MACROECONOMIC MODELS WITH ENDOGENOUS LEARNING

by

ERIC GAUS

A DISSERTATION

Presented to the Department of Economics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2010

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Eric Gaus

Title:

"Macroeconomic Models with Endogenous Learning"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Economics by:

George Evans, Co-Chairperson, Economics
Jeremy Piger, Co-Chairperson, Economics
Shankha Chakraborty, Member, Economics
Sergio Koreisha, Outside Member, Decision Sciences

and Richard Linton, Vice President for Research and Graduate Studies/Dean of the Graduate School for the University of Oregon.

June 14, 2010

Original approval signatures are on file with the Graduate School and the University of Oregon Libraries.

An Abstract of the Dissertation of

Eric Gaus for the degree of Doctor of Philosophy

in the Department of Economics to be taken June 2010

Title: MACROECONOMIC MODELS WITH ENDOGENOUS LEARNING

Accepted: ______________________________
Dr. George Evans, Co-Chair

Accepted: ______________________________
Dr. Jeremy Piger, Co-Chair

The behavior of the macroeconomy and monetary policy is heavily influenced by expectations. Recent research has explored how minor changes in expectation formation can change the stability properties of a model. One common way to alter expectation formation involves agents' use of econometrics to form forecasting equations. Agents update their forecasts based on new information that arises as the economy progresses through time. In this way agents "learn" about the economy.

Previous learning literature mostly focuses on agents using a fixed data size or increasing the amount of data they use. My research explores how agents might endogenously change the amount of data they use to update their forecast equations.

My first chapter explores how an established endogenous learning algorithm, proposed by Marcet and Nicolini, may influence monetary policy decisions. Under rational expectations (RE) determinacy serves as the main criterion for favoring a model or monetary policy rule. A determinant model need not result in stability under an alternative expectation formation process called learning. Researchers

appeal to stability under learning as a criterion for monetary policy rule selection.

This chapter provides a cautionary tale for policy makers and reinforces the importance of the role of expectations. Simulations appear stable for a prolonged interval of time but may suddenly deviate from the RE solution. This exotic behavior exhibits significantly higher volatility relative to RE yet over long simulations remains true to the RE equilibrium.

In the second chapter I address the effectiveness of endogenous gain learning algorithms in the presence of occasional structural breaks. Marcet and Nicolini's algorithm relies on agents reacting to forecast errors. I propose an alternative, which relies on agents using statistical information.

The third chapter uses standard macroeconomic data to find out whether a model that has non-rational expectations can outperform RE. I answer this question affirmatively and explore what learning means to the economy. In addition, I conduct a Monte Carlo exercise to investigate whether a simple learning model does, empirically, imbed an RE model. While theoretically a very small constant gain implies RE, empirically learning creates bias in coefficient estimates.

CURRICULUM VITAE

NAME OF AUTHOR: Eric Gaus

PLACE OF BIRTH: Norwalk, Connecticut

DATE OF BIRTH: June 8, 1982

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon
Boston University

DEGREES AWARDED:

Doctor of Philosophy, Economics, 2010, University of Oregon
Master of Arts, Economics, 2005, Boston University
Bachelor of Arts, Economics, 2005, Boston University

AREAS OF SPECIAL INTEREST:

Macroeconomics and Learning, Macroeconometrics, and Resource Economics

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellowship, University of Oregon Department of Economics 2006-2010
Graduate School Advisory Board, University of Oregon, 2008-present
Voluntary Research Assistant, Boston, Massachusetts, 2004-2005

GRANTS, AWARDS AND HONORS:

Co-Winner of the Miksell Prize, University of Oregon Department of Economics 2009
Outstanding Graduate Teaching Award, University of Oregon, 2009

Kleinsorge Summer Fellowship, University of Oregon Department of Economics 2009
Omicron Delta Epsilon, an Economic Honor Society, 2003-2005

ACKNOWLEDGMENTS

I would like to thank Shankha Chakraborty, George Evans, and Jeremy Piger for their generous time and help. Also, I would like to thank participants in the Macro Group at the University of Oregon for their helpful comments and suggestions. All remaining errors are my own.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table | Page

CHAPTER I

INTRODUCTION

Expectations form a cornerstone of modern macroeconomic research. For many decades a particular form of expectations, called rational expectations (RE), has served as the standard assumption. More recently researchers have considered ways to relax the RE assumption since it relies on agents acquiring more information than one might think reasonable.

One group of researchers have assumed that agents use econometrics to form forecasting equations instead of RE. In this framework as the economy generates new data, agents incorporate the information in their forecasting equations. In this way agents "learn" about the economy. There are many different assumptions one can make in a learning model, and this dissertation explores one set of those assumptions.

Researchers can differentiate learning models by changing the influence of new information. One particular formulation assumes that agents use all available information equally. Another assumes agents use only a fixed sample size of the newest data. In both of these cases the algorithm forces agents either to continually increase the sample size, or use a fixed sample size. My research explores the effects of allowing agents to endogenously change the amount of data they use. This dissertation finds that even small changes in expectation formation can lead to significant changes in the dynamics of the economy and estimation of macroeconomic models.

Marcet and Nicolini (2003) suggest an endogenous learning algorithm that has agents switching between a fixed sample size and letting the sample size grow. They argue that this type of algorithm would function well when the economy exhibits occasional structural breaks. My first chapter explores monetary policy and endogenous learning of the form of Marcet and Nicolini.

In macroeconomic theory a monetary policy rule closes the standard New Keynesian model. Under RE certain parameterizations of some policy rules can lead to indeterminacy. Under other parameterizations policy rules might result in determinacy, but might not be able to learn the equilibrium. Researchers have used stability under learning as a criterion for monetary policy rule selection. The endogenous learning algorithm produces some startling dynamics, and provides a word of caution in choosing a monetary policy rule.

In the second chapter I evaluate the forecasting ability of endogenous-gain algorithms. In addition to Marcet and Nicolini's version, I propose my own endogenous learning algorithm. I find little evidence that Marcet and Nicolini provides significant improvement to a standard gain, but my proposed algorithm seems to perform well under the circumstances for which it was designed.

My third chapter estimates a New Keynesian model and compares RE to two types of learning. I find that the data prefer the endogenous learning model and explore some implications of the learning behavior. This suggests that either RE does not accurately describe the expectation formation process in the economy or that the assumed learning rules capture some higher order dynamics that exist in the data. I also show that a learning estimation does not nest an RE model in an empirical sense.

1. Literature Review

In the wake of the formalization of rational expectations, by Muth (1961), economists attempted to ascertain whether rational expectations equilibria were "learnable." Blume *et al.* (1982) provides an overview of the literature, which essentially finds that if agents fail to make specification errors, the rational expectations equilibrium (REE) is "learnable" via econometric learning. Initially econometric learning assumed that agents have the correct model of the economy, but are uncertain of the parameter values. Consequently, agents estimate those coefficients using standard econometric techniques.

Naturally, the same resistance to rational expectations rose against the learning literature, since agents were still assumed to have a significant amount of information about the economy that economists themselves could not claim to have. Bray (1982), and Frydman (1982) provided the basis of the current learning literature by showing for specific cases that agents were capable of learning the REE even when their model was misspecified. Evans (1985) provided further analytical tools by developing a concept called expectational stability (E-stability). He defines a model as E-stable if the model returns to the REE when expectations are perturbed slightly.

The final component that provides the foundation for the contemporary learning literature is Evans and Honkapohja (1998), which defines the *E-stability principle.* The E-stability principle is a correspondence between the E-stability of an REE and its stability under adaptive learning.[1] This principle has guided much of the learning research in recent years and is the cornerstone of the learning literature.

The learning literature as a whole has remained mostly confined to the theoretical side of the discipline. One of the main theoretical contributions has been

[1]Evans and Honkapohja (2001) p41.

the ability to refine the number of equilibria in a rational expectations model by only considering those that are locally stable under learning. Given the large theoretical literature there have naturally been some forays into empirical estimation of models with learning, however, many of the early attempts were constrained by a lack of computing power. Thus many of these empirical works relied on calibration, that is, attempting to find parameter values that generate data that match the stylized facts of the data.

Current computing power has relaxed these constraints, which allows an econometrician to estimate DSGE models. These models allow for the researcher to impose some structure to the data and thereby identify the "deep" parameters of interest. Previously scholars used *ad hoc*, and sometimes contentious, identification strategies. The advance in technology has led to the ability for some researchers to relax the assumption of complete rationality and estimate DSGE models with learning behavior.

The component of learning that has econometric interest is the gain parameter, which governs the weight placed on updates to prior estimates. When agents use a constant-gain the econometrician can estimate it, whereas a decreasing-gain has no parameters to estimate since it is a function of time. A constant-gain applies more weight to the most recent data, which scholars associate with agents who are concerned with structural breaks. Considering the vast structural break and Markov switching literature, assuming agents worry about time varying parameters seems appropriate.

Milani (2007a) is the first to have estimated a small DSGE model that had agents updating estimates with a constant-gain. In particular, Milani (2007a) estimates a New Keynesian model whose stability properties have been well documented in Bullard and Mitra (2002). Now that econometricians are able to estimate these models, it is important for them to be cognizant of whether the model is stable. Stability of a model may provide natural parameter limits thus constraining the search space. In addition, previous identification techniques relied on the assumption of long-run stability, thus it seems unwise to estimate models that are unstable.

As mentioned above, a constant-gain can be associated with potential structural breaks. While this is beneficial on one hand, on the other, a constant-gain produces greater volatility. Consequently, Marcet and Nicolini (2003) suggest an *ad hoc* gain structure that switches between a constant and a decreasing-gain. Essentially, they assume that agents believe that parameters may be constant for a while, in which case a decreasing-gain would "stabilize" the economy, but every once in a while there might be a change in parameter value, in which case a constant-gain would allow agents to adjust to the new value more quickly.

Milani (2007b) shows that endogenous switching can have significant impacts on time-varying volatility, which he claims may help explain the Great Moderation. Recently debate has arisen over the way Milani incorporated Marcet and Nicolini (2003).[2] Specifically Marcet and Nicolini propose a rule for determining which type of gain to use that is based on recent forecast errors. While they choose to compare recent forecast errors to some arbitrary constant, Milani compares recent forecast errors to the historical average of forecast errors. Both of these is inherently backward looking, as it might take several periods for agents to realize a break has occurred, but the historical average adds further problems, which I discuss in Chapter III.

Murray (2008b) mimics Milani (2007b), but neither reports the stability properties of the switching-gain. Consequently, I contribute to this literature by

[2]See Bullard (2008).

examining the stability properties of this type of gain structure. Further, given some of the problems with the switching-gain, I suggest an alternative that is more forward-looking, and has some other favorable characteristics. Using the knowledge gained from examining these gain structures I analyze, in depth, what taking a learning model to data might reveal to the econometrician.

The rest of the dissertation unfolds as follows: the next chapter explores monetary policy and endogenous-gain learning. Chapter III introduces my alternative endogenous-gain and compares it and the Marcet and Nicolini gain to a constant-gain. The fourth chapter presents preliminary results from an estimation using my alternative endogenous-gain and compares it to the alternatives using Bayesian analysis. Chapter V concludes.

CHAPTER II

ADAPTIVE LEARNING, MONETARY POLICY AND INSTABILITY

Recent research on monetary policy has examined the stability of the rational expectations equilibrium (REE) of a standard New Keynesian (NK) model in the context of learning. Since determinacy does not guarantee stability under learning researchers suggest that policymaker should favor monetary policy rules that result in stability under learning. Learning relaxes the rational expectations assumption by allowing agents to use econometrics to forecast the variables of the economy. As new data appear over time agents "learn" by adjusting their forecast equations.

Most of the learning literature concerning monetary policy assumes that agents use a particular type of learning called decreasing-gain learning, which assumes agents utilize all data available. Evans and Honkapohja (2009) show that many of these interest-rate rules are unstable for plausible parameterizations of the so called constant-gain learning algorithm, which is akin to agents using a rolling window. It is common to assume constant-gain learning in the presence of unobserved structural breaks, since a decreasing-gain performs poorly under these circumstances.

In addressing the issue of hyperinflation, Marcet and Nicolini (2003) suggest a potential improvement to a constant-gain by creating hybrid of constant- and decreasing-gain. This "switching" gain seems reasonable when agents believe coefficients exhibit occasional structural breaks.[1] Marcet and Nicolini argue that during a hyperinflation episode the "tracking," or constant-gain, algorithm perform better than a decreasing-gain. This results from the constant-gain algorithms placing more weight on recent data instead of treating all data equally as is the case under a decreasing-gain. Since a decreasing-gain results in stability, the ability to switch to a decreasing-gain could plausibly allow monetary policy rules that were previously unstable at small values of the constant-gain.

In this chapter I show that this reasonable, switching-gain rule leads to some startling dynamics. I find that some values of the constant-gain portion of the switching mechanism can lead to a prolonged period of temporary deviations from the REE. In the very long-run these occurrences disappear and the REE is attained (a result also found by Marcet and Nicolini (2003) in the context of their hyperinflation model). I find that the switching-gain leads to significant increases in variance of the aggregate variables, specifically 4 to 6 times more output volatility, but does not lead to higher or lower means relative to the REE.

I consider several policy rules that have been suggested in the literature. I extend Evans and Honkapohja (2009) by exploring commitment rules that result from policymaker concerned with deviations from target interest-rates. Specifically, I compare an expectations based rule in the flavor of Evans and Honkapohja (2006) to a commitment rule from Duffy and Xiao (2007). In addition, I reexamine a Taylor-type discretionary policy and commitment rules suggested by Svensson and Woodford (2005) and McCallum and Nelson (2004). I find that the switching-gain has a slightly higher cutoff for stability for all the monetary policy rules. Only the expectations based rule with commitment is robustly stable.

[1] Miliani (2007b) provides support for the switching-gain with some empirical evidence that this type of switching-gain may help explain some of the great moderation.

The rest of the chapter proceeds as follows. First I present the standard NK model and introduces the learning framework. The second section describes the exotic dynamics present with the endogenously switching-gain under a Taylor-type rule. Then I examine various rules under commitment. The penultimate section explores multiple gains and how policymaker might make stability more likely. The fifth section concludes.

1. A New Keynesian Model

The following NK model, presented in section 3 of Evans and Honkapohja (2009), describes the economy,[2]

$$x_t = x^e_{t+1} - \phi(i_t - \pi^e_{t+1}) + g_t, \tag{1.1}$$

$$\pi_t = \beta\pi^e_{t+1} + \lambda x_t + u_t, \tag{1.2}$$

where x_t deviations of output from potential, π_t is inflation, and u_t and g_t are AR(1) processes. The following equations govern these processes:

$$u_t = \rho u_{t-1} + \tilde{u}_t, \text{ and } g_t = \mu g_{t-1} + \tilde{g}_t,$$

where $\tilde{g}_t \sim iid(0, \sigma_g^2)$, $\tilde{u}_t \sim iid(0, \sigma_g^2)$, and $0 < |\mu|, |\rho| < 1$.[3] The Euler equation for consumption generates the output equation (1.1), while (1.2) describes the New Keynesian Phillips Curve. The notation x^e_{t+1} refers to an expectational value, specifically $E^*_{t-1}x_{t+1}$, where the star indicates that expectations need not be rational.

[2]See Woodford (2003) for derivation.

[3]$\mu = \rho = 0.8$ for all simulations.

The model is closed by specifying an interest-rate rule. Duffy and Xiao (2007) suggest an optimal policy rule based on policymaker minimizing a loss function that includes interest-rate stabilization in addition to output and inflation stabilization. Specifically policymaker minimize the following loss function.[4]

$$E_0 \sum_{t=0}^{\infty} \beta^t [\pi_t^2 + \alpha_x x_t^2 + \alpha_i i_t^2], \tag{1.3}$$

where the relative weights of interest-rate and output stabilization are α_i and α_x, respectively. Using the first order conditions of this loss function Duffy and Xiao (2007) derive the following interest-rate rule,

$$i_t = \frac{\varphi\lambda}{\alpha_i}\pi_t + \frac{\varphi\alpha_x}{\alpha_i}x_t. \tag{1.4}$$

Throughout this chapter I maintain a focus on operational monetary policy rules in the sense of McCallum (1999). Since policy rules that use current values of endogenous aggregate variables are untenable, I follow Evans and Honkapohja (2009) by using expectations of contemporaneous (or future) values of output or inflation. This assumption changes the interest-rate rule slightly,

$$i_t = \frac{\varphi\lambda}{\alpha_i}\pi_t^e + \frac{\varphi\alpha_x}{\alpha_i}x_t^e. \tag{1.5}$$

By substituting (1.5) into (1.1) the model can be rewritten in matrix form as,

$$y_t = M_0 y_t^e + M_1 y_{t+1}^e + P\upsilon_t, \tag{1.6}$$

---

[4]All the targets have been set to zero for convenience.

where $y_t = (x_t, \pi_t)'$ and $\upsilon_t = (g_t, u_t)'$ and where,

$$M_0 = \begin{pmatrix} -\frac{\alpha_x \varphi^2}{\alpha_i} & -\frac{\varphi^2 \lambda}{\alpha_i} \\ -\frac{\alpha_x \varphi^2 \lambda}{\alpha_i} & -\frac{\varphi^2 \lambda^2}{\alpha_i} \end{pmatrix}, M_1 = \begin{pmatrix} 1 & \varphi \\ \lambda & \beta + \varphi\lambda \end{pmatrix} \text{ and } P = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}.$$

Rewriting the exogenous shocks in matrix form yields,

$$\upsilon_t = F\upsilon_{t-1} + \tilde{\upsilon}_t,$$

where

$$F = \begin{pmatrix} \mu & 0 \\ 0 & \rho \end{pmatrix}.$$

As in Evans and Honkapohja (2009) I obtain parameter values from Table 6.1 of Woodford (2003), with $\alpha_x = 0.048$, $\varphi = 1/0.157$, $\lambda = 0.024$, $\beta = 0.99$, and $\alpha_i = 0.077$. Further, since I assume an operational monetary policy, agents form expectations at time t, which means that they use current exogenous shocks to form expectations. Under these assumptions, agents' perceived law of motion (PLM) takes the form of the minimum state variable (MSV) solution,

$$y_t = a + c\upsilon_t \tag{1.7}$$

and expectations can be written as $y_t^e = a + c\upsilon_t$ and $y_{t+1}^e = a + cF\upsilon_t$. Substituting these expectations into (1.6) yields the actual law of motion (ALM),

$$y_t = (M_0 + M_1)a + (M_0 c + M_1 cF + P)\upsilon_t. \tag{1.8}$$

There exists a mapping of perceived coefficients to the actual coefficients, which the literature refers to as the T-map. In this particular case the T-mapping is,

$$T(a) = (M_0 + M_1)a,$$
$$T(c) = (M_0 c + M_1 cF + P).$$

The set up of this model implies that if agents start at the REE then they never deviate. However, a small, one time perturbation from the REE may cause agents' expectations to either diverge or reacquire the REE. E-stability exists if the model returns to the REE.

i. Recursive Least Squares

The learning literature uses a recursive form for two reasons. First, it satisfies certain properties for theoretical results and, second, it is mathematically convenient for time series simulations. The standard recursive least squares (RLS) updating equations for a regression of Y_t on X_t are,

$$\hat{\phi}_t = \hat{\phi}_{t-1} + \gamma R_t^{-1} X_t'(Y_t - X_t'\hat{\phi}_{t-1}),$$
$$R_t = R_{t-1} + \gamma(X_t X_t' - R_{t-1}),$$

where $\hat{\phi}_t$ are coefficient estimates and R_t is the moment matrix.

