Learning, information and heterogeneity*

Liam Graham†

20 August 2011

Abstract

Most DSGE models assume full information and model-consistent expectations. This paper relaxes both these assumptions in the context of the stochastic growth model with incomplete markets and heterogeneous agents. Households do not have direct knowledge of the structure of economy or the values of aggregate quantities; instead they form expectations by learning from the prices in their market-consistent information sets. The economy converges quickly to an equilibrium which is similar to the equilibrium with model-consistent expectations and market-consistent information. Learning does not introduce strong dynamics at the aggregate level, though more interesting things happen at the household level. At least in the context of this model, assumptions about information seem important for aggregates; assumptions about the ability to form model-consistent expectations less so.

JEL classification: D52; D84; E32.

Keywords: imperfect information; adaptive learning; dynamic general equilibrium; heterogeneity; expectations.

*For helpful comments, I thank Martin Ellison, Markus Knell, Francesco Lippi, Roland Meeks, Neil Rankin, Michael Reiter, Martin Summer, Stephen Wright and seminar participants at the Bank of England, Birkbeck College London the EIEF, Oxford, the Institute for Advanced Studies, Vienna. and the Oesterreichische Nationalbank. A previous version was circulated under the title "Learning from prices". I wrote the first draft of this paper while visiting the EIEF in Rome and would like to thank the Institute for giving me the opportunity to work in such a wonderful environment.

†Department of Economics, University College London, Gower Street, London WC1E 6BT, UK. Liam.Graham@ucl.ac.uk

1 Introduction

Most dynamic stochastic general equilibrium models assume agents can form model-consistent expectations and have full information. The learning literature relaxes the first assumption while typically retaining the second. The imperfect information literature relaxes the second while retaining the first. This paper relaxes both of these assumptions, in the context of the stochastic growth model with incomplete markets and heterogeneous agents.

Households are assumed to have "market-consistent" information sets (Graham and Wright, 2010). Incomplete markets will then lead to incomplete information, and with incomplete markets as in Krusell and Smith (1998) households' information sets will consist of the wage and the return to capital. To implement consumption, households need to forecast these prices and, since it is assumed they cannot form model-consistent expectations, they do so by estimating a vector autoregression (VAR) in the prices. Thus households in this model have no direct knowledge of the structure of the macroeconomy or the values of aggregate quantities. The model avoids the unrealistic cognitive demands of model-consistent expectations (which are even less realistic in models of heterogeneous information such as Nimark, 2007 or Graham and Wright, 2010 in which agents estimating infinite hierarchies of expectations), and the strong informational requirements of full information models. Conditional on their (in general non-model-consistent) beliefs, households are modelled as rational.[1]

Given households' VAR in prices (their "perceived law of motion", PLM), the paper derives a state space representation of the actual law of motion (ALM) in which the state is shown to expand to the full history of the economy[2]. Since households estimate a finite order VAR, any resulting equilibrium must therefore be a restricted perceptions equilibrium (RPE, in the sense of Evans and Honkapohja, 2001) and conditions are given for its stability and learnability.

The properties of the model are then studied numerically. In the stochastic steady state of the model, structural heterogeneity across agents creates heterogeneity of beliefs. To understand the impact of learning, a careful consideration of the steady state distribution of beliefs is necessary and this mechanism is clarified using a simple univariate example.

The main results are as follows

1. The economy converges to the restricted perceptions equilibrium from any stable prior. Under ordinary least squares learning, a standard theorem can be used to

[1] This is in contrast to the "Euler equation learning" approach in which agents have finite forecast horizons. For a discussion see Preston (2005), Evans et al, (2011) or Graham (2011).

[2] There is a close link with the "infinite hierarchy of expectations" (Townsend, 1983) that characterizes models with heterogeneous information sets.

show that there is $\sqrt{t}$ convergence. Under constant gain learning the speed of convergence depends on the chosen gain.

2. Constant gain learning only has small effects on the aggregate economy, changing the volatility of aggregates by at most a few percent from the model-consistent expectations equilibrium (MCEE). Learning can amplify or mute the effect of shocks, depending on the gain.

3. Household - level quantities are much more volatile than at the MCEE. Specifically, the volatility of household consumption growth is much closer to that found in the data.

4. If an intercept is included in the learning rule, the aggregate economy becomes so volatile that it is only stable with very small gains.

The first result is in contrast to existing studies (for example Carceles-Poveda and Giannitsarou, 2007) which show convergence is slow in a full information representative agent model. Graham (2011) shows that fast convergence is a consequence of modelling households as individually rational as opposed to the "Euler equation learning" approach (Honkapohja et al, 2011) of bounded individual rationality. The present paper shows that such fast convergence also occurs with restricted information sets.

