a comparison with the average expectation embodied in the interest rate. If she is more optimistic on emerging markets, in the sense of estimating a lower probability of devaluation than the average, then she will put all her wealth into emerging market debt; otherwise, she will put it all into U.S. assets. In this model, investor heterogeneity is key to determining the amount of emerging market assets held.

As in the above described representative agent model, a balance of payments identity relates the change in reserves to the trade balance (assumed for simplification to equal zero in all periods) plus the purchase of new debt by investors minus the principal and interest on maturing debt; assuming that there has been no devaluation or default:

$$R_t = R_{t-1} + D_t - (1 + r_{t-1})D_{t-1}. \tag{2.12}$$

Reserves earn no interest, but they could just as easily have been assumed to earn r^*.

Provided that R_t is above some threshold level (which we assume without loss of generality to be zero), there is no devaluation at t, i.e. $\delta_t = 0$ (absence of superscript indicates that this is the realized value of depreciation, not its expectation). However, if reserves would otherwise be negative, there is a devaluation or default which reduces the amount that will be repaid on borrowing undertaken at t. That is, the ex post return for the lender will be $(1 + r_t)/(1 + \delta_t)$, where the amount of the devaluation is equal to the shortfall in the balance of payments that would have pushed R_t negative, divided by D_t:

$$\delta_t = \frac{-R_t}{D_t} \tag{2.13}$$

or using the above equation for R_t

$$\delta_t = \frac{[(1 + r_{t-1})D_{t-1} - R_{t-1} - D_t]}{D_t} \tag{2.14}$$

Though the devaluation/default reduces the amount owed at $t + 1$, not t, we assume that, in this case, balance of payments arrears are accumulated within the period such that reserves at t do not go negative but instead equal zero.

2.3 Evolution of Heterogenous Beliefs

Next, we describe the evolution of beliefs about probability of devaluation in the context of social and individual evolutionary learning.

2.3.1 Social Learning - A Baseline Model

We first describe Arifovic and Massson's model of *social learning* with boundedly rational agents who acquire the experience and knowledge needed to improve their performance over time. This model imposes weak requirements on agents' computational abilities. In this paper, the model of social learning will be referred to as our *baseline* model. The learning algorithm describes imitation-based adaptation of the agents' expectational rules (here a rule is just a point estimate for π_t^i. Investors consider their own success and that of other investors and try to imitate those rules yielding above-average returns. In addition, they occasionally experiment with new expectational rules.

Realized rates of return determine measures of performance of the expectations used at time t that we call *fitness* values. Performance, μ_t^i, of each investor's rule is evaluated based on the ex post return on emerging market assets

$$\mu_t^i = (1 + r_t)/(1 + \delta_t) - 1 \tag{2.15}$$

if investor i invested her wealth in the emerging market and to

$$\mu_t^i = r^* \tag{2.16}$$

if she invested in the US market. In the case that due to devaluation the performance value of an expectational rule takes a negative value ($\delta_t > r_t$), it is truncated to zero. Thus all the expectations that resulted in $\lambda_t^i = 1$ receive the same performance value even though they may have different values of π_t^i. Similarly, all those that resulted in $\lambda_t^i = 0$ receive the same performance value even though they may have different π_t^i's.

Investors update their expectations of π_t^i at the end of each period by imitating rules that have proven to be relatively successful and by occasional experimentation with new expectational rules. These two aspects of expectations formation are described below.

Imitation At the beginning of each period t, investor i, $i \in [1, \ldots, n]$ compares her expectational rule to a rule of a probabilistically selected investor j. The probability, Pr_t^j, that an expectational rule j is selected for comparison is equal to the expectational rule's relative performance:

$$Pr_t^j = \frac{\mu_t^j}{\sum_{i=1}^{n} \mu_t^i}. \tag{2.17}$$

We can think of the selection of an expectational rule j as resulting from a spin of a roulette wheel where each expectational rule is assigned a slot proportionate to its relative performance value (proportional selection). Rules that performed better get larger slots than rules that did worse in the previous period, and thus well-performing rules have higher probability of being selected. Rules are selected with replacement. Once j is selected, investor i compares the performance of her own expectational rule to the performance of investor j's expectational rule. If the performance of her own rule is equal or higher, she keeps her own rule. Otherwise, investor i imitates (adopts) the expectational rule of investor j.

Note that in case of devaluation, if $\delta_t > r_t$, expectational rules of the investors who invested in the emerging market yield a negative return, which is truncated to zero. Thus expectations of all investors who invested in the emerging market will receive performance values equal to 0 and will not be imitated. Only the expectations of those investors who invested in the US market receive positive, equal probabilities of being selected in this case.

Imitation alone represents a type of herd behavior in that on average, over time, well-performing expectations will be imitated (followed) by a larger number of investors and on average, investors will encounter better-performing expectations more frequently.

Experimentation Once the imitation is completed, each investor, $i \in [1, ..., n]$, can experiment with her expectational rule. Experimentation takes place with probability p_{ex}. If the investor experiments with the expected probability of devaluation, a new expected probability of devaluation is determined by drawing a random number from the uniform distribution over the interval $[0, \pi_{max}]$.

The above describes the framework which is assumed to govern the interaction of the population of investors. If investors are not able to gather enough information to form reliable estimates of the future behavior of the markets, and based on that determine their optimal behavior, imitation of previously successful strategies seems a plausible behavioral assumption. This type of behavior is explicitly modeled in our framework using proportional selection such that expectational rules that yielded an above-average payoff tend to be used by more investors in the following period. Experimentation incorporates innovations by investors, done either on purpose or by chance.

2.3.2 Individual Evolutionary Learning - An Extended Model

Next we combine the currency crisis framework of Arifovic and Masson with the model of individual evolutionary learning used by Arifovic and Ledyard (2003). We describe the model and the way we are going to implement it in our simulations.

Agent behavior

At the beginning of period t, each investor, i, has a collection A_t^i of possible alternative expectational rules. Each expectational rule of investor i is given by a real number that represents $\pi_{j,t}^i$ at time t. A_t^i consists of J alternatives, $a_{j,t}^i$, for $j \in \{1, \dots, J\}$.[11] At each t, an investor selects an alternative randomly from A_t^i using a probability density Π_t^i on A_t^i.[12] This alternative then becomes the expectational rule that agent implements at time period t. We construct the initial set A_1^i by randomly selecting, with replacement, J expectational rules from the set of all possible rules within a predefined range. We construct the initial probability Π_1^i by letting $\Pi_1^i(a_{j,1}^i) = 1/J$.

After each investor chooses her expectational rule, we compute the emerging market interest rate, r_t. The next step is to determine the value of each investor's $\lambda_i(t)$. This is accomplished in the same manner as has already been described in the previous section. We use the rest of the model's equations to compute the level of reserves in the emerging market and extent of possible devaluation.

Based on the information obtained at t, each investor updates her collection of alternative expectational rules. This process consists of three pieces, computing foregone return, and performing experimentation and replication.

Foregone return

In updating A_t^i and Π_t^i, the first step is to calculate what we call *foregone* returns for each alternative expectational rule in the collection. This is the (expected) return, given the information at t, that the alternative $a_{j,t}^i$ would have received if it had been actually used, taking the behavior of other investors as given. We use the notation $r^i(a_j^i|s_t^i)$ to compute the hypothetical return of the alternative j that belongs to investor i's set of alternatives.

For each alternative j, we determine the value of hypothetical $\lambda_{j,t}^i$, given the value of $\pi_{j,t}^i$. Finally, using this value of $\lambda_{j,t}^i$, we compute the rules' foregone return. In this model, this represents their performance measure.

[11] J is a free parameter of the behavioral model that can be varied in the simulations. It can be loosely thought of as a measure of the processing and/or memory capacity of the agent.

[12] In essence the pair $(A_t^i.\Pi_t^i)$ is a mixed strategy for i at t.

Updating $\mathbf{A}_t^i$

We modify A_t^i with processes of experimentation and imitation analogous to the ones described above for social learning. Foregone returns play the role of fitness values. The process of imitation results in the increase in frequency of the better performing rules. In case of our extended model, it can be interpreted as a reinforcement of those expectational rules that resulted in higher foregone returns.

While algorithmically the process of experimentation is performed the same way in the two models, it has different interpretation and impact on the dynamics. In the baseline model (social learning) it is a trembling hand random mutation. However, in the extended model (individual learning), newly generated rules will not be automatically tried out when they are generated. They have first to increase their frequency, based on high foregone payoffs, in order increase their probability of actually being selected.

We refer to the above described model of individual evolutionary learning as our *extended* model in the subsequent analysis.[13]

2.4 Design of Simulations and Experiments with Human Subjects

2.4.1 Simulations

As mentioned earlier, we focus here on simulations in which δ_t^e is *not* allowed to evolve. This algorithm is referred to as the fixed - δ^e case by Arifovic and Masson. Here, the expectational rule is characterized by a single real number, π_t^i (the probability of devaluation), and it is assumed that the expected amount of devaluation, $\delta_t^{e,i}$, is equal across investors and time.

Agents (n) and Experimentation Rates (p_{ex}) We first simulate permutations over the rate of experimentation and number of agents for the baseline simulation (one rule per agent). Holding the experimentation rates at 0.33, 0.165, 0.0825, and 0.04 we simulate over population levels that include 100, 75, 50, 25, and 12. As total wealth remains constant throughout these simulations ($\overline{W}$), decreasing the number of agents has the effect of increasing the per period investment of each individual. Decreasing the experimentation rate has the effect of decreasing the amount of heterogeneity introduced in each period.

[13]Individual and social learning can be complimentary. It is feasible to incorporate both types of learning within a single model of adaptation. A model of individual learning can incorporate imitation *across* individual sets of J rules. This intra-individual imitation, occurring between randomly chosen pairs of individuals every t_i periods, allows individuals to mimic the strategies of other agents utilizing a fitness (payoff) criterion in order to determine the relative success of the two sets of rules. An individual imitates the other pair's rules if, and only if, this criterion is met.

Strategy Set Size - J In the model of learning in which agents have a set of alternative rules played probabilistically (A_t^i, Π_t^i), we simulate various permutations over the size of this strategy set J. We allow the strategy set size, J, to equal 45, 15, and 5. For each parameterization of J, we simulate over the various permutations of population levels according to 100, 75, 50, 25, and 12, and of experimentation rates according to 0.0825, and 0.04.

Simulations over very low specifications of the population of agents, $n = 12$, are used to gauge the impact of lower population levels on the simulations' dynamics. These are used in order to facilitate a comparison to experimental data where, due to constraints, population levels are below that which would be considered appropriate to approximate perfect competition. However, these levels may not be sufficient for ensuring the efficacy of the learning algorithm as diversity over rules reaches a critically low level. This is a concern, foremost, for social learning where diversity is a direct function of population levels. This direct relationship is not a characteristic of individual learning, as J allows for a break between the direct relationship between population and diversity over rules. For this reason we expect, *a priori*, the results of the low population individual learning parameterization of the model to be more rhobust with respect to decreases in the population and therefore offer a more favorable comparison with experimental data. Additionally, social learning entails knowledge of other individuals' rules which will not be a feature within the experimental environment.

Risk Averse Agents - b_i We extend the model of the portfolio choice of agents to one that includes a specification of risk averse investors. The equation that determines investment in the emerging market, as derived by Masson (2003), is

$$\lambda_t^i = \frac{b^i(r_t - \pi_t^i \delta_t^i - r^*)}{\pi_t^i(1 - \pi_t^i)(\delta_t^i)^2} \tag{2.18}$$

where, λ_t^i is set to unity if the above equation yields a result strictly greater than one, and zero if strictly less than zero. Here, b_i is a utility parameter negatively related to the degree of risk aversion of the particular investor. Risk neutrality is equivalent to setting this parameter to infinity. Each agent has the same measure of risk aversion ($b_i = b \ \forall \ i \in [1 \ldots n]$).

We maintain a parameterization of b_i equal to 1 and simulate the baseline model of expectations including the four population levels described above (100, 75, 50, 25, 12) and an experimentation rate equal to 0.0825. Using these parameterizations of population and exerimentation, the extended model incorporating b_i is simulated with 15 rules per agent.

Parameterization of Simulations As described above, the permutations over n, p_{ex}, J, and b_i include a total of 60 unique parameterizations of the simulations. All of the results of these simulations are presented in the Appendix.

2.4.2 Experiments with Human Subjects

Our experimental design follows closely that of our extended simulation design in which δ_t^i is equal to one for all investors and over all experimental periods.[14]

Subjects were economics SFU undergraduates, third and fourth year. They volunteered, i.e. none were participating for fulfillment of any course requirement, and were paid a "show-up" fee and awarded an additional payment dependent on performance.[15] We used Z-tree software for experimental economics developed by Urs Fischbacher to create our experimental environment.

Initial Conditions - Instructions Prior to the beginning of an experiment, subjects are given the following information: (1) the balance of payments identity that governs the currency reserves of the emerging economy's central bank in the following period; (2) the equation determining the rate of return in the emerging economy's asset market; (3) the fixed rate of return in the U.S. economy, r^*, and an initial value of the emerging market rate of return, r_0; (4) the initial level of investment in the emerging market, D_0; (5) The constant wealth available for investment, $\overline{W}$ in each period; (6) the equation governing their portfolio allocation; (7) and the method according to which experimental payoff is determined. This information is contained in a set of instructions read by, and to, participants of the experiment. Each experimental period proceeds in the following way:

Subjects' Assessment of π_t^i At the beginning of each period, subjects are asked to quantify the probability of devaluation. At any time may subjects view the report of variables described in the previous section or the *experiment parameters* and the history of relevant variables. Experimental subjects are prompted for their assessment of the probability of devaluation. In order to make this assessment more intuitive, they are asked to enter a probability over the span of $[0, 10]$ rather than $[0, .10] = [0, \pi_{max}]$. Their assessment is then converted to a $\pi_t^{e,i}$ by dividing by 100.[16] The rest of the calculations are performed following the equation presented earlier.[17]

Report of Results Subjects are shown their resulting portfolio and rate of return, and their experimental payoff for that period. Subjects are also informed of that periods' *ex ante* and *ex post*

[14]An alternative experimental design may be found in the work of Heinemann, Nagel and Ockenfels (2004). Their work tests the predictions of global game theory with respect to private information using a reduced form Morris and Shin (1998) model. However, as consecutive experimental periods are in no way related in terms of fundamentals, the work cannot focus on the recurrence or duration of devaluation and no-devaluation periods.

[15]The "show-up" fee was equal to 15 dollars. The performance dependent payment was calculated in a manner such that the average total payment across subjects amounted to approximately 25 dollars. Subjects were informed about the nature of the total payment prior to participation in the experiment.

[16]The parameterization of π_{max} is taken from the original work of Arifovic and Masson (2003) in order to maintain comparability of results. It's original specification was in order to align simulations' interest rate spreads with those of monthly emerging market data.

[17]Under the unlikely scenario that a subject's assessment equals the geometric mean of all assessments, the subject's wealth is invested wholly in the emerging market if $\pi_t^i < \pi_{max}/2$, wholly in the domestic market if $\pi_t^i > \pi_{max}/2$, and split equally between the emerging and domestic markets if $\pi_t^i = \pi_{max}/2$. However, these rules did not have to be implemented in any of the sessions.

rates of return in the emerging market (before and after any devaluation, r_t, δ_t and $(1+r_t)/(1+\delta_t)$), and of the total level of investment in the emerging market from the previous period, D_{t-1}.

Treatment Payoffs A per period payoff for each subject is based on earnings in excess of the per period investment. That is, a subject earns $r^*\frac{\overline{W}}{n}$ when invested in the domestic market, $r_t\frac{\overline{W}}{n}$ when invested in the emerging market, and $[(1+r_t)/(1+\delta_t)-1]\frac{\overline{W}}{n}$ when invested in the emerging market in periods in which a devaluation takes place. Wealth, $\overline{W}$, is not accumulating; each subject has the opportunity to invest a constant amount in each period that is not dependent on previous investment performance. Importantly, as was the case in the simulations' fitness functions, experimental profit is bounded below by zero. Cumulative experimental profit translates into cash payment via a conversion factor. Total payment to the subject is the sum of a "show-up fee" and the converted experimental profit.

Experimental subjects' information set It is important to emphasize which variables are in the participants' information set and which are excluded. Each participant knows the complete history of total foreign investment, the *ex ante* and *ex post* emerging market return, and the extent of devaluation. However, they do not have information on the following: (i) the current level of currency reserves of the emerging market's central bank, and (ii) the devaluation threshold. We assume that in reality, although reserve levels may be known by investors, the threshold under which devaluation occurs is unknown. We remove knowledge regarding the current level of reserves in order to avoid subjects' learning the devaluation threshold through repeated observation of devaluations.

2.5 Simulation Results

Initial Values The values of initial external debt, and reserves, US interest rate, as well as the value of total wealth were taken from Arifovic and Masson (2003). Thus, the initial values for external debt, and reserves were taken to be those prevailing in Argentina at the end of 1996. In these "fixed - δ^e" simulations, the trade balance does *not* evolve. Interest rates and flows are converted to monthly data. All stocks and flows are expressed as ratios to GDP, so the relevant interest rates are actually the difference between the nominal interest rate and the growth of nominal GDP. For r^*, the U.S. interest rate used was $(0.05-0.03)$, or 0.001666. Variables of interest include

$$D_1 = 412.8, R_1 = 73.2, T_1 = -0.3, n\overline{W} = 825.6 \tag{2.19}$$

where the value for total wealth, $n\overline{W}$, was arbitrarily chosen to be twice D_1, π_{max} was chosen as 0.1, and $\delta^e_{max} = \delta^{e,i}_t = 1$.

2.5.1 Spread Statistics

Masson (2003) studies empirical regularities within the returns on emerging market debt.[18] The data indicate that daily changes in spreads are definitely not normally distributed, exhibiting much fatter tails. The study also finds generally significant first-order autocorrelation coefficient.[19] Our intention is to compare our simulation and experimental results to these two regularities. It is worth emphasizing that these results are derived from daily (not monthly) observations.

First Difference in Interest Rate Spread
Summary Statistics - Masson (2003)

Standard Deviation	0.04832
Skewness	-0.305
Kurtosis	86.06
Jarque-Bera	8,004,456
Observations	27,842
AC(1) (EMBI+)	0.134

Table 2.1: First Difference in Interest Rate Spread - Summary Statistics - Masson (2003)

In Table 2.3, 2.4 and 2.5 of the Appendix, we include distribution statistics for the first difference in the emerging market's interest rate spread, $[(1+r_t)/(1+\delta_t) - (1+r^*)]$. We will compare the qualitative features of these distributions to those of Masson (2003).[20]

Standard Deviation - Second Moment The standard deviation of the first difference in interest rate spreads vary between permutations of the simulations. However, all simulations' standard deviation fall in the $[0.0242, 0.0982]$ range. It is somewhat striking that even for parameterizations originally considered extreme, the standard deviation falls in this relatively small range. Notably, in the baseline simulations (simulations 1 through 20), decreasing the population level has the effect of increasing this measured standard deviation.

[18] He uses a set of spreads on emerging market debt compiled by JP Morgan using daily data from 31 December 1993 to 19 July 2002. This data base comprises virtually the universe of all developing countries issuing Brady bonds and Eurobonds. The list of countries is the following (those included in JP Morgan's so-called EMBI+ index, see JP Morgan, 1995): Argentina, Brazil, Bulgaria, Colombia, Ecuador, Korea, Mexico, Morocco, Panama, Peru, Philippines, Poland, Qatar, Russia, South Africa, Turkey, Ukraine, and Venezuela. However, not all countries had bonds outstanding during the whole period 1993-2002; what observations existed were pooled to study the distribution of spreads.

[19] Masson notes that this could be due to market inefficiencies that allow arbitrage opportunities to exist, or could reflect lack of trading so that spreads quoted do not correspond to actual transactions.

[20] The data presented in the Appendix to this chapter represents a subset of 120 different parameterizations of the simulations. For brevity and parsimony, we exclude presenting parameterizations that yield results redundant to those considered herein. Distinct parameterizations within the population of simulations are associated with unique simulation numbers. Therein, the non-sequential numbering of simulations in the Appendix has been maintained to facilitate comparison with the entire sample utilized in other work.

Skewness - Third Moment From the distribution of the first difference in the emerging market interest rate spread for each permutation, we calculate the measure of skewness. In all of the simulations, the skewness statistic from this distribution measures positive falling on the range $[0.0742, 1.6117]$; this result does not appear to align itself well with the empirical findings based on daily data.

Kurtosis - Fourth Moment From the distribution of the first difference in the emerging market interest rate spread for each permutation, we calculate the measure of Kurtosis according to the following equation:

$$K = (\frac{1}{N})\Sigma_{i=1}^{N}(\frac{y_i - \overline{y}}{\widehat{\sigma}})^4 \tag{2.20}$$

Distributions with a kurtosis measure of 3 are referred to as mesokurtic, of which the normal distribution is a prime example. Those distributions with a kurtosis measure exceeding 3 are referred to as leptokurtic, and are characterized by slim or long-tails. Finally, those distributions with a kurtosis measure less than 3 are referred to as platykurtic (fat or short-tailed). Masson (2003) finds a high value of kurtosis over daily first difference in interest rate spreads. Over all data sets that they consider, this measure is in excess of 80. In most of our permutations, the kurtosis measure far exceeds that of a normal distribution, reaching a maximum of approximately 56 in the baseline simulation with 100 agents, experimentation with probability 0.0825, and with a risk aversion parameter equal to 1 (simulation number 96).

Although the values of kurtosis computed in our simulations do not reach the empirical measure of around 80, the measures are in excess of that associated with normal distribution (with the exception of three parameterizations).[21]

Jarque-Bera The normal distribution has a skewness and kurtosis measure of zero and three respectively. A simple test of normality is to find whether the computed values of skewness and kurtosis depart from the norms of 0 and 3. This is the logic behind the Jarque-Bera (JB) test of normality.

$$JB = N[\frac{S^2}{6} + \frac{[K-3]^2}{24}] \tag{2.21}$$

Where S refers to skewness and K, kurtosis. Under the null hypothesis of normality, JB is distributed as a Chi-square statistic with 2 degrees of freedom.

According to Masson (2003), daily change in spreads occur over a non-normal distribution. In all of our 60 permutations of the model, we reject the null hypothesis of normality using the Jarque-Bera

[21] Parameterizations that do not have Kurtosis measures in excess of 3 are contained in simulations 77 through 79, inclusive.

test.

Autocorrelation Coefficients We report the estimates of the first order autocorrelation coefficient from an autoregressive regression including the first difference in spread measures in Tables 2.3 through 2.5. The estimated first order autocorrelation coefficient is significantly negative in all of our simulations. This contrasts the positive correlation reported in Masson (2003). However, it is important to note that the positive correlation in Masson 's work is over daily changes in interest rate spreads, rather than the monthly changes expressed in the simulations of this paper. It is quite likely that the monthly first difference in spreads are negatively correlated empirically, while daily are positively; a result very common to financial data. However, this conjecture requires validation using data not available at this time.

Summary Overall, the regularities of the spread statistics are extremely robust over the permutations of the parameter choices of the simulations, both baseline and extended. The most important finding is the robustness *across* the models of learning. In sum, regardless of the choice of model and for its parameterization, the distribution of the first difference in interest rate spread is positively skewed with a Kurtosis measure well in excess of the normal and the interest rate spread is negatively autocorrelated. Although falling short of matching empirical data with respect to skewness and first order autocorrelation coefficients, standard deviation and kurtosis measures capture empirical regularities well.[22]

2.5.2 Duration Statistics over Parameter Permutations

The Baseline Model - Comparison with Arifovic and Masson (2003) In our simulations of the baseline model, the observed dynamics are identical to those reported by Arifovic and Masson. The model exhibits recurrent instances of devaluations. We now consider the average duration of devaluation and no-devaluation periods over the various permutations of parameter specifications, using our two models of learning. In each simulation, the baseline initial values described above are used.[23]

Tables 2.6 and 2.7 present the average duration of periods of devaluation and non-devaluation for each of the simulations. We differentiate between two definitions of devaluation. Our first definition corresponds to the standard definition of devaluation (the same was used in Arifovic and Masson). That is, a simulation is within a period of devaluation if δ_t is greater than zero (or, anytime reserves fall below their the threshold value). We refer to these as simply *devaluations.* They occur whenever

[22] However, we would like to point out that this comparison is made with qualification. The measures of kurtosis and skewness reported in Masson (2003) are those of daily data, while in our simulations, the generated data refers to monthly intervals.

[23] As in Arifovic and Masson, each simulation is run for 10,000 periods.

the emerging market's currency undergoes a depreciation against the domestic. The *ex post* emerging rate of return is lower than *ex ante* rate of return.

However, the fact that the emerging market's currency depreciated does not guarantee that the resulting rate of return earned from investing in the emerging market is lower than that of investing in the domestic market. A depreciation arising from reserves shortages may not be enough to make investing in the domestic market more attractive. Therefore, we also include a definition of devaluation periods that only include those in which the *ex post* rate of return in the emerging market is strictly lower than that of the domestic. We refer to these periods as *dynamically relevant devaluations*.

Why is this distinction important? The answer is related to the evaluation of the payoff (fitness) function used in the simulations and experiments with human subjects. Although a devaluation may have occurred in the previous simulation period, if it was not large enough to drive the *ex post* emerging market return below that of the domestic market, rules that translated into investment in the emerging market will propagate. Therefore, simulation dynamics are more likely to be based on the dynamically relevant devaluations rather than the standard definition of devaluation. We discuss the results across different types of simulations.

Baseline Simulations First, consider the baseline simulations (simulations 1 through 20). Consistent with the results of Arifovic and Masson (2003), holding the numbers of investors constant, decreasing the rate of experimentation (p_{ex}) decreases the average duration of periods of devaluation. Upon the onset of a devaluation, those investment rules associated with domestic investment earn higher rates of return than those associated with investment in the emerging economy. For a devaluation to continue, investment must favor the domestic market, therein pulling wealth out of the emerging economy. This occurs when those rules associated with domestic investment are imitated by investors; a process that is inherent in the social learning algorithm. However, with higher rates of experimentation, this imitation is not as effective and the favoring of the domestic economy is less prominent. Increased experimentation decreases the ability of imitation and therefore the swing towards domestic investment required for sustained devaluations is less probable.

Additionally, holding the rate of experimentation constant, lowering the population levels of the baseline simulations tends to decrease the average duration of periods of devaluation. However, this result does not hold for the two lowest specifications of p_{ex} where the duration measures for these parameterizations are already near their lower bound. As such, no decrease in the duration of devaluations is possible. This holds as well when considering periods without devaluations. Generally, decreasing the number of investors in the baseline simulation (ceteris paribus) has the effect of lowering durations of both devaluation and no-devaluation periods.

Extended Simulations Our extended simulations of individual evolutionary learning result in shorter duration of no-devaluation periods when the size of agents' collections of alternative rules is relatively small. In these simulations, we observe a more frequent switching between states of devaluation and those with no devaluation. Specifically, extended simulations in which agents have a collection of five rules and experimentation rates equal to 0.04 (simulations 76 through 80, inclusive) have average durations of successive periods without devaluation two to three times smaller than their baseline counterparts (simulations 16 through 20). This result holds across both specifications of the experimentation rate.

This decrease in duration measures from the baseline model does not hold when the number of rules in the investors' collections increases to it's largest specification ($J = 45$, simulations 61 through 65). Here, duration measures for no-devaluation periods are very comparable to the baseline model counterparts.

We conclude that decreasing the diversity of rules available for each agent is very important for decreasing the duration of of no-devaluation periods.[24] Smaller collections of rules are associated with shorter periods without devaluations.

Decreasing the size of each agent's collection has the effect of increasing the duration of devaluation periods. For both specifications of p_{ex}, the duration of devaluations is longest with the lowest specification of the number of rules in this collection (and with the number of investors, n equal to 100).

Holding the number of rules per agent and the rate of experimentation constant, decreasing the number of agents has the effect of lowering both the duration of devaluation and no-devaluation periods (consistent with the baseline results).

Decreasing the experimentation rate does not seem to have any general effects in the extended simulations with high numbers of rules in agents' subsets. However, when these subsets are quite low (5 rules), lowering the experimentation rate decreases the duration of devaluation and no-devaluation periods.

One could argue that some of the relatively smaller average durations of devaluation periods under the extended model of learning are empirically unrealistic. However, when we consider the simulations' duration statistics in light of our experimental data, these lower no-devaluation durations must be considered a success.

Risk Aversion - Baseline and Extended Model From the consideration of risk neutrality, we incorporate risk aversion by decreasing the risk aversion parameter (b_i) to a value of one.

Consistent with the conclusions for the baseline and extended simulations considered above, with

[24]Note that decreasing the number of agents in the baseline model would have the same effect on diversity. As described above, the resulting impact on duration statistics is the same.

risk aversion included in simulations, decreasing the number of agents (ceteris paribus) lowers both devaluation and no-devaluation duration measures.

In the baseline model, holding all parameters constant and decreasing the risk aversion parameter tends to increase the duration of devaluation and no-devaluation periods for simulations with larger numbers of agents (100, 75 and 50). For simulations in which the population is at one of its lowest two specifications, 12 and 25, decreasing the measure of risk aversion decreases the duration of no-devaluation periods considerably (duration measures for devaluation periods are already near their lower bound for these levels).