For a univariate system γ is usually a scalar. Frequently, the decreasing-gain version γ is the simply t^{-1}. Decreasing-gain implies that the weight of each observation is the same for a particular set of estimates, but decreases over time as the sample size grows. For the constant-gain version γ is fixed at some value between zero and one. Under this algorithm the oldest data has virtually no weight, which means one can think of constant-gain learning as a rolling window.[5]

A simple extension to this framework is to have multiple univariate RLS algorithms each with a distinct gain. A clever use of Kronecker products allows for a one step update of the coefficient estimates. Suppose the multivariate model of the economy can be written in matrix notation as,

$$w_t = \Psi z_t + \Omega w_t^e + \varepsilon_t, \tag{1.9}$$

where w_t represents an m x 1 vector of dependent variables, z_t represents an n x 1 vector of independent variables, Ψ, and Ω are m x m matrices of coefficients, and ε_t is a m x 1 vector of white noise error terms. As an example suppose that z_t is an exogenous VAR(1).[6] Following the MSV solution, agents would estimate the following model,

$$w_t = bz_t + \tilde{\varepsilon}_t, \tag{1.10}$$

where b is an m x n matrix of coefficients, and $\tilde{\varepsilon}_t$ are the corresponding error terms.

Examine the second equation in the RLS algorithm, which updates the moment matrix. Clearly, even though they have the same data creating the moment matrix, if each equation has a different gain parameter, then the moment matrices differ for each equation. With n different explanatory variables the moment matrix for a single equation is an n x n matrix. Since there are m equations, R_t is a mn x mn matrix. In order to create this matrix, define X_t as $I_m \otimes z_t$, where I_m is an m x m identity matrix, $\otimes$ denotes a Kronecker product, and z_t is an n x 1 vector data for period t.

[5]The window size can be found by taking the inverse of the value of the gain.

[6]This framework can be easily modified to accommodate alternative models.

The updating equation for the coefficient estimates must now conform to these matrix dimensions. Define $\hat{\phi}_t$ as $vec(b')$, which stacks the columns of b', and $Y_t = w_t$, where w_t is an m x 1 vector of data for period t. Assuming γ is a scalar, this setup generates the exact same results as setting up a different algorithm for each equation.

If one desires to incorporate multiple gain learning then γ must be redefined as an mn x mn diagonal matrix where the values on the diagonal are the gain on each coefficient. The first n are associated with the first equation the second n with the second equation and so on.

2. Endogenously Switching-Gains

Evans and Honkapohja show that under constant-gain learning this model achieves E-stablity if the constant-gain parameter takes values less than 0.024. Figures II.1 and II.2 depict a particular realization of the NK economy described above under constant-gain learning. Figure II.1 displays instability, while Figure II.2 displays stability.

Evans and Honkapohja refer to the result as not being "robustly stable," in the sense that it implies that if agents use less than 42 periods of data the model is unstable under learning. Most estimates of constant-gain values imply that agents use approximately 10 to 35 periods of data.[7] The next section allows agents to use an endogenous-gain, which potentially decreases the number of periods of data that results in stability.

The switch in Marcet and Nicolini's (2003) hybrid gain sequence is endogenously triggered by forecast errors. Large errors cause agents to suspect a structural break and therefore they would prefer to use a constant-gain to remove

[7]See Milani (2007a, 2007b) and Branch and Evans (2006).

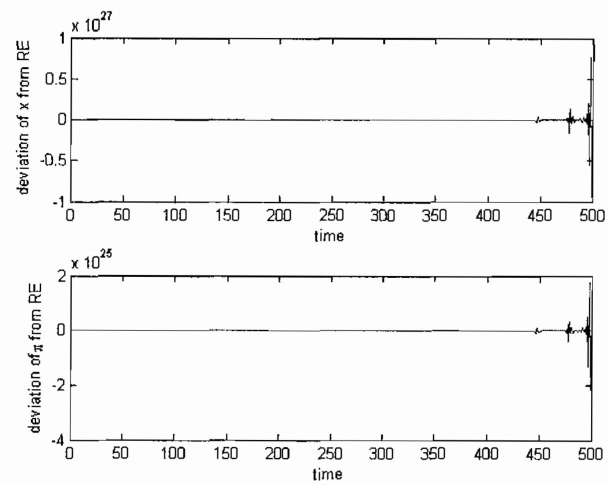

Figure II.1: Instability of optimal Taylor-type rule. $\gamma = 0.04$

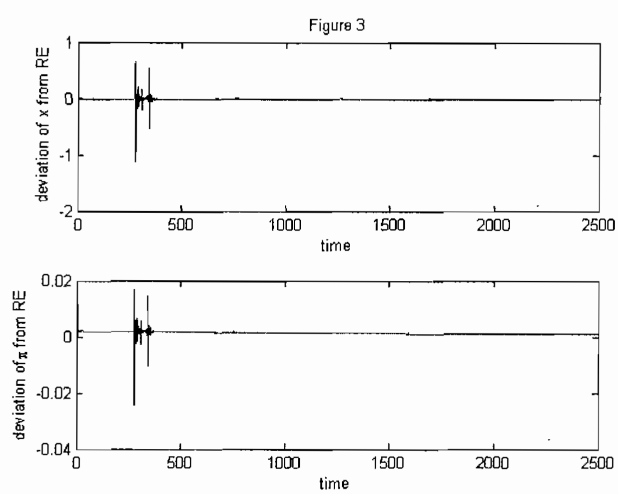

Figure II.2: Stability of optimal Taylor-type rule. $\gamma = 0.02$

the bias of the older data. Once forecast errors fall below a cutoff agents switch back to a decreasing-gain. In Milani (2007b) this cutoff is determined by the historical average of forecast errors. Here I use the Milani variety switching-gain where the historical volatility is a moving average,

$$\gamma_{z,t} = \begin{cases} \frac{1}{\bar{\gamma}_z^{-1}+k} & \text{if } \frac{\sum_{i=t-J}^{t}|z_i - z_i^e|}{J} < \frac{\sum_{i=t-W}^{t}|z_i - z_i^e|}{W}, \\ \bar{\gamma}_z & \text{if } \frac{\sum_{i=t-J}^{t}|z_i - z_i^e|}{J} \geq \frac{\sum_{i=t-W}^{t}|z_i - z_i^e|}{W}. \end{cases} \tag{2.1}$$

where k is the number of periods since the last switch to a decreasing-gain, z denotes a particular variable (π or y), J is the number of periods for recent calculations, and W is the number of periods for historical calculations.[8] Thus, the possibility for the output equation to have a decreasing-gain in the same period that the inflation equation has a constant-gain, and vice versa, exists.

In order to compare directly to Evans and Honkapohja (2009), I use the Woodford parameterization and use the same constant-gain value, $\bar{\gamma}_z$, for both equations, but I allow agents to use the switching-gain described above. In my baseline case I set the constant, $\bar{\gamma}_z$, equal to 0.025, which lies just outside the stable range found by Evans and Honkapohja (2009). I set the historical window length, W, to 35, which suggests that agents use about nine years of past data for the historical volatility indicator. The window length for recent data, J, is set to 4, which is the estimated value found by Milani (2007b).

The RLS algorithm requires a small burn in period to establish a history of error terms. In order to create a seamless transition, I set the burn-in to equal the inverse of the gain. During this period agents use the constant-gain. Given the constant-gain value of 0.025 this implies a burn in length of 40 periods. This ensures

---

[8]In Milani (2007b) W was set to 3000 for very long simulations.

no discontinuity at agent's first opportunity to switch; agents choose between keeping the constant-gain or allowing the value of the gain to decrease.

In the initialization period agent's expectations do not have an effect in the economy. Thus, the coefficients driving the simulation will be a small perturbation away from the RE values. When the initialization period ends agents use the switching-gain in (2.1).

Figure II.3 displays deviations of learning dynamics from the RE solution for a particular realization of the NK economy when agents use the switching-gain with $\bar{\gamma}_z = 0.025$ for $z = x, \pi$.[9] Convergence under learning typically occurs relatively rapidly, however, with a switching-gain large deviations occur for a prolonged period of time.[10] This behavior appears as a general characteristic of the stability of this type of model.

Figure II.4 provides the values of the gain at each point in time. Notice that right before the episode of instability, which occurs around 2400 in Figure II.3, both gains spend a significant period of time near 0.025, or the unstable constant-gain value. Also note that sequences do not mirror each other. In the penultimate section of this chapter I address the stability when gain values differ across equations.

The historical average suggested by Milani partially drives this result. Should one use an arbitrary value in the switching rule as suggested by Marcet and Nicolini (2003), then, for a given value of the constant-gain, there exists a value above which the model is explosive and below which the model rapidly settles into a continuous decreasing-gain regime.

---

[9]Though the last deviation may be in an indicator of instability, extending the simulation to 10,000 periods can show that this deviation is temporary and the future deviations remain close to the REE. This example is meant to show that relatively large deviations can occur later in the simulation.

[10]A Lucas model with a large impact of expectations may take an extended period of time to converge.

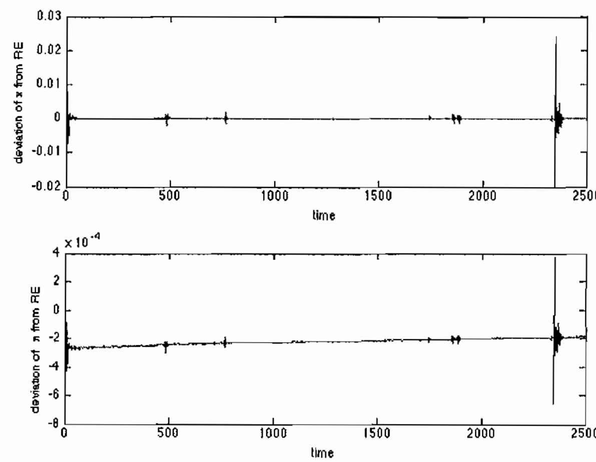

Figure II.3: Stability of optimal Taylor-type rule with endogenously switching-gain.

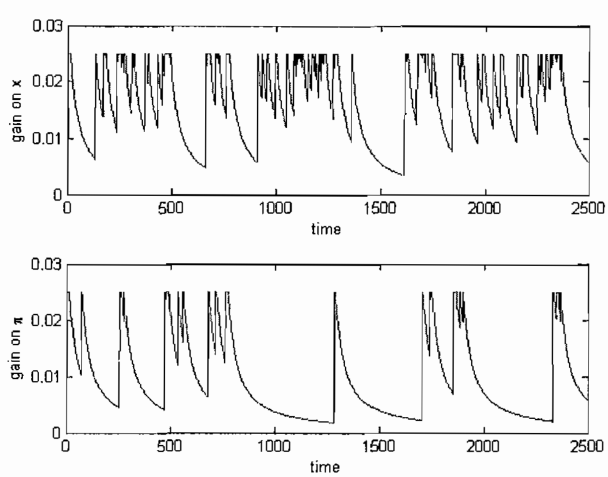

Figure II.4: Value of the endogenously switching-gain.

Table II.1: Examples of Temporary Deviations

	Mean		Variance	
	x	π	x	π
Total	0.9755	0.9757	1.0996	1.0000
100 Periods	0.9843	0.9978	4.8917	1.0020
Total	1.0082	1.0133	1.1178	1.0000
100 Periods	1.0489	0.9955	6.3118	1.0003

Shows the mean and variance of output and inflation of learning relative to RE. The Total row presents statistics for 5,000 period sample in which at least one episode of temporary instability occurs. The 100 Periods row presents statistics for the 100 periods surrounding the episode. I use the Woodford calibrated values and set W=35, J=4, and $\bar{\gamma}_z$=0.025

Table II.1 provides a comparison of the economic significance of the temporary deviations. I choose these examples from two independent simulations of 15,000 periods. After discarding the first 10,000 periods, I calculate the mean and variance of output and inflation relative to the REE. I calculate these statistics for the entire 5,000 periods and also for a 100 period window around the largest temporary deviation in that 5,000 period section.

These examples suggest that the exotic behavior leads to a large increase in variance relative to RE. The top example shows that both inflation and output may be lower than under RE, while the bottom example has both variables above RE. Thus, the only unambiguous result is the increase in variance.

Table II.2 displays stability results for several different historical window lengths.[11] These results are based on 5,000 simulations of 10,000 periods each. In order to evaluate stability I compare the last value of the estimated parameters and the T-map. If the coefficients lie within 2 percent I say that the particular simulation achieved stability. The values represent the percentage of simulations

[11]The constant-gain was set at 0.025, which is in the unstable region of a purely constant-gain regime.

Table II.2: Switching-Gain Stability: Varying the Window Size

Hist. Window	15	25	35	45	55	65
Percent Stable	38.32	68.00	78.18	83.66	85.46	86.86

Hist. Window	75	85	95	105	115	125
Percent Stable	87.32	86.92	87.28	86.54	86.84	85.12

Shows the percent of simulations in which the last value of the estimated parameters lie within 2 percent of the T-map. The historical window is the parameter the governs the number of periods used to calculate the historical average MSFE.

that achieve stability.[12] These results suggest that the relationship between the historical window length and stability exhibits some non-linearity. This non-linearity occurs because of the exogenous shocks that the economy happens to face in a particular simulation. Appendix A provides probabilities for a different set window lengths.

These results show that the intuition for stability under a Marcet and Nicolini or Milani switching-gain may depend on the detailed structure of the constant-gain portion of the algorithm. While Marcet and Nicolini (2003) find similar results in a model with multiple equilibria, I have documented exotic behavior in model with a single REE. Sargent (1999) uses a model in which agents temporarily escape a self-confirming equilibrium as well, but examines government beliefs, not beliefs of the entire economy. Cho *et al* (2002) examine the ordinary differential equations (ODEs) in the system and find that the "escape dynamics" include an additional ODE relative to the mean dynamics. policymaker should be concerned with the potential in this simple NK model for exotic behavior that temporarily strays from the REE.

[12]It should be noted that these percentages are most likely lower bounds since the arbitrary cutoff may happen to occur in the middle of one of the episodes of instability.

3. Optimal Policy with Commitment

Evans and Honkapohja (2009) postulate that the policy rule with commitment in Duffy and Xiao (2007) suffers from the same instability that arises under discretionary policy. As mentioned above, Evans and Honkapohja restrict their examination of commitment to rules where $\alpha_i = 0$, which leaves Duffy and Xiao's rule undefined. In this section, I evaluate the stability of Duffy and Xiao's commitment rule, compare it to the expectations based rule similar to Evans and Honkapohja (2003). I also investigate the cause of the temporary deviations by using policy rules suggested by McCallum and Nelson (2004) and Svensson and Woodford (2005).

As in Evans and Honkapohja (2009), I examine operational interest-rate rules, which requires expectations of contemporaneous variables, or nowcasts. Using nowcasts in the Duffy and Xiao optimal interest-rate rule under commitment (DX) results in,

$$i_t = \frac{\phi\lambda}{\alpha_i}\pi_t^e + \frac{\alpha_x\phi}{\alpha_i}(x_t^e - x_{t-1}) + \frac{\phi\lambda + \beta + 1}{\beta}i_{t-1} - \frac{1}{\beta}i_{t-2}. \tag{3.1}$$

The system under commitment can be written as,

$$y_t = M_0 y_t^e + M_1 y_{t+1}^e + N_0 y_{t-1} + N_1 w_{t-1} + P v_t, \tag{3.2}$$

where $w_t = (i_t, i_{t-1})'$ and the appropriate matrices for, M_0, M_1, N_0, N_1, and P. The MSV solution provides the PLM, which also supplies the form of the RE solution.

$$y_t = a + b_0 y_{t-1} + b_1 w_{t-1} + c v_t. \tag{3.3}$$

For this particular system, the ALM includes the interest-rate process. First, note that the law of motion governing the exogenous variables, w_t, can be written as,

$$w_t = Q_0 y_t^e + Q_1 y_{t-1} + Q_2 w_{t-1}. \tag{3.4}$$

Given this, it follows that the ALM has the following T-mapping,

$$\begin{aligned}
T(a) &= (M_0 + M_1(I + b_0 + b_1 Q_0))a, \\
T(b_0) &= M_0 b_0 + M_1(b_0^2 + b_1 Q_0 b_0 + b_1 Q_1) + N_0), \\
T(b_1) &= M_0 b_1 + M_1(b_0 b_1 + b_1 Q_0 b_1 + b_1 Q_2) + N_1), \\
T(c) &= M_0 c + M_1(b_0 c + b_1 Q_0 c + cF) + P.
\end{aligned}$$

Under the Woodford parameterization I find that the model achieves stability for values of the gain of 0.008 or less. Using Milani's switching-gain extends this region to 0.009, but it does not display the transitory exotic dynamics found under a Taylor-type rule. As predicted by Evans and Honkapohja (2009), the policy under commitment does not fare well under large gains. In fact the instability is so severe that even allowing for temporary switches to a decreasing-gain does not significantly extend the range of values that result in stability

For comparison I turn to the expectations based rule of Evans and Honkapohja (2009). Their rule applies when $\alpha_i = 0$, however, I generalize their expectations based rule to allow $\alpha_i > 0$.[13] This generalization results in the following interest-rate rule (EH),

$$i_t = \delta_1 x_{t-1} + \delta_2 x_{t+1}^e + \delta_3 \pi_{t+1}^e + \delta_4 u_t + \delta_5 g_t \tag{3.5}$$

[13]Note that when $\alpha_i = 0$, the EH rule is identical to the expectations based rule in Evans and Honkapohja (2009).

where,

$$\delta_1 = -\frac{\alpha_x}{\alpha_i \phi^{-1} + \lambda^2 \phi + \alpha_x \phi}, \quad \delta_2 = \frac{\lambda^2+\alpha_x}{\alpha_i \phi^{-1}+\lambda^2\phi+\alpha_x\phi}, \quad \delta_3 = \frac{\lambda\beta + \lambda^2\phi + \alpha_x\phi}{\alpha_i \phi^{-1} + \lambda^2 \phi + \alpha_x \phi},$$
$$\delta_4 = \frac{\lambda}{\alpha_i \phi^{-1} + \lambda^2 \phi + \alpha_x \phi}, \quad \delta_5 = \frac{\lambda^2+\alpha_x}{\alpha_i \phi^{-1}+\lambda^2\phi+\alpha_x\phi}.$$

If the monetary policymaker follow the EH rule, the matrix form of the model is,

$$y_t = My^e_{t+1} + Ny_{t-1} + Pv_t, \tag{3.6}$$

where M, N, and P are the appropriate matrices. The following MSV solution serves as the PLM,

$$y_t = a + by_{t-1} + cv_t. \tag{3.7}$$

Consequently the T-mapping is,

$$T(a) = M(I + b)a,$$
$$T(b) = Mb + N,$$
$$T(c) = Mbc + McF + P.$$

Much like the result in Evans and Honkapohja (2009) all the eigenvalues of the T-mapping under the Woodford parameterization lie within the unit circle. Though the expectations based rule satisfies the E-stability condition, the lagged endogenous variables imply that there exists a possibility for instability for sufficiently high values of the constant-gain.