Decreasing the risk aversion parameter in the extended model increases both the duration of devaluation and no-devaluation periods at all population levels.

2.5.3 Average Assessment (π_t^i) - Regression Analysis

Stylized facts regarding interest rate spreads leading up to and following currency devaluations are considered in the work of Tornell and Westermann (2001). In the consideration within this work, it is observed that interest rate spreads tend to increase in the period immediately preceding the onset of devaluation. This increase is estimated to be one percent. It is followed by a further increase in the period of devaluation of three and a half percent; a total increase of four and a half percent is observed leading up to currency devaluations. Following the onset of the devaluation, interest rate spreads tend to decrease.

This decrease in the interest rate spread following devaluation is considered in the recent work of Kasa and Cho (2003). Their work is motivated toward explaining the recession that appears to follow periods of currency devaluation. While third generation models of currency crises accounted for this observation through their inclusion of "balance sheet effects", currency crises are still the result of exogenous sunspot affects. Their application of a model of learning and adaptation to the beliefs of the policy-maker and the agents makes endogenous the onset of crises; the onset of currency crises may be linked to the stochastic properties of their model of learning and the structural features of the economy.[25] The fall in the interest rate spread may result from a mix of both risk premium effects and loose monetary policy. As noted by Kasa and Cho (2003), this loose monetary policy may be a concerted attempt to avoid the recession that follows devaluation. Of course, this policy tends to worsen the crises, deepening the impact of the initial devaluation of the value of the currency.

Our test on the first difference in interest rate spreads is related only to changes resulting from the increases of decreases in the risk premium. We attempt to find changes in this spread, derived from changes in the premium, that are not predicted by the change in the preceding period. Tables 2.8 through 2.14 include regressions on the first difference in average assessment ($\overline{\pi_t}$). There is no

[25]Notably, in the model considered herein, currency crises are also linked to the model of learning. However, as discussed above, currency crises are only a result of this adaptation on the part of agents; not to the economic fundamentals of the emerging economy.

constant term included in these estimations. We include six explanatory variables, including the first lag of difference in average assessment, and five dummy variables in two regressions per simulation. Each dummy controls for specific periods within the simulations. The details of our analysis are contained within Table 2.2.

Specification of Dummy Variables	
$D1$	$\delta_t = 0$ and $\delta_{t+1} > 0$
$D2$	$\delta_t > 0$ and $\delta_{t-1} = 0$
$D3$	$\delta_{t-1} > 0$ and $\delta_{t-2} = 0$
$D4$	$\delta_t = 0$ and $\delta_{t-1} > 0$
$D5$	$\delta_{t-1} = 0$ and $\delta_{t-2} > 0$
$D6$	$r_t - \delta_t \geq r^*$ and $r_{t+1} - \delta_{t+1} < r^*$
$D7$	$r_t - \delta_t < r^*$ and $r_{t-1} - \delta_{t-1} \geq r^*$
$D8$	$r_{t-1} - \delta_{t-1} < r^*$ and $r_{t-2} - \delta_{t-2} \geq r^*$
$D9$	$r_t - \delta_t \geq r^*$ and $r_{t-1} - \delta_{t-1} < r^*$
$D10$	$r_{t-1} - \delta_{t-1} \geq r^*$ and $r_{t-2} - \delta_{t-2} < r^*$

Table 2.2: Specification of Dummy Variables - Regression Analysis

We use $r_t - \delta_t$ as an approximation for $[(1 + r_t)/(1 + \delta_t) - 1]$. We can interpret the estimated coefficient on a dummy variable as the change in $\overline{\pi_t}$ in simulation periods with the characteristics as described in the above that is not explained with the lagged difference in $\overline{\pi_t}$. The dummy variables numbered one through five are associated with the standard definition of devaluations, and those numbered six through ten are associated with the stricter definition of dynamically relevant devaluation.

Our results for changes preceding devaluations coincide with that observed empirically. In the majority of parameterizations of the simulation, the period preceding the onset of a devaluation is characterized by a higher than expected interest rate spread. Evidence is found in the positive coefficient estimate of the $D6$ dummy variable. However, in the period in which a devaluation begins, $D7$, interest rate spreads are lower than would otherwise be predicted. Importantly, this measured effect is stronger than that inherent in the period preceding devaluation. As the lag of the first difference in interest rate spread has an estimated coefficient that is always less than one we may also conclude that there is an absolute fall in the interest rate spread these periods. This result stands in contrast to the data summarized by Tornwell and Westermann (2001).

Importantly, Kasa and Cho (2003) also find it difficult to model increases in the interest rate spread in the onset period of devaluation. Their conjecture is that policy makers "lose control" of a mild depreciation attempt.[26] In contrast, the fall in the spread we witness stems solely from a fall

[26] As noted by Kasa and Cho (2003), an empirical counterpart to this conjecture may be found in Britain's departure

in the mean investor sentiment regarding the likelihood of devaluation. Similarly, although following the onset of a period of devaluation we witness a very strong increase in the interest rate spread, $D8$, this effect is reflective of only an increase in the risk premium in isolation. Our model does not incorporate the potential of expansive monetary policy in an effort to alleviate economic recession.

2.6 Experimental Results and the Dynamics of Expectations

In this section we compare the results of our simulations to those obtained in the experiments with human subjects. We conducted a total of three experimental sessions. [27] We had 15 subjects in our first experimental session, and 11 subjects in the last two experimental sessions. The summary statistics are presented in Tables 2.5, 2.6 and 2.15.

In all experimental sessions we observe negative correlation between the first difference in spread statistics, kurtosis measures greater than that associated with the normal distribution, and positive skewness. These regularities match the simulations very well. In the sessions with 11 subjects, standard deviation measures are slightly larger than those of the session with 15 subjects. We noted above that smaller number of agents in the baseline simulations yielded larger standard deviation measures. The 15 subject session has a standard deviation measure within the range of those associated with the simulations. Kurtosis measures for all of the sessions fall within the range of those for the various permutations of the simulations.

The average duration of periods of devaluation and periods with no-devaluation are generally quite small when compared to those of the baseline Arifovic/Masson simulations (simulations 1, 6, 11, and 16). Although we conducted only one session with 15 subjects, it is noteworthy that the treatment with a larger number of subjects also has larger durations of devaluation and no-devaluation periods. These are not unexpected outcomes. Our discussion above refers to falling durations for specifications with a smaller number of agents; though these smaller durations are still larger than those of the treatments, especially with respect to no-devaluation periods. When we allow for smaller number of agents, the baseline simulations reasonably approximate the experimental results.

We have also noted that the simulations of our extended, individual learning model are sometimes associated with far more switching between devaluation and no-devaluation states. A final point with respect to durations is that simulations of the extended model match the experimental data very well. Consider, for example, simulation number 80: an extended, individual learning model with 12 agents, 5 rules each and an experimentation rate of 0.04. Its duration measures of 2.07 and 4.83 match the 11 subject sessions quite well with respect to duration of devaluation and no-devaluation periods, respectively. Similarly, simulations with slightly larger collections of rules (15 rules per

from the EMS following the September 1992 attack; interpreted as allowing Britain to embark on a policy of lower interest rates

[27]We set π_{max} 0.10 to match the number used in our simulations.

agent, simulations 71 through 75) perform well in matching the 15 subject session.

We now turn to the analysis of the behavior of the average assessment of devaluation in experiments with human subjects and in simulations of our baseline and extended model. Figures 2.1 - 2.3 plot the average assessment of devaluation ($\overline{\pi_t}$) and devaluation size (δ_t) over time. Data for a subset of periods of a baseline and extended model simulations are given in Figures 2.1 and 2.2, respectively. The results of one of the experimental session are contained in Figure 2.3.[28]

A defining characteristic of the plots of the experimental average assessment is the relatively small range in which these measures fall when compared to those of the standard baseline simulations. For example, in the final ninety experimental periods of the session presented in Figure 2.3, average assessment is never larger than 0.05, and in only a very few periods does it fall below 0.02.[29] A similar lower bound exists for the plots associated with the standard baseline simulations. However, in the majority of periods of devaluation, the average assessment climbs as high as 0.08. One may argue that the baseline simulation and the experimental results share a common lower bound for average assessment. It is important to note that within experiments, there are many situations wherein the onset of a devaluation is not associated with an average assessment close to the lower bound. This is rarely the case for the baseline simulation results. Additionally, the upper bound placed on assessment does not appear relevant for reversing these periods of devaluation in treatments, as average assessment rarely crosses the 0.05 level.

The plots of the average assessment for our extended model look much more like experimental data. Consider Figure 2.2, plotting the extended model simulation's results. Here, the plot of average assessment looks very much like those plotted for the experimental session. Periods of devaluation are not necessarily associated with the lower bound on assessment, and the reversal of these devaluation periods occurs far before average assessment can climb to its upper boundary. In this respect, extended,individual learning simulations appear to match the experimental dynamics much better than the baseline specification.

The extended, individual evolutionary learning simulations compare more favorably to the experimental results with respect to duration statistics. Specifically, they exhibit more frequent devaluation periods, and substantially shorter durations of no-devaluation periods. The range under which the average assessment occurs for the extended model simulations is quite smaller than that of the single-rule simulations. In addition to duration of devaluation and no-devaluation periods, this is a key characteristic the extended model simulations share with the experimental results.

Our examination of simulation and experimental data indicates that devaluations result from

[28] In order to facilitate comparison between simulation and experimental results, the following parameter choices are used for the baseline (Figure 2.1) and extended (Figure 2.2) simulation plots. The baseline simulation has 12 agents, one rule per agent, and a probability of experimentation set to 0.0825. The extended simulation is one in which 12 agents have 5 rules in their collections and experiment with a probability equal to 0.0825. Figure 2.3, that of the experimental data, is a session with 11 subjects.

[29] With respect to average assessment, the results of the other sessions are both qualitatively and quantitatively similar. Importantly, there is nothing particular to the specific experimental session we are discussing that cannot also be said of the other two sessions.

shifts in skewness of the distribution of $\pi_t^{e,i}$. Thus, the change of skewness plays crucial role in getting into periods of devaluation as well as getting out of them.[30] We estimate regressions over the time series of the first difference in average assessment. Included as independent variables is the first lag of differenced assessment, and the dummy variables specified above.

We define as sentiment reversal increases or decreases in average assessments that are *not otherwise predicted by the lag of the differenced assessment.* The estimated coefficients on the $D2, D4$ and $D7, D9$ dummy variables are reported in Table 2.8 through 2.15. The numbers show that there is quite some variation over the coefficient estimates across the different simulation permutations. There appears little consistency in the coefficient sign of the lagged difference in the assessment regressors. One thing to note is that in the simulations of the baseline model, both decreasing the rate of mutation, and decreasing the number of agents puts negative pressure on the $D2$ coefficient, often pushing it into negative territory.

2.7 Concluding Remarks

We study a model of currency crisis where the only source of volatility that contains potential for speculative attacks and devaluation of currency are agents' beliefs. The beliefs are heterogenous and evolve over time. We use two different frameworks, social learning and individual learning. As part of our methodology, we conduct a large number of simulations for different parameterization values to check for the robustness of the results.

One of the striking results is that most of the main features of the dynamics are present for the whole range of different parameter values and over a wide range of specifications. These include the 'fat tails', positive skewness, and negative correlation between the first difference in spread statistics. The 'fat tails' is also a feature that characterizes empirical data on the returns in the emerging markets.

We also conducted three experimental sessions with human subjects where we simulated the same type of the economy. The features of the exhibited dynamics coincide with those of our simulations, i.e. fat tails, positive skewness, and negative correlation between the first difference in spread statistics. Regarding the duration of devaluation and no-devaluation periods, and the range of values within which the assessment of devaluation varies, our extended, individual learning model matches the experimental data well.

[30]The model requires shifts in the skewness of the distribution over individual assessments in order to obtain variation in the flow of investment. Therefore, shifts in skewness are required for devaluations. Importantly, shifts in skewness are not necessarily associated with shifts in the average assessment, utilized to determine the interest rate in the emerging economy. Therein, there is no theoretical link in this model between the shifts in skewness required for devaluations and changes in the interest rate spread associated with the average assessment. Importantly, skewness in the average assessment over individuals does not necessarily translate into skewness of the interest rate spread. The two have no theoretical relation in the model considered herein.

References

Aghion, P., P. Bacchetta and A. Banerjee (2000). " Currency Crisis and Monetary Policy in an Economy with Credit Constraints", *European Economic Review*, 45: 1125-50.

Arifovic, J. and Ledyard, J. (2003) "Computer Testbeds and Mechanism Design: Application to the Class of Groves Ledyard Mechanisms for Provision of Public Goods," March 2003, manuscript.

Arifovic, J. and P. Masson (2003). "Heterogeneity and Evolution of Expectations in a Model of Currency Crisis," *Nonlinear Dynamics, Psychology and Life Sciences*, forthcoming.

Banerjee, A. (1992). "A Simple Model of Herd Behavior," *Quarterly Journal of Economics* 107 (August): 797-817.

Chari, V.V., and P. Kehoe (1998). "Hot Money," mimeo (Research Department, Federal Reserve Bank of Minneapolis, April).

Chamley, C. (2003), "Dynamic Speculative Attacks," *American Economic Review*, 93, 603-621.

Cho, In-Koo and K. Kasa, "Learning Dynamics and Endogenous Currency Crises", manusript.

Heinemann, F., R. Nagel, and P. Ockenfels (2004). "The Theory of Global Games on Test: Experimental Analysis of Coordination Games with Public and Private Information," *Econometrica*, Forthcoming.

Jeanne, O., and P. Masson (2000). "Currency Crisis, Sunspots, and Markov-Switching Regimes," *Journal of International Economics* 50:327-350.

Kasa, K. (2001) "Learning, Large Deviations and Recurrent Currency Crisis", forthcoming, *International Economic Review*

Krugman, P. (1979). "A Model of Balance of Payments Crises," *Journal of Money, Credit, and Banking* 11:311-25.

Masson, P. (2003). "Empirical Regularities in Emerging Market Bond Spreads," mimeo, October 14, Brookings Institution.

Morris, S, and H.S. Shin (1998). "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *American Economic Review* 88:587-597.

Obstfeld, M. (1997). "Destabilizing Effects of Exchange Rate Escape Clause", Journal of International Economics 43, 61-77.

2.8 Appendix

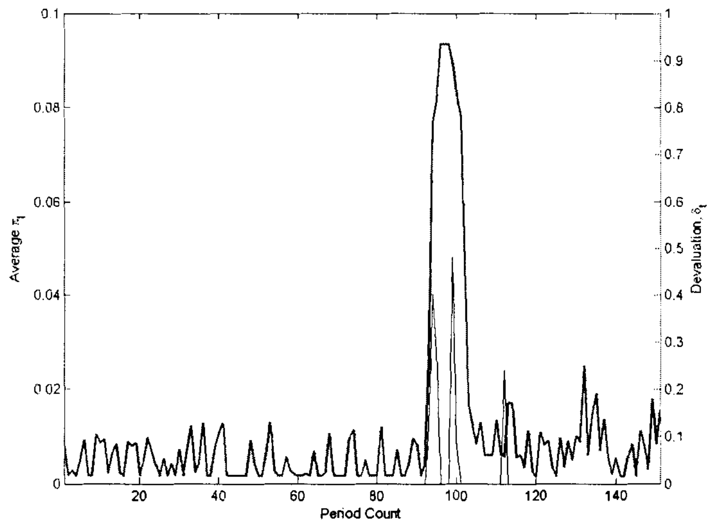

Figure 2.1: Baseline Simulation - 12 agents, 1 rule per agent, probability of mutation 0.0825

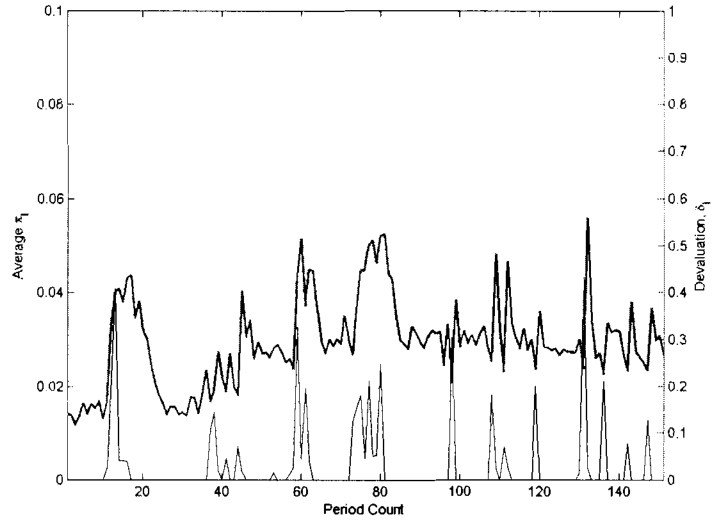

Figure 2.2: Extended Simulation - 12 agents, 5 rules per agent, probability of mutation 0.0825

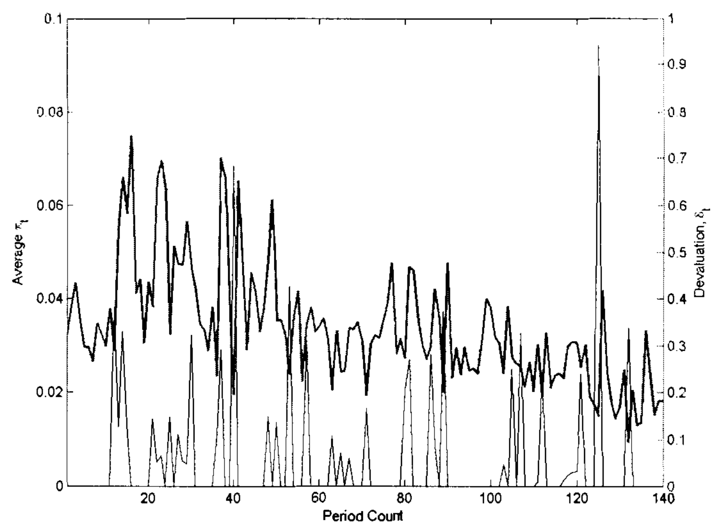

Figure 2.3: Treatment - 11 subjects

Simulation No.	Population	Rules	p_m	b	Std.Deviation	Skewness	Kurtosis	Jarque-Bera	AC(1) Coef.	Confidence Int.	
1	100	1	0.33	–	0.0699	1.0438	8.5243	32037.3309	-0.3265	-0.345	-0.3079
2	75	1	0.33	–	0.0753	1.0962	8.2406	30245.27	-0.393	-0.411	-0.375
3	50	1	0.33	–	0.0787	1.1404	7.904	28149.5239	-0.4139	-0.4317	-0.396
4	25	1	0.33	–	0.09	0.8078	6.8525	20615.9021	-0.4573	-0.4748	-0.4399
5	12	1	0.33	–	0.0982	0.5462	6.5964	18593.6343	-0.4945	-0.5115	-0.4775
6	100	1	0.165	–	0.0766	0.9016	12.9387	70990.1144	-0.5114	-0.5283	-0.4945
7	75	1	0.165	–	0.08	0.8307	13.2322	73981.4313	-0.5236	-0.5403	-0.5069
8	50	1	0.165	–	0.0797	0.8857	13.5166	77302.9049	-0.4883	-0.5054	-0.4712
9	25	1	0.165	–	0.0794	0.7467	13.8796	81062.1499	-0.478	-0.4952	-0.4608
10	12	1	0.165	–	0.077	0.4422	14.3286	85728.2631	-0.4991	-0.5161	-0.4821
11	100	1	0.0825	–	0.0798	0.3452	12.4875	65062.8705	-0.49	-0.5071	-0.4729
12	75	1	0.0825	–	0.0712	0.6327	17.3191	125442.2401	-0.4895	-0.5066	-0.4724
13	50	1	0.0825	–	0.0695	0.2827	17.2103	123345.5198	-0.5176	-0.5344	-0.5009
14	25	1	0.0825	–	0.0572	0.4267	28.7621	344448.0896	-0.4822	-0.4994	-0.465
15	12	1	0.0825	–	0.0658	0.1532	19.1618	152779.4133	-0.5041	-0.521	-0.4872
16	100	1	0.04	–	0.0638	0.0749	14.8354	91560.9185	-0.5026	-0.5196	-0.4857
17	75	1	0.04	–	0.0529	0.1593	19.4818	157927.693	-0.5048	-0.5217	-0.4879
18	50	1	0.04	–	0.0434	0.1856	34.4867	494835.243	-0.4865	-0.5036	-0.4693
19	25	1	0.04	–	0.0547	0.1293	30.0254	375070.3428	-0.4959	-0.5129	-0.4789
20	12	1	0.04	–	0.0829	0.1943	15.2031	96208.6146	-0.4899	-0.507	-0.4728
41	100	45	0.0825	–	0.0447	0.3807	24.4229	248376.0731	-0.5562	-0.5725	-0.5399
42	75	45	0.0825	–	0.054	0.4911	21.1627	186708.8063	-0.6947	-0.7088	-0.6806
43	50	45	0.0825	–	0.0552	0.5649	22.2541	206551.3147	-0.6779	-0.6923	-0.6635
44	25	45	0.0825	–	0.0597	0.3789	17.1078	121987.94	-0.7114	-0.7252	-0.6977
45	12	45	0.0825	–	0.056	0.4546	17.0052	120637.2949	-0.6358	-0.6509	-0.6206

Table 2.3: First Difference in Interest Rate Spread - Distribution Statistics

Simulation No.	Population	Rules	p_m	b	Std.Deviation	Skewness	Kurtosis	Jarque-Bera	AC(1) Coef.	Confidence Int.	
51	100	15	0.0825	–	0.0531	1.1422	13.1786	74415.9088	-0.5497	-0.566	-0.5333
52	75	15	0.0825	–	0.0551	0.8888	15.3142	98872.4335	-0.5256	-0.5423	-0.509
53	50	15	0.0825	–	0.0605	0.6319	11.8227	58806.3061	-0.568	-0.5842	-0.5519
54	25	15	0.0825	–	0.0622	0.5963	10.1505	43448.2459	-0.55	-0.5664	-0.5337
55	12	15	0.0825	–	0.0677	0.4301	9.902	41091.1048	-0.5885	-0.6044	-0.5727
56	100	5	0.0825	–	0.0464	0.5758	6.6146	18748.7372	-0.1635	-0.1829	-0.1442
57	75	5	0.0825	–	0.0521	0.6209	6.1288	16263.1434	-0.2912	-0.31	-0.2725
58	50	5	0.0825	–	0.0585	0.6048	5.4866	13127.9472	-0.3538	-0.3721	-0.3354
59	25	5	0.0825	–	0.0732	0.4029	5.2113	11563.8417	-0.4815	-0.4987	-0.4643
60	12	5	0.0825	–	0.0803	0.307	5.4475	12497.621	-0.5044	-0.5214	-0.4875
61	100	45	0.04	–	0.0534	0.8632	16.1728	110044.0002	-0.6442	-0.6592	-0.6292
62	75	45	0.04	–	0.052	0.5778	18.0145	135553.0267	-0.5669	-0.583	-0.5507
63	50	45	0.04	–	0.0553	0.5638	14.8698	92506.133	-0.6287	-0.644	-0.6135
64	25	45	0.04	–	0.0585	0.4676	14.7388	90727.2982	-0.6436	-0.6586	-0.6286
65	12	45	0.04	–	0.0605	0.5241	15.0211	94315.1026	-0.5907	-0.6066	-0.5749
71	100	15	0.04	–	0.0486	0.7894	9.6097	39448.3665	-0.319	-0.3376	-0.3005
72	75	15	0.04	–	0.0532	0.8449	9.0106	34958.8266	-0.3538	-0.3721	-0.3354
73	50	15	0.04	–	0.0628	0.5949	7.5715	24432.9029	-0.4578	-0.4753	-0.4404
74	25	15	0.04	–	0.0719	0.3839	7.4543	23355.7755	-0.5171	-0.5339	-0.5004
75	12	15	0.04	–	0.0735	0.4408	8.3718	29474.0369	-0.513	-0.5298	-0.4961

Table 2.4: First Difference in Interest Rate Spread - Distribution Statistics (Cont'd)

Simulation No.	Population	Rules	p_m	b	Std.Deviation	Skewness	Kurtosis	Jarque-Bera	AC(1) Coef.	Confidence Int.	
76	100	5	0.04	–	0.0538	0.2349	3.0786	4032.1203	-0.4344	-0.452	-0.4167
77	75	5	0.04	–	0.0637	0.233	2.2039	2109.0412	-0.508	-0.5249	-0.4911
78	50	5	0.04	–	0.0705	0.2113	1.8893	1557.5273	-0.5332	-0.5498	-0.5166
79	25	5	0.04	–	0.0787	0.2065	2.7314	3172.225	-0.5125	-0.5293	-0.4956
80	12	5	0.04	–	0.0846	0.17	3.8418	6184.8572	-0.5244	-0.5411	-0.5078
96	100	1	0.0825	1	0.0473	1.352	56.0627	1310620.723	-0.1491	-0.1685	-0.1297
97	75	1	0.0825	1	0.0527	1.6117	47.2117	931614.3243	-0.2821	-0.3009	-0.2633
98	50	1	0.0825	1	0.0569	0.9901	34.9589	510054.2031	-0.4244	-0.4422	-0.4067
99	25	1	0.0825	1	0.0367	0.0742	25.9926	281066.9713	-0.4975	-0.5145	-0.4805
100	12	1	0.0825	1	0.0242	0.4021	22.1091	203613.3823	-0.4771	-0.4942	-0.4599
116	100	15	0.0825	1	0.0434	0.5781	24.4355	248949.3194	-0.1512	-0.1706	-0.1318
117	75	15	0.0825	1	0.0464	0.7365	22.7333	215892.2112	-0.2455	-0.2645	-0.2265
118	50	15	0.0825	1	0.0514	1.0102	21.668	197011.1128	-0.2718	-0.2907	-0.253
119	25	15	0.0825	1	0.057	1.3648	20.4506	177082.8339	-0.299	-0.3177	-0.2803
120	12	15	0.0825	1	0.064	1.3874	19.2146	156789.7681	-0.3753	-0.3935	-0.3571
Treatment	15	–	–	–	0.065002	0.813709	10.7872	416.6527	-0.490306	(-0.0806)	[-6.08298]
Treatment	11	–	–	–	0.115439	0.330993	7.363626	104.7021	-0.823532	(-0.09155)	[-8.99530]
Treatment	11	–	–	–	0.140673	0.106237	5.862903	47.7312	-0.896168	(-0.08772)	[-10.2161]

Table 2.5: First Difference in Interest Rate Spread - Distribution Statistics (Cont'd)

Simulation No.	Population	Rules	p_m	b	Count(deval)	Ave.deval	Ave.non-deval
1	100	1	0.33	–	410	4.02	20.37
2	75	1	0.33	–	493	3.28	17.00
3	50	1	0.33	–	550	2.91	15.30
4	25	1	0.33	–	779	1.94	10.89
5	12	1	0.33	–	935	1.45	9.24
6	100	1	0.165	–	361	2.55	25.15
7	75	1	0.165	–	375	2.29	24.38
8	50	1	0.165	–	441	2.02	20.70
9	25	1	0.165	–	453	1.62	20.46
10	12	1	0.165	–	540	1.24	17.28
11	100	1	0.0825	–	464	1.24	20.31
12	75	1	0.0825	–	385	1.43	24.54
13	50	1	0.0825	–	427	1.11	22.31
14	25	1	0.0825	–	328	1.13	29.36
15	12	1	0.0825	–	503	1.13	18.75
16	100	1	0.04	–	526	1.02	17.98
17	75	1	0.04	–	516	1.01	18.37
18	50	1	0.04	–	338	1.03	28.55
19	25	1	0.04	–	391	1.08	24.49
20	12	1	0.04	–	621	1.15	14.95
41	100	45	0.0825	–	334	1.12	28.82
42	75	45	0.0825	–	460	1.24	20.50
43	50	45	0.0825	–	507	1.31	18.41
44	25	45	0.0825	–	583	1.19	15.96
45	12	45	0.0825	–	545	1.09	17.26
46	100	30	0.0825	–	417	1.48	22.50
47	75	30	0.0825	–	443	1.25	21.32
48	50	30	0.0825	–	481	1.38	19.41
49	25	30	0.0825	–	656	1.19	14.05
50	12	30	0.0825	–	605	1.14	15.41
51	100	15	0.0825	–	387	2.80	23.03
52	75	15	0.0825	–	404	2.07	22.68
53	50	15	0.0825	–	524	1.56	17.52
54	25	15	0.0825	–	664	1.32	13.74
55	12	15	0.0825	–	787	1.18	11.54