Similar to the expectations based rule when $\alpha_i = 0$, the EH rule is robustly stable. I find that values of the constant-gain equal to or larger than 0.184 result in the instability of the EH rule under interest-rate stabilization. Using the Milani switching-gain extends the stable range significantly. The EH rule remains stable until values of 0.252 or higher. In these simulations the distinctive dynamics of the Taylor-type rule simulations also does not occur.

i. Alternative Commitment Rules and Temporary Deviations

The exotic behavior that arose under a Taylor-type rule does not appear in the two forms of commitment rule found above. I also found that the range of values of the constant-gain that yield stability increases under the switching-gain. Evans and Honkapohja (2009) assess several other commitment rules, which may exhibit the exotic behavior, or become robustly stable under a switching-gain learning. To address these points, I set $\alpha_i = 0$ and check for exotic behavior and robust stability under the Svensson and Woodford (SW) and McCallum and Nelson (MN) rules.

McCallum and Nelson (2004) suggest a rule based on the optimality condition in the timeless-perspective. When ever this condition is above zero the interest-rate should be above the inflation rate. Using nowcasts the interest-rate rule can be written as,

$$i_t = \pi_t^e + \theta[\pi_t^e + \frac{\alpha_x}{\lambda}(x_t^e - x_{t-1})]. \tag{3.8}$$

Evans and Honkapohja (2009) establish the region of stability for this rule under constant-gain learning, which ends at 0.018. Under a Milani-type switching-gain the stable region extends to 0.019.[14] This value is not large enough to be considered robustly stable according to Evans and Honkapohja (2009).[15] In addition, this rule does not exhibit exotic behavior either.

Similar to the MN rule, Svensson and Woodford (2005) also use the timeless-perspective optimality condition, but also include a fundamentals based

[14]With Woodford parameterization except that $\alpha_i = 0$ and $\theta = 1.5$.

[15]They suggest that a reasonable value for a constant-gain is 0.1 or less.

term. This hybrid rule arose because fundamentals-based rules without interest-rate stabilization can result in indeterminacy and instability under learning.

It has been shown that under rational expectations one can obtain the following fundamentals-based reaction function,

$$i_t = \psi_x x_{t-1} + \psi_g g_t + \psi_u u_t. \tag{3.9}$$

where $\psi_x = b_x[\phi^{-1}(b_x - 1) + b_\pi]$, $\psi_g = \phi^{-1}$, and $\psi_u = [b_\pi + \phi^{-1}(b_x + \rho - 1)c_x + c_\pi \rho$. Additionally, $0 < b_x < 1$ is the unique solution to $\beta b_x^2 - (1 + \beta + \lambda^2/\alpha_x)b_x + 1 = 0$, $b_\pi = (\alpha_x/\lambda)(1 - b_x)$, $c_x = -[\lambda + \beta b_\pi + (1 - \beta\rho)(\alpha_x/\lambda)]^{-1}$, $c_\pi = -(\alpha_x/\lambda)c_x$.[16]

Svensson and Woodford modify the rule by adding the timeless-perspective optimality condition and introducing a new parameter, $\theta > 0$, that supplies the relative weight of the fundamentals versus interest-rate stabilization. The resulting interest-rate rule is,

$$i_t = \psi_x x_{t-1} + \psi_g g_t + \psi_u u_t + \theta[\pi_t^e + \frac{\alpha_x}{\lambda}(x_t^e - x_{t-1})]. \tag{3.10}$$

As found by Evans and Honkapohja (2009), the stability region of this rule ends at 0.018. Again, robust stability is not achieved under a Milani-type switching-gain since the stable region only extends to 0.02.[17] Despite this, I observe exotic behavior when the gain equals 0.02.

It turns out that only the EH rule is "robustly stable," that is, attains stability for plausible values of the the gain parameter. The switching mechanism increases the values of the gain parameter for all of the interest-rate rules yet that

---

[16] See Svensson and Woodford (2005) p34-35 for the derivation.

[17] Same parameretization as the MN rule.

increase is slight for most of them. This suggests that monetary policymaker should place greater emphasis on the role of expectations on interest-rates.

The appearance of the exotic behavior under two distinct interest-rate rules warrants further investigation. In order to establish the cause of this exotic dynamic Table II.3 displays the eigenvalues of the derivative of the T-mapping for each of the respective interest-rate rules.[18] From the table it is clear why the expectations based EH rule does not exhibit the exotic behavior; all of the eigenvalues on the EH rule T-mapping lie within the unit circle.

Evans and Honkapohja (2009) assert that the instability of large constant-gains arises from the large negative eigenvalues found in most of the interest-rate rules. However, this does not appear to be the case for the exotic dynamics. The DX rule has the largest negative values, yet does not display the same characteristics. In addition, the SW rule has smaller negative values and does display the exotic behavior.

The MN rule provides a suitable comparison to the SW rule, since the large negative numbers are approximately equal to each other. The clear pattern is that the eigenvalues that lie within the unit circle are much larger under the SW rule than under the MN rule. In order to test this hypothesis, I simulate the economy using different values for θ, and check for temporary deviations under the SW rule.

When $\theta = 1.75$, the eigenvalues for DT_a are -19.187, and 0.938, those for DT_b are -19.295, 0.833, -20.284, and -0.108, and those for DT_c are -19.405, and 0.727. In this case the exotic dynamics do not appear. However, when $\theta = 1$, the eigenvalues for DT_a are -9.570, and 0.990, those for DT_b are -9.672, 0.878, -10.604, and -0.118, and those for DT_c are -9.775, and 0.766. Under these circumstances the temporary deviations remain present. Even though the eigenvalues outside the unit circle are

[18] Appendix B supplies the analytical formulae for these derivatives.

Table II.3: Eigenvalues of the T-Maps

	Taylor	DX	EH	SW	MN
DT_a	-24.4349	-26.2047	0.0782	-15.9747	-16.1301
	0.9841	0.6446	0.9169	0.9485	0.873
DT_b	-	-25.3538	0.0350	-16.0818	-16.245
	-	-0.8794	0.9114	0.8423	0.762
	-	-26.0369	0	-17.0587	-17.2282
	-	-0.0000	0	-0.1104	-0.1718
DT_c	-24.6657	-25.6947	0.0715	-16.19	-16.3483
	0.7864	0.689	0.7192	0.7353	0.6626
DT_{b_1}	-	-26.3763	-	-	-
	-	-0.1009	-	-	-

Shows the Eigenvalues of the derivatives of the T-maps associated with each of the different monetary policy rules. I use Woodford parameterization, except that $\alpha_i = 0$, and $\theta = 1.5$ for SW and MN.

smaller, those within the unit circle are closer to one, which leads to the exotic behavior. Thus the evidence suggests that the exotic behavior arise when there exists a large negative root and a root within the unit circle, but close to one.

The eigenvalues result from the parameter values, which suggests that the exotic behavior may exist for different parameter settings and monetary policy rule combinations. Since policymaker have control over some of the parameters, α_i and α_x, they may be able to avoid these episodes. The next section provides some evidence for how policy may affect stability.

4. Multiple Constant-Gains

In this section I allow the gains to differ for each equation, but do not allow agents to endogenously switch between constant- and decreasing-gain. Recall that Evans and Honkapohja (2009) find that this particular model is unstable for a gain value at or above 0.024, and depict an example of the stability with a gain of 0.02, and instability with a gain of 0.04.

In order to remain consistent I use the same values for the constant-gain as bounds for the analysis. Instead of having the same constant-gain value for both equations, I allow the gain parameter used in updating the coefficients of the output equation to differ from that on the inflation equation. Using the Woodford parameterization, I simulate the economy 100 times for each set of gain parameters. Each simulation lasts 10,000 periods and I assess stability under learning at the end of each simulation. Stability exists when the change in all coefficients differ by less than two percent.[19]

Table II.4 displays the results of these simulations. I find that approximately 98 percent of the simulations were E-stable when the gain parameter on the output equation is 0.02 regardless of the values of the gain on the inflation equation. If the value of the gain on the output equations is greater than or equal to 0.024 then none of the simulations are stable. This finding is striking since it implies that stability depends on the pair of gain values and that one gain may exert greater influence on the stability under learning.

So far I have explored how different monetary policy rules may affect stability, however, policy parameters may also have a significant impact on stability under learning. In order to assess the effect of policy I let $\alpha_x = a_x 0.048$ and $\alpha_i = a_i 0.077$, and vary a_i and a_x. This allows for the examination of relative changes. Table II.5 presents the results for simulations, where a_i and a_x take on values between 0.5 and 1.5.

When the ratio between α_x and α_i remains the same, *i.e.* on the diagonal of the table, the greater the response by policymaker results in a greater probability of stability. The upper-right triangle shows that whenever α_x is larger than the Woodford parameter ratio none of the simulations achieve stability. The lower-left

[19]There exist several numerical techniques that assess stability under learning which would not significantly alter the results presented here.

Table II.4: Stability of Multiple Constant-Gains

		γ_x				
		0.020	0.021	0.022	0.023	0.024
γ_π	0.020	98	96	81	18	0
	0.022	98	97	82	18	0
	0.024	98	95	82	15	0
	0.026	98	95	82	14	0
	0.028	98	95	82	12	0
	0.030	98	95	79	8	0
	0.032	98	95	76	5	0
	0.034	98	95	74	4	0
	0.036	98	94	72	2	0
	0.038	98	93	70	2	0
	0.040	98	91	68	3	0

Values in the table are the percent of the simulations in which the difference in estimated coefficients from the T-map was less than two percent. γ_x is the gain associated with the output equation, and γ_π is the gain associated with the inflation equation.

triangle shows that the opposite is true when policy response to interest-rates is relatively stronger.

This may seem counterintuitive, since the stability of the model reacts more to changes to the gain on the output equation. A closer inspection of the set-up of the model shows that a stronger reaction to interest-rates results in a reduction in the effect of contemporaneous expectations, *i.e* the M_0 matrix. Specifically stronger interest-rate stabilization policy implies that policymaker react less to expected inflation and output. Therefore, this analysis reinforces the notion that policymaker should remain cognizant of agents' expectations.

5. Conclusion

Researchers have debated the merits of monetary policy rules under learning using two types of gain structures, decreasing and constant. A hybrid of these two

Table II.5: Policy Effect on Stability

		a_x										
		0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5
a_i	0.5	29	0	0	0	0	0	0	0	0	0	0
	0.6	100	41	0	0	0	0	0	0	0	0	0
	0.7	100	100	53	0	0	0	0	0	0	0	0
	0.8	100	100	100	63	0	0	0	0	0	0	0
	0.9	100	100	100	100	64	0	0	0	0	0	0
	1	100	100	100	100	100	68	0	0	0	0	0
	1.1	100	100	100	100	100	99	70	0	0	0	0
	1.2	100	100	100	100	100	100	99	75	0	0	0
	1.3	100	100	100	100	100	100	100	99	73	0	0
	1.4	100	100	100	100	100	100	100	100	98	76	0
	1.5	100	100	100	100	100	100	100	100	100	97	76

The numbers in the table are the percent of simulations in which the difference in estimated coefficients from the T-map was less than two percent. a_i and a_x scale the parameters of the monetary policymakers loss function, α_i and α_x, respectively. $\gamma_x = 0.022$ and $\gamma_\pi = 0.04$.

types of gains provides a cautionary tale for monetary policymaker. policymaker should realize the potential for a model that is stable in the very long run to experience 4 to 6 times more volatility for a particular length of time. In addition, though the switching-gain extended the stable region for all interest-rate rules, only the expectations based rule remains "robustly stable" in the sense suggested by Evans and Honkapohja (2009).

Up till now the importance of alternative gain sequences in the NK model has not been studied. While stability of traditional gain parameters abound, the stability results for alternatives, such as Marcet and Nicolini (2003) and Milani (2007b), have not been established. The analysis above shows that switching-gains result in stability, but potentially develop exotic dynamics. In addition, the analysis above provides evidence that under multiple gains the combination of gains determines stability of the model. Stability results may be more sensitive to a particular gain.

While the choice of the interest-rate rule is important, policymaker may also influence the stability of the model by changing policy parameters. The results above suggest that monetary policy maker should pay close attention to expectations and try to limit the impact of expectations on interest-rate fluctuations.

CHAPTER III

TIME VARYING PARAMETERS AND LEARNING

As models that incorporate learning behavior are brought to the data some researchers have used endogenous-gain parameters to approximate agent behavior under structural break uncertainty. A large empirical literature is devoted to determining structural breaks, which suggests that expectation formation should account for this potential strategy. Presumably when agents try to accomodate structural breaks specifically their forcasting ability should improve.

Though intuition suggests that the type of modeling proposed by Marcet and Nicolini (2003) should be superior to a constant-gain, the improvement has not been documented. I show, using a simple model, that the *ad hoc* endogenous-gain proposed by Marcet and Nicolini (2003) is indistinguishable from a constant-gain. In addition, I propose an alternative, which performs well compared to a constant-gain without expectational feedback. When expectational feedback exists my endogenous-gain is a Nash equilibrium.

The motivation for the Marcet and Nicolini switching-gain relies on the assumption that agents believe that coefficients may exhibit a structural break. Under time-varying parameters, the optimal estimation technique is Bayesian estimation using a filtering process. For example, random walk parameter variation is optimally estimated using the Kalman filter.[1] However, the learning literature

---

[1]Each filtering process depends on the exact type of time variation.

assumes that agents are boundedly rational, thus agents in these types of models may not have the optimal tools available to them. In many cases, lack of mathematical ability may be sufficient reason to dismiss agents' ability to use Bayesian techniques.

Several papers address time-varying parameters in a variety of contexts within learning. For example, Bray and Savin (1986) present a model where agents misspecify the model by assuming parameters are constant, when in fact expectational feedback causes parameters to vary over time. They find that agents, using a standard Durbin-Watson statistic, detect their misspecification error, which suggests that agents should adjust their specification appropriately. Bullard (1992) and McGough (2003) investigate the convergence properties of a model when agents correctly identify the time-varying nature of coefficients. The condition for convergence requires that agents believe that the conditional variance of the time varying parameter (TVP) declines toward zero.

The papers described above all assume the same TVP structure, namely, a random walk. Consequently, the use of a Kalman filter seems quite natural. Beck and Wieland (2002) and Wieland (2000) examine the performance of optimal Bayesian learning and alternative decision rules in a TVP world. However, alternative TVP processes are not compatible with the assumptions needed for using the Kalman filter.

Evans and Ramey (2006) use a model of TVP that is complex enough to make devising the optimal filter rather difficult. Thus, it is more natural for agents in such a setting to use an approximation or an *ad hoc* rule to capture the variation over time. Specifically Evans and Ramey (2006) examine agents choosing between a constant and a decreasing-gain.

My contribution extends and, in some ways, combines Carceles-Poveda and Giannitsarou (2007) and Beck and Wieland (2002). These papers assess the performance of various gain structures under different conditions. Carceles-Poveda and Giannitsarou (2007) examine initialization of decreasing-gain, constant-gain and stochastic gradient learning in models where the underlying parameters are constant. Beck and Wieland (2002) examine various decision rules in a model with TVP. I extend these papers by including endogenous-gain structures, and by investigating an alternative TVP process.

For the sake of simplicity, I use a Cobweb model with TVP, which I modify along two dimensions.[2] First, I compare a process similar to a random walk and an alternative TVP setting similar to Evans and Ramey (2006). Second, I examine the Nash equilibria in the presence of expectational feedback.

In addition to a constant and the Marcet and Nicolini *ad hoc* endogenous-gain, I propose an alternative endogenous-gain that includes the standard deviation of a potential estimate to derive the gain value. I use the mean squared forecast error as a benchmark.

Under occasional structural breaks I find that the Marcet and Nicolini gain rarely improves on a constant-gain. Statistically significant improvement is only a 0.5 percent improvement in mean squared forecast error over a constant-gain. My endogenous-gain has greater improvement, as much as 4 percent, and also is a Nash Equilibrium relative to a constant-gain. This 8 fold improvement over Marcet and Nicolini suggests that my endogenous-gain algorithm provides significant improvement. The endogenous-gain does not fair as well under a random walk type scenario.

---

[2]This model adopts the same framework as Bray and Savin (1986).

After describing the endogenous-gain processes used in the analysis, I demonstrate the stability of my endogenous-gain in a simple NK model. The third section presents the Cobweb model and results without expectational feedback. Section 4 examines the Nash equilibria of the model with expectational feedback. The fifth section concludes.

1. Endogenous-Gains

Econometric learning is a bounded rationality exercise that assumes that agents have some idea of what the economy looks like mathematically, but do not know the values of the parameters. Therefore, agents use recursively least squares (RLS) to update their coefficient estimates when they receive new information. The following is an example of a typical RLS algorithm:

$$\hat{\phi}_t = \hat{\phi}_{t-1} + \gamma R_t^{-1} X_t (Y_t - X_t' \hat{\phi}_{t-1}), \tag{1.1}$$

$$R_t = R_{t-1} + \gamma (X_t X_t' - R_{t-1}). \tag{1.2}$$

The first equation updates the coefficient estimate, $\hat{\phi}_t$, using the new data, Y_t, X_t, and the second equation updates the moment matrix, R_t. An important component in constructing the RLS algorithm is the gain parameter γ. This parameter governs the weight assigned to each observation. For example, the value of decreasing-gain decreases as each new observation is incorporated in the estimate. This means that recent data have little effect on the estimate value.

Another standard gain parameter choice, is a constant-gain. This type of gain gives each new observation the same weight. This means that the weight on each observation geometrically declines backwards in time. A constant-gain generally is used when one believes that the underlying parameter values vary over time.

Marcet and Nicolini's endogenous-gain combines these two concepts, allowing agents to switch between a decreasing-gain and a constant-gain. Agents take advantage of the lower volatility that a decreasing-gain provides when agents believe a parameter remains constant. Yet they realize that parameters might change by switching to a constant-gain when appropriate. While this makes sense in theory, the rule used to govern this choice is inherently backward looking.

Specifically agents will switch if an average of forecast errors over the last J periods are above some arbitrary value, v. If the recent forecast errors fall below that value, then agents decrease the gain gradually by keeping track of the number of periods, k, the recent forecast errors are below v.

$$\gamma_y = \begin{cases} \frac{1}{\bar{\gamma}_y^{-1}+k} & \text{if } \frac{\sum_{i=t-J}^{t} |y_i - y_i^e|}{J} < v, \\ \bar{\gamma}_y & \text{if } \frac{\sum_{i=t-J}^{t} |y_i - y_i^e|}{J} \geq v. \end{cases} \tag{1.3}$$

The rule Milani (2007b) uses is based on Marcet and Nicolini (2003), but sets the arbitrary value, v, equal to the historical average of absolute forecast errors. The historical average is updated as each period provides new information. The historical average can either use all information available or have window size larger than J.

These mechanisms rely on forecast errors of outcome variables when agents are concerned with coefficient movement. Therefore, I propose an alternative endogenous-gain. The motivation behind this gain is that agents should use coefficient estimates to determine whether there has been a change.

The alternative gain uses the standard deviation of a hypothetical estimate for the current parameter value using an average of the mean and variance of the past w periods. If the potential estimate lies several standard deviations away from the mean then agents should suspect a change in parameter values. Such a gain must increase with the standard deviation and have bounds of at least zero and at most one.

I propose using a function that also allows for modulation depending on the type of model an agent faces,

$$\gamma_b = \alpha_{lb} + \alpha_{sf} \frac{\left|\frac{\hat{b}_t - \bar{b}_t}{\bar{\sigma}_b}\right|}{1 + \left|\frac{\hat{b}_t - \bar{b}_t}{\bar{\sigma}_b}\right|}, \tag{1.4}$$

where $\hat{b}_t$ is the potential estimate, $\bar{b}_t$ is the average of the past w estimates, and $\bar{\sigma}_b$ is the average variance of the estimates.

The parameters, α_{lb} and α_{sf}, (that may be empirically testable) temper the potential values the endogenous-gain can take and adds to the generality of the model. The lower bound, α_{lb}, and the scaling factor, α_{sf}, define a range of gain values that the agent will use. Note that if agents set $\alpha_{sf} = 0$, this endogenous-gain becomes a constant-gain.

The following procedure updates agent's coefficient estimates. (1) use the value of the previous gain parameter to find the potential estimate value $(\hat{b}_t)$. (2) Then calculate, based on the potential estimate, the value of the current gain parameter. (3) Last, update the estimate using the gain parameter provided by step two.

There are several reasons why one might prefer this type of *ad hoc* endogenization of the gain parameter. First, it makes a clear distinction, statistically, why one might suspect that there is a structural break. Second, most macroeconomists would agree that agents do not only look backwards to form expectations. My endogenous-gain uses an initial estimate of the current parameter as the crucial piece of information for the decision mechanism. Lastly, this endogenous-gain nests the possibility of a constant-gain.

2. Stability of the Alternative

In the previous section I proposed an alternative endogenous-gain. In order to assess the stability properties I use the following NK model, presented in section 3 of Evans and Honkapohja (2009),[3]

$$x_t = x^e_{t+1} - \phi(i_t - \pi^e_{t+1}) + g_t, \tag{2.1}$$

$$\pi_t = \beta\pi^e_{t+1} + \lambda x_t + u_t, \tag{2.2}$$

$$i_t = \frac{\varphi\lambda}{\alpha_i}\pi^e_t + \frac{\varphi\alpha_x}{\alpha_i}x^e_t, \tag{2.3}$$

where u_t and g_t are AR(1) processes. The following equations govern these processes:

$$u_t = \rho u_{t-1} + \tilde{u}_t, \text{ and } g_t = \mu g_{t-1} + \tilde{g}_t,$$

where $\tilde{g}_t \sim iid(0, \sigma_g^2)$, $\tilde{u}_t \sim iid(0, \sigma_g^2)$, and $0 < |\mu|, |\rho| < 1$.

Note that if there are six coefficients there are six different gain processes. This means that the agents assess the potential for a structural break in each of the coefficients separately.

Since the endogenous-gain requires previous information, I begin all simulations with a 40 period burn-in. During the burn-in each equation receives an additional exogenous error each period for each equation. This allows for enough variability in the data to generate variance covariance matrices necessary for the construction of the endogenous-gain. After the burn-in, the additional exogenous variation shuts down and the simulation continues without any extraneous noise.

Similar to Evans and Honkapohja (2009) I use the calibrated parameter values from Table 6.1 of Woodford (2003), with $\alpha_x = 0.048$, $\varphi = 1/0.157$, $\lambda = 0.024$,

---

[3] See Woodford (2003) for derivation.

$\beta = 0.99$, and $\alpha_i = 0.077$. In Evans and Honkapohja (2009), they find that this particular parameterization results in instability if agents use a constant-gain greater than or equal to 0.024. In order to demonstrate the stability properties of my endogenous-gain I set $\alpha_{lb} = 0.005$ and $\alpha_{sf} = 0.035$.

Figure III.1 provides an example of the stability of the endogenous-gain. The temporary deviations from RE continue indefinitely. In addition, I find that two of the coefficients do not converge to their REE values. Since the temporary deviations from RE occur more frequently, the standard methodology for assessing stability does not provide useful results.

Therefore, in order to asses this type of stability I use the average deviations in the last 100 periods of the simulation to determine a model stability. I define a simulation as stable, if the average of the deviations in the last 100 periods are less than 0.01. I find stability in approximately 75 percent of 1000 simulations.

These results suggest that an endogenous-gain may provide several useful properties. This particular examples shows how an endogenous-gain can extend the range of gain values that result in stability, perhaps resulting in "robust" stability in the sense of Evans and Honkapohja (2009). In addition, the temporary deviations occur in a standard NK model, which suggests that recurrent irrational behavior may arise naturally with appropriate expectation formation modeling.

3. Evaluating the Endogenous-Gain

A Cobweb model serves as the basis for the analysis performed below. This type of model has been used from early on in the learning literature.[4] In order to examine the expectation formation in its purest setting, I first eliminate expectational feedback. I assume that agents know the structural process, except for

[4]See the pioneering work of Bray (1982), Bray and Savin(1986), and others.

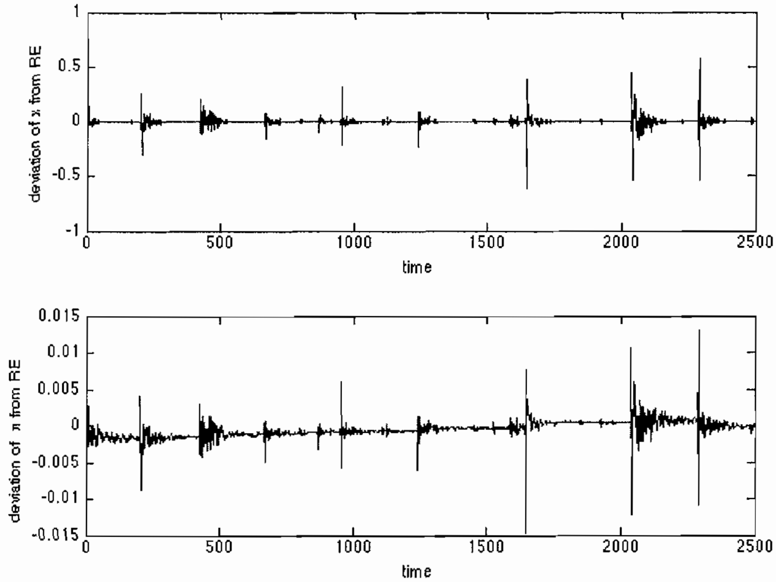

Figure III.1: Stability of optimal Taylor-type rule with alt. endogenous-gain.

the parameter values. Therefore, they estimate coefficients using RLS. For the analysis below I assume that agents have no knowledge of the process governing the TVP.

The most simple representation of this model would be one with exogenous data. In this case the data generating process is as follows:

$$y_t = \mu_t x_{t-1} + \phi y_t^e + \eta_t, \tag{3.1}$$

where y_t and x_t are data, μ_t is the TVP, ϕ is the coefficient on expectations, and η_t is an *iid*, mean zero, variance σ_η^2 white noise process. The superscript e represents the expected value of current data and is referred to as a "nowcast," *i.e.* $y_t^e = E_{t-1}y_t$. Nowcasting implies that people may not have access to current data. Since I first examine the case where there is no expectational feedback I set $\phi = 0$.

Setting $\phi = 0$ is desirable because it eliminates the possibility of game theoretic behavior. Note that if $\phi = 0$, this equation also represents the actual law of motion (ALM) since expectations are not included. The perceived law of motion (PLM) is very similar, reflecting agents basic knowledge of the system, and lack of knowledge of the parameter μ_t. Below, I discuss the relevance of the ALM and the PLM to E-stability.

The TVP process of the coefficient, μ_t, is based on Evans and Ramey (2006). The motivation for this particular process is that it suggests that parameter values are stable for the majority of time, but every once in a while they change.

$$\mu_t = \begin{cases} \mu_{t-1} & \text{with Prob. } (1\text{-}\varepsilon), \\ \upsilon_t & \text{with Prob. } \varepsilon. \end{cases} \tag{3.2}$$

where υ_t is an *iid*, mean zero, variance σ_υ white noise process, and ε is the probability of switching to a new parameter values. I choose this particular process because the initial intuition behind using a switching-gain such as Marcet and Nicolini was that agents believed there was potential for structural breaks.

While a parameter following a random walk changes value each period, the Evans and Ramey process exhibits properties akin to random structural breaks. Since the Bayesian filter associated with this process requires significant expertise, it stands to reason that agents would use an approximation in this case. Even the Kalman filter requires a fair amount of knowledge, thus an approximation of the Kalman filter when the parameter follows a random walk seems appropriate as well.

When the underlying parameter changes one would hope that the endogenous-gain places more weight on recent observations the switch to the new parameter value to occur more quickly. The Marcet and Nicolini variety only switches once the forecast errors were "bad enough," whereas my endogenous-gain reacts if statistically significant change in coefficient estimates occurs.

I consider a rigorous test of these gains. I generate rational expectations data, which then is used in algorithms for each type of gain. Each gain structure has a parameter or several parameters which can affect the forecast error. I optimize over these parameters to achieve the smallest mean squared forecast error (MSFE) for the given length of simulation.

For the standard constant-gain there is only one dimension with which to optimize over, namely, the value of the constant-gain itself, γ. Marcet and Nicolini requires three parameters, the value of the constant-gain $\bar{\gamma}_z$, the recent window length, J, and the arbitrary value, v. Finally, my endogenous-gain has three parameters, the lower bound α_{lb}, the scaling factor α_{sf}, and the window size w.

After optimization I conduct out-of-sample simulation using the optimized values for each gain. The optimization simulation lasts 50,000 periods. Using the optimized parameter values, I conduct 100 independent simulations of 40,000 periods. I drop the first 20,000 periods to eliminate any influence of initialization of the learning algorithm and any other initial conditions.

In performing this optimization routine on the Marcet and Nicolini gain, I found two surprising results. First, there were no cases where Marcet and Nicolini dominated a constant-gain, and second, the optimization routine performed inconsistently. In order to mitigate this I expanded my search of the parameter space and looked for optimal constant-gain values that could be improved by allowing for a switch.

For this process I first optimized for a constant-gain over 50,000 periods. Then conducted 100 independent simulations of 40,000 periods where agents follow the

Marcet and Nicolini rule for different values of the arbitrary cutoff. The results are provided in Table III.1.[5]

I find that over the space where Marcet and Nicolini theoretically should not perform well, it does not. Under conditions that one might expect Marcet and Nicolini to perform well, it does so only marginally. While the average MSFE of some of the simulations fall below one, only one is statistically different from one. In that one case two standard deviations away from the mean only results in a 0.7 percent improvement. The additional tables in the appendix show that other ratios of standard deviations do not show significant improvement.

Table III.2 displays the results for simulations with no expectational feedback, *i.e.* $\phi = 0$. The upper panel shows that the endogenous-gain has a lower MSFE when the ratio between the standard deviation is less than one half. The percentage improvement increases as the ratio gets smaller. At the same time the value of the optimal constant-gain gets larger. The estimates of the relative MSFE is two standard deviations away from one for all the chosen parameter settings, except when the ratio is two-thirds.

In the lower panel the structural breaks occur less frequently, and consequently there is less improvement on the constant-gain. It takes a much larger ratio (one-eighth) for the relative MSFE to be more than two standard deviations from one. The optimal constant-gain values are significantly higher when the structural breaks occur more frequently.

Finally, looking at the optimal endogenous-gain parameters a particular pattern appears. In each case the optimal constant-gain lies between the lower and upper bounds of the endogenous-gain. The upper bound of the endogenous-gain

[5]Additional tables for other parameter settings are included in the Appendix.

Table III.1: Forecast Ability of the Marcet and Nicolini Switching-Gain

v	$1-\varepsilon$					
	0.01	0.05	0.1	0.9	0.95	0.99
0.5	1.0000	1.0000	1.0000	0.9999	0.9999	0.9994
	4.71E-08	1.37E-07	4.13E-08	0.0004252	0.000177	0.0004043
1	1.0000	1.0000	1.0000	0.9997	0.9997	0.9987
	1.43E-07	9.91E-08	1.20E-07	0.001217	0.001272	0.000554
1.5	1.0000	1.0000	1.0000	1.0000	0.9998	0.9980
	2.24E-07	1.54E-07	1.68E-07	0.0006277	0.000665	0.001087
2	1.0000	1.0000	1.0000	1.0030	1.0010	0.9965
	2.37E-07	1.61E-07	2.53E-07	0.002593	0.002526	0.001466
2.5	1.0000	1.0000	1.0000	1.0070	1.0020	0.9967
	3.51E-07	4.32E-07	3.03E-07	0.002864	0.002661	0.002924
3	1.0000	1.0000	1.0000	1.0100	1.0050	0.9951
	4.67E-07	2.83E-07	3.57E-07	0.003248	0.003667	0.003743
3.5	1.0000	1.0000	1.0000	1.0150	1.0080	0.9948
	6.03E-07	4.20E-07	3.97E-07	0.003901	0.004053	0.003219
4	1.0000	1.0000	1.0000	1.0210	1.0120	0.9950
	1.06E-06	4.66E-07	2.83E-06	0.005149	0.005139	0.003146
10	1.0000	1.0000	1.0000	1.1910	1.2020	1.1060
	5.55E-06	3.08E-06	4.57E-06	0.01644	0.01195	0.01321

The first row indicates for each value of v indicates mean of the relative MSFE. The second row reports the standard deviation. v is the value of forecast error above which agents switch to a constant gain. ε is the probability of structural break in the underlying coefficient. The variance of the exogenous shock to the forecasted variable $\sigma_\nu = 4$ and the variance of the coefficient $\sigma_\eta = 2$.

increases as the ratio of the standard deviations decreases. In addition, as the frequency of structural breaks increases the optimal window length decreases.

i. An Alternative TVP Process

While some empirical work relies on structural breaks, it is also common to assume that parameters follow a random walk. This assumption works in practice because the data sets tend to be fairly short and/or the standard deviation is restricted to be quite small. The intuition behind assuming a random walk is that a coefficient drifts around, but in all likelihood remains bounded given the small data set and restrictions on the standard deviation.

Table III.2: Forecast Ability of the Endogenous-Gain in a Model with Occasional Structural Breaks

			$\varepsilon = 0.01$			
σ_η		$\sigma_\nu = 4$			$\sigma_\nu = 3$	
2	0.9878	(0.0048)	0.1981	1	(0)	0.1536
	0.1149	0.3749	19	0.1536	0	20
1	0.9723	(0.0119)	0.3411	0.9793	(0.0087)	0.2756
	0.1131	0.5350	19	0.1088	0.4440	19
0.5	0.9569	(0.0180)	0.5243	0.9685	(0.0134)	0.4457
	0.1159	0.6524	18	0.0976	0.5578	20
			$\varepsilon = 0.05$			
σ_η		$\sigma_\nu = 4$			$\sigma_\nu = 3$	
2	0.9971	(0.0027)	0.3707	0.9995	(0.0014)	0.2920
	0.2881	0.2856	11	0.2500	0.1704	11
1	0.9880	(0.0066)	0.5801	0.9927	(0.0046)	0.4939
	0.3224	0.5043	12	0.3027	0.4081	11
0.5	0.9778	(0.0107)	0.7632	0.9831	(0.0099)	0.6922
	0.3666	0.5559	11	0.3017	0.5703	11

The first row in each box displays the relative MSFE, the standard deviation of the relative MSFE, and the optimal constant-gain value. The second row displays the optimal lower bound, scaling factor and window size of the endogenous-gain. v is the value of forecast error above which agents switch to a constant gain. ε is the probability of structural break in the underlying coefficient. σ_ν and σ_v are the variances of the exogenous shock to the forecasted variable and the coefficient, respectively.

Since a random walk by definition is not bounded, and since I perform long simulations, I assume that μ_t is an AR(1) with a normally distributed error term.

$$\mu_t = \lambda \mu_{t-1} + \omega_t, \tag{3.3}$$

where ω_t is an *iid*, mean zero, variance σ_ω white noise process and $0 < \lambda < 1$. I will assign a large value to λ so as to come close to a random walk.

I simulate the model under four different parameter settings. The results can be found in Table III.3. The simulations show that the endogenous-gain does not improve over the optimal constant-gain in any economically meaningful simulations.

Table III.3: Forecast Ability of the Endogenous-Gain with a Random Walk Time-Varying Coefficient

σ_η		$\sigma_\nu = 1$			$\sigma_\nu = 0.5$	
1	0.9958	(0.0019)	0.5447	0.9992	(0.0006)	0.3502
	0.4334	0.2252	17	0.3109	0.0938	14
0.5	0.9899	(0.0046)	0.7308	0.9961	(0.0020)	0.5447
	0.5039	0.3361	17	0.4007	0.2297	18

The first row in each box displays the relative MSFE, the standard deviation of the relative MSFE, and the optimal constant-gain value. The second row displays the optimal lower bound, scaling factor and window size of the endogenous-gain. σ_ν and σ_v are the variance of the exogenous shock to the forecasted variable and the innovations to the time-varying coefficient, respectively.