Table 2.6: Count and Duration Measures - Dynamically Relevant Devaluations

Simulation No.	Population	Rules	p_m	b	Count(deval)	Ave.deval	Ave.non-deval
56	100	5	0.0825	–	433	5.02	18.08
57	75	5	0.0825	–	532	3.93	14.86
58	50	5	0.0825	–	697	3.02	11.33
59	25	5	0.0825	–	1001	1.79	8.19
60	12	5	0.0825	–	1198	1.37	6.98
61	100	45	0.04	–	404	1.97	22.78
62	75	45	0.04	–	422	1.36	22.33
63	50	45	0.04	–	539	1.33	17.22
64	25	45	0.04	–	580	1.19	16.04
65	12	45	0.04	–	615	1.13	15.13
71	100	15	0.04	–	405	3.92	20.77
72	75	15	0.04	–	458	3.45	18.39
73	50	15	0.04	–	635	2.30	13.44
74	25	15	0.04	–	827	1.51	10.59
75	12	15	0.04	–	859	1.31	10.34
76	100	5	0.04	–	961	3.70	6.71
77	75	5	0.04	–	1360	2.57	4.78
78	50	5	0.04	–	1525	2.21	4.35
79	25	5	0.04	–	1547	1.80	4.66
80	12	5	0.04	–	1545	1.49	4.98
96	100	1	0.0825	1	92	3.76	104.92
97	75	1	0.0825	1	131	2.70	73.63
98	50	1	0.0825	1	228	1.39	42.46
99	25	1	0.0825	1	872	1.01	10.46
100	12	1	0.0825	1	2130	1.00	3.69
116	100	15	0.0825	1	165	4.54	56.40
117	75	15	0.0825	1	189	4.12	48.79
118	50	15	0.0825	1	243	3.18	37.97
119	25	15	0.0825	1	335	2.50	27.35
120	12	15	0.0825	1	450	1.90	20.32
Treatment	15	–	–	–		2.73	10.75
Treatment	11	–	–	–		1.27	4.64
Treatment	11	–	–	–		1.32	4.12

Table 2.7: Count and Duration Measures - Dynamically Relevant Devaluations (Cont'd)

No.	Pop.	Rules	p_m	b	$L(1)\overline{\pi_t}$	$D6$	$D7$	$D8$	$D9$	$D10$
					$\overline{\pi_t} = \beta_1(L(1)\overline{\pi_t}) + \beta_2(D6) + \beta_3(D7) + \beta_4(D8) + \beta_5(D9) + \beta_6(D10)$					
0	–	–	–	–	β_1	β_2	β_3	β_4	β_5	β_6
					$\beta_1 + z\sigma_{\beta_1}$	$\beta_2 + z\sigma_{\beta_2}$	$\beta_3 + z\sigma_{\beta_3}$	$\beta_4 + z\sigma_{\beta_4}$	$\beta_5 + z\sigma_{\beta_5}$	$\beta_6 + z\sigma_{\beta_6}$
					$\beta_1 - z\sigma_{\beta_1}$	$\beta_2 - z\sigma_{\beta_2}$	$\beta_3 - z\sigma_{\beta_3}$	$\beta_4 - z\sigma_{\beta_4}$	$\beta_5 - z\sigma_{\beta_5}$	$\beta_6 - z\sigma_{\beta_6}$
					R-square					
1	100	1	0.33	–	0.3486	-0.0001	0.0023	0.0276	-0.0009	-0.0243
					0.3388	-0.0006	0.0019	0.0271	-0.0014	-0.0248
					0.3584	0.0003	0.0028	0.028	-0.0005	-0.0239
					0.7537					
2	75	1	0.33	–	0.3421	0.0002	0.0023	0.0254	-0.0014	-0.0228
					0.3311	-0.0003	0.0019	0.0249	-0.0019	-0.0233
					0.3531	0.0006	0.0028	0.0259	-0.0009	-0.0223
					0.6962					
3	50	1	0.33	–	0.2988	0.0006	0.0025	0.0232	-0.002	-0.0208
					0.2861	0.0001	0.002	0.0227	-0.0025	-0.0213
					0.3114	0.0012	0.0031	0.0237	-0.0014	-0.0202
					0.5987					
4	25	1	0.33	–	0.1633	0.0013	-0.0011	0.0198	-0.0018	-0.014
					0.1475	0.0007	-0.0017	0.0191	-0.0025	-0.0146
					0.1791	0.0019	-0.0006	0.0204	-0.0012	-0.0134
					0.4145					
5	12	1	0.33	–	-0.1057	0.0017	-0.0034	0.0135	0.0008	-0.007
					-0.1238	0.0011	-0.004	0.0126	-0.0001	-0.0077
					-0.0875	0.0024	-0.0027	0.0145	0.0018	-0.0063
					0.2529					

Table 2.8: Regression Analysis - First Difference in Average π - Dynamically Relevant Devaluations

No.	Pop.	Rules	p_m	b	$L(1)\overline{\pi_t}$	$D6$	$D7$	$D8$	$D9$	$D10$
6	100	1	0.165	–	0.4591	0.0005	0.0014	0.024	-0.0051	-0.0178
					0.4457	-0.0001	0.0008	0.0235	-0.0057	-0.0184
					0.4725	0.0011	0.0019	0.0246	-0.0045	-0.0172
					0.5687					
7	75	1	0.165	–	0.4205	0.0006	0.0005	0.0238	-0.0056	-0.0165
					0.4061	0	-0.0001	0.0232	-0.0062	-0.0171
					0.4348	0.0012	0.001	0.0245	-0.0049	-0.0159
					0.5194					
8	50	1	0.165	–	0.3473	0.0013	-0.0015	0.0204	-0.0053	-0.0126
					0.3314	0.0007	-0.0021	0.0197	-0.0059	-0.0132
					0.3632	0.0019	-0.001	0.021	-0.0046	-0.012
					0.4043					
9	25	1	0.165	–	0.095	0.0012	-0.003	0.0176	-0.0049	-0.0071
					0.0769	0.0006	-0.0036	0.0167	-0.0058	-0.0077
					0.1131	0.0019	-0.0023	0.0185	-0.0041	-0.0064
					0.2049					
10	12	1	0.165	–	-0.2596	0.0016	-0.0052	0.0104	-0.0022	-0.0009
					-0.2781	0.0009	-0.0059	0.0091	-0.0035	-0.0016
					-0.241	0.0023	-0.0045	0.0117	-0.0009	-0.0002
					0.1653					
11	100	1	0.0825	–	0.3552	0.0004	-0.0022	0.0166	-0.0098	-0.0044
					0.3379	0.0001	-0.0025	0.016	-0.0105	-0.0047
					0.3725	0.0007	-0.0019	0.0173	-0.0092	-0.0041
					0.3187					
12	75	1	0.0825	–	0.335	0.0009	-0.0021	0.0178	-0.01	-0.0058
					0.3181	0.0005	-0.0025	0.0171	-0.0107	-0.0062
					0.352	0.0014	-0.0017	0.0185	-0.0093	-0.0054
					0.3019					
13	50	1	0.0825	–	-0.0062	0.0005	-0.0031	0.0106	-0.0054	-0.0013
					-0.0257	0.0002	-0.0034	0.0097	-0.0063	-0.0017
					0.0132	0.0009	-0.0027	0.0116	-0.0044	-0.001
					0.1396					
14	25	1	0.0825	–	-0.2739	0.0003	-0.0037	0.0086	-0.0049	0
					-0.2926	-0.0002	-0.0041	0.0073	-0.0062	-0.0005
					-0.2553	0.0008	-0.0032	0.01	-0.0036	0.0005
					0.1351					
15	12	1	0.0825	–	-0.3447	0.0004	-0.0038	0.0035	0.0001	0.0007
					-0.3631	-0.0001	-0.0043	0.0023	-0.001	0.0002
					-0.3264	0.0009	-0.0033	0.0046	0.0012	0.0012
					0.1647					

Table 2.9: Regression Analysis - First Difference in Average π - Dynamically Relevant Devaluations (Cont'd)

No.	Pop.	Rules	p_m	b	$L(1)\overline{\pi_t}$	$D6$	$D7$	$D8$	$D9$	$D10$
16	100	1	0.04	–	-0.1418	0	-0.0016	0.0021	-0.0002	0
					-0.161	-0.0002	-0.0018	0.0012	-0.0012	-0.0001
					-0.1225	0.0001	-0.0015	0.003	0.0007	0.0001
					0.1403					
17	75	1	0.04	–	-0.3487	0.0002	-0.0018	0.0027	-0.0014	0.0006
					-0.367	0	-0.002	0.0017	-0.0024	0.0005
					-0.3304	0.0003	-0.0017	0.0037	-0.0004	0.0008
					0.2303					
18	50	1	0.04	–	-0.2761	-0.0001	-0.002	0.0056	-0.004	0.0004
					-0.2948	-0.0003	-0.0023	0.0045	-0.005	0.0002
					-0.2574	0.0002	-0.0018	0.0066	-0.0029	0.0006
					0.133					
19	25	1	0.04	–	-0.2914	0.0003	-0.0022	0.002	-0.0003	0.0002
					-0.3102	0	-0.0025	0.0011	-0.0012	-0.0001
					-0.2727	0.0006	-0.0019	0.0029	0.0006	0.0005
					0.1234					
20	12	1	0.04	–	-0.1773	0.0002	-0.0027	0.003	-0.0001	0
					-0.1965	-0.0001	-0.0031	0.0023	-0.0009	-0.0004
					-0.1581	0.0006	-0.0024	0.0038	0.0007	0.0004
					0.0873					
41	100	45	0.0825	–	-0.0282	0.0013	-0.0025	0.0169	-0.0041	-0.01
					-0.044	0.001	-0.0028	0.0156	-0.0053	-0.0104
					-0.0125	0.0016	-0.0022	0.0181	-0.0028	-0.0097
					0.5738					
42	75	45	0.0825	–	0.2315	0.0014	-0.0024	0.0201	-0.0026	-0.0162
					0.2153	0.001	-0.0028	0.0191	-0.0036	-0.0167
					0.2476	0.0019	-0.0019	0.0211	-0.0016	-0.0157
					0.5489					
43	50	45	0.0825	–	0.2021	0.0016	-0.0025	0.0207	-0.0028	-0.0163
					0.1856	0.0011	-0.0029	0.0197	-0.0038	-0.0168
					0.2185	0.0021	-0.002	0.0217	-0.0018	-0.0158
					0.5162					
44	25	45	0.0825	–	-0.0286	0.0011	-0.003	0.0153	0.003	-0.0123
					-0.0463	0.0006	-0.0035	0.0141	0.0019	-0.013
					-0.0109	0.0017	-0.0024	0.0165	0.0042	-0.0117
					0.4626					
45	12	45	0.0825	–	-0.3512	0.0006	-0.0038	0.0068	0.0083	-0.0058
					-0.3684	0.0001	-0.0044	0.0052	0.0067	-0.0064
					-0.3339	0.0012	-0.0033	0.0084	0.0099	-0.0052
					0.4222					

Table 2.10: Regression Analysis - First Difference in Average π - Dynamically Relevant Devaluations (Cont'd)

No.	Pop.	Rules	p_m	b	$L(1)\overline{\pi_t}$	$D6$	$D7$	$D8$	$D9$	$D10$
51	100	15	0.0825	–	0.5554	0.0002	-0.0016	0.0286	-0.002	-0.0248
					0.5457	-0.0002	-0.002	0.0281	-0.0025	-0.0253
					0.5652	0.0006	-0.0012	0.0291	-0.0016	-0.0244
					0.7867					
52	75	15	0.0825	–	0.4519	0.0013	-0.003	0.0264	-0.0036	-0.0224
					0.4405	0.0009	-0.0034	0.0258	-0.0042	-0.0228
					0.4634	0.0017	-0.0026	0.0269	-0.003	-0.022
					0.7299					
53	50	15	0.0825	–	0.3606	0.0011	-0.0029	0.025	-0.0037	-0.0196
					0.3476	0.0007	-0.0033	0.0243	-0.0043	-0.0201
					0.3737	0.0015	-0.0026	0.0256	-0.003	-0.0192
					0.6928					
54	25	15	0.0825	–	0.1171	0.0016	-0.0042	0.0222	-0.0022	-0.0151
					0.1017	0.0012	-0.0046	0.0213	-0.0031	-0.0156
					0.1325	0.002	-0.0038	0.023	-0.0014	-0.0146
					0.6015					
55	12	15	0.0825	–	-0.1283	0.0016	-0.005	0.0165	0.0022	-0.01
					-0.1452	0.0011	-0.0055	0.0153	0.0011	-0.0105
					-0.1113	0.0021	-0.0045	0.0176	0.0033	-0.0094
					0.5288					
56	100	5	0.0825	–	0.504	-0.0001	-0.0021	0.0201	0.0009	-0.0158
					0.4956	-0.0003	-0.0023	0.0199	0.0006	-0.0161
					0.5124	0.0001	-0.0018	0.0204	0.0011	-0.0156
					0.8278					
57	75	5	0.0825	–	0.473	-0.0005	-0.0022	0.0198	0.0006	-0.0156
					0.4638	-0.0008	-0.0024	0.0195	0.0003	-0.0159
					0.4823	-0.0003	-0.0019	0.02	0.0009	-0.0154
					0.8					
58	50	5	0.0825	–	0.3996	-0.0002	-0.0024	0.0186	0.0005	-0.0149
					0.3889	-0.0004	-0.0027	0.0183	0.0002	-0.0152
					0.4102	0.0001	-0.0021	0.0189	0.0008	-0.0146
					0.7558					
59	25	5	0.0825	–	0.1876	0	-0.0035	0.0169	0.0005	-0.0122
					0.1741	-0.0002	-0.0038	0.0165	0.0001	-0.0125
					0.2012	0.0003	-0.0033	0.0173	0.0009	-0.0118
					0.6888					
60	12	5	0.0825	–	-0.0671	-0.0001	-0.0049	0.0123	0.0029	-0.0072
					-0.0834	-0.0004	-0.0052	0.0118	0.0023	-0.0076
					-0.0507	0.0003	-0.0046	0.0129	0.0034	-0.0068
					0.5856					

Table 2.11: Regression Analysis - First Difference in Average π - Dynamically Relevant Devaluations (Cont'd)

No.	Pop.	Rules	p_m	b	$L(1)\overline{\pi_t}$	$D6$	$D7$	$D8$	$D9$	$D10$
61	100	45	0.04	–	0.5592	0.0011	-0.0016	0.0266	-0.0044	-0.0226
					0.5481	0.0008	-0.002	0.026	-0.005	-0.023
					0.5702	0.0015	-0.0013	0.0271	-0.0038	-0.0222
					0.7513					
62	75	45	0.04	–	0.3356	0.0014	-0.0026	0.0215	-0.0056	-0.0142
					0.3206	0.0011	-0.0029	0.0207	-0.0064	-0.0146
					0.3505	0.0018	-0.0022	0.0223	-0.0048	-0.0138
					0.5874					
63	50	45	0.04	–	0.3152	0.0014	-0.0028	0.0213	-0.0047	-0.0148
					0.3	0.001	-0.0032	0.0205	-0.0055	-0.0153
					0.3303	0.0017	-0.0025	0.0221	-0.0039	-0.0144
					0.5811					
64	25	45	0.04	–	0.1147	0.0016	-0.0034	0.0181	-0.0012	-0.012
					0.0975	0.0011	-0.0038	0.0171	-0.0022	-0.0125
					0.1318	0.002	-0.003	0.0191	-0.0002	-0.0115
					0.5142					
65	12	45	0.04	–	-0.1427	0.0006	-0.0041	0.0134	0.0028	-0.0076
					-0.1606	0.0001	-0.0046	0.0121	0.0015	-0.0081
					-0.1247	0.0011	-0.0037	0.0146	0.0041	-0.007
					0.429					
71	100	15	0.04	–	0.5516	-0.0004	-0.0013	0.0232	-0.0001	-0.0195
					0.5431	-0.0006	-0.0016	0.0229	-0.0004	-0.0198
					0.5601	-0.0001	-0.001	0.0236	0.0002	-0.0192
					0.8267					
72	75	15	0.04	–	0.5302	0.0001	-0.0017	0.0228	-0.0001	-0.0196
					0.5214	-0.0002	-0.002	0.0225	-0.0004	-0.0199
					0.539	0.0004	-0.0014	0.0232	0.0002	-0.0193
					0.819					
73	50	15	0.04	–	0.4639	0.0003	-0.0022	0.0215	-0.0012	-0.0176
					0.4533	0	-0.0024	0.0211	-0.0016	-0.0179
					0.4745	0.0006	-0.0019	0.0219	-0.0009	-0.0173
					0.7725					
74	25	15	0.04	–	0.2378	0.0004	-0.0033	0.0185	-0.0012	-0.0127
					0.2234	0.0001	-0.0036	0.018	-0.0018	-0.0131
					0.2521	0.0007	-0.003	0.0191	-0.0007	-0.0123
					0.6537					
75	12	15	0.04	–	-0.0471	0.0003	-0.0049	0.0156	0.0001	-0.0079
					-0.0639	0	-0.0053	0.0148	-0.0007	-0.0083
					-0.0303	0.0007	-0.0046	0.0164	0.0008	-0.0074
					0.5316					

Table 2.12: Regression Analysis - First Difference in Average π - Dynamically Relevant Devaluations (Cont'd)

No.	Pop.	Rules	p_m	b	$L(1)\overline{\pi_t}$	$D6$	$D7$	$D8$	$D9$	$D10$
76	100	5	0.04	–	0.4222	-0.0003	-0.0013	0.0093	0.001	-0.0081
					0.411	-0.0004	-0.0014	0.0091	0.0008	-0.0083
					0.4335	-0.0001	-0.0011	0.0094	0.0011	-0.008
					0.7464					
77	75	5	0.04	–	0.3334	-0.0003	-0.0016	0.0089	0.001	-0.0073
					0.3211	-0.0005	-0.0017	0.0087	0.0008	-0.0075
					0.3457	-0.0002	-0.0014	0.009	0.0012	-0.0072
					0.7383					
78	50	5	0.04	–	0.246	-0.0002	-0.002	0.0082	0.0016	-0.0071
					0.2324	-0.0003	-0.0021	0.008	0.0013	-0.0073
					0.2596	0	-0.0018	0.0084	0.0018	-0.0069
					0.7067					
79	25	5	0.04	–	0.0513	-0.0002	-0.0029	0.007	0.0025	-0.0052
					0.0352	-0.0004	-0.003	0.0068	0.0022	-0.0054
					0.0673	0	-0.0027	0.0073	0.0027	-0.005
					0.6015					
80	12	5	0.04	–	-0.1553	0.0001	-0.0042	0.0048	0.0043	-0.003
					-0.1727	-0.0002	-0.0044	0.0044	0.0039	-0.0033
					-0.1378	0.0003	-0.004	0.0051	0.0046	-0.0027
					0.5372					
96	100	1	0.0825	1	0.3745	0.0007	0.0002	0.0374	-0.0052	-0.034
					0.3627	-0.0001	-0.0006	0.0366	-0.006	-0.0347
					0.3862	0.0015	0.0009	0.0382	-0.0044	-0.0332
					0.6479					
97	75	1	0.0825	1	0.3173	0.0017	-0.0011	0.0291	-0.0064	-0.0232
					0.3026	0.0009	-0.0019	0.0282	-0.0073	-0.024
					0.3321	0.0025	-0.0003	0.03	-0.0055	-0.0224
					0.4567					
98	50	1	0.0825	1	0.0218	0.001	-0.0041	0.0149	-0.0043	-0.0064
					0.003	0.0004	-0.0047	0.0139	-0.0053	-0.007
					0.0406	0.0016	-0.0035	0.016	-0.0032	-0.0058
					0.1766					
99	25	1	0.0825	1	-0.48	0	-0.0046	-0.0018	0.0039	0.0026
					-0.4971	-0.0003	-0.0048	-0.0043	0.0013	0.0023
					-0.4628	0.0002	-0.0043	0.0008	0.0064	0.0028
					0.3534					
100	12	1	0.0825	1	-0.4539	0.0008	-0.0059	0.0022	-0.001	0.0031
					-0.4712	0.0006	-0.0061	-0.0046	-0.0078	0.0029
					-0.4365	0.001	-0.0057	0.009	0.0057	0.0033
					0.4291					

Table 2.13: Regression Analysis - First Difference in Average π - Dynamically Relevant Devaluations (Cont'd)

No.	Pop.	Rules	p_m	b	$L(1)\overline{\pi_t}$	$D6$	$D7$	$D8$	$D9$	$D10$
116	100	15	0.0825	1	0.3808	0.0011	-0.0034	0.0376	-0.0031	-0.0318
					0.3717	0.0006	-0.0038	0.0371	-0.0036	-0.0323
					0.3899	0.0016	-0.0029	0.0381	-0.0025	-0.0313
					0.7984					
117	75	15	0.0825	1	0.3556	0.002	-0.0043	0.0371	-0.0027	-0.0325
					0.346	0.0014	-0.0048	0.0365	-0.0033	-0.033
					0.3652	0.0025	-0.0038	0.0377	-0.0022	-0.032
					0.7784					
118	50	15	0.0825	1	0.3079	0.0024	-0.0048	0.0368	-0.0031	-0.0307
					0.2972	0.0018	-0.0053	0.0362	-0.0037	-0.0312
					0.3186	0.0029	-0.0042	0.0375	-0.0024	-0.0301
					0.7414					
119	25	15	0.0825	1	0.1632	0.0014	-0.0046	0.0352	-0.0032	-0.026
					0.1503	0.0008	-0.0052	0.0344	-0.004	-0.0267
					0.1762	0.002	-0.004	0.036	-0.0024	-0.0254
					0.6374					
120	12	15	0.0825	1	-0.0145	0.0018	-0.0062	0.0327	-0.0033	-0.0202
					-0.0295	0.0011	-0.0068	0.0316	-0.0044	-0.021
					0.0006	0.0025	-0.0055	0.0338	-0.0023	-0.0194
					0.5432					

Table 2.14: Regression Analysis - First Difference in Average π - Dynamically Relevant Devaluations (Cont'd)

Population	Rules	p_m	b	$L(1)\overline{\pi_t}$	$D1$	$D2$	$D3$	$D4$	$D5$
				$\overline{\pi_t} = \beta_1(L(1)\overline{\pi_t}) + \beta_2(D1) + \beta_3(D2) + \beta_4(D3) + \beta_5(D4) + \beta_6(D5)$					
–	–	–	–	β_1	β_2	β_3	β_4	β_5	β_6
				t_1	t_2	t_3	t_4	t_5	t_6
				R-square					
15	–	–	–	-0.118476	-0.004642	-0.000609	0.037645	-0.003965	-0.031612
				-3.113637	-2.188383	-0.287191	17.79897	-1.874918	-14.90833
				0.784863					
11	–	–	–	-0.325833	0.000645	0.005392	-0.004772	-0.001404	-0.005087
				-3.774778	0.236037	1.989777	-1.374776	-0.390415	-1.821604
				0.178759					
11	–	–	–	-0.216998	-0.002724	-0.00406	0.012254	0.000447	-0.007116
				-2.979921	-1.24158	-1.869841	4.386003	0.158368	-3.111756
				0.380659					

Population	Rules	p_m	b	$L(1)\overline{\pi_t}$	$D6$	$D7$	$D8$	$D9$	$D10$
				$\overline{\pi_t} = \beta_1(L(1)\overline{\pi_t}) + \beta_2(D6) + \beta_3(D7) + \beta_4(D8) + \beta_5(D9) + \beta_6(D10)$					
–	–	–	–	β_1	β_2	β_3	β_4	β_5	β_6
				t_1	t_2	t_3	t_4	t_5	t_6
				R-square					
15	–	–	–	-0.057469	-0.004096	0.007844	0.031718	0.000465	-0.025934
				-1.105364	-1.521912	2.822085	11.21665	0.164437	-9.737971
				0.606854					
11	–	–	–	-0.319874	0.001194	0.00262	-0.007453	0.002164	-0.004763
				-3.691546	0.413502	0.89769	-1.684557	0.479884	-1.626236
				0.16009					
11	–	–	–	-0.196566	-0.001454	-0.005393	0.011502	0.000851	-0.005358
				-2.509163	-0.667641	-2.51832	3.555103	0.26349	-2.325622
				0.363499					

Table 2.15: Regression Analysis - First Difference in Average π - Treatment Results

Chapter 3

Economic Growth

Modelling economic growth with endogenous transition through technological paradigms

3.1 Introduction

The purpose of this paper is to model the process of technological transition at the firm level, and to investigate the implications of this model for macroeconomic aggregates. It will be argued that the slow diffusion of new technologies results from the fact that, although potentially more productive in the long run, these new technologies initially lack the accrued incremental innovations of their predecessor. They therefore are less productive during their infancy and diffusion of this new technology is slow. Additionally, this productivity gap between old and new technologies may cause temporary negative aggregate output shocks if a sufficient number of firms adopt the new technology simultaneously.

Long-run growth in income per capita requires increases in productivity. Some productivity improvements result from incremental innovations within a given technology being employed. They are a result of what is referred to as learning-by-doing, or learning-by-using. Improvements to productivity through incremental innovations occur with diminishing returns. As the return to productivity of learning-by-doing within a given technological paradigm diminishes, the only way to achieve further improvements is through technological advancement. Technological advancement refers to firm appropriation of newly discovered technologies with higher productivity potential. This is often referred to as radical innovation.

In the pursuit of profit maximization, firm's adopt new technologies only in order to appropriate the potential productivity improvements inherent within them. However, this potential level of productivity may only be achieved after considerable accumulation of incremental innovations. Prior to the accumulation of these innovations, a newly adopted technology may be less productive than its predecessor despite the fact that it has a higher overall productivity potential. Therefore, a firm may be required to accept a short run decrease in productivity in order to earn the higher productivity level of the new technology through learning-by-doing.

The manner in which innovation and technological progression is modelled within this work is consistent with some key macroeconomic phenomena. In particular, negative shocks to economic growth are a possible characteristic of paradigm transition at the firm level. Falls in productivity associated with technological progression result in periods of negative economic growth. These negative shocks to productivity are not exogenous, as is assumed in many works for simplification, but instead a result of the process of technological progression at the firm level.

Previous considerations of the importance of learning-by-doing include Parente (1994), Lucas (1993), and Jovanovic and Nyarko (1996). Each of these works examine the firms' decision regarding technological upgrades in light of it's expertise in it's current and potential technological grade. In order to highlight the differences between this work and it's predecessors, consider the work of Jovanovic and Nyarko (1996). Their work contains a one-agent, Bayesian model of learning-by-doing and technological choice. The firm is myopic, and maximizes current period return in each period by production utilizing a single technological grade. Experience yields information which raises productivity and improves decisions. This is modelled through Bayesian updating. Importantly, this information applies not only to the current grade of technology in use by the firm, but also superior grades; information is transferable. Finally, there is no recall of old technologies, and the size of the upgrade is limited.

Although firms within the model of this work are myopic, each is heterogeneous in their productivity characteristics. Experience yields information which raises productivity. However, transferability of this information is not a characteristic of the model herein. This characteristic stems from consideration of an important question. Is the process governing the transfer of knowledge on prior variance of newer grades of technology in the information set of the firm? In previous literature, it is assumed that firm's know the process and the manner in which it changes through progression of technological grades. This information pertaining to newer grades of technology is available even before the firm has devoted any resources toward it. The assumption maintained within this work is that the only manner in which a firm may learn about a newer grade of technology is to devote resources toward production within it. These resources take the form of productive assets, labour and capital. Firm's may contribute a portion of their resources toward a newer technological grade while maintaining production in it's predecessor. Therefore, while firms may not gain information regarding a newer technology without using it in the production process, it does not need to fully commit to a single technology at any given point in time.

The manner in which technological progression and learning is modelled in this work makes it definitively evolutionary. It is a model characterised by heterogeneity, experimentation, and selection. It has important advantages over other models in the literature in that it offers a natural model of experimentation by agents and allows consistency with Nelson and Winter's (1982) interpretation of Schumpetarian competition.

Nelson and Winter's (N&W) conception regarding the metaphorical evolutionary process of Schumpetarian competition yields model characteristics that are distinct from their interpretation of those inherent within "orthodox" economic considerations. First, they emphasise a population perspective wherein an 'industry' or 'economy' is seen as a taxonomic class incorporating a certain degree of variety of processes and/or products. This variety must, in principle, be transferable between different firms or agents. A certain similarity of the search spaces of firms is required for this to be possible. However, there may be major differences with respect to the 'distance' between different sources of knowledge (Andersen 1996).

Second, N&W heavily emphasise the importance of the natural introduction of variety and the economic selection over this variety. This variety introduction is founded on the individual's pursuit of non-normal profit. However, N&W consider this variety and selection only within an economic pattern. In other words, N&W emphasise change which follows "natural trajectories" within given "technological regimes" (N&W 1982, 258-262) rather than radical change.

According to Nelson and Winter, "a vast array of particular models can be constructed within the broad limits of the theoretical schema" but the "enormous generality" of the schema cannot be exploited immediately (N&W, 1982, 19). In order to obtain real understanding about how to handle their powerful family of models, N&W prefer to concentrate on "very simple examples" and to "distinguish sharply between the power and generality of the theoretical ideas we employ and the much more limited results that our specific efforts have yielded thus far." (N&W, 1982, 20). The model presented herein extends the complexity of the N&W examples in a manner that also extends what they refer to as their "limited" results.

The model is built from a simplified version of the MOSES training-and-innovation model proposed in the work of Ballot and Taymaz (1994, 1996, 1997, 1998). The MOSES model is a complete Micro-to-Macro simulation model of the (Swedish) economy. Their model highlights the interaction between human capital and innovation. Here, firms decide on the allocation of funds between training, R&D and production. These decisions in turn affect macroeconomic growth.

Our simplification of the model is for reasons of parsimony. In order to model technological transition in a manner that is consistent with negative macroeconomic output shocks, firms' investment in R&D is not strictly required. Nor, it will be argued, is the emphasis placed on R&D warranted when one looks at a significant technological transitions of the past; specifically, the Industrial Revolution.