The specific values chosen highlight that the relative standard errors provide the appropriate comparison and show the general pattern under these conditions.

When the endogenous-gain does significantly improve on the constant-gain, the parameter values imply agents use too little data. Changing the ratio of standard deviations to reduce the value of the constant-gain causes the improvement of the endogenous-gain to vanish.

This result should not be surprising since the motivation for the endogenous-gain was occasional structural breaks. This experiment, shows that the endogenous-gain is tailored toward structural breaks and does not necessarily improve on all time varying parameter possibilities.

4. When Expectations Matter

An obvious criticism of the models described above, is that agents form expectations but do not use them and therefore are not included in the data generating process. Most learning models have some sort of expectational feedback, which leads me to generalize the basic model to include expectations. I do this by allowing ϕ to take values other than zero.

Restating the Cobweb model with purely exogenous data from above,

$$y_t = \mu_t x_{t-1} + \phi y_t^e + \eta_t, \tag{4.1}$$

where $\phi \neq 0$. With this formulation it is important to assess E-stability, so that parameter values are chosen appropriately. In general, E-stability imposes restrictions on some, but not all the parameters of the model.

Assuming that agents observe the parameter values at all points in time, standard rational expectations yield the following result. This is not necessary for determining E-stability, but it can provide guidance.

$$y_t^{RE} = \frac{\mu_t}{1-\phi} x_{t-1} + \frac{1}{1-\phi}\eta_t. \tag{4.2}$$

In order to assess E-stability, one starts with an assumption regarding agents' beliefs over the data generating process. Following the literature, I assume that agents use the minimum state variable (MSV) solution in determining their PLM,

$$y_t = b_t x_{t-1} + e_t. \tag{4.3}$$

With this assumption agents expectations are $y_t^e = b_t x_{t-1}$, which yields the following ALM,

$$y_t = (\mu_t + \phi b_t) x_{t-1} + \eta_t. \tag{4.4}$$

Clearly the ALM is quite different than the PLM, however, if agents are able to learn the REE then the coefficients on x_{t-1} must be equal to each other. Thus, there should be a correspondence between the PLM and the ALM, which is referred to as a T-mapping. Given this particular ALM and PLM the T-map is,

$$T(b_t) = (\mu_t + \phi b_t). \tag{4.5}$$

For a univariate model assessing E-stability is straightforward. E-stability requires that the derivative of $T(b_t) - b_t$ be less than zero. Thus, the following condition must hold for stability: $\phi < 1$. The intuition behind this result is that in order to learn the feedback of expectations must be self reinforcing.

Table III.4 displays the results of the simulations with expectational feedback. Across the board I find that my endogenous-gain performs better with expectational feedback. The same general patterns appear as without expectational feedback, which suggests that smaller ratios of standard deviations maintain corresponding improvements.

One cause for concern is the relatively high values of the optimal constant-gain. Even the smallest optimal constant-gain value implies that agents use three periods of data in forming their estimates. Most empirical work suggests that agents have smaller gains, or use more data. When agents use the endogenous-gain, the lower bound appears more plausible.

i. Nash Equilibrium

Up till this point have assumed that agent's use the social optimal gain. However, in the presence of expectational feedback a single agent may do better, relative to the rest of the agents, using an different gain value. Finding the Nash equilibrium for a constant-gain is straightforward since there agents only control one parameter. Under the endogenous-gain agents have two parameters to choose which makes the process more difficult.

Table III.4: Forecast Ability of the Endogenous-Gain in a Occasional Structural Break Model with Expectational Feedback

			$\varepsilon = 0.01$			
σ_η		$\sigma_\nu = 4$			$\sigma_\nu = 3$	
2	0.9805	(0.0057)	0.3721	0.9881	(0.0040)	0.2972
	0.1797	0.5685	18	0.1574	0.4756	15
1	0.9564	(0.0125)	0.5733	0.9701	(0.0099)	0.4894
	0.1474	0.7672	17	0.1357	0.6847	19
0.5	0.9315	(0.0159)	0.7551	0.9418	(0.0139)	0.6834
	0.1288	0.8286	19	0.1297	0.7975	20
			$\varepsilon = 0.05$			
σ_η		$\sigma_\nu = 4$			$\sigma_\nu = 3$	
2	0.9885	(0.0046)	0.6014	0.9943	(0.0030)	0.5065
	0.3892	0.5163	9	0.3482	0.4258	12
1	0.9709	(0.0072)	0.7927	0.9790	(0.0060)	0.7208
	0.4188	0.5610	14	0.3878	0.5551	11
0.5	0.9619	(0.0090)	0.9126	0.9645	(0.0079)	0.8715
	0.5414	0.4457	14	0.4620	0.5179	14

The first row in each box displays the relative MSFE, the standard deviation of the relative MSFE, and the optimal constant-gain value. The second row displays the optimal lower bound, scaling factor and window size of the endogenous-gain. v is the value of forecast error above which agents switch to a constant gain. ε is the probability of structural break in the underlying coefficient. σ_ν and σ_v are the variances of the exogenous shock to the forecasted variable and the coefficient, respectively. I set the coefficient on the expectational term $\phi = 0.5$

I focus on the economically relevant cases where the ratio between standard deviations are one-half and two-thirds. I find that the Nash equilibrium of the constant-gain almost matches the social optimal. Figures III.2 and III.3 displays the fixed point of the equilibrium constant-gain value and the MSFE of the best responses.

The exact intersection when $\sigma_\nu = 4$, $\sigma_\eta = 2$ occurs at 0.3733 and at that point the optimal response of the endogenous-gain results in a 1.5 percent improvement in MSFE. For $\sigma_\nu = 3$, $\sigma_\eta = 2$, the intersection occurs at, 0.2977 and at the point the optimal response of the endogenous-gain results in a 1 percent improvement in MSFE.

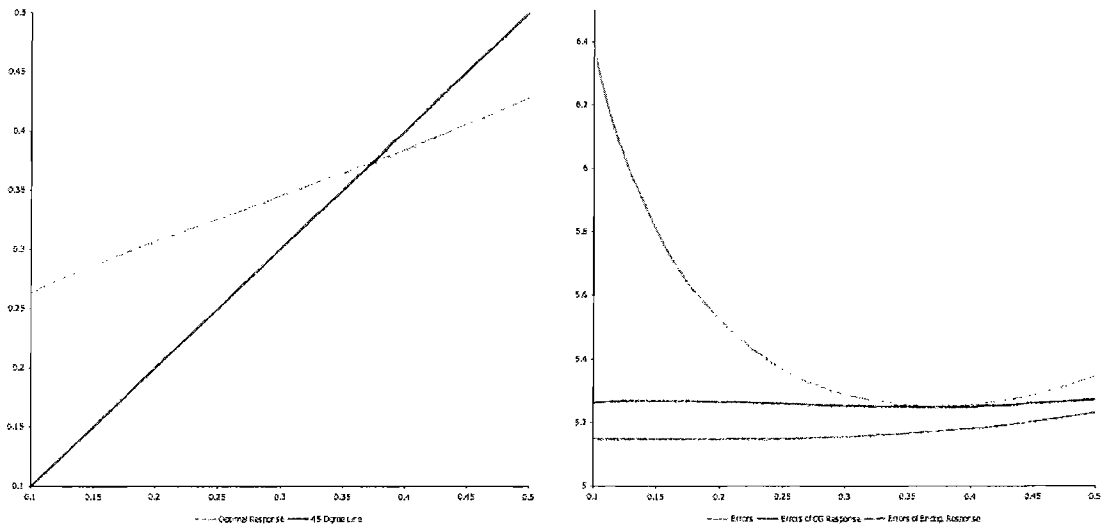

Figure III.2: The Nash equilibrium constant-gain value and MSFE of best responses. $\sigma_\nu = 4$, $\sigma_\eta = 2$, and $\varepsilon = 0.01$.

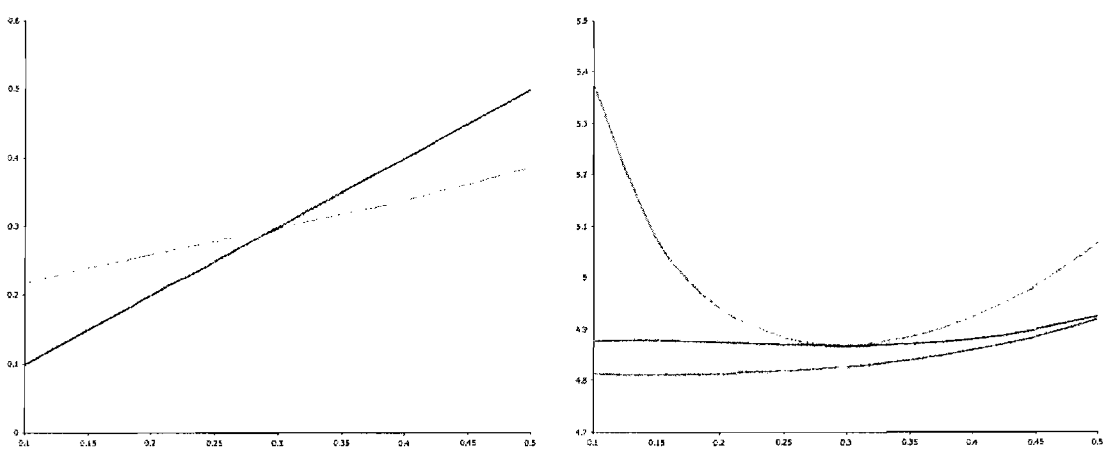

Figure III.3: The Nash equilibrium constant-gain value and MSFE of best responses. $\sigma_\nu = 3$, $\sigma_\eta = 2$, and $\varepsilon = 0.01$.

In addition if I generate data with agents using the endogenous-gain response to the constant-gain, then a single agent using a constant-gain cannot improve. That is, given what everyone else is doing the endogenous-gain is preferred.

These results illustrate that the socially optimal constant-gain value and the Nash constant-gain are identical in this particular model. Since this result does not hold generally, the movement due to the structural break process must dominate the learning dynamics. Game theoretic behavior serves no purpose because the agents' concern themselves with tracking the structural break process.

5. Conclusion

Given the explosion of empirical research involving assumptions of learning behavior by agents, an in depth comparison of different gain structures seems appropriate. While a constant-gain is frequently used when the underlying parameters vary over time, alternatives have been postulate to more accurately describe agents' behavior. The switching-gain of Marcet and Nicolini intuitively seems designed for potential structural breaks. I find that the switching-gain rarely offers significant improvement over a constant-gain in the presence of structural breaks.

As an alternative I have proposed an endogenous-gain that relies on statistical information to determine the value of the gain. The stability properties of my endogenous-gain closely relate to the dynamics found in Marcet and Nicolini (2003) and warrant further research. I find that my endogenous-gain performs well under occasional structural breaks, reducing the MSFE by at most 4 percent. In addition, when agent's expectations feedback into outcome variables, my endogenous-gain dominates the Nash equilibrium constant-gain.

CHAPTER IV

DSGE ESTIMATION WITH LEARNING

Empirical macroeconomists spend a lot of time trying to explain, and in some cases predict, output, inflation and interest-rates. Most estimations have used reduced form time-series models. Bayesian techniques have allowed econometricians to estimate structural parameters of dynamic stochastic general equilibrium (DSGE) models, such as a simple NK model.[1] More recently, scholars have relaxed the assumption of RE, by allowing agents to "learn" about the economy.[2]

A natural consequence of a learning model is that the reduced form coefficients vary over time. This allows for a better fit of the data, which is why learning models tend to dominate RE. In this chapter, I present my preferred NK model, discuss the estimation procedure, and present some comparative results. These results are based on data from 1989-2007, which one might suspect has no clear structural breaks. I also perform a Monte Carlo exercise that shows that the learning estimation does not capture an RE equilibrium.

Milani (2007a) provides one of the first DSGE estimations with learning.[3] The model I present more closely resembles Milani(2007b), but differs in several

---

[1]See, for example, An and Schorfheide (2007), Fernandez-Villaverde and Rubio-Rameriz (2007), and Justiniano and Primiceri (2008).

[2]See, for example, Milani (2007a,2007b, 2008), Murray (2007a, 2007b, 2008), and Solobodyan and Wouters (2008).

[3]There have been other attempts to empirically test for learning behavior using other identification techniques. See, for example, Branch and Evans (2006), and Chevillon *et al.* (2010).

dimmensions. First, I assume that monetary policymaker use "nowcasts" to inform their decision over policy, whereas in Milani policymakers use lagged output and inflation. Second, Milani's agents fail to account for autocorrelation. I also assume that agents perceive past structural shocks.

Many empirical projects , most notably Fuhrer and Moore (1995), have shown that there is persistence in interest/inflation rates. Thus many researchers, including Milani, include a lagged interest-rate term in the monetary policy rule. However, Cogley and Sbordone (2008), have shown that time varying trend inflation can provide acceptable persistence. Therefore, instead of using lagged interest-rates I assume that there is trend component to interest-rates that follows a random walk.

My estimation strategy innovates on another dimension since most papers with forward looking RE models do not solve for the RE solution to estimate the coefficients. Instead researchers have incorporated the RE forecast errors as part of the error terms.[4] Actual data is used as an estimate of the RE value. Applying the same technique to learning would marginalize the contribution of incorrect expectation formation.

My results suggest that my endogenous-gain learning rule does the best job of describing the data even though there is no apparent time-variation of the reduced form coefficients. Using three model comparison strategies constant-gain learning clearly outperforms RE, and endogenous-gain learning outperforms both. Examination of the actual time path of the reduced form coefficients shows little time variation, which suggests that although learning allows for time variation in coefficients this is not what causes the improvement.

I dive straight into the model in the next section. I explain in further detail the Bayesian estimation strategy employed in the second section. The third section

[4]See, for example, Clarida *et al.* (2000) and Kim and Nelson (2006).

presents the results. The fourth examines ALM implied by agents learning to a simple TVP estimation of a VAR. Section five concludes.

1. Incorporating Trend Interest-Rates and Learning

As noted earlier, I depart from Milani (2007b) on several dimensions. Milani assumes that the monetary policy rule is backward looking. I favor contemporaneous expectations model of Evans and Honkapohja (2009) since I have shown the stability properties of this model.

Thus the economy is described by the following NK model[5]

$$x_t = x^e_{t+1} - \phi(i_t - \pi^e_{t+1}) + g_t, \tag{1.1}$$

$$\pi_t = \beta\pi^e_{t+1} + \lambda x_t + u_t, \tag{1.2}$$

$$i_t = \iota_t + \theta_\pi \pi^e_t + \theta_x x^e_t + \varepsilon_t, \tag{1.3}$$

where u_t and g_t are AR(1) processes and ι_t is a time varying trend that follows a random walk.[6] This addition follows the same line of reasoning of Cogley and Sbordone (2008) and results in a model that does not involve lagged endogenous variables. The following equations govern these processes, $u_t = \rho_u u_{t-1} + \nu_{u,t}$, and $g_t = \rho_g g_{t-1} + \nu_{g,t}$.

Having a model without lagged endogenous variables allows for quick calculation of the REE. Thus instead of pushing the RE into the error term, I can solve the model at each step. Assuming agents observe lagged shocks is critical. I

---

[5] See Woodford (2003) for derivation.

[6] The AR(1) assumption could be expanded to an AR(2) for future work. The decision whether agents then know the lag structure of the exogenous shocks becomes important. Misspecification of agents may not just lie in what data agents use, but also in their beliefs over autocorrelation.

propose two different information sets for agents to use, one which favors the rational expectations agents and one that favors learning agents.

In this section I demonstrate the formulation of the model when agents do not have access to the time varying trend. This specification benefits learning, since I assume the learning agents estimate an intercept. This allows agents to capture some of the trend.

I assume that agents perceive the autocorrelation and that they have some way of estimating those coefficients of the autocorrelation and distinguishing past shocks. Thus, they know the following equation and its coefficients:

$$\upsilon_t = F\upsilon_{t-1} + \tilde{\upsilon}_t, \tag{1.4}$$

where $\upsilon_t = (g_t, u_t, \varepsilon_t)'$, $\tilde{\upsilon}_t$ are the respective *iid* errors, and

$$F = \begin{pmatrix} \rho_g & 0 & 0 \\ 0 & \rho_u & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

I assume that agents see lagged exogenous shocks and that agents estimate an intercept term. Thus, the PLM takes the following form,

$$Z_t = a_t + c_t\upsilon_{t-1} + \tilde{\upsilon}_t, \tag{1.5}$$

where $Z_t = (x_t, \pi_t, i_t)'$, and a_t and c_t are coefficient vector and matrix of appropriate dimensions, respectively. The agents learn the model coefficients according to the following RLS formulae:

$$\hat{\psi}_t = \hat{\psi}_{t-1} + \gamma_{t,y} R_t^{-1} X_t (Z_t - X_t' \hat{\psi}_{t-1}) \tag{1.6}$$

$$R_t = R_{t-1} + \gamma_{t,y}(X_t X_t' - R_{t-1}) \tag{1.7}$$

where $\hat{\psi}_t = (a_{1t}, c_{11t}, c_{12t}, c_{13t}, a_{2t}, c_{22t}, c_{22t}, c_{23t}, a_{3t}, c_{33t}, c_{32t}, c_{33t})'$ is a vector of the estimated coefficients, $X_t = I_3 \otimes (1, v_{t-1}')'$ is a matrix of the stacked regressors, and $\gamma_{t,y}$ is a matrix with the gain parameters on the diagonal. Using the PLM (1.7) and the RLS equations, (1.6) and (1.7), we find the agents expectations:

$$\hat{E}_{t-1} Z_t = a_{t-1} + c_{t-1} v_{t-1}, \tag{1.8}$$

$$\hat{E}_{t-1} Z_{t+1} = a_{t-1} + c_{t-1} F v_{t-1}, \tag{1.9}$$

where I_3 indicates a 3x3 identity matrix. Rewriting equations (1.2) to (4.1) in matrix form:

$$AZ_t = Trnd + B\hat{E}_{t-1} Z_t + C\hat{E}_{t-1} Z_{t+1} + v_t \tag{1.10}$$

where

$$A = \begin{pmatrix} 1 & 0 & \phi \\ -\lambda & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \theta_x & \theta_\pi & 0 \end{pmatrix} \quad C = \begin{pmatrix} 1 & \phi & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad Trnd = \begin{pmatrix} 0 \\ 0 \\ \iota_t \end{pmatrix}$$

Substitution of expectations equations (1.8) and (1.9) into (1.1) and (1.2) results in the Actual Law of Motion (ALM):

$$Z_t = A^{-1} Trnd + A^{-1}(B + C) a_{t-1} + A^{-1}(Bc_{t-1} + Cc_{t-1} F) v_{t-1} + A^{-1} v_t. \tag{1.11}$$

When considering RE with agents not having the trend interest-rate information I assume that agents make an error. Since the agents do not have access to the trend interest-rate I calculate the RE coefficients assuming the trend does not exist and then reincorporate it in the actual law of motion. I refer to this as Pseudo-RE, and it can be written as,

$$Z_t = A^{-1}Trnd + \bar{c}v_{t-1} + A^{-1}v_t, \tag{1.12}$$

where $\bar{c}$ is the MSV RE solution assuming agents do not see the time varying trend. $\bar{c}$ results from the vector, $(I_9 - (I_3 \otimes M_0 + F' \otimes M_1))^{-1}(F' \otimes I_3)vec(P)$.

i. Gain Structures

In my empirical analysis I differentiate the model on one dimension, namely the expectation formation process. I compare rational expectations to two learning processes, a single constant-gain, and my alternative gain. The implementation of a single constant-gain is straight forward. All that needs to be done is set $\gamma_{t,y} = \gamma$ in (1.6) and (1.7).

For my alternative endogenous-gain recall the formula,

$$\gamma_{b,t} = \alpha_{lb} + \alpha_{sf}\frac{\left|\frac{\hat{b}_t - \bar{b}_t}{\bar{\sigma}_b}\right|}{1 + \left|\frac{\hat{b}_t - \bar{b}_t}{\bar{\sigma}_b}\right|}, \tag{1.13}$$

where b refers to a particular coefficient. There are three parameters that can be estimated, α_{lb}, α_{sf}, and w, but for simplicity I assume the $w = 10$ and only estimate the other two.[7] While this setup is certainly less restrictive than the constant-gains

[7]This choice reflects the results from Chapter III, which suggest that for the different data generating processes the average optimal value of w is around 10 when changes in structure occur every 5 years.

it still forces the lower and upper bounds to be the same across all parameters. Relaxing this assumption would nearly triple the number of estimated parameters.

While learning may appeal to some because it does not require as much knowledge, on the part of the agents, as RE, learning also provides an additional advantage. The ALM provides the DSGE structure of the estimation and it contains agents coefficient estimates, which change over time. This means that constant-gain learning incorporates time variation in the reduced form of the model.

In addition, Branch and Evans (2007) shows that learning can cause endogenous volatility. Milani (2007b) shows that endogenous learning can help explain some of the changes in volatility of the macroeconomy. Thus, my endogenous-gain incorporates time variation in both coefficients and volatility while only adding two additional parameters.[8]

2. Bayesian Estimation

In recent years there has been an abundance of papers that use Bayesian methods to estimate DSGE models. An and Schorfheide (2007) provide general guidelines for Bayesian estimation of DSGE models. Milani (2007a and 2007b) uses this technique to estimate a model with learning similar to the one above. In contrast, Murray (2008a, 2008b, and 2008c) all rely on a maximum likelihood approach. I favor the Bayesian approach, because it provides clearer model comparison. Specifically, I use a single block Random Walk Metropolis-Hastings (RW-MH) algorithm to sample the posterior distribution of model parameters.

[8]The time variation allowed is heavily restricted to the functional form of my endogenous-gain and the learning algorithm. This restriction might improve forecasting ability.

In order to conduct Bayesian estimation I rewrite the economy in state space form:

$$\xi_t = D_t + F_t\xi_{t-1} + G\omega_t \tag{2.1}$$

$$Y_t = H\xi_t \tag{2.2}$$

where $\xi_t = (Z_t, u_t, g_t, e_t, \iota_t)'$, $\omega_t \sim N(0, Q)$, and D_t, F_t, G and H are the appropriate matrices. Under RE F_t remains constant, since the deep parameters are constant, and $D_t = 0$, since the RE solution has no intercept. Under learning, D_t contains $A^{-1}(B + C)a_{t-1}$ and zeros, and F_t contains $A^{-1}(Bc_{t-1} + Cc_{t-1}F)$, and updates the unobservables.

Once in state space form, the procedure is straightforward. Let the vector Ω contain the structural parameters of the model,

$$\Omega = \{\beta, \phi, \lambda, \theta_\pi, \theta_x, \mu, \rho, \sigma_g, \sigma_u, \sigma_e, \sigma_\iota, g\}. \tag{2.3}$$

To form the posterior requires evaluating the likelihood function, $L(\bullet)$, at the candidate parameter vector. The Kalman filter combined with the state space described above produces the likelihood value. Multiplying the likelihood by the priors, $p(\bullet)$ detailed below, results in the posterior distribution.

I use the Metropolis-Hastings algorithm to generate 1,250,000 draws from the posterior distribution. The first 250,000 are discarded as burn-in values. The Metropolis-Hastings algorithm relies on a high volume of draws from a candidate distribution. These draws are accepted or rejected based on the ratio of the posterior of the candidate to the previous draw.

Suppose that the previous draw is defined as Ω, and Ω^* is defined as the candidate. A standard candidate distribution is a random walk through the parameter space,

$$\Omega^* = \Omega + c\Sigma, \tag{2.4}$$

where c is a scaling term, and Σ is a covariance matrix. For certain algorithms $\Sigma = I$ for simplicity. I opt for simplicity, but modify some of the diagonal elements to match the scale of the parameters.

In order to determine the acceptance probability, α, for each draw I use the following equation,

$$\alpha = \min\left\{\frac{p(\Omega^*)L(\Omega^*)}{p(\Omega)L(\Omega)}, 1\right\}. \tag{2.5}$$

Thus if a candidate draw has a higher posterior value, then the previous draw the algorithm accepts the candidate with probability 1. If the candidate has a much lower posterior value, then the probability of acceptance is low. This ensures that the algorithm ranges over some of the unlikely parameter values while fully exploring the peak of the posterior.

Averaging the acceptance probabilities of each draw over all the draws results in the acceptance rate. Geweke (1999) suggests calibrating the candidate distribution to achieve acceptance rates between 25-40 percent.

i. Priors

Table IV.1 reports the prior distributions for each of the structural parameters in the model. I use the analysis above to form the prior over the constant-gain. Similar to Milani (2007b), I impose a dogmatic prior on β, namely, I set β equal to 0.99. Gamma distributions form the priors for all the standard deviations and the slope of the Phillips Curve. The monetary policy parameters have a normal prior, and the correlation coefficients of the autocorrelated errors have uniform prior.

Table IV.1: Prior Distributions

Description	Param	Distr.	Stats.	Mean
Discount Rate	β	–	–	0.99
Elas. of Subs.	ϕ	IG	1.5,1	1.5
Slope of PC	λ	IG	0.25,1	0.25
Feedback to Infl.	θ_π	N	1.5,0.5	1.5
Feedback to Output	θ_x	N	0.5,0.25	0.5
Corr. of g_t	μ	U	0,0.97	0.485
Corr. of u_t	ρ	U	0,0.97	0.485
Std. g_t	σ_g	IG	0.5,3	0.5
Std. u_t	σ_u	IG	0.5,3	0.5
Std. e_t	σ_e	IG	0.5,3	0.5
Std. ι_t	σ_ι	IG	0.5,3	0.5
Gain Params	γ	U	0,0.4	0.2
Alt. Params	α_{lb},α_{sf}	U	0, 0.8	0.4

Note: IG stands for Inverse-Gamma with scale and shape values given, N stands for Normal with mean and variance values given, and U stands for Uniform with upper and lower bound values given.

ii. Data

The data come directly from the Federal Reserve Bank of St. Louis economic database, FRED, and the Congressional Budget Office, CBO. The quarterly data begin in 1984:III and ends in 2007:IV. I use the first twenty periods to initialize the learning algorithm, thus the 1989:III-2007IV sample is used for the actual estimation. Previous literature has shown that structural changes occurred prior to 1984, which would lead to further complexity of the model and estimation technique.[9]

I define inflation as the annualized quarterly rate of change of the GDP deflator. The output gap is the log difference between GDP and potential GDP (as defined by the CBO). And finally, for the interest-rate I use the federal funds rate.

[9] For example see Fernndez-Villaverde and Rubio-Ramirez. (2007), Fernndez-Villaverde *et al.* (2009), Justiniano and Primiceri (2008), and Milani (2007).

Table IV.2: Estimation Results of a Model with Limited Information

Description	Param	Pseudo-RE	Constant-Gain	Endog-Gain
Elas. of Subs.	ϕ	0.674	0.0344	0.0987
Slope of PC	λ	0.200	0.0986	0.238
Feedback to Infl.	θ_π	1.134	1.0897	0.499
Feedback to Output	θ_x	0.0172	0.299	1.198
Corr. of g_t	μ	0.945	0.789	0.469
Corr. of u_t	ρ	0.954	0.722	0.823
Std. g_t	σ_g	0.0932	0.968	0.843
Std. u_t	σ_u	0.243	0.651	0.628
Std. e_t	σ_e	0.0985	0.133	0.1289
Std. ι_t	σ_ι	0.856	0.671	0.618
Gain	γ	–	0.000698	–
Lower Bound	α_{lb}	–	–	0.000349
Scale Factor	α_{sf}	–	–	0.0496
ML(CJ)		-332.2	-303.9	-262.7
ML(Harmonic)		-300.7	-283.0	-264.2
BIC		774.2	739.4	667.3

Results presented here represent the mean of the posterior distribution of each parameter. I calibrate the values of c to ensure acceptance rates between 25-40%. All of these estimations have the same starting values, except for the learning parameters. ML(CJ) is the marginal likelihood value calculated according to Chib and Jeliaskov (2001). ML(Harmonic) is the marginal likelihood using the modified harmonic mean, as suggested by Adolfson *et al.* (2007). BIC is calculated at the median value of the posterior distributions.

3. Results

Table IV.2 displays the results of estimations that differ in the assumptions over expectations. I use three different criteria for model selection. Bayesian model comparison relies on obtaining the marginal likelihood. Chib and Jeliazkov (2001) provide one method of approximation that relies on the same Markov Chain Monte Carlo methodology used in sampling from the posterior. However, Adolfson *et al.* (2007) suggest that in conjunction with RW-MH a modified harmonic mean reaches the approximate marginal likelihood value more quickly. In addition, I use the

Bayesian Information Criterion (BIC).[10] In each case numbers closer to zero indicate a better model.

Looking at the estimates of the RE model the results seem consistent with other literature. The data suggest that the model does not obey the Taylor rule, but over this time period Fernndez-Villaverde *et al.* (2009) find similar results. However, under learning these estimates get even smaller.

According to each of model comparison methods the data clearly favors the learning models. Of the learning models the endogenous-gain version still provides significant improvement even though only one additional parameter is estimated. This comparison does not do the RE model justice, since under the learning assumption the ALM coefficients vary over time.

A simple reduced form TVP would increase the number of parameters estimated by five, which makes it unlikely to outperform a learning model. A preferred model would allow the deep parameters to vary over time. Unfortunately, allowing the combination of agents using the shocks as data and time varying deep parameters seriously complicates the estimation. While the non-linearity in the reduced form could be managed with a block sampling method, an estimation of the structural equations would require non-linear techniques suggested by Fernndez-Villaverde, J. and J. Rubio-Ramirez (2008). Therefore, I save this comparison for future work.

While the results from the RE model are consistent with previous research, the data seem to favor the learning models. The estimation results from the learning models, however, run counter to past empirical work.

Figure IV.1 displays the posterior distributions of each of the endogenous-gain parameters. The data is clearly informative as the posterior distribution appear

---

[10]In calculate the BIC I use the likelihood calculated at the median values of the posterior distribution.

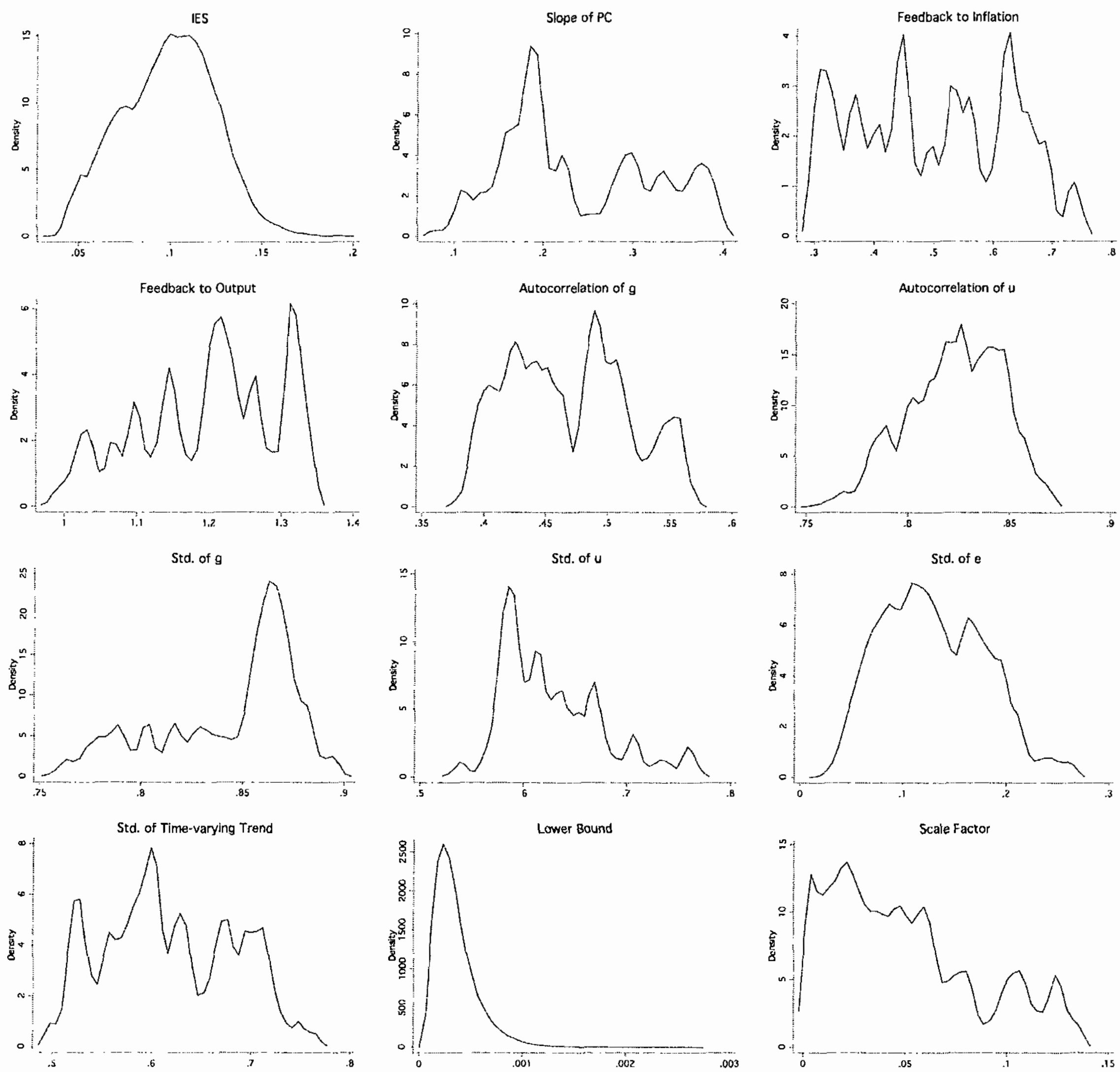

Figure IV.1: Posterior Distributions of the Endogenous-Gain Estimation Parameters quite different than the prior distributions. Some were so different that I chose not to include representations of the prior distributions in these graphs.

Clearly the inter-temporal elasticity of substitution, the lower bound, and the exogenous structural shocks clearly favor certain values of the distribution. The other parameters have narrowed the prior distributions to a range of the parameter space, but the data does not speak clearly for a particular value. The inter-temporal elasticity of substitution is centered on 0.1, which is quite different than the prior. The posterior of the lower bound parameter, which governs the lowest value that the endogenous-gain can take, places a lot of weight near zero, which matches the constant-gain estimation.

i. Expanding the Information Set

As noted earlier, the model above favors learning. By including information about trend inflation in agents data set swings the favor toward rational expectations. The estimation strategy remains the same, but the underlying matrices change.

I assume that agents observe the random walk time varying trend in much the same way they observe the autocorrelated errors. This implies that agents know the following,

$$v_t = Fv_{t-1} + \tilde{v}_t, \tag{3.1}$$

where $v_t = (g_t, u_t, \varepsilon_t, \iota_t)'$, $\tilde{v}_t$ are the respective *iid* errors, and

$$F = \begin{pmatrix} \rho_g & 0 & 0 & 0 \\ 0 & \rho_u & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Rewriting the NK model in a convenient form,

$$Z_t = M_0 \hat{E}_{t-1} Z_t + M_1 \hat{E}_{t-1} Z_{t+1} + P v_t, \tag{3.2}$$

where $M_0 = A^{-1}B$, $M_1 = A^{-1}C$ and

$$P = A^{-1} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

This change in formulation implies that the coefficient matrix in the MSV solution is now a (3x4), the intercept terms, if any, remain the same.

$$Z_t = (M_0 + M_1)a_{t-1} + (M_0 c_{t-1} + M_1 c_{t-1} F + PF)v_{t-1} + P\tilde{v}_t. \tag{3.3}$$

In contrast, the Law of Motion under rational expectations is,

$$Z_t = \bar{c} v_{t-1} + P\tilde{v}_t, \tag{3.4}$$

where $\bar{c}$ results from the vector, $(I_{12} - (I_4 \otimes M_0 + F' \otimes M_1))^{-1}(F' \otimes I_3)vec(P)$.

Table IV.3 presents similar results to when agents had less information. Both learning models outperform the rational expectations model, and the endogenous-gain model improves upon the constant-gain specification. In this case, the improvement on rational expectations appears to be much greater.