In the following section, we highlight the fact that a key element of technological advancement, and therein economic growth, is learning and adaptation on the part of economic agents employing the new technology. Section 3.3 of this paper will explore the previous examinations by Ballot and

Taymaz, and highlight the motivation for the proposed simplification. Section 3.4 will examine in detail the algorithm used in order to model economic growth in this piece. Section 3.5 examines some of the key results of the model's simulation, and Section 3.6 concludes.

3.2 Background - The Industrial Revolution as Transition Through Technological Paradigms

3.2.1 Some empirical regularities and the slow transition to the factory system

Five explanations for the slow transition to the factory system are summarized by Pereira (2002): (1) the competitiveness of the putting-out system, (2) interest groups, (3) the low margin of efficiency of the new factories, (4) social learning and technological spillovers, and (5) the "bandwagon" or "gold rush" effect. Herein, our concern will be in capturing the latter three of these five in a model of technological progression. We examine these three below.[1]

Low margin of efficiency of the new technologies of factories - Experimentation As argued by Pereira, at the inception of the Industrial Revolution, early factories were not more effective than historic industries. Technologies were crude and took time to become fully operational and productive. Furthermore, there were many technical difficulties associated with the development of some technologies which prevented their earlier diffusion. Consequently, the efficiency of the new technologies was initially low, but slowly improved during a period of highly intensive learning-by-doing. After inventors, technicians and factory workers solved these initial technical problems, the productivity of the sectors associated with the new technology increased rapidly.[2]

Social Learning - Imitation Dissemination of knowledge among potential industrialists was crucial for the diffusion of not only new machines, but also of the factory system. As Aghion and Howitt argue:

> The way that a firm typically learns to use a new technology is not to discover everything on its own but to learn from the experience of other firms in a similar situation, namely other firms for whom the problems that must be solved before the new technology can be successfully be implemented bear enough resemblance to the problems that must be solved in this firm. (Alghion and Howitt, 1988, pp.129)

[1] Atkenson and Kehoe (2003) examine similar characteristics in their consideration of the slow adoption of technologies associated with the Second Industrial Revolution

[2] For an overview of the evidence, See Pereira (2002)

Critical Mass and the "bandwagon" effect - Selection Early examples of proto-factories were not totally uncommon before the Industrial Revolution. Industrial success, however, was a phenomenon that began mostly after the Industrial Revolution. Most proto-factories did not manage to survive for considerable periods of time (Crouzet, 1985). Furthermore, most of them did not employ mechanical machines (Landes, 1986). Nonetheless, proto-industrialization shows that there was a long trajectory of mechanization that stretches back to earlier decades and, in some cases, centuries (Bekar and Lipsey, 2001).

What changed? After early industrialists such as Arkwright and Watt obtained spectacular profits with the new factories, a bandwagon effect ensued and factories of all sizes sprung up. As argued by Rosenberg (1996), pervasive uncertainties are often the norm in the development and application of new technologies. As argued above, several technical problems complicated an entrepreneur's decision of whether or not to invest in new technologies. Investing in new technologies was an expensive and risky business in which the distribution of incomes was truly uncertain. However, this distribution of incomes was likely skewed to the lower end, as can be attested by the relatively high number of bankruptcies during the end of the process of proto-industrialization and the early stages of the Industrial Revolution. Many investors preferred either to invest elsewhere, or to delay their investments, rather to engage in the risky endeavor (Crouzet, 1985; Pollard, 1965). These problems were eventually solved by social learning and the achievement of a critical mass in the new technology, as emulation could now occur more prominently and profitably allowing the survival rate to increase.

Intermediate Adoption In their consideration of the Second Industrial Revolution, the adoption of the modern technology of electricity, Devine (1983) and David (1990, 1991) stress the complete redesign of the manufacturing process that accompanied this transition. Technology and the organization of the manufacturing process are two sides of the same coin. While the manufacturing processes associated with old and the new technology were radically different, the transition occurred through a process of evolution. Between the old manufacturing setup and that associated with electricity were two intermediate stages in which the higher technology was mixed with older styles of the manufacturing process. These intermediate stages represent periods in which firms were not fully committed to the newer technology and its ideal form of organization. Their production process shared technologies from both the new technology and its predecessor.

3.2.2 Summarizing the qualitative features of technological transition

Summarizing the above discussion pertaining to the Industrial Revolution, the technologies encompassing the Industrial Revolution were available long before they were broadly appropriated by firms. Firms did not invest in new technology, as it was inferior to its predecessors in terms of productivity. Those firms that did invest in the new technology were failures. Social learning was an important factor bringing the productivity of new technologies in line with, then surpassing that of old technologies. Once this occurred, the profits earned by those firms utilizing new technologies drew other firms into this new technological paradigm.[3]

The Industrial Revolution has been the best documented and arguably the most important technological transition. Our goal throughout the remainder of this work will be to model the process of technological transition at higher frequencies in a manner consistent with these qualitative features. Therein, it is assumed that these important regularities are present in all transitions, though possibly in smaller magnitudes depending on the nature of the newer technology.

3.3 Background - The theoretical framework

3.3.1 The general approach of Ballot and Taymaz

A model that captures some of the aforementioned historical regularities of technological transition and incorporates the importance of learning is contained in various works by Ballot and Taymaz. In these works, long run growth in productivity is achieved only through transition towards superior technological paradigms. Each paradigm has an upper bound with respect to improvements in productivity obtained through learning-by-doing.

A technology is represented in the model by a set of *techniques*. Each technique is assumed to take only one of two possible values: 0, or 1. For each technological paradigm, there is an optimal organization of techniques that guarantees maximum performance in terms of capital and labour productivity. This is referred to as the *global technology*. The technological level of the firm within a paradigm is measured by its closeness to the global technology. A particular paradigm's optimal organization of techniques is distinct from that of any other paradigm's. Therein, each paradigm may be characterised by the organization of this given set of techniques and the resulting productivity that this optimal organization of techniques results in.

Genetic algorithms are used as a tool to generate new technologies within a paradigm. Firms recombine their own sets of techniques to obtain new ones, recombine their sets with those of other firms, or invent new sets entirely. These analogies pertain to the genetic opererators of recombination, imitation and mutation, respectively. Only innovations that improve productivity are adopted (the genetic election operator).

Firms must allocate available resources towards different uses. They must invest in physical

[3]Alcaly (2003) also considers previous cycles of technological change and economic reaction, such as the invention of steam power and later electric power, the development of the internal combustion engine and adoption of mass production techniques in automobiles and steel. Comparing innovations in semiconductors, software and communications technology with those of earlier periods suggests the traumas of the last few years, including the Internet boom and crash, are predictable growing pains. He argues that such changes do create new economies that are qualitatively better than the economies they replace, but more slowly and erratically than people expect at the time and with bigger problems along the way.

assets since they embody new technology and because they depreciate. They must invest in specific and general human capital in order to facilitate incremental innovations and imitation of other firms, respectively. Contrary to specific human capital, general human capital is transferable and not a direct factor in production. Finally, they should invest in R&D in order to facilitate radical innovations. Profits result from the market process. It is assumed that the precise relation of the above expenditures is far too complex to be fully understood by the firm. Consequently the firms' decisions must be modelled as boundedly rational rules with integrated learning.

3.3.2 The transition between technologies

It is important to note that for paradigm transition in the Ballot and Taymaz framework requires each firm to devote resources directly into accumulation of general human capital and R&D, neither of which contributes directly to production. This pulls resources away from the actual production process. If enough firms engage in such investment, there is the theoretical possibility of a negative output shock. This possibility interferes with the examination of learning-by-doing for creating falls in productivity sufficient for negative output growth.

Ballot and Taymaz assume that progression requires that at least one firm engages in sufficient R&D investment to facilitate a radical innovation. Once this radical innovation occurs, the firm in which it took place is forced into producing only in the new technological paradigm. Only then may other firms imitate the transition to the new paradigm.[4] Contrary to the Ballot and Taymaz approach, in the model presented in this work, all firms have available the opportunity to produce in the newer paradigm. Firms are not required to invest in R&D in order to facilitate production in the new paradigm.

Radically innovative firms still face the same difficulties as in the Ballot and Taymaz setup. The first firms to implement a radical innovation may not be very successful since it may be less productive, as the technology in the lower paradigm has been improved through incremental innovations. Notice that the potential for negative macroeconomic output shocks is still inherent. Although a new technological paradigm has more productivity potential, it has not undergone the incremental innovations that the older paradigm has. It therefore may be less productive. If this is the scenario, a negative output shock may be observed if a sufficient number of firms shift production into the newer paradigm. Importantly, this captures an aforementioned feature of the Industrial Revolution; transition into the new technologies was hampered due to the fact that these were temporarily less productive than older ones. Additionally, firms did not require an outright investment in R&D in order to take advantage of new technologies. Firms had the technology at there disposal and chose to continue using its predecessors.

[4]Note that if the firm that originally observed the radical innovation fails, radical imitation is no longer possible. A new firm must go through the same process of radical innovation in order to facilitate progression. Although the radical innovation has already taken place, because there is no firm producing in the newer paradigm another radical innovation must take place before imitation is available.

3.3.3 Technological transition without R&D

In dropping the assumption that R&D is required by firms prior to adoption, we require a new process for the transition through paradigms. In the model presented here, it is assumed that firms may devote a chosen percentage of their labour force to production in the newer paradigm. This simplifying assumption allows for incremental innovation by all firms, regardless of scale. Again, they need not discover this paradigm through radical innovation or imitation. However, in order to maintain production in a paradigm, they must devote a minimum percentage of their labour towards it.

Each paradigm has a minimal level of labour required in order to make it available as a viable production technology. If at any time the firm does not devote sufficient labour to a technology, it becomes unavailable for use in the production process. Importantly, if a firm splits its labour between two technological paradigms, and the minimum labour investment is maintained in each, the firm may produce using both technologies. This is distinctly different from the Ballot and Taymaz setup. In the model herein, firms devote labour to technological paradigms in a continuous manner. In their work, once a firm makes a radical innovation or radically imitates, it is forced into producing with the newer technology in full. Firms in their model cannot adopt a new technology incrementally.

As in the Ballot and Taymaz framework, firms make incremental innovations through learning-by-doing and imitation of other firms that are producing using the same technological paradigm. This is modelled using genetic learning operators. However, firms learn only according to the relative division of labour. That is, their ability to achieve incremental innovations within a paradigm is directly correlated to their choice regarding the division of labour applied to production using this technology.

This setup has an important characteristic. A motivation of this work is in detailing the possible sufficiency of learning-by-doing for negative macroeconomic output shocks. This model of learning makes it more unlikely for these to occur. As noted above, a barrier for firms' progression into newer paradigms is the possibility that these technologies are less productive, as they have not accumulated the incremental innovations that the preceding paradigms have. If firms are given the ability to engage in production using the new technology continuously rather than discretely, they may remove this barrier by accumulating incremental innovations in the newer paradigm while not devoting all of their production capabilities towards it. Essentially, firms have the ability to lessen the overall productivity effects on their production process by slowly moving into the newer paradigm rather than shifting all of their resources in whole.

It is important to note that the model presented here differs in another fundamental manner from that of Ballot and Taymaz. As discussed below, some markets are simplified from that of the Ballot and Taymaz simulations; particularly the labour market and the removal of intermediate goods for production. Ballot and Taymaz rely on the MOSES model of the Swedish economy; a model that is beautiful in its complexity.[5] The work presented in this paper outlines a stand-alone, self-contained simulation that is accessible and easily adapted for other considerations.

3.4 The Model

3.4.1 Setup

Firms' Variables - Initialization

Prior to simulation, firm specific variables and characteristics are initialized. This initialization occurs in what is referred to as period zero. N firms are created and for each firm, $2x$ technology sets are drawn at random. The technology employed by a firm can be represented by a number of "techniques", $F^P = \{f_1^P, f_2^P, ..., f_k^P\}$, where F^P is the technology used by the firm in paradigm P, and f_i^P is the i'th technique. A technique is assumed to have only one of two possible values, $f_i^P \in \{f_i^{1P}, f_i^{2P}\} = \{0, 1\}$. In simulations, the firm and global technology will be represented by a k element binary vector. We refer to the global technology as T^P. Notice that technology sets are paradigm specific, P. The firm carries with it x sets per paradigm. A firm may only produce in two paradigms at a time. Therefore, they must carry with them $2x$ total sets.

Each firm is endowed with a paradigm variable, p_j $(j = 1..N)$, which defines the lowest paradigm it is currently producing in. This variable is set to 1 for all firms in period zero. Additionally, each firm is endowed with a "switch gene". This gene is a binary string that will be converted into a real number in each period in order to determine the relative use of a firm's labor between two relevant paradigms. The relevant paradigms are p_j and $p_j + 1$. These are the paradigms that, at any given point in time, a firm may devote labor and capital for production. For all firms, in period zero it is set to a string that when converted is equal to zero; i.e. the binary null. We refer to the real value equivalent of the "switch gene" for firm j as ψ_j.

All firms undergo selection every m periods (see below). Selection is the process in which firms update their paradigm variable in order to mimic more successful firms, or experiment with new values in order to capture non-normal profits. Firms are separated into groups of equal size, N/m. A firm undergoes selection with the other members of its group, each group due for selection sequentially.

Global Variables - Initialization

In addition to initializing firm specific characteristics, the global variables of the simulation require initialization. By global variable, we refer to variables shared by each firm throughout the entire simulation.

[5] See Eliasson (1991), and Taymaz (1991) for a description of the MOSES model and data set

First, an upper bound is placed on the number of paradigms, $\overline{P}$.[6] For each paradigm $p < \overline{P}$, the global technology set, T^P, is drawn by randomly choosing the bits contained in each binary string.

Second, a minimal labor investment is created for each paradigm, $m_p \in [0, 1] \quad \forall \quad p \quad | \quad 1 \leq p \leq P$. If at any time a firm does not invest at least m_p percent of its labor in paradigm p, it cannot produce in that paradigm. A firm produces in two paradigms only if $m_{p_j+1} < \psi_j < 1 - m_{p_j}$.

3.4.2 A simulation period

Following initialization, the simulation cycles through a predetermined number of periods, T. Each simulative period is characterised by each firm and household undergoing the period stages outlined below.

Modelling Households - Aggregate Input Supply

Households are modelled using an overlapping generations framework where the total population in every period equals $2L$, equally divided between young and old. We use the notation that subscripts denote birthdates and parentheses denote real time. Individuals born within a generation are indexed by a superscript $i \in (1, 2, ..., L)$. Utility for the individual is defined over consumption in each period of life according to the following equation.

$$U_i = ln(c_t^i(t)) + ln(c_t^i(t + 1)) \tag{3.1}$$

Only young individuals have the opportunity and ability to work. Each young individual is endowed with a normalized unit of time with which she may engage in labour in order to earn wages. We abstract from the labour-leisure choice potentially facing individuals and simply assume that each agent enjoys no disutility from working. As such, each young individual supplies this normalized unit of labour inelastically, earning the wage $w(t)$ when young.

Other than this normalized labour unit, agents are born with no endowment. The single, perishable good produced in this economy may either be consumed or used as an input into production. Therefore, agents have the following lifetime budget constraint over consumption.

$$c_t^i(t) + \frac{c_t^i(t+1)}{1 + r(t+1)} \leq w(t) \tag{3.2}$$

Given the lifetime budget constraint and utility function of the individual, maximization yields the following equations for aggregate labour and capital supply.

$$\overline{L}^s(t) = L \tag{3.3}$$

[6] In no simulations is this upper bound binding.

$$\overline{K}^s(t) = \frac{1}{2}Lw(t-1) + \frac{1}{2}\sum_{j=1}^{J}\pi_j(t-1) \quad (3.4)$$

In the above equation, it is assumed that profits, $\pi(t-1)$, are re-invested at the same ratio as wages, $w(t-1)$.

Modelling Firms

Pre-production For each firm, the following steps are taken prior to production. First, each firms' "switch gene" is converted into its real number equivalent, ψ_j. Labour is assigned to produce in the paradigm $p_j + 1$ according to the ratio given by ψ_j, the rest of the labour available to the firm $(1 - \psi_j)$ is assigned to production in paradigm p_j.[7]

The degree of correspondence (DC) is calculated for the paradigms in which a firm is producing $(i = p_j, p_j + 1)$ according to the equation

$$DC_{j,i} = \sum_{z=1}^{k+i-1} a_z w_z. \quad (3.5)$$

$$a_z = 0 \quad if \quad t_z \neq f_z, \quad a_z = 1 \quad if \quad t_z = f_z. \quad (3.6)$$

where w_z is the weight for the technique z; t_z^P and f_z^P denote techniques of T^P (the global technology set) and F^P (the firm's technology set) respectively. The parameter k denotes the size of technology sets relevant for the first paradigm of production. Note that the size of technology sets is incremented for paradigms subsequent to the first. This incrementing is intended to capture the assumption that higher, more productive technological paradigms are more difficult for firms to master.[8]

Of each firms' x possible technology sets per relevant paradigm, only that with the highest degree of correspondence is used in order to determine the technological level. Others are carried in order to capture firms' technological memory, the importance of which will be clear after a discussion of the learning process firms' undergo.

The technological level of the firm is now computed by an exponential function of the DC value, according to

$$A_{j,i} = \Lambda^i exp(\lambda^i DC_{j,i}). \quad (3.7)$$

where Λ and λ are free parameters of the model.

[7] Note that this is determined prior to the firms' labour market activities.

[8] The increasing size of the string representing a firms technology is not a characteristic of the original Ballot and Taymaz works.

The production function Firm production occurs in each of the two relevant paradigms (p_j,p_j+1) according to the following equations.

$$Y_{j,p_j} = (K_{j,p_j})^{\alpha}(A_{j,p_j}(1-\psi_j)L_j)^{\beta} \tag{3.8}$$

$$Y_{j,p_j+1} = (K_{j,p_j+1})^{\alpha}(A_{j,p_j+1}\psi_j L_j)^{\beta} \tag{3.9}$$

Total production for the firm is simply the sum of firm production in the two relevant paradigms.

$$Y_j = Y_{j,p_j} + Y_{j,p_j+1} = ((K_{j,p_j})^{\alpha}(A_{j,p_j}(1-\psi_j))^{\beta} + (K_{j,p_j+1})^{\alpha}(A_{j,p_j+1}\psi_j)^{\beta})L_j^{\beta} \tag{3.10}$$

Profit maximization occurs myopically with respect to time. That is, it is assumed that each firm maximizes intra-period profits at all points in time. Using parentheses to denote discrete time, in each period firms maximize the following profit equation.

$$\pi_j(t) = Y_j(t) - r(t)(K_{j,p_j}(t) + K_{j,p_j+1}(t)) - w(t)L_j(t) \tag{3.11}$$

The price of the output is normalized to unity. Therefore, $r(t)$ and $w(t)$ denote both nominal and real rental and wage rates respectively.

Aggregate Input Demand - Profit Maximization For each firm, labour demand is determined according to the following unconstrained maximizing demand equation.

$$L_j^d = (\frac{w}{\beta})^{(\frac{\alpha-1}{1-\alpha-\beta})}(\frac{r}{\alpha})^{(\frac{-\alpha}{1-\alpha-\beta})}[(A_{j,p_j}(1-\psi_j))^{(\frac{-\beta}{\alpha-1})} + (A_{j,p_j+1}\psi_j)^{(\frac{-\beta}{\alpha-1})}]^{(\frac{1-\alpha}{1-\alpha-\beta})} \tag{3.12}$$

Total capital investment for firm j, $(K_{j,p_j} + K_{j,p_j+1} \equiv K_j)$, is determined by the following profit maximizing equation.

$$K_j^d = (\frac{w}{\beta})^{(\frac{-\beta}{1-\alpha-\beta})}(\frac{r}{\alpha})^{(\frac{\beta-1}{1-\alpha-\beta})}[(A_{j,p_j}(1-\psi_j))^{(\frac{-\beta}{\alpha-1})} + (A_{j,p_j+1}\psi_j)^{(\frac{-\beta}{\alpha-1})}]^{(\frac{1-\alpha}{1-\alpha-\beta})} \tag{3.13}$$

Each firm is identical in the production function parameters α and β. However, firms are heterogeneous in their technological levels (A_{j,p_j},A_{j,p_j+1}) and switch gene (ψ_j) characteristics.

For each firm j, we define and calculate the following.

$$\chi_j = [(A_{j,p_j}(1-\psi_j))^{(\frac{-\beta}{\alpha-1})} + (A_{j,p_j+1}\psi_j)^{(\frac{-\beta}{\alpha-1})}]^{(\frac{1-\alpha}{1-\alpha-\beta})} \tag{3.14}$$

Summing total labour and capital demand over all firms, we now have the following aggregate demand for each input.

$$\overline{L}^d(t) = \sum_{j=1}^{J} L_j^d(t) = (\frac{w(t)}{\beta})^{(\frac{\alpha-1}{1-\alpha-\beta})}(\frac{r(t)}{\alpha})^{(\frac{-\alpha}{1-\alpha-\beta})}\sum_{j=1}^{J}\chi_j(t) \tag{3.15}$$

$$\overline{K}^d(t) = \sum_{j=1}^{J} K_j^d(t) = (\frac{w(t)}{\beta})^{(\frac{-\beta}{1-\alpha-\beta})}(\frac{r(t)}{\alpha})^{(\frac{\beta-1}{1-\alpha-\beta})}\sum_{j=1}^{J}\chi_j(t) \tag{3.16}$$

Input Market Clearing

Wages and rental rates are determined through the simultaneous solution to the following two equations.

$$\overline{L}^d(t) = \overline{L}^s(t) \tag{3.17}$$

$$\overline{K}^d(t) = \overline{K}^s(t) \tag{3.18}$$

There is no unemployment in this model. The simultaneous solution to these two equations yields market clearing wage and rental rates, $w^*(t)$ and $r^*(t)$.

Production

Each firm hires a profit maximizing quantity of labour and capital, $L_j^d(t)^*$ and $K_j^d(t)^*$, determined by the substituting $w^*(t)$ and $r^*(t)$ into equations (3.12) and (3.13). Given this profit maximizing labour demand, each firm divides its total capital demanded into production in the two relevant paradigms according to the following equations.

$$K_{j,p_j}^d(t)^* = (\frac{r^*(t)}{w^*(t)})^{\frac{1}{\alpha-1}}(A_{j,p_j}(t)(1-\psi_j(t))L_j^d(t)^*)^{\frac{-\beta}{\alpha-1}} \tag{3.19}$$

$$K_{j,p_j+1}^d(t)^* = (\frac{r^*(t)}{w^*(t)})^{\frac{1}{\alpha-1}}(A_{j,p_j+1}(t)\psi_j(t)L_j^d(t)^*)^{\frac{-\beta}{\alpha-1}} \tag{3.20}$$

A firm's total production and profit are now determined by substituting the optimal labour and capital demands into the firm specific production function.

Selection

In every period, N/m firms consider altering a key characteristic. This characteristic is their division of labour between paradigms; their "switch gene". This process is referred to as selection. As described above, a firm undergoes selection only every m periods. The liklihood of selecting a different characteristic is dependent not only on a firm's "fitness", but also that of other firms considering selection.

Fitness The fitness of firms is determined by the weighted sum of historical profits during the previous m periods.[9] The current period's profit level is included in this history.

Therefore, fitness for firm j at time t is determined by

$$\mu_{j,t} = \sum_{q=1}^{m} [\Phi_q \pi_{j,t-q+1}] \tag{3.21}$$

where Φ_q refers to the weight placed on each individual element of the summation and $\pi_{j,t}$ refers to the profits of firm j in period t. Note that during selection, firms evaluate the performance of only the current fitness level, based on the preceding m periods, not the complete history of firm profits.

Although only a subset of N/m firms consider selection, the fitness value for all firms is calculated in every period. Fitness values of all firms are required to enable imitation of technology sets and "switch genes". In the first m periods of the simulation, those firms that are due for selection have not lived a full m periods. In these periods, fitness is based on the summation over only the number of periods they have lived.

Rank and Selection After the fitness values of all firms are calculated, the fitness values of those firms considering selection are ranked. The bottom ϕ percent of these firms' will replace their current "switch gene".[10] This selection procedure is analogous to the (μ, λ)-selection process described in the literature of genetic algorithms. While all firms currently considering selection are eligible to alter their "switch gene", those firms under-performing relative to the whole necessarily attempt to imitate the characteristic from one of the more successful firms. However, this does not preclude other firms considering selection from also altering their division of labour. This process is modelled utilizing an evolutionary algorithm outlined below.

[9]Note that there is no reason these weights need necessarily sum to 1

[10]Note that the replacement rate at this point is constant. However, I will be able to track the distribution of labor between paradigms of those firms falling to the selection process. That is, the ψ_j of bankrupt firms. An alternative to this rather simple selection procedure will be outlined below.

Evolutionary Algorithm

The simulation now progresses into the "genetic-learning" phase in which firms attempt to improve their profitability by improving their technological capabilities (technology sets) and their choice of relative production between their two relevant technological paradigms ("switch gene"). This is accomplished by experimenting with new characteristics, or imitating those of other firms. Imitation is not done blindly; a firm will only attempt to copy the characteristics of another if it perceives the comparison firm as more profitable. In each period, every firm attempts to improve their technological capabilities. However, only firms currently considering selection attempt to improve their division of labour ("switch gene").

Given a firm is attempting to imitate another's characteristics, it must first select a firm for comparison. The probability of firm j is selected for comparison is equal to that firm's relative fitness, computed using the following equation

$$Pr_t^j = \frac{\mu_t^j}{\sum_N^{i=1} \mu_t^i} \tag{3.22}$$

Technology sets For each firm, learning within a paradigm is a function of the proportion of labour that is devoted to production within that paradigm; equal to the real value equivalent of the "switch gene", ψ_j. A random number is drawn from the uniform distribution over the interval $[0, 1]$. This number is compared to the real value equivalent of the switch gene of each firm. If the number is greater than the gene, learning occurs only in paradigm p_j. If it is less than or equal to the gene, a firm's learning occurs only in paradigm $p_j + 1$.[11]

Once the paradigm in which learning may take place is determined, a firm is drawn from the set of all firms that have capital devoted to production in this paradigm. The probability of firm j is selected for comparison is equal to that firm's relative fitness to all other firms with production in the paradigm in which learning is to occur.[12] We refer to this firm as the *comparison firm*.

The firm has a choice to either spend time working with its own technology sets, or to look to other firms' sets. The firm compares its fitness value to that of its comparison firm. If its fitness value is higher than that of the comparison firm it works solely with its own technology sets; *recombining* them in an attempt to increase productive efficiency. If the comparison firm's fitness is higher, the firm will attempt to *imitate* the comparison firm.

Notice that the use of relative fitness is an imperfect signal of the comparison firm's degree of correspondence. A comparison firm may have a superior degree of correspondence in the paradigm in which learning is to occur, but a lower relative fitness if, for example, they are producing heavily

[11] This is not to literally imply that learning may only occur in one paradigm at a time; only that, over many periods the amount of learning that may take place in a specific paradigm is a function of the amount of labor devoted to that paradigm

[12] Notice that the summation in the above equation will not necessarily be over all N firms

in a newer paradigm in which accumulated learning is small. We will assume, however, that this is the best signal available to the firm. This assumption requires that it is very costly to determine the actual production practices (other than the relevant paradigm) of another firm with respect to the division of labour. Although firms can achieve full information regarding another firm's production practices, costs are high enough that they are constrained to knowledge pertaining to only a single firm in every period. It selects the rival for comparison by comparing relative fitness.

Recombination. If a firm determines it is to work with its own technology sets, it selects two of the x sets randomly from those specific to the paradigm it is learning in. It selects randomly a crossover point for one of the technology sets. A bit is selected as a crossover point with probability of $1/(k+p-1)$; each bit is equally likely for selection.[13] Each of the technology sets is broken at this crossover point, yielding four subsets. One of the subsets for each technology set is switched with that of the other, giving two new and distinct technology sets. Next, mutation occurs. Each binary bit from the two new technology sets is inverted with probability $Pr^{DC}_{m|r}$.

The firm now has $x+2$ technology sets. These sets are ranked by the magnitude of their degree of correspondence, DC. The bottom two sets in this ranking are dropped, leaving once again only x sets in the paradigm. This final process replaces an "election operator".

Imitation. If the firm is attempting to imitate the production process of the comparison firm, it takes the technology set used by the comparison firm in whole and adds it to the x sets relevant to the paradigm of learning. This set then undergoes mutation. Each bit in the binary string of the imitated technology set flips with probability $Pr^{DC}_{m|i}$.[14] The $x+1$ technology sets are then ranked according to their DC, and the lowest one is dropped, leaving once again x technology sets relevant to the paradigm in which learning occurred.

Importantly, *Recombination* or *Imitation* with respect to the techniques a firm employs occurs before that of the "switch gene" (using last periods switch gene) as they are intended to capture learning by doing and imitation in the previous period.

"Switch Gene" Only firms that underwent selection may alter their "switch gene". This is in order to capture the assumption that research, development, and learning is a medium to long-term agenda. Firms will not make large decisions in technological focus on an inter-period basis, rather they commit m periods to a new technology in order to reap payoffs; understanding that there is a great deal of experimentation and learning-by-doing that must occur before the potential technology becomes productive.

A firm that is eligible to alter its switch gene may try something drastic or revolutionary (experimentation, create a new, totally random switch gene); or try to imitate another firm's; or do nothing.