The information used clearly has an effect on policy parameters. Assuming RE, monetary policy followed the Taylor rule without the time varying trend information, but did not with the information. Constant-gain learning obeyed the

Table IV.3: Estimation Results of a Model with Trend Interest-Rate Information

Description	Param	RE	Constant-Gain	Endog-Gain
Elas. of Subs.	ϕ	0.456	1.0516	0.942
Slope of PC	λ	0.000144	0.0724	0.133
Feedback to Infl.	θ_π	0.519	1.475	1.449
Feedback to Output	θ_x	0.612	0.482	0.465
Corr. of g_t	μ	0.956	0.964	0.961
Corr. of u_t	ρ	0.782	0.854	0.886
Std. g_t	σ_g	0.454	0.616	0.558
Std. u_t	σ_u	0.262	0.568	0.588
Std. e_t	σ_e	0.787	0.545	0.512
Std. ι_t	σ_ι	0.631	0.527	0.521
Gain	γ	–	0.0000398	–
Lower Bound	α_{lb}	–	–	0.000177
Scale Factor	α_{sf}	–	–	0.0974
ML(CJ)		-350.4	-267.8	-256.2
ML(Harmonic)		-343.5	-342.9	-315.6
BIC		827.5	701.7	674.0

Results presented here represent the mean of the posterior distribution of each parameter. I calibrate the values of c to ensure acceptance rates between 25-40%. All of these estimations have the same starting values, except for the learning parameters. ML(CJ) is the marginal likelihood value calculated according to Chib and Jeliaskov (2001). ML(Harmonic) is the marginal likelihood using the modified harmonic mean, as suggested by Adolfson *et al.* (2007). BIC is calculated at the median value of the posterior distributions

Taylor rule in both scenarios, and endogenous-gain learning followed the Taylor rule with the information.

In terms of comparing across the endogenous-gain specifications, I find that the scale factor, α_{sf}, doubles when agents incorporate the time varying trend in their information set. This might result from agents using smaller time windows to follow the random walk behavior of the trend.

One concern one might have is that some of the posterior means (specifically, θ_π, θ_x, σ_g, σ_u, σ_e and σ_ι) remain close to the prior means. Figure IV.2 shows that this is not the case. Only in one case does the mean of the prior did receive any weight in the posterior distribution.

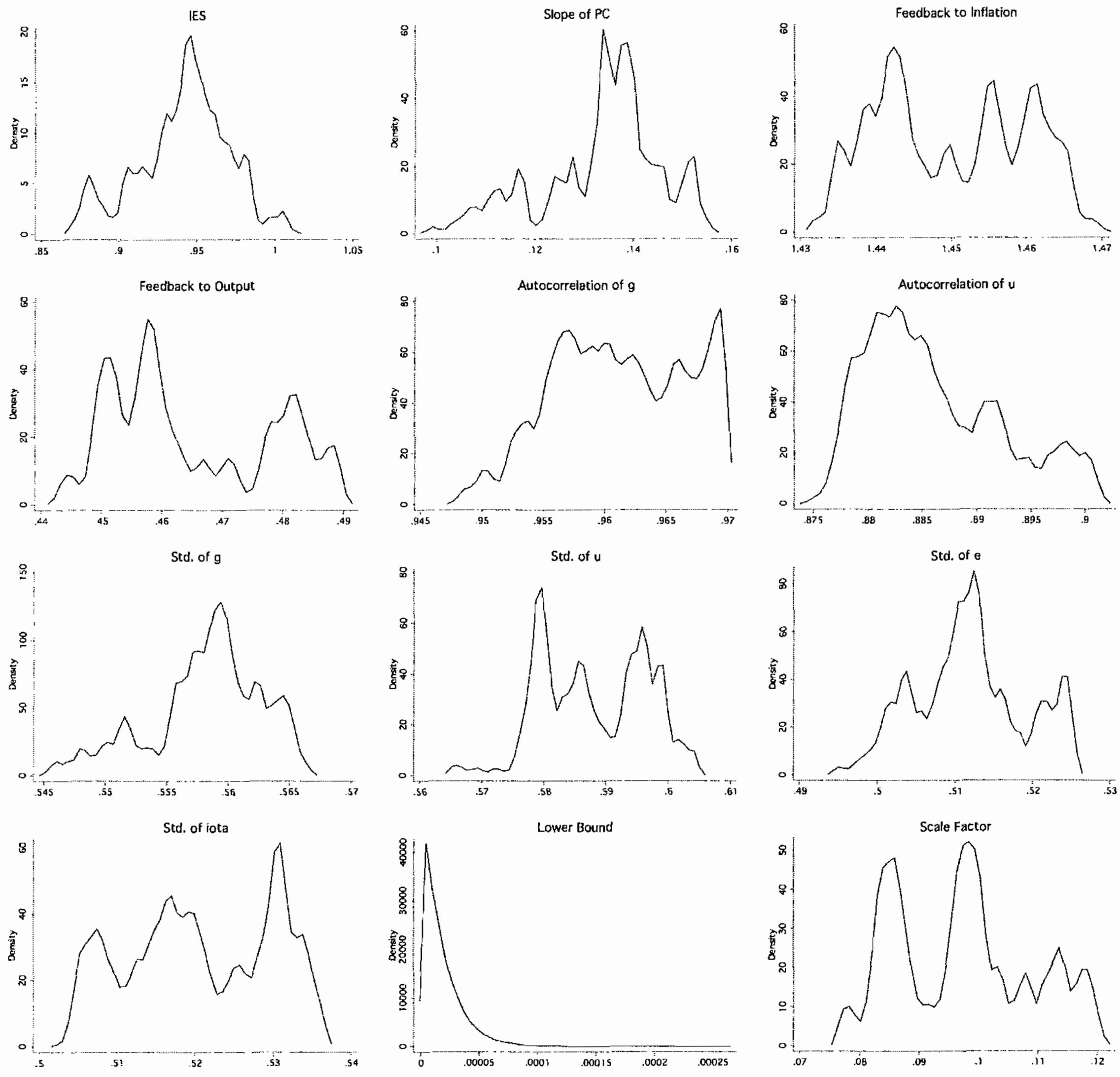

Figure IV.2: Posterior Distributions of the Endogenous-Gain with Trend Information *ii. Perceptions vs. Reality*

If one takes the learning hypothesis seriously then the learning estimation provides a convenient by product, the agents' perceptions. By backing out the PLM and the ALM for the coefficients of the reduced form model that agents estimate, one can interpret what agents react to, and how their reactions change over time. Since, in terms of model comparison, the endogenous-gain estimation is preferred the analysis below uses the results from the endogenous-gain learning estimation.

Figure IV.3 displays graphs of all twelve coefficients that agents estimate when they do not have information on the trend of inflation. Each column represents each equation for output, inflation and interest-rates respectively. The first row illustrates the constant component of the forecasting equation, the following rows represent the coefficients on the errors, g_t,u_t, and e_t, respectively. The black line in each graph represents the ALM and the gray line the PLM.

Looking at the ALM I find evidence of structural breaks in four of the coefficients, which justifies the use of the endogenous-gain. The break in these four coefficients indicate that something the inflation process has changed. Though agents do not follow the ALM very closely, they do react to the break. The structural break occurs at the beginning of the new millennium, right before the 2001 recession.

This period also saw a change in perceptions about reactions to interest-rate shocks. Prior to the break agents perceived that output and interest-rates responded to past interest-rate shocks. After the break the perceived inflation responded the most to past interest-rate shocks.

While agents, in general, do not perceive the ALM well, they do the worst job following the intercept for inflation. Recall that I hypothesized that not allowing agents to have trend interest-rate information would benefit the learning model.

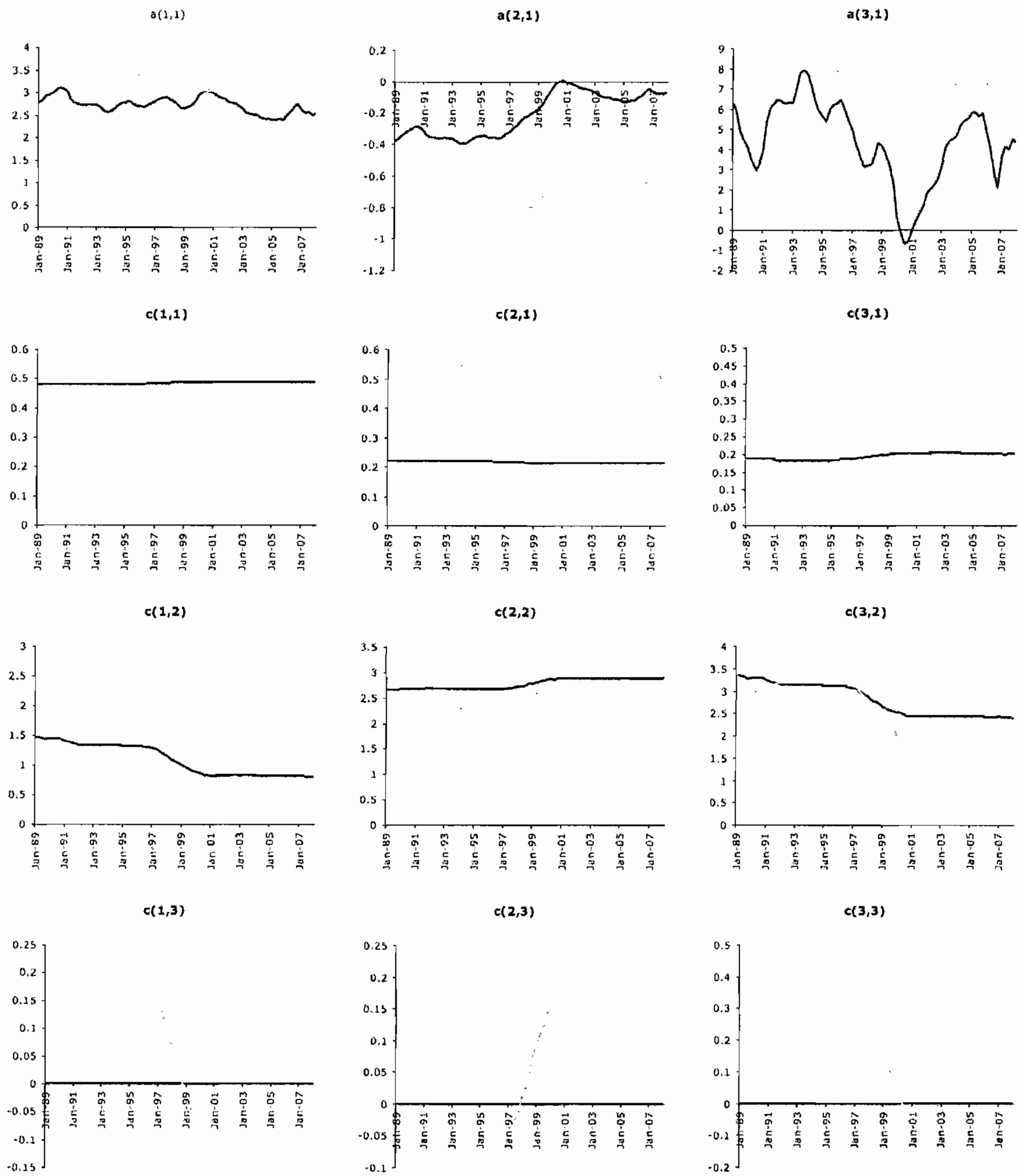

Figure IV.3: ALM (black line) and PLM (gray line) of Median Parameter Values without Trend Information

This result shows that learning does not pick up the trend inflation as a part of the intercept. This inability to track the trend of inflation probably causes the differences between PLM and ALM.

Figure IV.4 displays the graphs of all 15 coefficients that agents estimate when they do have information on the trend of inflation. The extra row supplies the coefficients on the lagged trend of inflation.

In these graphs we see much less movement of the actual and perceived coefficients. There are still a few indications of a structural break around 2000, but not nearly as significant as when agents use less information. In addition, agents do much worse in following the ALM. It does not appear that the large gain in BIC by learning results from time variation of the parameters, since coefficients exhibit fairly stable dynamics.

4. Rational Expectations Data and Learning

One result in Milani (2007b) asserts that if agents are learning and there is no conditional heteroscedasticity then an econometrician may be fooled into estimating ARCH/GARCH models. However, no research to date has investigated the converse: would a researcher observe learning dynamics when agents use rational expectations (RE)?

Chevillon *et al.* (2010) investigate a similar question, but focus on identification. They also use classical inference as opposed to the Bayesian techniques favored here. Specifically, Chevillon *et al.* show that the Anderson-Rubin statistic, with appropriate choice of instruments, can result in valid inference.[11]

[11] Appropriate instruments usually are predetermined variables.

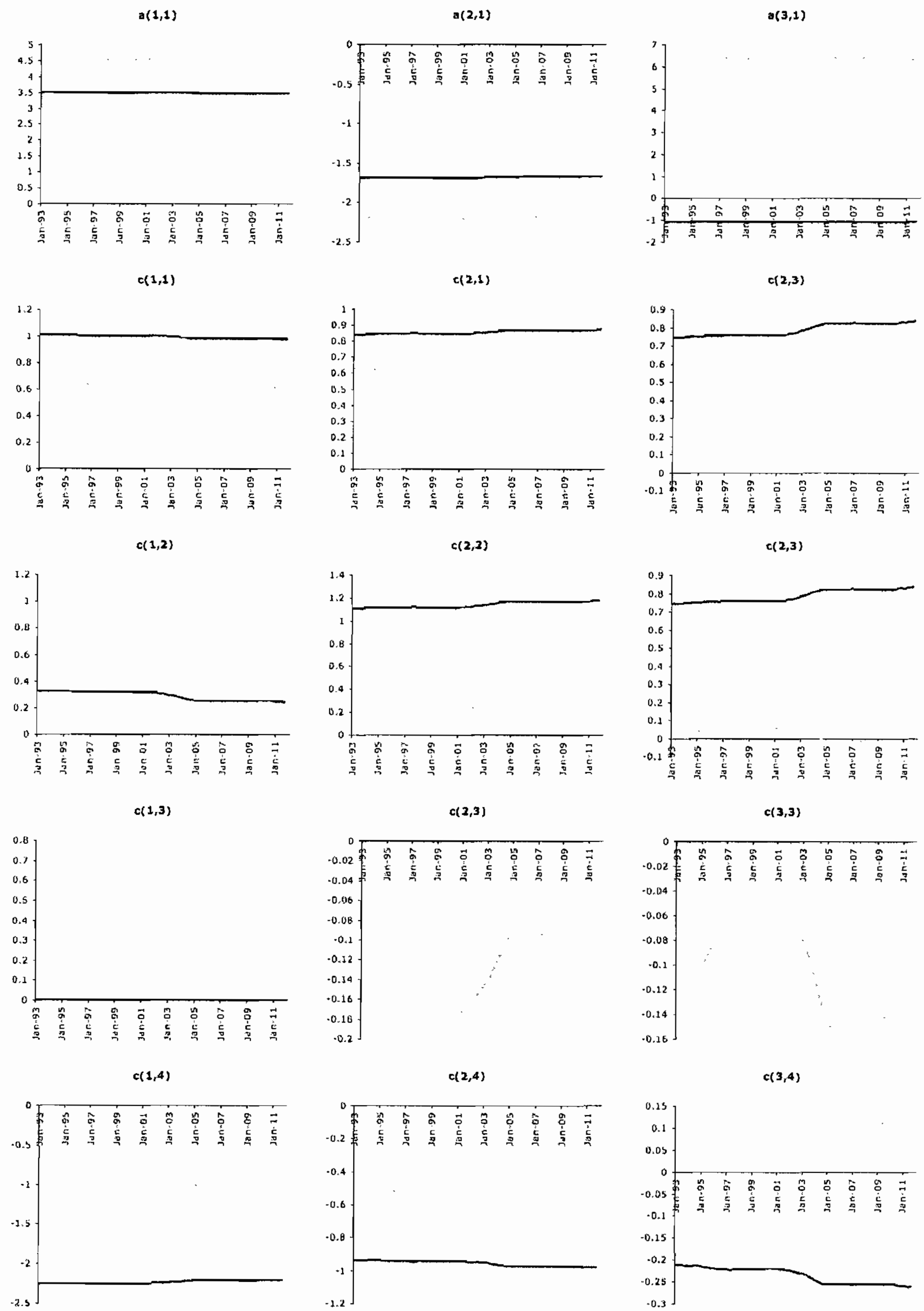

Figure IV.4: ALM (black line) and PLM (gray line) of Median Parameter Values with Trend Information

The economy is described by a similar NK model as above except I remove the time varying trend of the interest-rate from equation (2.3).

$$i_t = \theta_\pi \pi_t^e + \theta_x x_t^e + \varepsilon_t, \tag{4.1}$$

In order to derive the rational expectations solution used for the simulations I rewrite the NK model in matrix notation,

$$y_t = M_0 y_t^e + M_1 y_{t+1}^e + P v_t, \tag{4.2}$$

where $y_t = (x_t, \pi_t, i_t)'$, $v_t = (g_t, u_t, \varepsilon_t)'$, and M_0, M_1, and P are the appropriate matrices. Assuming the MSV solution $y_t^{RE} = \bar{c} v_{t-1}$ one can substitute in and solve for the RE coefficients $\bar{c}$. The substitution yields,

$$\bar{c} = (M_0 \bar{c} + M_1 \bar{c} F + PF). \tag{4.3}$$

Using the following identity $vec(ABC) = (C' \otimes A) vec(B)$, one can easily show that $\bar{c}$ results from the vector, $(I_9 - (I_3 \otimes M_0 + F' \otimes M_1))^{-1}(F' \otimes I_3) vec(P)$. Thus, the RE law of motion is,

$$y_t = \bar{c} v_{t-1} + P \tilde{v}_t. \tag{4.4}$$

For the learning estimation procedure I follow the same steps as above to obtain the following ALM:

$$Z_t = A^{-1}(B + C) a_{t-1} + A^{-1}(B c_{t-1} + C c_{t-1} F) F v_{t-1} + A^{-1} v_t. \tag{4.5}$$

i. Monte Carlo Experiment

For simulations of the rational expectations model I use the same values for the NK parameters as the previous chapter. Finally, I calibrate the parameters of the error terms as $\mu = \rho = 0.8$ and $\sigma_g = \sigma_u = \sigma_e = 0.2$. I conduct 100 simulations of RE data of 120 periods each. This means that each estimation relies on 100 periods of data.

In order to make an accurate portrayal I estimate the model assuming rational expectations and assuming learning. Table IV.4 displays the results for the Monte Carlo experiment.

The RE estimation naturally pins down all the parameter estimates within a single standard deviations. This holds within each estimation and across the estimations. The learning model, however, has small standard errors within each estimation, and relatively large standard errors across estimations. This suggests that particular realizations of the rational expectations model can fool a researcher into believing that learning exists in the model. Not surprisingly, the model comparison values overwhelmingly favor the rational expectations model.

Turning to the parameter estimates of learning estimation a striking pattern emerges. The learning assumption cause the researcher to underestimate the deep parameters and the correlations and overestimate the standard errors of the shocks. The learning process subsumes some of the autocorrelation, and as a byproduct it alters the parameter estimates.