[13]The length of these binary strings is $k+p-1$, where p is the paradigm in which learning takes place

[14]Mutation may be interpreted here as either directed experimentation on the part of the firm trying to imitate, or as representing an imperfect process of mimicking the comparison firm

The probability that it will attempt to imitate another firm's switch gene is equal to Pr_i^{ψ}. Otherwise, with probability $(1 - Pr_i^{\psi})$, it randomly selects a new switch gene. If the firm is under-performing relative to others considering selection (i.e., a firm ranked in the bottom ϕ percent), it may only work with imitation. These firms may not experiment with its switch characteristic; it must imitate one of a more successful firm; that is, $Pr_i^{\psi} = 1$.

Imitation. A firm selects another from the *whole set of firms* for comparison. A firm has a probability of being selected according to its relative fitness value described above. The firm then compares its fitness to that of the randomly selected firm. If the fitness of the comparison firm is higher, it imitates the comparison firm's gene with mutation. That is, it takes their gene in whole. Each binary bit of this gene has a probability of mutating (binary switching) equal to $Pr_{m|i}^{\psi}$. Importantly, if the firm being imitated does not have the same value for p_j, the switch characteristic being imitated is set to binary null if $p_j > p_i$, and unity if $p_j < p_i$; where firm i is being imitated by firm j. If the firm being compared has a lower fitness value, the comparer does not imitate. If the relative fitness of the comparison firm is lower than the firm attempting to imitate, nothing is done to the imitator's switch gene. If the imitating firm is one that is under-performing, it does not compare fitness values; it simply imitates with mutation the switch characteristic of the comparison firm.

Experimentation If the firm does not attempt to improve via imitation of a more fit firm, it experiments with a brand new switch gene. A new switch gene is drawn randomly, with each bit of the gene having a probability of 0.5 of taking the value of unity. This has the effect of drawing a random ψ_j from the uniform distribution over [0,1], according to the precision dependent on the size of the string representing the gene.[15]

Finally, if after altering its switch gene, the real value of this gene does not maintain the minimum level of labour investment in the lowest applicable paradigm, p_j; i.e. $1 - \psi_j < m_j$, the switch characteristic is set to zero and the firm specific paradigm parameter, p_j, is incremented by one. A new set of x technology sets are drawn randomly for the paradigm $p_j + 2$, and those technology sets for the paradigm p_j are dropped. The firm may now only produce in paradigms $p_j + 1$ and $p_j + 2$.

After a firm falls below the minimum investment for a specific paradigm, it can no longer produce in that paradigm. It has progressed fully into a newer paradigm. The firm will never produce in the older paradigm again, it may only progress to newer ones. This is because if it attempts to imitate a more successful firm that is still producing in the older paradigm, it will imitate a switch value of zero, limiting its progression into the next paradigm, but never pulling it back into an older one.

Following the evolutionary learning facet of the simulation, a period ends. The simulation progresses by moving into the aggregate input supply stage of the algorithm.

[15]Notably, an election operator would not quite fit with such experimentation; since we are swimming in true uncertainty. The idea is that the firm is taking a risk; wagering short run losses against future gains. Any election operator would be required to look into the future; a future no firm could predict.

Selection - Firm Bankruptcy

As modelled, a fixed and exogenous number of firms undergoing the selection process are forced to imitate another's switch gene each period. An alternative specification of the model could entail endowing each firm (at its inception) a given level of cash or wealth. As firms accumulate profits and losses, this level of wealth would be adjusted accordingly. A firm would only be forced to imitate during selection if, and only if, its level of wealth fell below a given level; most intuitively this level would be zero.

While a bankruptcy criterion for selection has some intuitive characteristics, it lacks a notion of relative performance. Firms are not only evaluated with respect to their wealth position, but also in their performance with respect to similar firms. A firm with positive net equity that is under-performing with respect to other firms is ripe for takeovers and mergers. The selection criterion modelled within this work captures this idea well. That said, the equity specification outlined above has merit and should be evaluated in future work.

Furthermore, a sufficient level of selection is necessary for evolutionary algorithms to perform in a satisfactory manner. The selection process outlined within the model above maintains a level of selection regardless of absolute performance. Allowing selection to be based on absolute performance may not invoke sufficient selective pressure required for the efficacy of the learning algorithm.

3.5 Simulation Results

We begin with consideration of a single parameterization of the model referred to as the reference simulations (Simulation 1). The parameter choices for these reference simulations are contained in Table 3.2 and Table 3.3. Using the reference simulations' learning parameters specified in Table 3.2, the remaining free parameters contained in Table 3.3 are selected in a manner that yields results approximating U.S. quarterly growth rates. As there is no empirical basis for the choice of the minimum labour requirements appropriate for each paradigm (m_j), the reference simulations' parameters are utilized over thirty distinct and randomly chosen specifications of minimum investment.[16] All simulations occur over 500 periods, or 125 years.

In order to facilitate a comparison with actual economic data, in Table 3.1 we present summary statistics of a sub-sample of the entire simulations' duration. This sub-sample contains only the first 190 quarters in order to limit the time frame to that of available data. Table 3.1 presents a summary of the average quarterly growth rate in real income per capita and the ratio of quarters in which this growth is negative. These figures represent averages over the entire set of randomly chosen specifications for the minimum labour requirements appropriate for each paradigm. Average

[16]While each simulation shares an identical process of random number generation, a unique initial seed value for this process is chosen for each in order to ensure results are robust to different sequences of random numbers.

standard deviations are presented in parentheses. The standard deviation of measures *across* each simulation are presented in italics. The reference simulations (Simulation 1) are characterized by an average quarterly growth rate of 0.5814 percent; only 0.0312 percent higher than that of the U.S. data (Simulation 0).[17] Of course, as mentioned above, given the reference simulations' learning parameters (Table 3.2), other simulation parameters have been selected such that the quarterly growth rate matches the data well. Although the quarterly growth rate is very similar to that observed in actual data, the ratio of quarters in which negative growth occurs is somewhat lower; 9.96 percent versus 22.53. Although periods of negative growth are a characteristic of the reference simulations, they occur with less frequency than observed in actual U.S. data.

Growth statistics over simulations' entire duration are presented in Table 3.5. Presented in Table 3.6 are the growth statistics for average paradigm.[18] We emphasize that this table represents rates of growth in average paradigm, not absolute changes. Table 3.7 contains statistics summarizing the ratio of periods with negative output growth to those in which output growth is non-negative. Again, these figures represent average values across the randomly determined specifications for the minimum labour requirements identical to those appropriate to Table 3.1.

An important feature of these reference simulations is that despite firms' ability to adopt new technologies incrementally, periods of negative growth occur. While the average rate of growth over the entire time series equals 1.07 percent, in 13.07 percent of simulative periods negative growth to total output occurs (See Tables 3.5 and 3.7, respectively).

The results of a single simulation of the reference parameters are contained in Figures 3.1 through 3.3. The minimum labour requirements for this single simulation are presented in Table 3.4. Figure 3.1 contains a plot of the per period growth rate in total production. As there is no growth rate in the population or change in the price level, this figure also represents real per period growth rates per capita. The time series for the log of total production is presented in Figure 3.2 and the average paradigm of production is presented in Figure 3.3.

There is a decisive co-movement between aggregate output growth and transition towards newer technological paradigms. The transition towards adopting new technologies is gradual. This highlights the important implications of learning-by-doing and inter-firm imitation in the process of technological appropriation, as discussed above.

Periods preceding adoption are characterized by low levels of output growth; in some scenarios these low levels of growth border on stagnate. Here, we define periods of stagnate growth as those characterized by low levels of aggregate output growth, in absolute terms. Additionally, these are periods in which positive levels of growth appear as often as negative and in equal magnitude on average. Periods of stagnate growth yield no significant trend in aggregate output.

[17] Real income per capita data is calculated utilizing population and real gross domestic product figures from the U.S. department of Commerce: Bureau of Economic Analysis.

[18] To calculate the average paradigm of production, we utilize the lower of the two relevant paradigms for the firm. As production may occur in the lower and upper relevant paradigms, this measure will be negatively biased

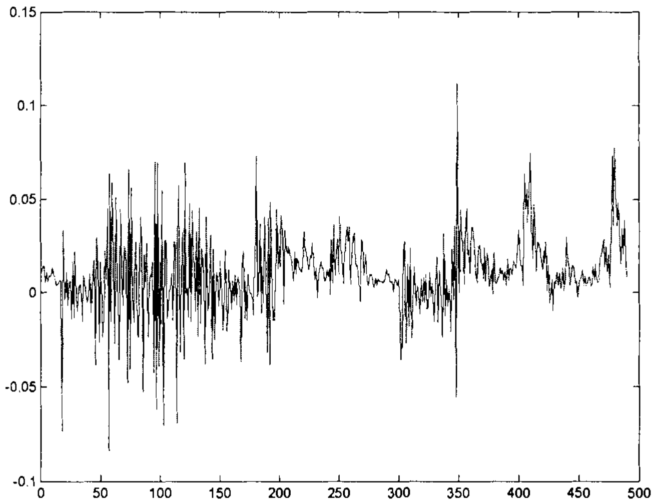

Figure 3.1: Growth Rate in Total Production Per-Capita.

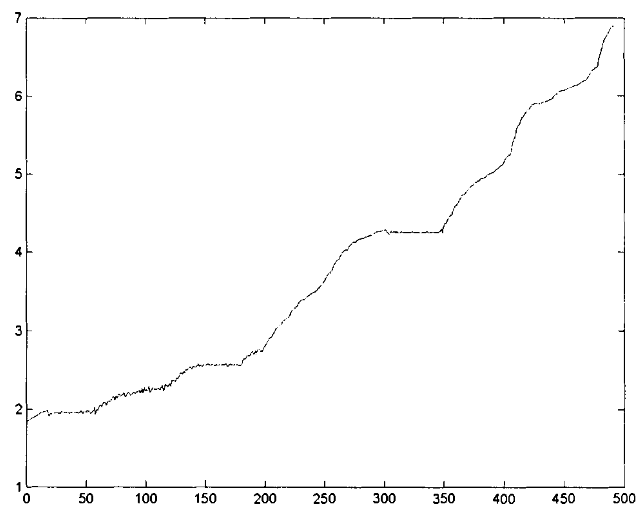

Figure 3.2: Log of Total Per-Capita Production.

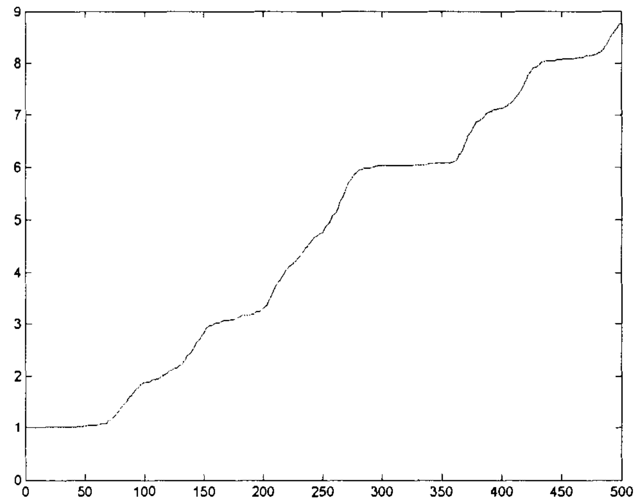

Figure 3.3: Average Paradigm.

Importantly, prolonged growth stagnation occurs in scenarios where a newer, more productive paradigm requires a high level of labour devoted to it in order to make it available to a firm. Notice that the stagnation in growth in the log of total output that occurs between the simulative periods 300 and 350 is associated with a similar stagnation in the level of average paradigm (Figure 3.3). In Table 3.4, the minimum labour requirements for each paradigm are presented. The stagnation in paradigm growth occurs around the sixth paradigm. For availability of the next paradigm, firms must devote 49.46 percent of it's labour's time to production in it.[19] The relatively high requirement makes it very difficult for firms to enter into production in this paradigm. High minimum requirements (m_j) have two important growth effects. First, they make entering into production utilizing the technology associated with high minimum labour requirement difficult. As mentioned, this may cause stagnation in growth as transition *into this technology* occurs very slowly. However, technologies with high minimum requirements are easier for firms to abandon. As such, transition *out of this technology* and into its successor is more likely to occur, having a positive impact on growth. Therein, the transition between technologies required for long run growth in per capita incomes will be dependent not only on the minimum labour requirement of the successive technology, m_{j+1}, but also on its predecessor, m_j.

An important question remains. Are these negative growth periods a result of negative shocks to productivity? In order to answer this, consider the model simulated without the ability of paradigm progression. That is, each firm's value of ψ is set at zero and does not undergo any changes.

[19]The upper bound on these requirements, m_j, is 50 percent

Furthermore, the degree of correspondence is not dynamic. Regardless of the parameterization, it should be clear that wages will instantaneously adjust, unemployment will settle at a level of zero, capital stocks and interest rates will remain constant and no growth will occur. Now, if we add the ability for firms to increase their labour devoted to the subsequent paradigm (ψ) and maintain that the technological level of the firm in the following paradigm (A_{j,p_j+1}) is greater than the latter (A_{j,p_j}), the overall productivity of the firm increases, both labour and capital demand increase, and there is no possibility for total output to decrease. The only fundamental cause of negative shocks to economic growth available in this model is through negative productivity shocks at the firm level. A firm progressing into a paradigm that it is less proficient in will observe an overall productivity decline, a decline in its demand for labour and capital, and its output will decline. Of course, the fall in labour demand will put downward pressure on wages, therein dampening this immediate effect. Essentially, there are no exogenous shocks or effects inherent in the labour or capital market that may give rise to negative economic growth. Effects observed in these markets must be in response to productivity fundamentals.

3.5.1 Genetic Parameter Effects

The genetic parameters specific to the model may have important effects on the dynamics of the simulation that are unintuitive. As such, the model is simulated with 11 different permutations of parameter choices over the reference simulations. These permutations include parameter variation over the probability of bit mutation given a firm is recombining their technology sets ($Pr^{DC}_{m|r}$), variation over the probability of bit mutation given a firm is imitating another's technology sets ($Pr^{DC}_{m|i}$), variation over the probability a firm will imitate another's switch gene (Pr^{ψ}_{i}), and variation over the likelihood of mutation given the firm is imitating another's switch gene ($Pr^{\psi}_{m|i}$). The parameter choices for each of these variables is given in Table 3.2. All other simulation parameters are identical to those utilized in the reference simulations. Importantly, each simulation shares an identical set of minimum labour investment requirements for each paradigm and an identical process of random number generation. The results of these simulations are contained in Table 3.1 (sub-sample) and Tables 3.5 through 3.7 (entire simulation). Importantly, in every permutation of the genetic parameters, regardless of the resulting average rate of growth, each simulation is afflicted with periods of negative growth in total output (See Table 3.7).

Variation in the genetic parameters results in important variation in the level of growth in real output per capita. Over the entire simulations' sample, against the reference simulations with an average quarterly growth rate of 1.07 percent, permutations yield a maximum average growth rate of 5.13 percent (simulation number 7). Over the sub-sample comparable to U.S. data presented in Table 3.1, all permutations over the learning parameters yield simulations with average quarterly growth rates within a single standard deviation of empirical observations. However, in none of these simulations is the ratio of periods characterized by negative growth similar to that of the data. In

Simulation No.	Average Growth	Ratio Negative
0	**0.5502 (0.8714)**	**22.53%**
1	0.5814 (0.6151) *(0.4002)*	9.96% *(7.53)*
2	0.5133 (0.5242) *(0.2835)*	11.76% *(8.18)*
3	0.6399 (0.6990) *(0.3150)*	9.98% *(6.63)*
4	0.6704 (0.6528) *(0.3071)*	9.58% *(8.73)*
5	0.8565 (0.7837) *(0.3190)*	6.15% *(4.63)*
6	1.0884 (1.1674) *(0.3689)*	6.08% *(5.40)*
7	1.3307 (1.4892) *(0.5579)*	4.87% *(4.45)*
8	0.5095 (0.6171) *(0.2611)*	16.75% *(8.21)*
9	0.6052 (0.5737) *(0.2533)*	6.22% *(6.75)*
10	0.5999 (0.5490) *(0.2427)*	5.31% *(5.25)*
11	0.5909 (0.6028) *(0.2600)*	10.77% *(5.73)*
12	0.5795 (0.5414) *(0.2157)*	7.88% *(6.01)*

Table 3.1: Simulation Sub-Sample versus Real World Data

simulation 8, although this ratio increases from that of the reference (from 9.96 to 16.75), it continues to fall short of that observed in actual data (22.53).

With respect to the baseline simulation, there are some notable effects derived from changes in these genetic parameters. First, increasing the probability of mutation given a firm is imitating another's switch gene ($Pr^{\psi}_{m|i}$) appears to increase the level of growth (simulations 2 through 4). From the baseline value of 0.05, decreasing this rate to 0.025 also decreases the average rate of quarterly growth from 1.07 to 1.02. Increasing this parameter to 0.10 and 0.15 has positive impacts on the overall level of average aggregate growth, increasing it from 1.07 to 1.49 and 2.07, respectively. We conclude that there is a positive relation between the probability of mutation given the firm is imitating another's switch gene and average growth.

There is intuition behind this result. Remember, in every period a given number of firms have the ability to alter their switch genes, and choose to imitate another firm's gene with a probability determined by Pr^{ψ}_{i}. In fact, of these with the potential to alter this gene, those that fail the selection process must alter their gene through imitation. There are two effects mutation has with respect to this imitation.

First, the firm is more likely to imitate firms that are very successful. As such, mutation is a terrible process that destroys the attempt to mimic the successful firm. Increasing mutation leaves these firms with labour divisions that are not as productive in the short run. This causes decreases in average growth rates.

However, increasing this mutation rate also has the effect of pushing more and more firms into the newer technological paradigm even when it is less productive. It forces these firms to begin building a competence in the new technology even before it is profitable to do so. The more firms that are pushed into this new paradigm, the faster will incremental innovations occur, as firms may take advantage of imitation. While these firms will perform terribly with respect to other firms that choose not to produce in this paradigm, their competence is available for imitation once these firms do progress. In the long run, the ability to imitate this competence has the effect of increasing growth rates.

Although the first of these affects will decrease growth rates in the short run, when considering average growth rates over a long horizon, the second affect dominates.

Decreases in the probability of imitating another's switch gene, (Pr^{ψ}_{i}), greatly increases the overall growth of the economy (simulations 5 through 7). It is the strongest of the genetic parameter affects discussed in this section. The intuition behind this effect is exactly the same as that behind the long run effect of increasing the size of the $Pr^{\psi}_{m|i}$ parameter. Forcing firms to experiment with new switch genes has the effect of pushing a greater number of firms towards adopting the new technology, regardless of the productivity of that new technology. While these firms suffer low levels of productivity, they build a competence in the paradigm that may be imitated by other firms in the future. The more firms that experiment with this new technology, the faster the level of this competence can increase through social learning.

Although not simulated, we would expect that further decreases in the probability of imitating another firm's switch gene would eventually lead to decreases in the average level of economic growth. This stems from the fact that too many firms would be progressing into new technological paradigms for the economy to ever achieve competency in any one of them. Although the economy would progress into more potentially productive technological paradigms, it would never spend enough time in any particular paradigm to achieve a level of competence that enabled capturing the potential productivity of the technology it uses at any given point in time.

Similar to the positive relation between the probability of mutation given the firm is imitating another's switch gene, it appears that increases in the probability of mutation given imitation of another's technology ($Pr^{DC}_{m|i}$) has positive impacts on average levels of quarterly growth. For the parameter values simulated (0.05, 0.10, 0.10 and 0.15), average growth increases directly with larger values of mutation (1.04, 1.07, 1.28 and 1.47, respectively). Although mutation decreases the ability of firms to imitate the technologies of more profitable firms, it appears that the negative impacts on growth this effect has is outweighed by the benefits stemming from the increases in diversity associated with high rates of mutation over technologies. Interestingly, this relationship does not hold with respect to mutation when a firm is recombining it's own technology sets. Here, a convex relation between $Pr^{DC}_{m|r}$ and growth rates exists, where growth is maximized at the reference simulations' parameterization of mutation. However, the strength of this relationship is weak. Although both increases and decreases in this parameter yield a fall in average growth, their magnitude is smaller than any other learning parameter effect.

These results allow for the interpretation of economies characterized by extended periods of stagnate growth in terms of the model considered within this work. Differences in growth are assumed to be the sole result of differences in the pace of technological appropriation. Two things are required for appropriation, experimentation within the new technology and imitation of firms that find success within this technology. If an economy is not fostered by an environment conducive to either of these two effects, transition through technologies occurs at a slower rate, and growth in per capita income is lower than that of other economies. If the conditions relevant for experimentation and imitation in an economy suffering from such effects are altered, growth within the economy will increase, correlated with an increase in the pace of technological appropriation.

3.6 Concluding Remarks

The model examined in this work allows for the investigation of the sufficiency of learning-by-doing for explaining negative macroeconomic output shocks in an evolutionary model of technological transition. It has been argued that the productivity gap between old and new technologies causes temporary negative aggregate output shocks despite the ability of firms to adopt these new technologies in a non-discrete manner. The productivity gap between old and new technologies results from the lack of accrued incremental innovations in the newer technologies. These results appear highly robust with respect to changes in the underlying parameters of the evolutionary algorithm of firm adaptation. While periods of negative growth in real output per capita are a regularity of the simulations, they occur with less frequency than found in actual U.S. data. The framework outlined within this work has important advantages over other models in the literature in that it offers a natural model of experimentation by agents and allows consistency with Nelson and Winter's (1982) interpretation of Schumpetarian competition. The manner in which innovation and technological progression is modelled within this work is also consistent with key observations of significant technological transitions of the past; specifically, the Industrial Revolution.

References

Aghion, P., Howitt P. (1998) **Endogenous Growth Theory**, Cambridge, Massachusetts: The MIT Press.

Alcaly R. (2003) **The New Economy: And What It Means For America's Future**, New York: Farrar Straus and Giroux.

Andersen E.S with Anne K. Jensen, Lars Madsen and Martin Jrgensen (1996) "The Nelson and Winter Models Revisited: Prototypes for Computer-Based Reconstruction of Schumpeterian Competition", Working Paper, Danish Research Unit For Industrial Dynamics (DRUID).

Atkeson, A., Kehoe, P. (2003) "The Transition to a New Economy After the Second Industrial Revolution", Federal Reserve Bank of Minneapolis, Research Department Staff Report 296.

Ballot G., Taymaz E. (1999) "Technological change, learning and macro-economic coordination: An evolutionary model", *Journal of Artificial Societies and Social Simulation*, 2(2).

Ballot G., Taymaz E. (1998) "Human Capital, Technological Lock-in and Evolutionary Dynamics" in **Microfoundations of Economic Growth: A Schumpeterian Perspective**, Eliasson G. and Green C. (eds.), Ann Arbor: The University of Michigan Press, 301-330.

Ballot G., Taymaz E. (1997) "The dynamics of firms in a micro-to-macro model: The role of training, learning and innovation", *Journal of Evolutionary Economics*, 7, 435-457.

Ballot G., Taymaz E. (1996) "Firm sponsored training, technical change and aggregate performance in a micro-macro model" in **Microsimulation and public policy**, A. Harding (ed.) North-Holland, Amsterdam.

Ballot G., Taymaz E. (1994) "Training, learning and innovation. A micro-to-macro model of evolutionary growth", in **International J.A. Schumpeter Society Conference on Economic Dynamisism: Analysis and Policy**, Munster, Germany, August 17-20, 1994.

Becker G.S. (1964) **Human Capital**. New York: NBER.

Bekar, C., Lipsey R. (2001) "Science, Institutions and the Industrial Revolution", Working Paper, Simon Fraser University.

Cohen W.M., Levinthal D.A. (1990) "Absorptive capacity: a new perspective on learning and innovation", *Administration Science Quarterly*, 3, 128-152.

Crouzet, F. (1985) **The First Industrialists**, Cambridge University Press.

David, P.A. (1990) "The dynamo and the computer: An historical perspective on the modern productivity paradox" *American Economic Review*, 80, 355-361.

David, P.A. (1991) "Computer and dynamo: The modern productivity paradox in a not-too-distant mirror" in **Technology and Productivity: The challenge for economic policy**, Paris: Organisation for Economic Co-operation and Development.

Devine, W.D. (1983) "From shafts to wires: Historical perspective on electrification" *Journal of Economic History*, 43, 347-372.

Eliasson G. (1994) "General purpose technologies, industrial competence and economic growth" Working Paper, Department of Industrial Economics and Management, The Royal Institute of Technology, Stockholm.

Eliasson G., (1991) "Modelling the experimentally organized economy", *Journal of Economic Bevavior and Organization*, 16, 163-182.

Jovanovic, B., Nyarko Y. (1996) "Learning by Doing and the Choice of Technology", *Econometrica*, 64(6), 1299-1310.

Landes, D. (1986) "What Do Bosses Really Do?" *Journal of Economic History*, 46(3), 585-623.

Lucas, R.E. (1993) "Making a Miracle" *Econometrica*, 61(2), 251-272.

Nelson, R.R., and Winter, S.G. (1982), **An Evolutionary Theory of Economic Change**, Cambridge, Massachusetts: Belknap Press.

Parente, S. (1994) "Technology Adoption, Learning-by-Doing, and Economic Growth" *Journal of Economic Theory*, 63(2), 346-369.

Pereira, A. (2002) "The Industrial Revolution as an Organizational Revolution", Working Paper, Simon Fraser University.

Polland, S. (1965) **The Genesis of Modern Management**, Cambridge: Harvard University Press.

Rosenberg, N. (1996) "Uncertainty and Technological Progress", in **The Mosaic of Economic Growth**, Ralph Landau, T. Taylor and G. Wright (eds.), Stanford: Stanford University Press.

Taymaz E. (1991) **MOSES on PC: Manual, initialization and calibration**, IUI, Stockholm.