Another interesting point is that the learning estimation does not nest the rational expectations solution like one might suspect. In the theoretical learning literature an extremely small constant-gain is typically considered consistent with rational expectations. Even though the constant-gain term is not statistically

Table IV.4: RE and Learning Monte Carlo Experiment Results

Description	Param	Actual	RE-Est	Learning-Est
Elas. of Subs.	ϕ	6.369	6.355	6.425
			(0.210)	0.243
			(0.252)	*(0.335)*
Slope of PC	λ	0.024	0.0234	0.022
			(0.0050)	(0.0037)
			(0.0060)	*(0.0002)*
Feedback to Infl.	θ_π	1.5	1.810	1.130
			(0.333)	(0.0752)
			(0.358)	*(0.0628)*
Feedback to Output	θ_x	0.5	0.736	0.252
			(0.253)	(0.0574)
			(0.267)	*(0.0348)*
Corr. of g_t	μ	0.8	0.799	0.783
			(0.0064)	(0.0169)
			(0.0073)	*(0.0032)*
Corr. of u_t	ρ	0.8	0.793	0.852
			(0.0580)	(0.0450)
			(0.0521)	*(0.0109)*
Std. g_t	σ_g	0.2	0.206	0.253
			(0.0170)	(0.0257)
			(0.0174)	*(0.0065)*
Std. u_t	σ_u	0.2	0.201	0.467
			(0.0163)	(0.0637)
			(0.0157)	*(0.0710)*
Std. e_t	σ_e	0.2	0.201	0.243
			(0.0149)	(0.0217)
			(0.0150)	*(0.0095)*
Constant-Gain	γ			0.010
				(0.0033)
				(0.0001)
ML(CJ)			-146.1	-286.2
ML(Harmonic)			-46.3	-118.5
BIC			372.1	662.5

Note: Results presented here represent the mean of each parameter. Regular parentheses indicate average standard deviation within estimations. Italicized parentheses indicate standard deviation across estimations. ML(CJ) is the marginal likelihood value calculated according to Chib and Jeliaskov (2001). ML(Harmonic) is the marginal likelihood using the modified harmonic mean, as suggested by Adolfson *et al.* (2007). BIC is calculated at the median value of the posterior distributions different from zero, all the parameter estimates of the learning estimation do not contain the actual parameter values in a 95% confidence interval.

5. Conclusion

Using a simple NK model I show that endogenous-gain learning provides significant improvement on both RE and constant-gain learning. I use a different approach than other DSGE estimations by using the lagged, filtered estimates of the residuals as the regressors. I find that, conditional on the specification, agents perceptions do not align with the actual path of the reduced form coefficients.

One reason why learning might fit the data better than RE is because it allows for time-variation in the reduced form coefficients, however, analysis of the reduced form coefficients shows little variation over time. In addition, I have shown that if the underlying data generating process resulted from RE learning would perform poorly. The Monte Carlo experiment also underscores that a learning estimation does not nest the rational expectations result. Further research is necessary to determine why this is the case.

CHAPTER V

CONCLUSION

Macroeconomic models are heavily influenced by expectations. The learning literature provides an excellent opportunity to relax the assumption of rational expectations. My research explores a particular version of learning where agents are allowed to change the amount of data they use to forecast future variables.

Marcet and Nicolini (2003) provides one of the first learning rules in economics that allows agents to adjust the size of the data set they are using. The second chapter showed that this particular form of learning exhibits some exotic dynamics. These dynamics are a direct result of the tension between the instability of learning under "large" constant-gain values and stability of decreasing-gain learning. I find that during the episodes of temporary instability learning results in 4 to 6 times more output volatility.

Chapter three focused on the forecasting ability of the Marcet and Nicolini switching-gain and an alternative endogenous-gain. I find little evidence that the Marcet and Nicolini switching-gain provides significant improvement over a constant-gain. The alternative endogenous-gain can have up to eight times the best improvement of the Marcet and Nicolini switching-gain.

The penultimate chapter uses a Bayesian estimation strategy of a NK DSGE model to compare RE, constant-gain learning, and endogenous-gain learning. I find compelling evidence that the DSGE model with endogenous-gain learning fits macroeconomic data better than RE and constant-gain learning.

The Monte Carlo exercise from the fourth chapter suggests that the properties of a learning estimation deserve further investigation. Specifically, exploration of how learning reacts to ARCH/GARCH dynamics and time-varying parameters. Constant-gain learning with fixed deep parameters may be able to capture unknown time-variation in parameters. Endogenous-gain learning might fit a model with ARCH/GARCH dynamics, since that learning process generates endogenous volatility.

Other future research should include looking at alternative information sets for different agents. For example Bullard and Mitra (2006) provides theoretic results when the Central Bank has a different information set than the private sector. This would be straightforward to implement in a learning estimation and may improve prediction.

Continuing on the line of information sets, most structural break models assume that the underlying coefficients change values. An alternative for DSGE models could be to assume that the structural break occurs in the information set. Agents might use lagged endogenous variables in certain circumstances and exogenous shocks in others.

APPENDIX A

ADDITIONAL TABLES

Table A.1: Switching-Gain Stability: Varying Historical Window Length

Hist. Window	35	85	135	185	235	285
Percent Stable	78.18	86.92	84.18	78.58	72.26	67.18

Hist. Window	335	385	435	485	535	585
Percent Stable	62.06	58.44	54.12	52.50	50.46	49.00

Shows the percent of simulations in which the last value of the estimated parameters lie within 2 percent of the T-map. The historical window is the parameter the governs the number of periods used to calculate the historical average MSFE.

Table A.2: Forecast Ability of the Marcet and Nicolini Switching-Gain: $\sigma_\nu = 4$ and $\sigma_\eta = 0.5$

v	$1-\varepsilon$					
	0.01	0.05	0.1	0.9	0.95	0.99
0.5	1.0000	1.0000	1.0000	1.0640	1.0870	1.0310
	6.73E-08	1.61E-05	1.40E-07	0.0146	0.01981	0.02547
1	1.0000	1.0000	1.0000	1.2380	1.2700	1.1380
	2.93E-07	1.08E-05	2.36E-07	0.02539	0.03364	0.04417
1.5	1.0000	1.0000	1.0000	1.3370	1.3860	1.2410
	4.23E-07	2.14E-06	4.15E-07	0.03095	0.04325	0.05617
2	1.0000	1.0000	1.0000	1.3880	1.4510	1.2970
	6.82E-07	9.31E-07	5.54E-07	0.03356	0.04578	0.0535
2.5	1.0000	1.0000	1.0000	1.4310	1.5070	1.3610
	8.26E-07	6.83E-05	7.55E-07	0.04187	0.0448	0.05789
3	1.0000	1.0000	1.0000	1.4880	1.5640	1.4220
	1.15E-06	1.39E-06	8.24E-07	0.03864	0.04678	0.05412
3.5	1.0000	1.0000	1.0000	1.5390	1.6350	1.5110
	1.23E-06	2.97E-05	1.80E-06	0.03726	0.05384	0.05477
4	1.0000	1.0000	1.0000	1.5990	1.7000	1.5970
	1.29E-06	3.09E-05	1.23E-06	0.04544	0.06223	0.08049
10	1.0000	1.0000	1.0000	2.4670	3.1310	4.1860
	1.08E-05	8.89E-06	8.23E-06	0.1092	0.159	0.4163

The first row indicates for each value of v indicates mean of the relative MSFE. The second row reports the standard deviation. v is the value of forecast error above which agents switch to a constant gain. ε is the probability of structural break in the underlying coefficient.

Table A.3: Forecast Ability of the Marcet and Nicolini Switching-Gain: $\sigma_\nu = 2$ and $\sigma_\eta = 2$

v	$1-\varepsilon$					
	0.01	0.05	0.1	0.9	0.95	0.99
0.5	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	6.88E-08	8.14E-08	1.05E-07	8.61E-05	0.0002599	0.0003398
1	1.0000	1.0000	1.0000	1.0000	0.9995	0.9989
	2.03E-07	1.80E-07	1.51E-07	0.001294	0.0008205	0.0004679
1.5	1.0000	1.0000	1.0000	1.0030	1.0000	0.9973
	3.06E-07	2.08E-07	2.70E-07	0.001996	0.001665	0.001421
2	1.0000	1.0000	1.0000	1.0050	1.0020	0.9966
	4.47E-07	2.73E-07	4.09E-07	0.002419	0.002391	0.001941
2.5	1.0000	1.0000	1.0000	1.0100	1.0060	0.9976
	6.25E-07	5.24E-07	6.16E-07	0.003472	0.002709	0.002289
3	1.0000	1.0000	1.0000	1.0180	1.0130	0.9978
	9.65E-07	6.60E-07	6.15E-07	0.003438	0.004086	0.002763
3.5	1.0000	1.0000	1.0000	1.0290	1.0210	0.9993
	1.59E-06	1.05E-06	1.12E-06	0.004924	0.00509	0.003423
4	1.0000	1.0000	1.0000	1.0410	1.0350	1.0030
	3.58E-06	1.40E-06	1.33E-06	0.005826	0.006158	0.004386
10	0.9999	0.9999	0.9999	1.2390	1.3470	1.3880
	5.09E-05	3.46E-05	4.10E-05	0.02088	0.02318	0.04167

The first row indicates for each value of v indicates mean of the relative MSFE. The second row reports the standard deviation. v is the value of forecast error above which agents switch to a constant gain. ε is the probability of structural break in the underlying coefficient.

APPENDIX B

DERIVATIVES OF THE T-MAPS

The derivatives of the Taylor-type discretionary policy rule T-mapping are as follows,

$$DT_a = M_0 + M_1,$$
$$DT_c = F' \otimes M_1 + I \otimes M_0.$$

The derivatives of the Duffy and Xiao commitment rule are as follows,

$$DT_a = M_0 + M_1(I + \bar{b}) + M_1\bar{b}_0 Q_0,$$
$$DT_{b_0} = I \otimes M_0 + \bar{b}_0' \otimes M_1 + I \otimes M_1\bar{b}_1 Q_0,$$
$$DT_{b_1} = I \otimes M_0 + I \otimes M_1\bar{b}_0 + I \otimes M_1\bar{b}_1 Q_0,$$
$$DT_c = F' \otimes M_1 + I \otimes M_0 + I \otimes M_1\bar{b}_0 + I \otimes M_1\bar{b}_1 Q_0.$$

The derivatives of the Evans and Honkapohja expectations based commitment rule are as follows,

$$DT_a = M(I + \bar{b}),$$
$$DT_b = \bar{b}' \otimes M,$$
$$DT_c = F' \otimes M + I \otimes M\bar{b}.$$

The derivatives of the Svensson and Woodford and the McCallum and Nelson commitment rules are as follows,

$$DT_a = M_0 + M_1(I + \bar{b}),$$

$$DT_b = \bar{b}' \otimes M_1 + I \otimes M_1\bar{b} + I \otimes M_0,$$

$$DT_c = F' \otimes M_1 + I \otimes M_1\bar{b} + I \otimes M_0.$$

BIBLIOGRAPHY

[1] Adolfson, M, J Lind and M Villani. "Bayesian Analysis of DSGE Models-Some Comments," *Econometric Reviews* 26.2 (2007):173-185.

[2] An, S and F Schorfheide. "Bayesian Estimation of DSGE Models," *Econometric Reviews* 26.2-4 (2007).

[3] Beck, GW and V Wieland. "Learning and Control in a Changing Economic Environment," *Journal of Economic Dynamics and Control* 26.9-10 (2002):1359-1377.

[4] Blanchard, O and C Kahn. "The Solution of Linear Difference Models Under Rational Expectations," *Econometrica* (1980):1305-1311

[5] Blume, L, M Bray and D Easley. "Introduction to the Stability of Rational Expectations Equilibrium.," *Journal of Economic Theory* 26.2 (1982):313-317.

[6] Branch, W and B McGough. "Consistent Expectations and Misspecification in Stochastic Non-Linear Economies," *Journal of Economic Dynamics and Control* 29.4 (2005):659-676.

[7] Branch, W and G Evans. "A Simple Recursive Forecasting Model," *Economics Letters* 91.2 (2006):158-166.

[8] ——. "Model Uncertainty and Endogenous Volatility," *Review of Economic Dynamics* 10.2 (2007):207-237.

[9] Bray, M. "Learning, Estimation, and the Stability of Rational Expectations," *Journal of Economic Theory* 26.2 (1982):318-339.

[10] Bray, M and N Savin. "Rational Expectations Equilibria, Learning and Model Specification.," *Econometrica* 54.5 (1986):1129-1160.

[11] Bullard, J. "Time-Varying Parameters and Non-Convergence to Rational Expectations Under Least Square Learning," *Economic Letters* 40.2 (1992):159-166.

[12] ——. (2008) "Discussion of Fabio Milani, Learning and Time-Varying Macroeconomic Volatility." *International Research Forum on Monetary Policy*, 2008, Frankfurt am Main.

[13] Bullard, J and K Mitra. "Learning About Monetary Policy Rules," *Journal of Monetary Economics* 49.6 (2002):1105-1129.

[14] Carceles-Poveda, E and C Giannitsarou. "Adaptive Learning in Practice," *Journal of Economic Dynamics and Control* 31.8 (2007):2659-2697.

[15] Chevillon, G, M Massmann and S Mavroeidis. "Inference in Models With Adaptive Learning," *Journal of Monetary Economics* (2010).

[16] Chib, S and I Jeliazkov. "Marginal Likelihood From the Metropolis-Hastings Output," *Journal of the American Statistical Association* 96.453 (2001):270-281.

[17] Cho, I, N Williams and T Sargent. "Escaping Nash Inflation," *The Review of Economic Studies* 69.1 (2002):1-40.

[18] Clarida, R, J Gali and M Gertler. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory*," *Quarterly Journal of Economics* 115.1 (2000):147-180.

[19] Cogley, T and A Sbordone. "Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve," *American Economic Review* 98.5 (2008):2101-2126.

[20] Duffy, J and W Xiao. "The Value of Interest Rate Stabilization Policies When Agents Are Learning," *Journal of Money, Credit and Banking* 39.8 (2007):2041-2056.

[21] Evans, G. "Expectational Stability and the Multiple Equilibria Problem in Linear Rational Expectations Models," *Quarterly Journal of Economics* 100.4 (1985):1217-1233.

[22] Evans, GW and S Honkapohja. "Economic Dynamics With Learning: New Stability Results," *Review of Economic Studies* (1998):23-44.

[23] ——. *Learning and Expectations in Macroeconomics.* Princeton Univ Pr, 2001.

[24] Evans, GW and Honkapohja, S. "Robust Learning Stability with Operational Monetary Policy Rules" in *Monetary Policy under Uncertainty and Learning.* Eds. Schmidt-Hebbel, K and Walsh, C. Santiago: Central Bank of Chile, (2009). 145-170.

[25] Evans, G and B McGough. "Monetary Policy, Indeterminacy and Learning," *Journal of Economic Dynamics and Control* 29.11 (2005):1809-1840.

[26] ——. "Implementing Optimal Monetary Policy in New-Keynesian Models With Inertia," *mimeo* (2006).

[27] Fernndez-Villaverde, J and JF Rubio-Ramrez. "Estimating Dynamic Equilibrium Economies: Linear Versus Nonlinear Likelihood," *Journal of Applied Econometrics* 20.7 (2005):891-910.

[28] ——. "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies* 74.4 (2007):1059-1087.

[29] Fernndez-Villaverde, J, P Guerron and JF Rubio-Ramirez. "Fortune Or Virtue: Time-Variant Volatilities Versus Parameter Drifting in US," *mimeo* (2009).

[30] Frydman, R. "Towards an Understanding of Market Processes: Individual Expectations, Learning, and Convergence to Rational Expectations Equilibrium," *The American Economic Review* 72.4 (1982):652-668.

[31] Fuhrer, J and G Moore. "Inflation Persistence," *The Quarterly Journal of Economics* (1995):127-159.

[32] Geweke, J. "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews* 18 (1999):173.

[33] Honkapohja, S and K Mitra. "Learning Stability in Economies With Heterogeneous Agents," *Review of Economic Dynamics* 9.2 (2006):284-309.

[34] Justiniano, A and G Primiceri. "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review* 98.3 (2008):604-641.

[35] Kim, C and C Nelson. "Estimation of a Forward-Looking Monetary Policy Rule: A Time-Varying Parameter Model Using Ex Post Data," *Journal of Monetary Economics* 53.8 (2006):1949-1966.

[36] Marcet, A and JP Nicolini. "Recurrent Hyperinflations and Learning," *American Economic Review* 93.5 (2003):1476-1498.

[37] McCallum, B and E Nelson. "Timeless Perspective Vs. Discretionary Policy in Forwardlooking Models," *Federal Reserve Bank of St. Louis Review* 86 (2004):43-56.

[38] McGough, B. "Statistical Learning With Time-Varying Parameters," *Macroeconomic Dynamics* 7.01 (2003):119-139.

[39] Milani, F. "Learning and Time-Varying Macroeconomic Volatility," *Manuscript, UC-Irvine* (2007).

[40] ——. "Expectations, Learning and Macroeconomic Persistence," *Journal of Monetary Economics* 54.7 (2007):2065-2082.

[41] ——. "Learning, Monetary Policy Rules, and Macroeconomic Stability," *Journal of Economic Dynamics and Control* 32.10 (2008).

[42] ——. "Empirical Significance of Learning in a New Keynesian Model With Firm-Specific Capital," *mimeo* (2008).

[43] ——. "Regime Switching, Learning, and the Great Moderation," *mimeo* (2008).

[44] ——. "Initial Expectations in New Keynesian Models With Learning," *mimeo* (2008).

[45] Muth, J. "Rational Expectations and the Theory of Price Movements," *Econometrica* 29.3 (1961):315-335.

[46] Sargent, T. *The Conquest of American Inflation.* Princeton University Press, 1999.

[47] Slobodyan, S and R Wouters. "Learning in an Estimated Medium-Scale DSGE Model.," *mimeo* (2007).

[48] ——. "Estimating a Medium-Scale Dsge Model With Expectations Based on Small Forecasting Models," *mimeo* (2008).

[49] Svensson, L and Woodford, M. "Implementing Optimal Policy Through Inflation-Forecast Targeting" in *The Inflation-Targeting Debate.* Eds. Bernanke, B and Woodford, M. Chap 2. Chicago: University of Chicago Press, (2005). 19-83.

[50] Wieland, V. "Monetary Policy, Parameter Uncertainty and Optimal Learning," *Journal of Monetary Economics* 46 (2000):199-228.

[51] Woodford, M. *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press, 2003.