3.7 Appendix

Simulation No.	$Pr^{DC}_{m\|r}$	$Pr^{DC}_{m\|i}$	Pr^{ψ}_{i}	$Pr^{\psi}_{m\|i}$
1	0.10	0.10	0.95	0.05
2	0.10	0.10	0.95	0.025
3	0.10	0.10	0.95	0.10
4	0.10	0.10	0.95	0.15
5	0.10	0.10	0.90	0.05
6	0.10	0.10	0.85	0.05
7	0.10	0.10	0.80	0.05
8	0.10	0.05	0.95	0.05
9	0.10	0.15	0.95	0.05
10	0.10	0.20	0.95	0.05
11	0.05	0.10	0.95	0.05
12	0.15	0.10	0.95	0.05

Table 3.2: Genetic Parameter Specification

T - Simulation Periods	500
N - Number of Firms	100
L_l - Total Initial Population	200
n_l - Growth Rate in Population	0.0
K_0 - Initial Capital Supply	50
δ - Depreciation Rate of Capital	0.0
x - Technology sets per Paradigm	3
k - Bit Length of First Paradigm's Technology Sets	18
l_s - Bit Length of Switch Gene	8
m - Number of Periods Used in Fitness Calculation	10
ϕ - Percent of New Firms From Those Under Selection	0.20
λ - Parameter in Equation Linking DC to $A_{j,i}$	1.15
Λ - Parameter in Equation Linking DC to $A_{j,i}$	1
α - Capital's Share of Output	0.33
β - Labour's Share of Output	0.60

Table 3.3: Baseline Parameter Specification

Paradigm (j)	Minimum Labour Investment (m_j)
1	0.1639
2	0.3314
3	0.2443
4	0.0581
5	0.3819
6	0.1725
7	0.4946
8	0.2414
9	0.1337
10	0.3534
11	0.3289
12	0.2374

Table 3.4: Minimum Labour Investment

Simulation	10-500	10-100	101-200	201-300	301-400	401-500
1	1.07	0.44	0.71	0.88	1.44	1.82
	(2.19)	(0.39)	(0.82)	(1.21)	(2.27)	(3.32)
2	1.02	0.44	0.58	0.95	1.20	1.86
	(1.99)	(0.43)	(0.61)	(1.16)	(1.87)	(3.21)
3	1.49	0.53	0.74	1.15	1.98	2.97
	(3.14)	(0.50)	(0.88)	(1.28)	(2.84)	(5.29)
4	2.07	0.55	0.78	1.50	2.72	4.67
	(4.61)	(0.48)	(0.81)	(1.96)	(4.19)	(7.92)
5	2.32	0.60	1.09	1.79	3.01	5.01
	(5.47)	(0.48)	(1.06)	(2.47)	(4.91)	(9.58)
6	3.73	0.79	1.36	2.72	4.61	8.97
	(9.30)	(0.67)	(1.62)	(3.72)	(7.60)	(16.82)
7	5.13	0.87	1.75	3.59	7.18	11.93
	(13.51)	(0.73)	(2.18)	(5.28)	(11.82)	(24.26)
8	1.04	0.41	0.60	0.96	1.25	1.95
	(2.20)	(0.46)	(0.76)	(1.42)	(2.15)	(3.45)
9	1.28	0.49	0.71	1.17	1.76	2.21
	(2.68)	(0.38)	(0.75)	(1.49)	(2.72)	(4.10)
10	1.47	0.49	0.70	1.25	1.90	2.95
	(3.16)	(0.35)	(0.73)	(1.57)	(2.91)	(5.25)
11	0.99	0.46	0.71	0.80	1.29	1.67
	(1.88)	(0.43)	(0.76)	(1.02)	(1.93)	(2.89)
12	1.03	0.48	0.67	1.04	1.26	1.66
	(2.04)	(0.40)	(0.67)	(1.30)	(2.12)	(3.05)

Table 3.5: Average Aggregate Growth Rate

Simulation	10-500	10-100	101-200	201-300	301-400	401-500
1	0.42	0.65	0.53	0.45	0.26	0.22
	(0.67)	(0.83)	(0.63)	(0.52)	(0.36)	(0.30)
2	0.42	0.65	0.53	0.45	0.26	0.22
	(0.67)	(0.83)	(0.63)	(0.52)	(0.36)	(0.30)
3	0.45	0.77	0.60	0.42	0.26	0.23
	(0.65)	(0.88)	(0.60)	(0.42)	(0.27)	(0.24)
4	0.48	0.82	0.59	0.43	0.34	0.23
	(0.61)	(0.84)	(0.53)	(0.38)	(0.28)	(0.22)
5	0.49	0.90	0.70	0.38	0.29	0.20
	(0.65)	(0.91)	(0.60)	(0.37)	(0.30)	(0.23)
6	0.52	1.12	0.66	0.39	0.26	0.22
	(0.68)	(1.01)	(0.55)	(0.34)	(0.25)	(0.22)
7	0.54	1.15	0.71	0.41	0.29	0.20
	(0.67)	(0.98)	(0.55)	(0.31)	(0.24)	(0.18)
8	0.42	0.66	0.56	0.43	0.26	0.22
	(0.66)	(0.82)	(0.65)	(0.51)	(0.31)	(0.28)
9	0.43	0.76	0.58	0.43	0.26	0.18
	(0.66)	(0.91)	(0.61)	(0.47)	(0.29)	(0.24)
10	0.45	0.79	0.58	0.41	0.31	0.19
	(0.64)	(0.89)	(0.56)	(0.43)	(0.30)	(0.22)
11	0.42	0.72	0.60	0.35	0.28	0.19
	(0.66)	(0.90)	(0.63)	(0.41)	(0.35)	(0.25)
12	0.42	0.74	0.59	0.39	0.24	0.17
	(0.65)	(0.89)	(0.61)	(0.43)	(0.29)	(0.23)

Table 3.6: Average Paradigm Growth Rate

Simulation	10-500	10-100	101-200	201-300	301-400	401-500
1	13.07	7.44	12.26	16.13	15.52	14.01
2	12.54	9.33	13.97	10.88	11.21	17.59
3	11.45	8.52	11.31	9.83	15.99	11.80
4	8.58	6.70	12.19	9.49	6.20	8.47
5	8.11	5.44	6.80	8.92	8.59	10.92
6	7.86	4.22	7.78	8.05	9.39	9.86
7	6.49	3.85	5.79	6.70	7.17	8.98
8	18.88	15.33	18.05	17.71	24.38	19.35
9	8.84	4.59	7.71	8.75	12.09	11.05
10	6.48	3.93	6.57	6.40	6.36	9.22
11	12.92	8.26	13.06	14.21	14.07	15.10
12	12.15	6.56	9.09	11.38	17.47	16.26

Table 3.7: Ratio of Periods with Negative Growth Rate

Chapter 4

The Muth Model

Intelligent Mutation Rate Control in an Economic Application of Genetic Algorithms

4.1 Introduction

Genetic Algorithms are the best known representation of a class of direct random search methods called evolutionary algorithms which are widely used to solve complex optimization and adaptation problems. Their use within economics is grounded on their ability to represent the adaptation of individuals to the underlying parameters of their economic environment. They facilitate a departure from the rational expectations hypothesis, which requires in it's place a model of learning employed in order to describe the manner in which agents make decisions about their economic behavior.

Genetic algorithms describe the evolution of a population of rules, representing different possible beliefs, in response to experience. A population of n individual rules are represented by binary vectors, $x^i = (x_1^i, x_2^i, ..., x_k^i) \ \varepsilon \ \{0,1\}^k$, of fixed length k. This population of rules may represent different agents interacting, referred to as *social learning*, or a single agent's mutually competing ideas, referred to as *individual learning.* In each of these representations, the frequency with which a given rule is represented in the population indicates the degree to which it is accepted in a population of agents, or the degree of credence attached to it, respectively. The success of a particular rule is referred to as it's *fitness* and is determined according to a specific *fitness function.* Rules whose application has been more successful are more likely to become represented in the population. This occurs according to a classical probabilistic proportional selection operator that uses the relative fitness to serve as selection probability.

Heterogeneity is introduced into the population through two evolutionary operators, *crossover* and *mutation*. The crossover operator works by first randomly assigning each rule in the population to a pair. For each pair, the crossover operator exchanges information between different individual rules with a given probability, p_c.

This paper focuses on the mutation operator, which introduces innovation into the population by inverting bits of the binary vectors. Each bit has a small and independent probability of inversion, p_m. This operator is typically assessed as a secondary one which is of little importance in comparison to crossover. Most applications of genetic algorithms work with small, constant settings of $p_m \in [0.001, 0.01]$. While there are logical and mathematical bounds on the choice of the magnitude of p_m, it remains a free parameter of the algorithms implementation.

In a simple, non-economic directed search implementation of genetic algorithms, the choice of the mutation rate is of concern in a very practical sense. Many applications favor larger or non-constant (though deterministic) settings of the mutation rate for increasing the speed at which the algorithm converges on the solution. This practical importance is also a concern for implementation in economic settings.

In an economic system of constant change, there is likely a requirement for constant introduction of innovation. It is likely, however, that there is an optimal rate at which this innovation occurs and that it is dependent on the underlying stochastic nature of the system in question. More dynamic, or stochastic, environments may call for a higher level of *maintained* experimentation. Importantly, this rate of experimentation is likely linked to the economic system and not an exogenous parameter of human learning.

In addition, genetic algorithms are used in comparison to an actual human learning process.[1] In these settings, fixing the rate of mutation may be problematic for other, non-practical reasons. If bitwise mutation is to be analogous to some actual human learning operator, fixing the rate of this action seems inappropriate, *a priori*. Learning agents are likely to adapt the rate with which they experiment with new rules as the perceived benefit of this experimentation decreases. Modelling economic choices using a genetic algorithm with a fixed mutation rate may introduce a biased amount of innovation over the population of rules.

Ex post, this bias precludes the use of fixed mutation rates for reasons of parsimony. The use of fixed mutation rates allows for constant introduction of innovation over a population of rules and therefore the ability to adapt to an economic environment. However, in many cases, fixed mutation rates preclude the system's true convergence to an economic equilibrium. Even as the system converges to equilibrium levels, the rate at which individuals experiment remains the same. When compared to actual human behavior, this outcome is problematic. If the system is to converge, mutation rates must fall to zero or an *election operator* must be used.

The use of an *election* operator limits the introduction of innovation to situations in which the mutated rule is associated with an "expected" increase in fitness (Arifovic, 1994). After the crossover

[1] Examples of such work include Arifovic and Ledyard (2004), and Arifovic and Maschek (2004).

and mutation operators have generated potential new rules, the election operator tests these rules before they are permitted to become members of the population. The fitness that each new rule would have attained is determined, holding all variables relevant for the calculation fixed at the previous period's values. This value is referred to as *potential fitness*. The potential fitness of the new rules is compared to the actual fitness associated with the rule's parents (parents are the pairs of rules that are used in the application of the crossover operator). A new rule may only replace a parent rule if it's potential fitness is higher than that of a parent's actual fitness.

The economic use of genetic algorithms is becoming popular for their possible representation of a learning process. The use of an election operator is analogous to the acceptance of "simple" expectations, and may not capture actual human behavior in a satisfactory manner. If multiple equilibria exist, switching between these equilibria is an impossibility with the adoption of the election operator in genetic algorithm models of agent behavior. The election operator's efficacy is grounded on the control of innovation over rules in the population. The introduction of innovation is achieved through the crossover and mutation operators. A more realistic control of the introduction of innovation over the population may be found in the control of these operators directly, rather than the adoption of simple expectations.

This work acknowledges the importance of the introduction of innovation, but maintains the level with which innovation is introduced should be determined within the framework of the model rather than being exogenously imposed or limited through the use of simple expectations. Herein, the focus is limited to the mutation operator. An alternative mechanism for controlling mutation lies in the *on-line learning*, or *self-adaptation* of this parameter. We consider this mechanism below in Section 4.4. We wish to consider how this mechanism affects an economic application of genetic algorithms in terms of variation in aggregate outcomes and convergence. With respect to the latter, we will examine the potential for self-adaptation to replace the election operator as sufficient for yielding convergent results. The economic environment in which we consider these questions is presented in Section 4.2, and the application of genetic algorithm adaption in this environment without self-adaptation is presented in Section 4.3. Results of the analysis are presented in Section 4.5, and conclusions follow.

4.2 Proposed Environment for Analysis

We wish to compare the performance of simple genetic algorithms to ones in which the election operator is included and those characterized by self-adaptation in an economic setting, or environment. The cobweb environment examined by Arifovic (1994) is proposed for the analysis.[2] Its choice is

[2]The rational expectations version of the model has been considered in the work of Muth (1961). Versions of the model with alternative formulations of learning have been presented in the works of Nerlove (1958), Carlson (1969), Townsend (1978), DeCanio (1979), Frydman (1981), Brandenburger (1984), Bray and Savin (1986), Marcet and Sargent (1987), and Nyarko (1990). The model has been simulated in an experimental setting by Holt and Williamil (1986) and Welford (1989). While divergent behavior characterizes most of the above algorithms in the

motivated by the following considerations. First, it is a simple environment that lends itself well for comparing results of a simple genetic approach, an extended approach including the election operator, and the approach of this work (inclusion of self-adaptation). Additionally, the cobweb model is easily adapted to one of constant change. Though considered constant in this work, the underlying parameters of the model which determine the rational expectations solution could easily change according to some Markov-switching process, or encompass stochastic exogenous shocks.

In the work of Arifovic, a genetic algorithm is used to update the firms' decision rules determining production in the following period. Her results show that genetic algorithms in this setting are characterized by convergence to the rational-expectations equilibrium for a much wider range of parameter values than other algorithms.

4.2.1 Description of the cobweb model

The model contains n firms in a competitive market. Firms produce the same good and each is a price taker. Each firm has an identical cost function given by

$$C_{i,t} = xq_{i,t} + \frac{1}{2}ynq_{i,t}^2 \tag{4.1}$$

where $C_{i,t}$ is firm i's cost of production for sale at time t and quantity $q_{i,t}$. Since the production of goods takes time, quantities produced must be decided before a market price is observed. Expected profit of an individual firm, $\Pi_{i,t}^e$, is

$$\Pi_{i,t}^e = P_t^e q_{i,t} - xq_{i,t} - \frac{1}{2}ynq_{i,t}^2 \tag{4.2}$$

where P_t^e is the expected price of the good at time t. Each firm chooses a quantity $q_{i,t}$ to maximize its expected profit $\Pi_{i,t}^e$ on the basis of its expectations regarding the prevailing price P_t^e. The first order condition for profit maximization with respect to $q_{i,t}$ is given by the following equation

$$q_{i,t} = \frac{1}{yn}(P_t^e - x) \tag{4.3}$$

where the price P_t that clears the market at time t is determined by the demand curve

$$P_t = A - B\sum_{i=1}^{n} q_{i,t} \tag{4.4}$$

The *rational expectations equilibrium* is characterized by $P_t^e = P_t$ and $q_{t,i} = q_t$ for all i. By rearranging equation x, this may be expressed as

unstable case, this was not observed in the experimental settings with human subjects.

$$x + ynq_t = A - Bnq_t$$

Solving the above expression we arrive at the following characterization of the *rational expectations equilibrium* per firm quantity

$$q_t = q^* = \frac{A - x}{n(B + y)} \tag{4.5}$$

A key objective of the Arifovic (1994) work is to determine whether quantities produced by firms that are using a genetic algorithm as their learning scheme will converge to this constant *rational expectations* quantity and how these results compare to the results of other learning and experimental behavior (See Arifovic (1994) pp.07). Her application of the genetic algorithm to the above economic framework is described in the proceeding sub-section.

4.3 Application of the basic genetic learning algorithm

A population of binary strings, A_t, represents a collection of firms' decision rules at time period t. These binary strings are of fixed length, k, written over the {0,1} alphabet. These strings are decoded into their integer equivalent and normalized in order to give their production level equivalent. For a string i of length k the decoding works in the following manner:

$$x_{i,t} = \sum_{j=1}^{k} a_{i,t}^{j} 2^{j-1}$$

where $a_{i,t}^{j}$ is the value (0,1) taken at the jth position in the string.

After a string is decoded, its integer value is normalized in order to obtain a real number value $q_{i,t}$ that represents production levels at time t for firm i:

$$q_{i,t} = x_{i,t}/\overline{K}$$

where $\overline{K}$ is the normalizing coefficient.

Fitness of a rule i at time t, $\mu_{i,t}$, is determined by the value of firms' profit earned in the period.

$$\mu_{i,t} = \Pi_{i,t} = P_t q_{i,t} - C_{i,t}$$

Firms' decision rules are updated using three genetic operators: reproduction, crossover, and mutation. Reproduction makes the copies of individual chromosomes according to their relative fitness. The probability that a chromosome $A_{i,t}$ will get a copy $C_{i,t}$ is given by

$$P(C_{i,t}) = \frac{\mu_{i,t}}{\sum_{i=1}^{n} \mu_{i,t}}$$

In every period, n rules are created from the pool of rules utilized in the preceding period. The algorithmic form of the reproduction operator is like a biased roulette wheel where each string is allocated a slot sized in proportion to its relative fitness. Thus, rules with a higher fitness value in the preceding period having a higher probability of existing in the subsequent period (or existing in higher numbers). These copies represent the pool of rules which then undergoes crossover and mutation.

Crossover exchanges the parts of pairs of randomly selected strings. It operates in two stages. First, two strings are drawn from the pool of copies at random. Then, a random integer is drawn, $b \in [1, k-1]$. Two new strings are formed by swapping the set of binary values to the right of the position b. In each of the $n/2$ randomly determined pairs, this crossover occurs with probability p_c.

Mutation is the process of a random change in the value of a position within a string. Each position has a probability, p_m, of being altered by mutation, independent of other positions.

After the members of the new population are determined, the quantity that will be produced and offered for sale at time t is computed for each firm. Individual quantities are summed up and the market price, P_t, is determined. Next, costs associated with each firm's production level are calculated and each firm's fitness level is then determined.

The above described steps are applied iteratively for T generations. The population of chromosomes at time period 0 is randomly determined.

4.4 Two-Level Learning in Genetic Algorithms

The self-adaptation principle incorporates certain strategy parameters into the representation of each individual. The strategy parameter set of an individual provides a parameter setting for mutation when applied to this particular individual, and strategy parameters evolve by means of mutation (and recombination) just as the object variables do.

The genealogy of on-line learning may be traced to its origins in the work of Schwefel in the context of multimembered evolution strategies (1987, 1992, 1995). Independently of this, Fogel et al. (1991) developed an almost identical procedure for evolutionary programming.[3]

The specific introduction of self-adaptation considered here has been proposed and tested in specific environments by Bäck and Schütz. They propose a self-adaptive mechanism of a single mutation rate per individual such that the following requirements are fulfilled:

[3] See also Fogel (1995).

- (1) Mutation of the mutation rate $p_m \; \varepsilon \;]0,1[$ yields a mutation rate $p'_m \; \varepsilon \;]0,1[$.
- (2) The expected change of p_m by repeatedly mutating it equals zero. That is, $E[p'_m] = p_m$.
- (3) Small changes are more likely than large ones.
- (4) A modification by a factor c occurs with the same probability as a modification by $1/c$.

The first requirement simply maintains that after mutation, the new individual mutation rate remains in the mathematical bounds appropriate. Requiring that repeated mutations yield an expected change equal to zero is done to ensure that the only force driving the direction of these mutations is selection. There is to be no drift in mutation rates not associated with higher levels of fitness. The final two requirements give structure to the distribution of changes in the individual mutation rates. The distribution of potential changes to p_m is positively skewed (according to the fourth requirement) making a given factor increase in the mutation rate more likely than a decrease of equal factor.

Based on these requirements, a logistic transformation of the form

$$p'_m = (1 + \frac{1-p_m}{p_m} exp(-\gamma N(0,1)))^{-1} \tag{4.6}$$

such that p'_m is distributed according to a logistic normal distribution with the given probability density function

$$f_{p'_m}(x) = \frac{1}{(2\pi)^{(1/2)}\gamma(1-x)x} exp(\frac{-(ln\frac{x}{1-x} - \zeta)^2}{2\gamma^2}) \tag{4.7}$$

where $\zeta = ln\frac{p_m}{1-p_m}$. We refer to p'_m as the *mutated mutation rate.* The *learning rate*, γ, allows for a control of the adaptation speed evolutionary strategies' mutation rate. For a given rate of mutation, p_m, the variance over the mutated mutation rates, p'_m, increases for higher values of the rate of learning, γ. As such, more heterogeneity over individual mutation rates is introduced in each period.

The algorithms works as follows. The genotype of an individual consists of a bitstring of length k and an individual mutation rate p^i_m that controls the bitwise mutation of $(x^i_1, x^i_2, ..., x^i_k)$ according to the mutated mutation rate $p^{i'}_m$.

$$x^i = (x^i_1, x^i_2, ..., x^i_k, p^i_m) \; \varepsilon \; \{0,1\}^k \times \;]0,1[$$

The mutation process yields a new individual $x^{i'} = (x^{i'}_1, x^{i'}_2, ..., x^{i'}_k, p^{i'}_m)$. Crossover is applied only to the binary vector and has no impact on the mutation rate, but this is certainly an area of future research. This algorithms allows for incorporation and exclusion of the election operator.

In their original application of the algorithm, Back and Schutz use the counting-ones problem, $f(x) = \sum_{i=1}^{k} x_i \rightarrow min$, to build a modified continuous optimization problem in which switching occurs every 250 periods between f and $f'(x) = k - f(x) \rightarrow min$. A "cycle" is referred to as the period between switches of the optimization problem. Their results show that during a given cycle of the optimization problem, mutation rates decrease drastically from values close to 0.1 to near their lower bound. Within only a few generations of the switching, mutation rates increase back to this higher value. Convergence velocity increases, driving an improvement in the objective function value over simulations in which self-adaptation was not included.[4]

4.4.1 Selective pressure

Mutation Rates

As emphasized by Bäck and Schütz, self adaptation works by means of the selective advantage or disadvantage of mutation rates. This advantage is expressed by its impact on the fitness function of the rule it is associated with. Bäck and Schütz argue that the self-adaptation mechanism can only work effectively if at least one bit per binary string is mutated on average. As such, they impose a lower bound on the rate of mutation, $\underline{p_m}$, that is equal to $1/k$.

This argument is particularly sound with respect to convergence velocity, especially in the counting-ones framework in which they are working. Here, the problem is altered every 250 periods; the population of rules must adapt to a new solution every 250 periods. Maintaining a sufficient level of mutation is required in order to allow sufficient diversity so as to converge on an alternative solution. With respect to velocity of convergence, if self adaptation is to outperform algorithms with constant mutation rates in their environment, this lower bound of $1/k$ is required.[5].

In the framework considered herein, the parameterization of the model for which these rules are required to adapt to is held constant over the duration of the simulation. Maintaining a level of diversity as high as $1/k$ may not be required for the self adaptation mechanism to work effectively. Additionally, many economic works utilizing the genetic algorithm have enjoyed success with mutation rates lower than that required for average mutation of one bit per binary string.

Regardless, if given a sufficient duration of simulative periods, selection of better performing rules with respect to fitness will occur, even if the rate of mutation is lower than $1/k$. As the environment within which these rules are functioning is non-changing, even slight selective pressure will allow convergence to rules that will, over the long run, outperform the average.

[4] While under optimal circumstances the modified problem requires oscillating behavior of the mutation rate, the problem does not contain complementarity between rules with respect to their fitness; a characteristic of most economic environments. This work is the first examination of the performance of self adaptation in an environment in which a rule's fitness is also a function of the other $n - 1$ rules.

[5] It will be argued that even in such a framework, placing a lower bound on the rate of mutation is not strictly required if one incorporates *fitness dependent mutation modificators* (see below)

We therefore incorporate a lower bound on mutation, $\underline{p_m}$, but include values that are less than $1/k$. If lower bound values around $1/k$ are required for effective self adaptation in this particular environment, then simulation results with lower values will provide evidence supporting such. We emphasize that this may not be appropriate in frameworks with model parameters that are changing over some period within the simulation.

Selection

For the self adaptation principle to work, Schwefel (1987, 1992) has demonstrated that a relatively strong selective pressure, such as that provided by (μ, λ)-selection, is required.[6] As such, Bäck and Schütz utilize simulations incorporating both proportional selection (that which is used within this work) and the stronger (μ, λ)-selection process.

In their work, Back and Schutz conclude that the only difference between proportional and (μ, λ)-selection consists in the fact that smaller selective pressure of proportional selection allows for a larger diversity of the mutation rates and implies a slightly slower convergence velocity.

In this work, as in the original Arifovic (1994) work, we utilize proportional selection. However, in order to increase the selective pressure slightly, we alter the determination of rules' fitness from the original. Here, we apply a monotonic transformation to the rules' profit in order to calculate its fitness.

$$F_{i,t}^{eff} = \frac{\mu_{i,t} - \mu_{min}}{\mu_{max} - \mu_{min}} \tag{4.8}$$

In the above equation, $F_{i,t}^{eff}$ denotes the *effective fitness* of individual rule i in period t. The maximal and minimal fitness value in the population at the current generation are denoted by μ_{max} and μ_{min}, respectively. The conversion simply rescales the fitness values to the [0,1] interval.

The conversion to effective fitness has the effect of allowing for diverse fitness values even when the absolute, or raw, fitness levels of the rules are quite close together. This is likely to occur when a simulation approaches convergence to the economic equilibrium of competitive output. This has the effect of favoring replication of rules that perform better in terms of profits, even if the improvements in profits are slight in absolute, or raw terms. In order to evaluate the effect of this increase in selective pressure, simulations will be run with, and without the effective fitness modification inherent in the above equation.

The above formulation of effective fitness allows an extension of the algorithm in which alterations of individual mutation rates beyond a global rate are associated with an energy cost to fitness.

[6]This selection functions according to μ parent individuals creating $\lambda > \mu$ offspring by recombination and mutation. Only the best μ offspring individuals are selected as parents for the next generation

4.4.2 Fitness dependent mutation modificators

In an extension to the evolutionary algorithm described above, we assume that changes in individual mutation rates may not occur without cost. In a genetic model of plant evolution, Kim (1998) extends the basic modelling concept in a manner that includes an energy cost of mutation rate adaptation. This energy cost manifests within the rules' fitness function and is dependent on the rules' individual *mutation modificator*. Herein, the adoption of Kim's mutation modificator allows for a link between mutation rate adaptations and fitness penalties.

An individual rule's mutation modificator is an implicit function of it's particular mutation rate:

$$p_{i,t} = m \cdot q^{\kappa_{i,t}} \tag{4.9}$$

In the above equation, $p_{i,t}$ denotes the individual mutation rate of rule i in generation t. The mutation modificator, $\kappa_{i,t}$, is a characteristic of each individual rule that is linked to its particular mutation rate by the parameters m and q, referred to as the *global mutation rate* and the *mutation rate modificator factor*, respectively.

Solving for κ_i yields the following equation:

$$\kappa_{i,t} = \frac{\ln(p_{i,t}) - \ln(m)}{\ln(q)} \tag{4.10}$$

The mutation modificator is a positive function of the percent difference between the individual and the global mutation rate. The fitness cost associated with per unit changes in the absolute value of the mutation modificator is determined by the parameter ρ in the following alternative effective fitness equation.

$$F_{i,t}^{eff} = \frac{\mu_{i,t} - \mu_{min}}{\mu_{max} - \mu_{min}} - \rho|\kappa_{i,t}| \tag{4.11}$$

Assuming q is greater than one, the effective cost of mutation rate deviations from the global rate is determined by the size of $\rho/\ln(q)$.

$$F_{i,t}^{eff} = \frac{\mu_{i,t} - \mu_{min}}{\mu_{max} - \mu_{min}} - \frac{\rho}{\ln(q)}|\ln(p_{i,t}) - \ln(m)| \tag{4.12}$$

The rescaling of raw profit levels into effective fitness, $F_{i,t}^{eff}$, allows for application of the penalty, ρ, that is independent of the absolute raw fitness levels. Setting the mutation rate modification penalty to zero permits accordance with the genetic algorithm described in previous sections.

There exists an economic interpretation for total fitness costs that result when individual mutation deviates from the global rate. Total fitness penalties are the result of two qualitatively distinct costs associated with different levels of individual mutation.

First, reductions in mutation rates are likely associated with increased effort costs as they would require avoiding trembling hand perturbations of the binary encoded rule. Such costs would be a decreasing function of the individual mutation rate. That is, such costs can be avoided by adopting higher levels for the likelihood of mutation.

However, adaptation of higher mutation rates requires concerted effort on the part of the individual to continually search out new rules for adoption. Additionally, these larger mutation rates increase in frequency in which new rules are adopted and are therefore associated with relatively higher effort costs. This second effort cost is an increasing function of the individual mutation rate.

The addition of these two qualitative costs yields the total fitness cost associated with a particular level of mutation. At the global mutation rate, these costs are minimized and equal zero. The absolute value of the effort cost at the global mutation rate is not important as only relative fitness determines evolutionary dynamics. The only important characteristic of the global mutation rate is that at this value fitness costs associated with mutation are minimized.

4.5 Simulation Results

4.5.1 Replicating the Arifovic (1994) Results

We begin by simulating the original algorithm specified by Arifovic (1994) in order to hold it as a benchmark for the alternative of adaptive mutation.[7] This is simply a simulation of the algorithm incorporating adaptive mutation where the learning rate, γ, takes the value of 0. For all individuals, the mutation rate is initialized at a value of 0.025. As the learning rate is null, in these baseline simulations the mutation rate does not deviate from this value.

In addition to the binary encoding process utilized in the original work considered above, we also incorporate a framework utilizing an encoding process in which adjacent integers differ by only a single bit (the *hamming distance* between adjacent integers equals one). We refer to the two encoding processes as *Binary-Coded* and *Gray-Coded* integers, respectively. *Gray codes* are a group of alternative encoding methods in which this adjacency property holds. Their use in the implementation of genetic algorithms has been shown to improve the performance over implementation utilizing binary encoding. This performance improvement is grounded on the increased potential for small perturbations through successive single mutations of the encoded string.[8]

Every simulation is run for a duration of 10,000 periods and contains 100 individual rules (n)

[7]Arifovic (1994) also considers an application of *individual* learning (refer to section 2.3.2). While feasible through adjustment of the fitness criterion, adaptive mutation has not yet been investigated in models of *individual* learning. Additionally, interpreting the algorithm as a reduced form description of human adaptation is more problematic due to the variation of mutation rates within an individual's set of potential rules. We leave this analysis for future work.

[8]See Hollstein (1971) for a consideration of genetic algorithm performance utilizing Gray-coded integers in a pure mathematical optimization problem.

with 30 bits per encoded rule (k). They share identical demand and cost parameters fundamental to the rational expectations outcome for market price and individual quantity. These include the cost parameters, x and y, which are set at 0.00 and 0.016, respectively. Parameters specific to market demand, A and B, take the values of 2.296 and 0.0168, respectively. According to the equation determining the rational expectations outcome for market price presented in the preceding section, the perfectly competitive market price, P^*, is equal to 1.12.

Summary statistics for the time series of price for the baseline Arifovic (1994) algorithm are contained within Table 4.1. Simulations occur with a probability of crossover (p_c) equal to zero ("Without Crossover") and equal to 0.60 ("With Crossover"). For each parameterization of the probability of crossover, we utilize the algorithm incorporating the election operator, and the algorithm without its presence. For each simulation, the average price ($\bar{P}^*$) and standard deviation of price (δ) are reported. We also calculate the standard deviation of price over sub-periods of the entire simulation equal to 25. The average of these sub-period standard deviations ($\bar{\delta}$) is reported for each simulation.[9]

Price Statistics

	Without Crossover		With Crossover	
	No Election	Election	No Election	Election
Binary-Coded Integers				
$\bar{P}^*$	1.1400	1.1200	1.1402	1.1200
δ	0.0442	0.0033	0.0436	0.0033
$\bar{\delta}$	0.0383	0.0002	0.0380	0.0003
Gray-Coded Integers				
$\bar{P}^*$	1.1364	1.1200	1.1395	1.1200
δ	0.0435	0.0040	0.0438	0.0039
$\bar{\delta}$	0.0391	0.0003	0.0397	0.0003

Table 4.1: GA simulations of the Cobweb model - Arifovic (1994)

The results presented in Table 4.1 are consistent with those reported in the original Arifovic (1994) work. Although true convergence does not occur in the simulations not utilizing the election operator, the average price over each simulation is within one standard deviation of the rational expectations outcome. For each simulation without the election operator, the standard deviation of price over the 25 period sub-samples is only slightly smaller than that calculated over the entire

[9] In order to ensure that the results are robust to different sequences of random numbers, all simulations are conducted over multiple runs using different initializing seed values for the random number generator. The set of initializing seeds for the random number generating process is identical between simulation frameworks.

simulation. Plots of the time series of price for the simulations are contained within Figure 4.1 and 4.2. As the strength of the similarity between the simulation with and without the crossover operator precludes yielding any additional insight, we do not present the analogous plot for the simulation that does not incorporate the crossover operator.

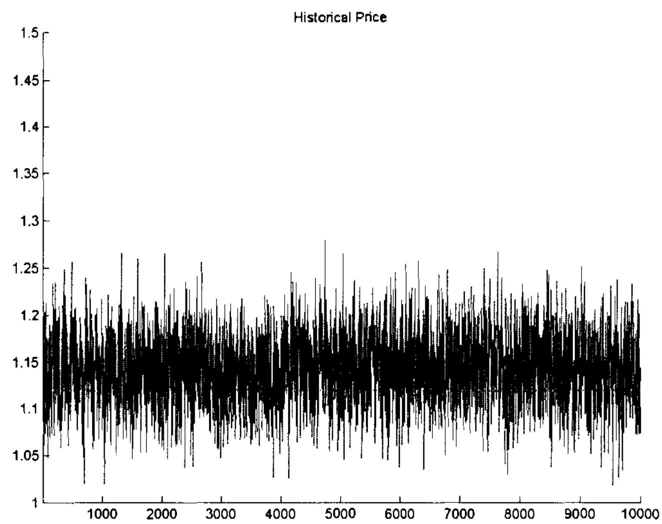

Figure 4.1: Market Price (With Crossover) - Arifovic (1994).

The use of Gray-coded integers improves the performance of the simulations only marginally. Average price statistics are slightly closer to their rational expectations level. However, the standard deviation of prices does not differ from their binary-coded counterparts in any significant or consistent manner. While Gray encoded strings may increase the performance of genetic algorithms in some contexts, they do not appear to do so in the framework considered within this work. We proceed using binary-coded integers in the remainder of this work.

4.5.2 Adaptive Mutation - Baseline Fitness Function

Against these baseline simulations of Arifovic (1994), we consider the results for the algorithm incorporating adaptive mutation in Table 4.2 and 4.3.

The dynamics of the simulation are investigated over various permutations of the parameters specific to the evolution of individuals' mutation rates. These will include the rate of learning, γ, and the minimum allowable value for the individual mutation rate, $\underline{p_m}$. This minimum allowable value functions as a strict lower bound; any alteration of the individuals' mutation rate that leaves it below this threshold is not allowed. Mutation rates that fall below this threshold are reset at the

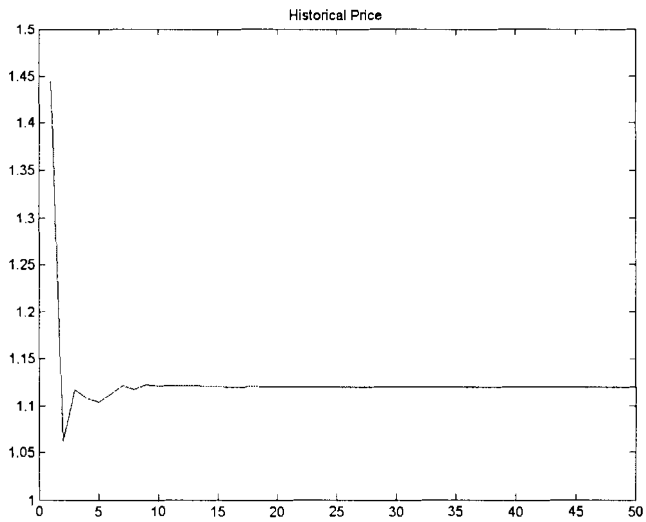

Figure 4.2: Market Price (With Crossover) - Election Operator - Arifovic (1994).

specific value of this lower bound. This lower bound is motivated towards maintaining a minimum level of heterogeneity and innovation within the simulation.

In total, there are thirty-five permutations of these two parameters. We allow the learning rate, γ, to take five different values;

$$\gamma \ \varepsilon \ [0.05, 0.10, 0.15, 0.20, 0.25]$$

The lower bound on individual mutation rates takes seven different values,

$$\underline{p_m} \ \varepsilon \ [0.00, 0.003\bar{3}, 0.006\bar{6}, 0.010, 0.015, 0.020, 0.025]$$

Over the permutations of the learning rate (γ) and the lower bound on mutation ($\underline{p_m}$), we simulate algorithms including the crossover operator ($p_c = 0.6$) and without ($p_c = 0.0$). Price summary statistics and summary statistics for the distribution of average mutation rates for the simulation without incorporating the crossover operator are contained in Table 4.2 and Table 4.3, respectively. These statistics for the simulations including the crossover operator are included in Table 4.8 and Table 4.9 of the Appendix.

The summary statistics for the distribution of average mutation rates include the sample mean ($\bar{p_i'}$), standard deviation ($\delta_{p_i'}$), and skewness ($\gamma_{p_i'}$). Each of these summary statistics is based on the per-period average mutation rate across all individual rules.

Price Statistics

γ		Lower bound on mutation rates 0.00	0.00$\bar{3}$	0.00$\bar{6}$	0.010	0.015	0.020	0.025
	$\bar{P}^*$	1.1233	1.1364	1.1337	1.1320	1.1348	1.1389	1.1452
0.05	δ	0.0363	0.0390	0.0450	0.0450	0.0426	0.0455	0.0456
	$\bar{\delta}$	0.0132	0.0269	0.0329	0.0368	0.0376	0.0405	0.0413
	$\bar{P}^*$	1.0952	1.1370	1.1389	1.1356	1.1391	1.1444	1.1494
0.10	δ	0.0345	0.0393	0.0425	0.0441	0.0451	0.0444	0.0472
	$\bar{\delta}$	0.0075	0.0291	0.0342	0.0364	0.0394	0.0400	0.0427
	$\bar{P}^*$	1.0887	1.1316	1.1400	1.1454	1.1422	1.1459	1.1510
0.15	δ	0.0223	0.0474	0.0434	0.0427	0.0449	0.0446	0.0468
	$\bar{\delta}$	0.0041	0.0318	0.0351	0.0370	0.0393	0.0413	0.0431
	$\bar{P}^*$	1.1150	1.1396	1.1368	1.1404	1.1454	1.1489	1.1571
0.20	δ	0.0295	0.0430	0.0464	0.0454	0.0467	0.0467	0.0476
	$\bar{\delta}$	0.0030	0.0321	0.0369	0.0390	0.0420	0.0421	0.0439
	$\bar{P}^*$	1.1123	1.1384	1.1414	1.1452	1.1492	1.1544	1.1640
0.25	δ	0.0324	0.0419	0.0448	0.0460	0.0471	0.0472	0.0476
	$\bar{\delta}$	0.0031	0.0328	0.0370	0.0389	0.0423	0.0432	0.0446

Table 4.2: GA simulations of the Cobweb model - Adaptive Mutation - *Baseline Fitness Function* - **Price Statistics**

Distribution Statistics - p'_i (Mutation Rate)

γ		Lower bound on mutation rates 0.00	$0.00\bar{3}$	$0.00\bar{6}$	0.010	0.015	0.020	0.025
	$\bar{p}'_i$	0.0031	0.0075	0.0112	0.0158	0.0229	0.0313	0.0372
0.05	$\delta_{p'_i}$	0.0063	0.0054	0.0044	0.0040	0.0044	0.0071	0.0067
	$\gamma_{p'_i}$	3.7419	3.1814	3.3711	2.3457	1.4544	1.6680	1.5604
	$\bar{p}'_i$	0.0011	0.0082	0.0160	0.0198	0.0294	0.0354	0.0431
0.10	$\delta_{p'_i}$	0.0035	0.0042	0.0065	0.0060	0.0081	0.0067	0.0085
	$\gamma_{p'_i}$	7.2641	3.8987	2.2407	2.5686	2.0058	1.2180	1.7087
	$\bar{p}'_i$	0.0008	0.0113	0.0180	0.0246	0.0329	0.0420	0.0503
0.15	$\delta_{p'_i}$	0.0036	0.0061	0.0066	0.0078	0.0084	0.0096	0.0115
	$\gamma_{p'_i}$	8.2036	2.4786	1.5607	1.8421	1.6887	1.5670	2.0373
	$\bar{p}'_i$	0.0006	0.0131	0.0213	0.0286	0.0390	0.0489	0.0562
0.20	$\delta_{p'_i}$	0.0034	0.0060	0.0082	0.0093	0.0112	0.0121	0.0119
	$\gamma_{p'_i}$	10.236	2.0066	1.8230	1.7424	1.7525	1.3166	1.1583
	$\bar{p}'_i$	0.0006	0.0154	0.0253	0.0332	0.0454	0.0551	0.0662
0.25	$\delta_{p'_i}$	0.0033	0.0073	0.0097	0.0100	0.0123	0.0137	0.0159
	$\gamma_{p'_i}$	11.7072	2.2108	1.7286	1.2271	1.2037	1.2425	1.3045

Table 4.3: GA simulations of the Cobweb model - Adaptive Mutation - *Baseline Fitness Function* - **Distribution Statistics - p'_i (Mutation Rate)**

For comparison with baseline simulation results in Figure 4.1, plots of the time series of price and average mutation rates are included in Figures 4.3 through 4.7.

Figures 4.3 and 4.4 contain simulation output in which $(\gamma, \underline{p_m})$ is set equal to $(0.05, 0.03\bar{3})$ and the likelihood of crossover, p_c, equals 0.0 and 0.60, respectively. Figures 4.5 through 4.7 present simulations without crossover for the following $(\gamma, \underline{p_m})$ pairs - (0.05, 0.00), (0.15, 0.0), and (0.25, 0.0). Plots of the average mutation rate are accompanied by the time series of their distributions' minimum and maximum values.

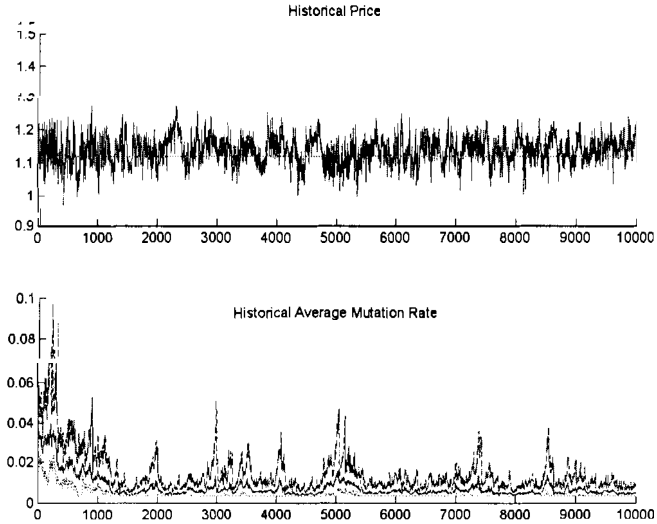

Figure 4.3: Market Price (Without Crossover) - Learning Rate (γ) 0.05, Lower bound on Mutation $(\underline{p_m})$ $0.03\bar{3}$.

Interestingly, in the majority of simulations, calculated over the simulations' entire duration the average price is above the rational expectations outcome. However, these mean prices are all within a single standard deviation of this equilibrium outcome. That is, the majority of simulations have mean prices that are above the competitive equilibrium outcome, though not significantly.[10]

Importantly, this is an expected result. Consider the rational expectations outcome for individual quantity and price, 0.70 and 1.12, respectively. Assume that all rules are currently consistent with this rational expectations outcome for individual quantity and that over this population mutation occurs. Furthermore, assume that for each mutation that increases an individual quantity by x percent, there is an associated mutation over a rule that decreases it by this same x percent. As such, the distribution of rules following mutation is centered around the rational expectations outcome and is symmetric. Average individual output is unchanged, and the price in the following period will

[10]Notably, we find that the Arifovic (1994) framework suffers from the same consistent positive discrepancy in average prices over the duration of the simulation (refer to Table 4.1).

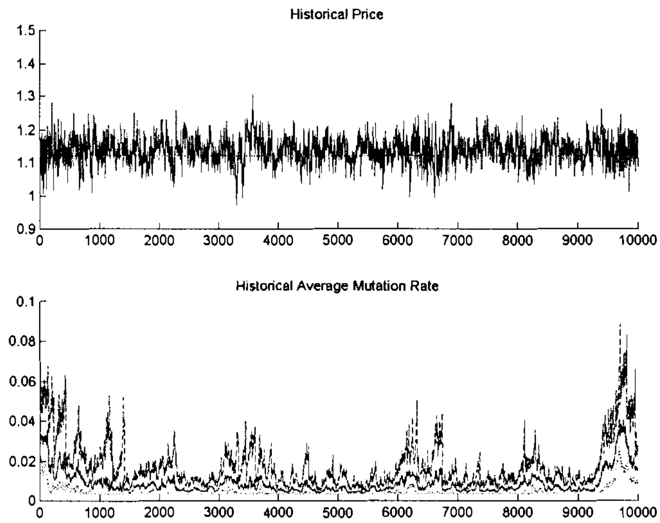

Figure 4.4: Market Price (With Crossover) - Learning Rate (γ) 0.05, Lower bound on Mutation (p_m) $0.00\bar{3}$.

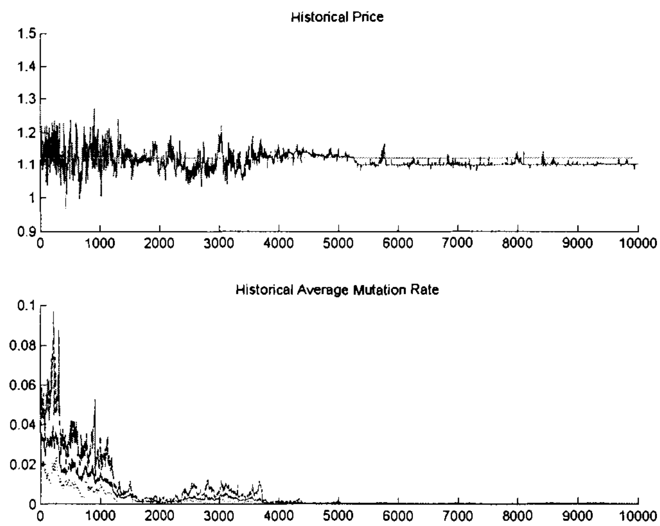

Figure 4.5: Market Price (Without Crossover) - Learning Rate (γ) 0.05, Lower bound on Mutation (p_m) 0.0.

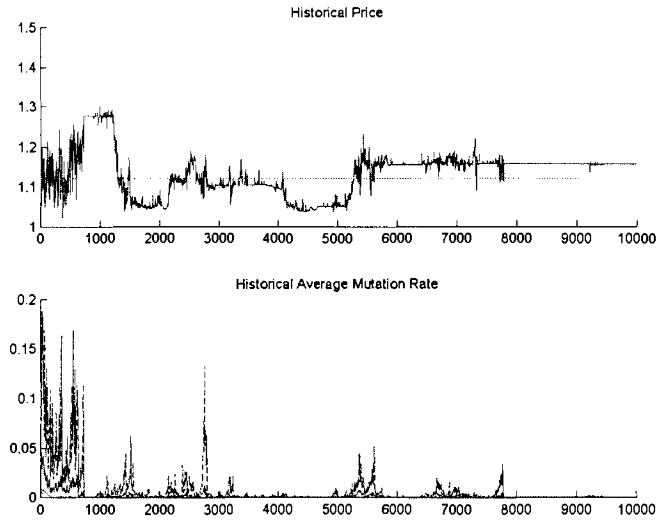

Figure 4.6: Market Price (Without Crossover) - Learning Rate (γ) 0.15, Lower bound on Mutation (p_m) 0.0.

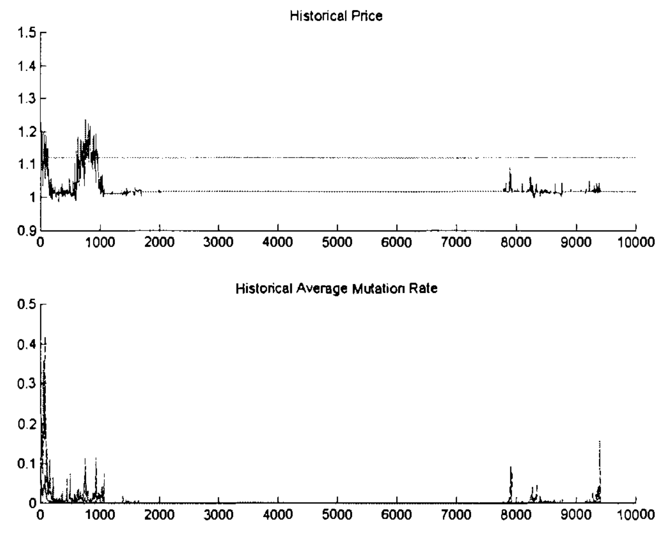

Figure 4.7: Market Price (Without Crossover) - Learning Rate (γ) 0.25, Lower bound on Mutation (p_m) 0.0.

remain at its rational expectations value.

All rules that are associated with a quantity that is not equal to the rational expectations level are associated with lower levels of profit, and therefore lower levels of fitness. However, those rules that deviated *below* the rational expectations level of output will be associated with a higher fitness than those that deviated *above* this level by an equal percent. This stems from the specification of the cost function. It's second derivative is positive, implying the changes in profit levels from equal x percent changes in output will not equal (holding the price constant at its rational expectations level). Rules associated with the rational expectations outcome are the most likely to be selected during the subsequent replication process. However, deviations below this level are more likely to be selected than their associated increase, as their fitness values are superior. As such, in the following period, replication results in a distribution of rules skewed towards lower values of production. This, in turn, favors positive deviations in price from its rational expectations outcome. Eventually, assuming no further mutations occur, the continuing favor for rules associated with output levels close to the rational expectations solution in the replication process eliminates all mutated rules from the population, returning it to the rational expectations level. However, this re-convergence occurs only after a positive deviation in the price.

Discussion

With respect to the comparison between the baseline and self adaptive simulations, four important regularities warrant discussion.

- Result (1) - For certain parameterizations of the self adaptation mechanism, $(\gamma, \underline{p_m})$, sample distribution statistics pertaining to market price are indistinguishable between simulations utilizing fixed and self adaptive mutation. Statistical convergence is unaffected by the adoption of endogenous mutation rates.

That is, for intermediate values of the parameters of the self adaptation mechanism, the mean and standard deviation of price over the full duration of the simulation are indistinguishable from those of the baseline Arifovic (1994) simulations. If one is only concerned with statistical convergence over the entire duration of simulations spanning many generations, including the complexity of self adaptation may not be parsimonious. This important regularity, however, holds only over the entire simulation sample. Contrarily, over smaller sub-samples of the simulations, these self adaptation distribution statistics look quite different from those of the baseline simulations.

- Result (2) - Adaptive mutation lowers small duration deviation measures from their population equivalent and from those of the baseline simulations.

Sub-sample deviation measures do decrease from their population counterparts in the baseline simulations. However, this decrease is of the magnitude of only 13.3 and 12.8 percent in the simulations with and without crossover, respectively. These magnitudes are far smaller from those of simulations with self adaptation, even in those utilizing intermediate values for the $(\gamma, \underline{p_m})$ parameters. For example, consider in Table 2 those simulations in which the rate of learning, γ, takes the value of 0.10. When the lower bound on mutation, $\underline{p_m}$, is set at 0.025 the difference between the full sample standard deviation, δ, and the sub-sample standard deviation, $\overline{\delta}$, is roughly 9.5 percent. Despite the fact that $\underline{p_m}$ is set equal to the rate of mutation in the simulation of the baseline Arifovic framework, the difference between full and sub-sample standard deviation measures is slightly lower. This stems from the fact that $\underline{p_m}$ acts as a lower bound on mutation rates and that the average rate of mutation over the course of a simulation will always be larger than this value. As the lower bound on mutation falls, so to will the average rate of mutation (see Table 4.3). In simulations in which the average rate of mutation is lower, the difference between the full and sub-sample standard deviation of price will increase. As evidence, consider the simulation for which the rate of learning, γ, is set equal to 0.10 and the lower bound on mutation, $\underline{p_m}$, is allowed to be 0.00. Here the difference between full and sub-sample standard deviation of price is approximately 78 percent.

The substantial decrease in the sub-sample deviation measures from their population counterparts is indicative of an important serial autocorrelation inherent within the time series for price that is not a factor in the baseline simulations.

- Result (3) - When comparing baseline versus simulations incorporating self adaptation, the short duration dynamics of price look substantially different. Autocorrelation relationships in price become important for determining the intra-period dynamics of price.

This is clearly a phenomenon in the figures presented above. In the baseline simulation, deviations from the rational expectations outcome appear to have no constructive relationship with each other. That is, these deviations appear to be simply white-noise. This is not the case for deviations associated with simulations incorporating self adaptation. In these simulations, negative deviations from the rational expectations outcome are very likely to be followed be a subsequent negative deviation, indicative of a increasingly significant autoregressive relationship. For research concerning short run dynamics, self adaptation may no longer be excluded on the basis of parsimony.

A very important point is worth noting. Adaptive mutation is not necessary for the relationship described above in results (2) and (3); nor is it sufficient. These results are derived from the fact that average mutation rates are allowed to fall to an extremely low level. Any application of genetic algorithms in which mutation rates are quite low will be characterized in the same manner as we have described above. This stems from the fact that for mutation rates below $1/kn$, less than a single mutation is expected in every period across the entire population of rules. For periods in which no mutation occurs, those rules played in the preceding period are no different from those played in the current. The outcomes between these two periods do not differ; contributing to a autocorrelation in results. The results are dependent on the very low mutation rates that result from the adaptive mutation mechanism, and not the mechanism itself.[11]

- Result (4) - The progression towards a significant autocorrelation relationship in outcomes inherent with adaptive mutation may be at the expense of lower convergence reliability.

If true convergence is to be attained in simulations involving genetic algorithm models without utilizing the election operator, mutation rates must fall to levels approaching zero. That is, the lower bound on mutation rates, $\underline{p_m}$, must be equal to zero. As stressed in the preceding paragraphs, this will lead to significant autocorrelation in the time series of simulated outcomes. The timing in which this autoregressive relationship is attained has important implications for the convergence reliability of simulations incorporating adaptive mutation. This timing is critically related to the rate of learning, γ. As evidence of this claim, consider Figures 4.5 through 4.7 in which the lower bound on mutation, $\underline{p_m}$, is equal to zero and the rate of learning varies between the values 0.05, 0.15, and 0.25.

High values of γ cause high speeds of adaptation with respect to the individual rate of mutation. When the lower bound on mutation rates is too low, this large rate of learning causes a very high, and possibly premature adaptation of very low mutation rates. This premature adoption has the effect of removing the introduction of diversity into the population prior to the widespread adoption of a rule consistent with rational expectations. As such, the convergence reliability of the simulations is quite low. That is, the reliability of the rational expectations outcome in these simulations is somewhat low when compared to those simulations with lower learning rates and/or larger lower bounds on the rate of mutation. Though the reliability with respect to the rational expectations outcome is somewhat low, the velocity with which these simulations approach a non-rational outcome is very high. In comparison, the relative success of simulations with lower learning rates is likely driven by the fact that decreases in the mutation rate are much slower, therein avoiding simulative traps characterized by very low rates of mutation and non-rational price levels. Mutation rates in such simulations stay sufficiently high for a long enough progression of generations so as to allow rules to adopt a rational expectations strategy.

Importantly, such traps never theoretically dismiss the possibility of convergence. As mutation rates may never take a value of zero, there will always be some innovation introduced into the environment, though this innovation may not occur in every period.[12] All that is required for convergence to eventually occur is for an innovation to be introduced that has a high enough relative fitness so as to begin the process of replication. The low mutation, non-rational situation is referred

[11] The sufficiency of adaptive learning for results (2) and (3) will be examined alongside *fitness dependent mutation modificators* considered later in this work.

[12] Note that a deterministic mutation rate of $1/k = 0.033$ leads to an expected mutation of one binary bit per individual in every period. A deterministic mutation rate of $1/(kn) = 0.00033$ leads to an expected mutation of one binary bit over the entire population, therein guaranteeing the expectation of at least one innovation in every period. Any deterministic mutation rate lower than $1/(kn)$ will introduce innovation, though not, in expectation, every period.

to as a trap only to emphasize that simulations *will* escape this trap, though it will occur with very low probability in every period.

In brief, autocorrelated outcomes associated with very low rates of mutation are required for convergence to the rational expectations outcome. As such, the lower bound on mutation must be set at or very near zero. However, if you approach these low mutation rates too quickly, you are likely converging to a non-rational outcome. Even for the lowest parameterization of the rate of learning, the reliability of the outcome with respect to the rational expectations equilibrium is far from anything warranting implementation of adaptive mutation.

4.5.3 Adaptive Mutation - Extended Fitness Function

The importance of selective pressure has already been discussed in Section 4.1. Bäck and Schütz have claimed that for the self adaptation mechanism to work effectively at least one bit per binary string must be mutated on average. Of course, as discussed above, this would preclude convergence without an election operator. However, the purpose of their lower bound on mutation is to guarantee enough diversity to ensure significant selective advantage. This need not be the only manner to attain selective pressure. Schwefel (1987, 1992) demonstrated that strong selective pressure is necessary for the self adaptation principle to work. He proposed (μ, λ)-selection to attain such selective pressure, though Bäck and Schütz have shown only a smaller convergence velocity is associated with proportional selection, not a lack of convergence. In any case, as demonstrated in the preceding section, as the lower bound on mutation is allowed to fall below the levels proposed by Bäck and Schütz the necessity for significant selective pressure increases beyond that which may be supplied by the baseline fitness calculation considered in previous literature.

There is the possibility that low-mutation pitfalls may be avoided if one strengthens the selective pressure within the algorithm. This stronger selective pressure may preclude the premature adoption of very low levels of mutation. With a stronger selection pressure, rules adopting a critically low level of mutation before reaching the rational expectations outcome will have a lower likelihood of replication. Their propagation is less likely when selection is more strict, as other rules with only slightly higher fitness values have a higher likelihood of replication.

In order to assess this conjecture, we simulate the same parameterizations of the framework considered above. However, instead of the baseline fitness function we adopt the transformed effective fitness function of equation (8). Table 4.4 and 4.5 contain price and mutation rate sample statistics for the effective fitness simulations without the crossover operator. These tables are directly comparable to Table 4.2 and 4.3 in which the baseline fitness function was utilized. The appendix contains Tables 4.10 and 4.11 in which the effective fitness framework is simulated with the crossover operator. These tables are directly comparable to Tables 4.8 and 4.9 of the appendix.

Time series data for select simulations are contained in Figures 4.8 and 4.9. In each of these

Price Statistics

γ		Lower bound on mutation rates 0.00	$0.00\bar{3}$	$0.00\bar{6}$	0.010	0.015	0.020	0.025
	$\bar{P}^*$	1.1241	1.1237	1.1329	1.1401	1.1438	1.1415	1.1417
0.05	δ	0.0248	0.0135	0.0224	0.0250	0.0322	0.0348	0.0383
	$\bar{\delta}$	0.0063	0.0108	0.0182	0.0219	0.0288	0.0317	0.0349
	$\bar{P}^*$	1.1261	1.1235	1.1349	1.1424	1.1426	1.1437	1.1440
0.10	δ	0.0212	0.0148	0.0236	0.0291	0.0345	0.0386	0.0405
	$\bar{\delta}$	0.0046	0.0123	0.0198	0.0251	0.0311	0.0350	0.0378
	$\bar{P}^*$	1.1164	1.1272	1.1408	1.1438	1.1462	1.1444	1.1472
0.15	δ	0.0155	0.0193	0.0264	0.0306	0.0368	0.0395	0.0416
	$\bar{\delta}$	0.0032	0.0148	0.0223	0.0270	0.0333	0.0369	0.0386
	$\bar{P}^*$	1.1323	1.1300	1.1391	1.1450	1.1428	1.1472	1.1517
0.20	δ	0.0178	0.0215	0.0285	0.0338	0.0392	0.0402	0.0425
	$\bar{\delta}$	0.0024	0.0173	0.0242	0.0297	0.0346	0.0369	0.0393
	$\bar{P}^*$	1.1202	1.1331	1.1424	1.1444	1.1449	1.1494	1.1552
0.25	δ	0.0160	0.0232	0.0299	0.0357	0.0398	0.0415	0.0440
	$\bar{\delta}$	0.0025	0.0187	0.0256	0.0309	0.0361	0.0382	0.0409

Table 4.4: GA simulations of the Cobweb model - Adaptive Mutation - *Extended Fitness Function* - **Price Statistics**

Distribution Statistics - p_i' (Mutation Rate)

γ		Lower bound on mutation rates 0.00	$0.00\bar{3}$	$0.00\bar{6}$	0.010	0.015	0.020	0.025
0.05	$\bar{p_i'}$	0.0013	0.0060	0.0109	0.0144	0.0218	0.0273	0.0352
	$\delta_{p_i'}$	0.0037	0.0032	0.0032	0.0028	0.0037	0.0035	0.0051
	$\gamma_{p_i'}$	6.4844	6.6422	3.9720	3.6932	1.4744	1.9281	1.1864
0.10	$\bar{p_i'}$	0.0009	0.0070	0.0125	0.0183	0.0264	0.0346	0.0416
	$\delta_{p_i'}$	0.0031	0.0027	0.0030	0.0041	0.0057	0.0073	0.0073
	$\gamma_{p_i'}$	9.0291	4.7813	2.1136	1.6081	2.2835	2.0052	2.0025
0.15	$\bar{p_i'}$	0.0007	0.0086	0.0154	0.0217	0.0301	0.0405	0.0477
	$\delta_{p_i'}$	0.0030	0.0033	0.0042	0.0056	0.0063	0.0092	0.0088
	$\gamma_{p_i'}$	8.5024	3.1278	1.5962	1.9415	1.4696	1.5192	1.2669
0.20	$\bar{p_i'}$	0.0006	0.0105	0.0180	0.0257	0.0364	0.0461	0.0549
	$\delta_{p_i'}$	0.0029	0.0044	0.0059	0.0076	0.0090	0.0099	0.0116
	$\gamma_{p_i'}$	10.047	2.7600	2.1765	1.6540	1.2425	1.0442	1.3947
0.25	$\bar{p_i'}$	0.0006	0.0124	0.0213	0.0297	0.0412	0.0526	0.0626
	$\delta_{p_i'}$	0.0032	0.0048	0.0073	0.0082	0.0112	0.0138	0.0147
	$\gamma_{p_i'}$	10.437	2.0891	2.0598	1.1134	2.4710	1.8112	1.2425

Table 4.5: GA simulations of the Cobweb model - Adaptive Mutation - *Extended Fitness Function* - **Distribution Statistics - p_i' (Mutation Rate)**

figures, the lower bound on mutation, $\underline{p_m}$, is zero. The rate of learning, γ, is 0.15 in Figure 4.8 and 0.25 in Figure 4.9. These two figures have baseline fitness counterparts presented in Figures 4.6 and 4.7, respectively.

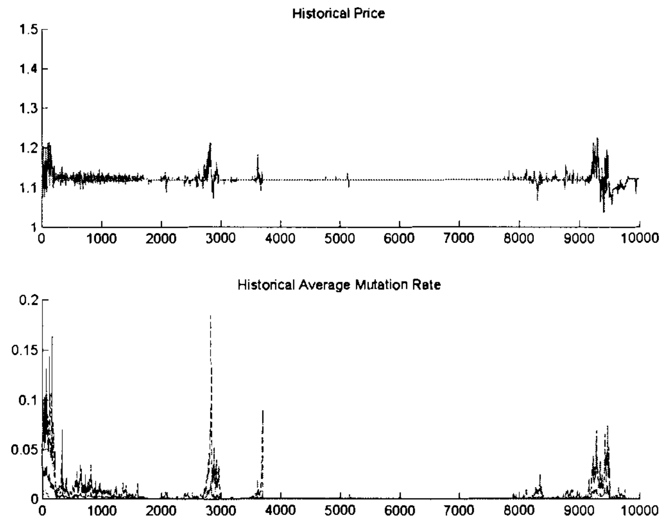

Figure 4.8: Market Price (Without Crossover) - Learning Rate (γ) 0.15, Lower bound on Mutation (p_m) 0.0, Effective Fitness Transformation.

Discussion

Comparing Tables 4.4 and 4.5 with their baseline fitness function counterparts, Tables 4.2 and 4.3, provides insight into the importance of having very strict selective pressure, especially when one considers simulations in which the lower bound on mutation is allowed to be zero.

- Result (5) - In algorithms utilizing adaptive mutation with a lower bound below that proposed by Bäck and Schütz, a high degree of selective pressure is required in order to avoid low-mutation non-convergence traps.

This result becomes even more apparent when one compares Figures 4.7 and 4.8 with their baseline fitness function counterparts. Situations in which mutation rates fall to a critically low level preceding convergence to the rational expectations outcome are much less likely to occur under the effective fitness transformation. While these traps become far less likely, they are still not an impossibility. However, the convergence reliability of the adaptive mutation algorithm with the effective fitness transformation represents a significant improvement over the constant mutation algorithm baseline. Simulations with lower bounds of mutation set at zero are characterized by long

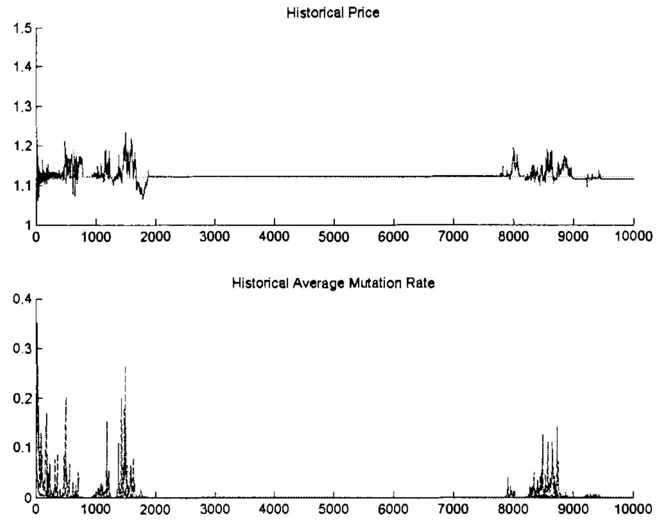

Figure 4.9: Market Price (Without Crossover) - Learning Rate (γ) 0.25, Lower bound on Mutation ($\underline{p_m}$) 0.0, Effective Fitness Transformation.

stretches of convergent behavior with periodic deviations. These deviations are relatively short lived, with behavior converging with the rational expectations outcome afterwards.

The constant introduction of diversity proposed by Bäck and Schütz is not a necessity if one can introduce selective pressure in other ways. Here, the transformation of raw profits into the effective fitness function is sufficient for providing such pressure.

4.5.4 Adaptive Mutation - Fitness *dependent* mutation modificators ($\rho > 0$)

While the effective fitness transformation is sufficient in the environment considered within this work, it may not necessarily be sufficient in all simulation contexts. However, imposing a limitation on mutation rates may not be required if one is willing to incorporate a fitness dependent mutation modificator penalty. Such penalties may replace the need for a lower bound by limiting the reduction in mutation rates only to those that increase raw fitness values above the cost associated with their adoption.

As it has been shown that there is no empirical distinction between simulations with, and without the crossover operator, the introduction of fitness dependent mutation modificators will be limited to simulations with a likelihood of crossover (p_c) equal to zero. The lower bound on mutation rates ($\underline{p_m}$) is held at 0. For each simulation, the global mutation rate (m) and the mutation rate modificator factor (q) are set equal to 2 and 0.033, respectively.[13] Various permutations of the remaining underlying simulation parameters are considered. The mutation rate modificator penalty (ρ) is drawn from the following set.

$$\rho \ \varepsilon \ [0.0001, 0.001, 0.01, 0.05, 0.1, 0.2]$$

The parameter governing the rate of learning, γ, is drawn from the same set as within the above analysis.

$$\gamma \ \varepsilon \ [0.05, 0.10, 0.15, 0.20, 0.25]$$

As such, there are thirty permutations of these two parameters. Results for each parameterization are contained in Table 4.6 and Table 4.7. Varying the mutation rate modificator penalty (ρ), these tables are directly comparable to the first column of Table 4.2 and 4.3 where the value of ρ is implicitly equal to zero.[14] Plots associated with two specific parameterizations are contained in Figure 4.10 and 4.11. Figure 4.11 is directly comparable to Figure 4.9 in which the mutation modificator penalty is implicitly zero.

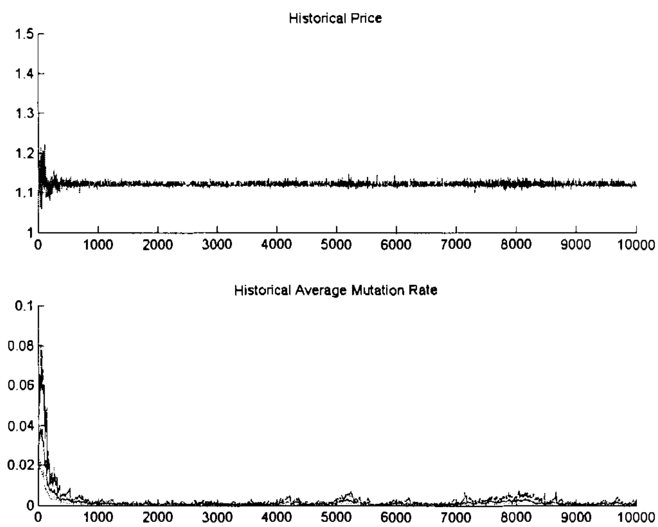

Figure 4.10: Market Price (Without Crossover) - Learning Rate (γ) 0.05, Lower bound on Mutation (p_m) 0.0, Fitness Penalty (ρ) 0.0001.

[13] Refer to equation (4.9) and (4.12).

[14] By way of illustration, compare the two equations determining relative fitness in the fitness independent and dependent settings; equation (4.8) and equation (4.12), respectively

Price Statistics

γ		Mutation Rate Modificator Penalty (ρ) 0.0001	0.001	0.01	0.05	0.1	0.2
	$\bar{P}^*$	1.1210	1.1211	1.1230	1.1427	1.1438	1.1418
0.05	δ	0.0073	0.0086	0.0113	0.0342	0.0354	0.0365
	$\bar{\delta}$	0.0039	0.0040	0.0067	0.0310	0.0328	0.0336
	$\bar{P}^*$	1.1213	1.1234	1.1238	1.1436	1.1436	1.1412
0.10	δ	0.0077	0.0099	0.0125	0.0349	0.0360	0.0378
	$\bar{\delta}$	0.0042	0.0038	0.0079	0.0311	0.0333	0.0348
	$\bar{P}^*$	1.1214	1.1213	1.1234	1.1432	1.1435	1.1426
0.15	δ	0.0082	0.0080	0.0133	0.0352	0.0373	0.0369
	$\bar{\delta}$	0.0045	0.0041	0.0092	0.0315	0.0343	0.0339
	$\bar{P}^*$	1.1215	1.1213	1.1242	1.1417	1.1417	1.1436
0.20	δ	0.0079	0.0080	0.0136	0.0358	0.0378	0.0364
	$\bar{\delta}$	0.0042	0.0050	0.0100	0.0317	0.0341	0.0340
	$\bar{P}^*$	1.1218	1.1214	1.1251	1.1427	1.1425	1.1415
0.25	δ	0.0077	0.0073	0.0137	0.0351	0.0369	0.0368
	$\bar{\delta}$	0.0049	0.0047	0.0101	0.0316	0.0339	0.0343

Table 4.6: GA simulations of the Cobweb model - Adaptive Mutation - *Mutation Modificator Penalty* - **Price Statistics**

Distribution Statistics - p_i' (Mutation Rate)

γ		Mutation Rate Modificator Penalty (ρ) 0.0001	0.001	0.01	0.05	0.1	0.2
	$\bar{p_i'}$	0.0016	0.0016	0.0035	0.0271	0.0317	0.0325
0.05	$\delta_{p_i'}$	0.0036	0.0039	0.0047	0.0056	0.0034	0.0023
	$\gamma_{p_i'}$	7.8388	6.3624	3.5069	-0.0615	0.1929	-0.0157
	$\bar{p_i'}$	0.0017	0.0014	0.0041	0.0267	0.0317	0.0336
0.10	$\delta_{p_i'}$	0.0029	0.0027	0.0044	0.0060	0.0042	0.0028
	$\gamma_{p_i'}$	6.8804	7.8137	3.6015	0.4071	0.1903	0.2773
	$\bar{p_i'}$	0.0019	0.0016	0.0048	0.0280	0.0331	0.0342
0.15	$\delta_{p_i'}$	0.0033	0.0032	0.0045	0.0066	0.0049	0.0032
	$\gamma_{p_i'}$	5.4898	7.0157	2.1214	0.4927	0.3914	0.2984
	$\bar{p_i'}$	0.0017	0.0020	0.0055	0.0296	0.0344	0.0352
0.20	$\delta_{p_i'}$	0.0032	0.0041	0.0052	0.0074	0.0057	0.0037
	$\gamma_{p_i'}$	8.7904	10.2387	3.0695	0.4085	0.4068	0.4679
	$\bar{p_i'}$	0.0021	0.0020	0.0057	0.0304	0.0354	0.0365
0.25	$\delta_{p_i'}$	0.0033	0.0031	0.0046	0.0086	0.0061	0.0042
	$\gamma_{p_i'}$	7.9113	9.1336	1.8007	0.6882	0.4141	0.3604

Table 4.7: GA simulations of the Cobweb model - Adaptive Mutation - *Mutation Modificator Penalty* - **Distribution Statistics - p_i' (Mutation Rate)**

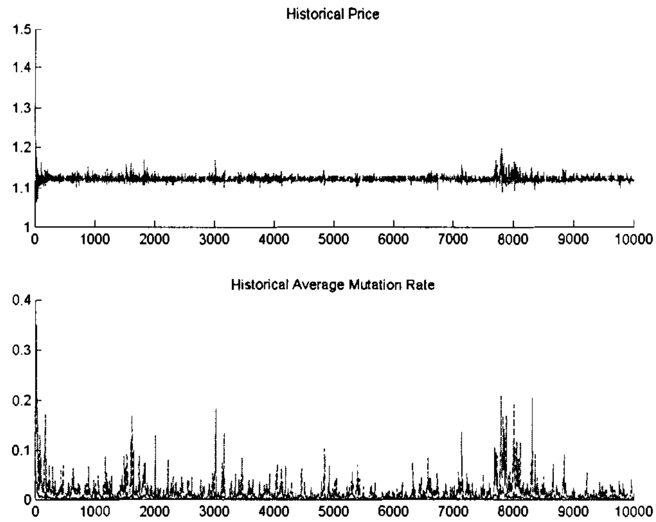

Figure 4.11: Market Price (Without Crossover) - Learning Rate (γ) 0.25, Lower bound on Mutation (p_m) 0.0, Fitness Penalty (ρ) 0.0001.

Discussion

The addition of fitness penalties has important impacts on the resulting dynamics of the simulations. Although true convergence is not associated with any of the parameterizations, a notable result warrants discussion.

- Result (6) - Maintaining diversity without invoking a lower bound on mutation rates is possible through the introduction of even very small fitness dependent mutation modificator penalties.

As already noted, low mutation rates are attained through very high levels for the parameter of learning (γ) in conjunction with a level for the lower bound of mutation ($\underline{p_m}$) set at zero. The introduction of the fitness penalty assures rules with mutation rates below the level associated with long run fitness improvements greater in absolute value than their respective fitness penalty are not proliferated. The higher the mutation rate modificator penalty (ρ), the larger must be any fitness improvements associated with mutation rate deviations for their proliferation among the population. As evidence of this fact, in Table 4.7, for high levels of the rate of learning, γ, as the mutation rate modificator penalty increases (ρ), so to does the average rate of mutation across all individuals, $\bar{p}'_i$. For sufficiently punitive levels of ρ, average rates of mutation approach the parameterization of m, the global mutation rate.[15]

[15]Refer to equation (4.9) and (4.12).

Therefore, even in simulative environments in which an effective fitness transformation is insufficient to avoid non-convergent low-mutation rate traps, imposing a lower bound on mutation is not necessary for maintaining diversity. Utilizing a modificator penalty guarantees a level of diversity that balances the potential fitness benefits of lower mutation rates against their cost. Invoking a lower bound on mutation rates is equivalent to placing a very punitive fitness penalty on mutation adaptation. While this special case is sufficient for maintaining a predetermined level of diversity, it is not necessary. Even for the very smallest fitness penalty, mutation rates are unlikely to fall to levels associated with low mutation rate traps, regardless of the simulative context. While these fitness penalties are not required in all environments, in those where an insufficient level of selective pressure presents a problem for convergence, introducing such penalties may serve to replace the parameterization of a lower bound on mutation.

4.6 Conclusion

Limiting its focus to the mutation operator, this work acknowledges the importance of the introduction of innovation, but maintains the level with which innovation is introduced should be determined within the framework of the model rather than being exogenously imposed or limited through the use of an election operator. The mechanism proposed for determining the rate at which innovation is introduced is based on the idea of *on-line learning*, or *self-adaptation*.

The performance of simple genetic algorithms to ones in which the election operator is included and those characterized by *self-adaptation* in an economic setting is compared. The cobweb environment examined by Arifovic (1994) is utilized for the analysis.

It is demonstrated that for *self-adaptation* to yield results consistent with convergence to the rational expectations equilibrium, a high degree of selective pressure is required. In the framework considered, a simple fitness transformation is sufficient for providing this required selective pressure. In simulations utilizing this fitness transformation, those with lower bounds of mutation set at zero are characterized by long stretches of convergent behavior with periodic deviations. These deviations are relatively short lived, with behavior converging with the rational expectations outcome afterwards.

Though this fitness transformation is sufficient in the context of this work, it is argued that in environments for which it is insufficient, utilizing a lower bound on mutation above zero may not be required if the model incorporates *fitness dependent mutation modificators.* Utilizing a *modificator penalty* guarantees a level of diversity that balances the potential fitness benefits of lower mutation rates against their cost. Even for the very smallest fitness penalty, mutation rates are unlikely to fall to levels associated with low mutation rate traps, regardless of the simulative context.

In an economic system of constant change, there is likely a requirement for constant introduction of innovation. It is likely, however, that there is an optimal rate at which this innovation occurs and that it is dependent on the underlying stochastic nature of the system in question. This work lends itself to extensions in which the economic environment is characterized by constant change; it's consideration is left for future work.

References

Arifovic, J. (1994), "Genetic algorithm learning and the cobweb model," *Journal of Economic Dynamics and Control*, 18, 3-28.

Arifovic, J. and J. Ledyard, (2004) - Forthcoming, "Computer testbeds: the dynamics of the Groves-Ledyard mechanism."

Arifovic, J. and M.K. Maschek, (2004), "Expectations and currency crisis - an experimental approach." Manuscript (Simon Fraser University).

Bäck, T. and M. Schütz, "Intelligent Mutation Rate Control in Canonical Genetic Algorithms."

Brandenburger, A. (1984), "Information and learning in market games," Unpublished manuscript (Stanford University, Stnaford, CA).

Bray, M.M. and N.E. Savin (1986), "Rational Expectations Equilibria, Learning, and Model Specification," *Econometrica*, 54, 1129-1160.

Carlson, J. (1969), "An invariably stable cobweb model," *Review of Economic Studies*, 69, 360-362.

DeCanio, S.J. (1979), "Rational Expectations and Learning from Experience," *Quarterly Journal of Economics*, 93, 47-57.

Frydman, R. (1982), "Towards and understanding of market processes," *American Economic Review*, 72, 652-668.

Fogel, D.B. (1995), **Evolutionary Computation: Toward a New Philosophy of Machine Intelligence**. (IEEE Press, Piscataway, NJ).

Fogel, D.B, L.J. Fogel and W. Atmar (1991), "Meta-evolutionary programming." In R.R. Chen, editor, **Proc. 25th Asilomar Conference on Signals, Systems and Computers**. (Pacific Grove, CA) 540-545.

Hollstien, R.B. (1971), "Artificial genetic adaptation in computer control systems." (Doctoral dissertation, University of Michigan).

Holt, C. and A. Williamil (1986), "A laboratory experiment with a single-person cobweb," *Atlantic Economic Journal*, 219, 51-54.

Kim, J.T. (1998), "Energy Dependent Adaptation of Mutation Rates in Computer Models of Evolution." In C. Adami, R. Below, H. Kitano and C.E. Taylor, editors, **Artificial Life VI: Proceedings of the Sixth International Conference on Artificial Life** (The MIT Press, Cambridge, Massachusetts) 248-255.

Marcet, A. and T.J. Sargent (1987), "Convergence of least-squares learning in environments with hidden state variables and private information," *Journal of Political Economy*, 97, 337-368.

Muth, J.F. (1961), "Rational Expectations and the Theory of Price Movements," *Econometrica*, 29, 315-335.

Nerlove, M. (1958), "Adaptive Expectations and Cobweb Phenomena," *Quarterly Journal of Economics*, 72, 227-240.

Nyarko, Y. (1990), "Bayesian Rationality and Learning Without Common Priors," Unpublished manuscript (New York University, New York, N.Y.).

Schwefel, H.-P. (1987), "Collective intelligence in evolving systems." In W. Wolff, C.-J. Soeder and F.R. Drepper, editors, **Ecodynamics, Contributions to Theoretical Ecology** (Springer, Berlin) 95-100.

Schwefel, H.-P. (1992), "Imitating evolution: Collective, two-level learning processes." In U.Witt, editor, **Explaining Process and Change - Approaches to evolutionary economics** (The University of Michigan Press, Ann Arbor, MI.) 49-63.

Schwefel, H.-P. (1995), **Evolution and Optimum Seeking.** Sixth-Generation Computer Technology Series (Wiley, New York, N.Y.)

Townsend, R.M. (1978), "Market anticipations, rational expectations, and Bayesian analysis," *International Economic Review*, 19, 481-494.

Wellford, C.P. (1989), "A laboratory analysis of price dynamics and expectations in the cobweb model," Discussion paper 89-15 (University of Arizona, Tucson, AZ).

4.7 Appendix

Price Statistics

γ		Lower bound on mutation rates						
		0.00	$0.00\bar{3}$	$0.00\bar{6}$	0.010	0.015	0.020	0.025
	$\bar{P}^*$	1.1435	1.1376	1.1364	1.1353	1.1370	1.1404	1.1467
0.05	δ	0.0407	0.0381	0.0410	0.0418	0.0445	0.0462	0.0451
	$\bar{\delta}$	0.0162	0.0294	0.0330	0.0362	0.0396	0.0408	0.0411
	$\bar{P}^*$	1.1215	1.1403	1.1397	1.1397	1.1389	1.1459	1.1510
0.10	δ	0.0360	0.0425	0.0430	0.0436	0.0445	0.0446	0.0458
	$\bar{\delta}$	0.0117	0.0332	0.0360	0.0383	0.0405	0.0411	0.0425
	$\bar{P}^*$	1.1163	1.1380	1.1351	1.1396	1.1440	1.1500	1.1561
0.15	δ	0.0317	0.0411	0.0446	0.0439	0.0450	0.0455	0.0463
	$\bar{\delta}$	0.0091	0.0340	0.0380	0.0387	0.0408	0.0425	0.0433
	$\bar{P}^*$	1.1190	1.1385	1.1364	1.1403	1.1490	1.1557	1.1637
0.20	δ	0.0344	0.0429	0.0455	0.0450	0.0448	0.0472	0.0477
	$\bar{\delta}$	0.0082	0.0360	0.0398	0.0402	0.0412	0.0438	0.0446
	$\bar{P}^*$	1.1277	1.1380	1.1409	1.1442	1.1557	1.1628	1.1699
0.25	δ	0.0269	0.0446	0.0464	0.0468	0.0482	0.0478	0.0497
	$\bar{\delta}$	0.0068	0.0382	0.0408	0.0422	0.0441	0.0446	0.0465

Table 4.8: GA simulations of the Cobweb model (with genetic crossover) - Adaptive Mutation - *Baseline Fitness Function* - **Price Statistics**

Distribution Statistics - p_i' (Mutation Rate)

γ		Lower bound on mutation rates 0.00	$0.00\bar{3}$	$0.00\bar{6}$	0.010	0.015	0.020	0.025
	$\bar{p_i'}$	0.0036	0.0087	0.0143	0.0193	0.0266	0.0339	0.0399
0.05	$\delta_{p_i'}$	0.0059	0.0059	0.0068	0.0077	0.0075	0.0113	0.0095
	$\gamma_{p_i'}$	3.2003	2.7056	1.8773	1.5539	1.7262	2.4885	2.1294
	$\bar{p_i'}$	0.0027	0.0124	0.0190	0.0241	0.0316	0.0391	0.0479
0.10	$\delta_{p_i'}$	0.0057	0.0077	0.0088	0.0097	0.0087	0.0091	0.0122
	$\gamma_{p_i'}$	4.4553	1.5181	1.6488	2.6054	1.5770	1.3193	1.4336
	$\bar{p_i'}$	0.0018	0.0146	0.0207	0.0285	0.0369	0.0480	0.0554
0.15	$\delta_{p_i'}$	0.0049	0.0085	0.0077	0.0103	0.0101	0.0136	0.0131
	$\gamma_{p_i'}$	6.5463	2.0391	1.5639	1.5339	1.3211	1.5530	1.0230
	$\bar{p_i'}$	0.0017	0.0182	0.0275	0.0324	0.0452	0.0550	0.0649
0.20	$\delta_{p_i'}$	0.0051	0.0111	0.0116	0.0110	0.0150	0.0160	0.0164
	$\gamma_{p_i'}$	5.8095	1.9103	1.7343	1.2378	1.5272	1.3641	1.1479
	$\bar{p_i'}$	0.0015	0.0208	0.0318	0.0399	0.0521	0.0622	0.0731
0.25	$\delta_{p_i'}$	0.0051	0.0112	0.0136	0.0140	0.0164	0.0183	0.0191
	$\gamma_{p_i'}$	6.5630	1.7534	1.7123	1.2275	1.2931	1.2306	1.2149

Table 4.9: GA simulations of the Cobweb model (with genetic crossover) - Adaptive Mutation - *Baseline Fitness Function* - **Distribution Statistics - p_i' (Mutation Rate)**

Price Statistics

γ		Lower bound on mutation rates 0.00	$0.00\bar{3}$	$0.00\bar{6}$	0.010	0.015	0.020	0.025
	$\bar{P}^*$	1.1125	1.1260	1.1347	1.1403	1.1428	1.1448	1.1434
0.05	δ	0.0327	0.0178	0.0235	0.0290	0.0329	0.0374	0.0387
	$\bar{\delta}$	0.0097	0.0133	0.0195	0.0255	0.0301	0.0343	0.0359
	$\bar{P}^*$	1.1216	1.1277	1.1366	1.1442	1.1436	1.1420	1.1458
0.10	δ	0.0262	0.0193	0.0243	0.0302	0.0358	0.0386	0.0411
	$\bar{\delta}$	0.0103	0.0149	0.0213	0.0273	0.0330	0.0359	0.0383
	$\bar{P}^*$	1.1148	1.1321	1.1416	1.1430	1.1461	1.1455	1.1497
0.15	δ	0.0371	0.0216	0.0283	0.0326	0.0366	0.0396	0.0422
	$\bar{\delta}$	0.0072	0.0174	0.0250	0.0298	0.0342	0.0368	0.0399
	$\bar{P}^*$	1.1314	1.1350	1.1435	1.1450	1.1454	1.1473	1.1530
0.20	δ	0.0224	0.0246	0.0312	0.0360	0.0397	0.0416	0.0424
	$\bar{\delta}$	0.0051	0.0205	0.0283	0.0327	0.0369	0.0393	0.0404
	$\bar{P}^*$	1.1060	1.1371	1.1443	1.1437	1.1470	1.1503	1.1571
0.25	δ	0.0300	0.0260	0.0331	0.0376	0.0406	0.0428	0.0448
	$\bar{\delta}$	0.0061	0.0219	0.0302	0.0347	0.0380	0.0404	0.0425

Table 4.10: GA simulations of the Cobweb model (with genetic crossover) - Adaptive Mutation - *Extended Fitness Function* - **Price Statistics**

Distribution Statistics - p_i' (Mutation Rate)

γ		Lower bound on mutation rates 0.00	0.00$\bar{3}$	0.00$\bar{6}$	0.010	0.015	0.020	0.025
	$\bar{p}_i'$	0.0026	0.0075	0.0121	0.0181	0.0246	0.0310	0.0363
0.05	$\delta_{p_i'}$	0.0058	0.0055	0.0052	0.0078	0.0093	0.0081	0.0060
	$\gamma_{p_i'}$	3.7898	3.3326	2.6625	2.5162	2.9381	2.9008	1.1570
	$\bar{p}_i'$	0.0022	0.0083	0.0139	0.0206	0.0296	0.0374	0.0441
0.10	$\delta_{p_i'}$	0.0044	0.0040	0.0042	0.0061	0.0081	0.0083	0.0090
	$\gamma_{p_i'}$	4.5306	3.1532	2.1559	1.8259	1.8191	1.4788	1.4476
	$\bar{p}_i'$	0.0017	0.0102	0.0177	0.0247	0.0340	0.0435	0.0523
0.15	$\delta_{p_i'}$	0.0051	0.0053	0.0063	0.0085	0.0085	0.0108	0.0121
	$\gamma_{p_i'}$	5.3465	3.0264	1.6788	2.4378	1.2273	1.5013	1.3626
	$\bar{p}_i'$	0.0014	0.0131	0.0225	0.0293	0.0414	0.0496	0.0594
0.20	$\delta_{p_i'}$	0.0047	0.0067	0.0083	0.0093	0.0120	0.0122	0.0144
	$\gamma_{p_i'}$	7.0063	2.2829	1.6435	2.1173	1.7315	1.2523	1.2234
	$\bar{p}_i'$	0.0012	0.0146	0.0256	0.0336	0.0455	0.0568	0.0661
0.25	$\delta_{p_i'}$	0.0042	0.0068	0.0105	0.0109	0.0139	0.0149	0.0155
	$\gamma_{p_i'}$	6.3664	1.9054	1.6776	1.3594	1.6579	1.1313	1.1527

Table 4.11: GA simulations of the Cobweb model (with genetic crossover) - Adaptive Mutation - *Extended Fitness Function* - **Distribution Statistics - p_i' (Mutation Rate)**