The speed of convergence is important for reasons of informational plausibility. To avoid arbitrary dynamics arising along the convergence path, models with learning are typically initialized with learning rules that have already converged. If convergence is slow this becomes a very strong assumption, endowing agents with the knowledge they are supposed to be learning. If convergence is fast this assumption becomes much more plausible.

The second result goes against the simple intuition drawn from representative agent models that constant gain learning, by increasing the volatility of expectations, increases the volatility of the economy. The economy with learning is characterized by a (stochastic) steady state distribution of beliefs across households. The impact of this distribution of beliefs on aggregates depends on its shape (in general it will not be symmetrical or centred on the restricted perceptions equilibrium) and the non-linearity of consumption to beliefs. The overall effect on the macroeconomy is modest, with a reduction in the impact effect of shocks but an increase in persistence combining to give an increase in the volatility of aggregates of at most a few percent over the equilibrium with model-consistent expectations. Higher gains can result in *lower* volatilities. In practice, given the data typically available, it would be difficult to distinguish the aggregate economy with learning from one with model-consistent expectations.

The combination of the first two results allows a model-consistent expectations equilibrium to be interpreted as the outcome of a learning process that has already converged

(Grandmont, 1998, makes this point). However, the equilibrium that the economy converges to is that with market-consistent information, described in Graham and Wright (2010), which differs in significant ways from the equilibrium with full information. At least in the context of this model, assumptions on information seem important, assumptions on the ability of households to form model-consistent expectations less so.

This is particularly striking given that households are making decision under very limited information. The rules they use to forecast their income are misspecified in three ways: they do not know the true law of motion for the economy; they cannot observe the state variables and whereas the true law of motion will be infinite-dimensional they are restricted to using a finite number of lags. Yet still the economy converges quickly to an equilibrium that would in practice be hard to distinguish from one in which agents form model-consistent expectations.

From a modelling point of view, the approach of this paper has a clear advantage. Papers in the learning literature typically makes assumptions, often tacitly about what information agents have and how they use it. Some papers assume that agents use one information set when choosing how much to consume and another when they form expectations[3]. Some papers assume that agents only use a subset of the information at their disposal to form expectations[4]. Many papers model a representative agent (for example, Carceles-Poveda and Giannitsarou, 2007 or Milani, 2007, 2011), but need to then assume that the agent does not realize that solving for their own consumption is the same as solving for the law of motion of the economy as a whole[5]. Other examples can be found throughout the learning literature. Such assumptions may be perfectly valid, but they make it hard to identify the extent to which results are due to these specific informational assumptions and the extent to which they are due to the central issue of the learning literature - the inability of agents to form model-specific expectations. This paper avoids such assumptions by on the one hand modelling heterogeneity explicitly and on the other by simply assuming that households know nothing about the macroeconomy apart from their market-consistent information sets.

Excellent overviews of the literature on imperfect information and learning can be found in Hellwig (2006) and Carceles-Poveda and Giannitsarou (2007) respectively. Although most learning models consider a representative agent, there is a growing literature on learning with heterogeneity. The bulk of this literature, for example Giannitsarou (2003), Branch and McGough (2004) or Branch and Evans (2006) address the problem

[3] Fout and Francis (2011) study this issue and coin the term "information-consistent learning". Another example is Eusepi and Preston (2011) who assume that agents observe the innovation to technology for the purposes of calculating their consumption but do not use it in their learning rule.

[4] An example of this is Evans et al (2009) which assumes agents forecast interest rates using information only on lagged interest rates.

[5] Justifying this approach by assuming an economy with many identical agents who do not know they are identical again involves an artificial restriction - that agents do not run a simple regression of individual quantities on aggregates which would immediately reveal a perfect correlation.

of heterogeneity in learning rules or initial conditions. An exception is Honkapohja and Mitra (2006) which addresses a general model with both learning and structural heterogeneity, which under full information, would nest the model in this paper. However under imperfect information this is not the case and this paper extends the techniques introduced by Honkapohja and Mitra (2006) to an economy with incomplete and heterogeneous information sets.

The remainder of this paper is structured as follows. Section 2 presents the model. Section 3 states the PLM, derives the ALM and gives stability conditions. Section 4 investigates the convergence properties of the model under ordinary least squares learning and section 5 studies the economy under constant gain learning. Section 6 concludes. Derivations and proofs are in the Appendix.

2 The model

This section presents a model of the type that is becoming standard in the dynamic general equilibrium literature[6]. There are a large number of households and a large number of firms, divided across S islands. There are shocks to aggregate and island-specific labour productivity. Markets are incomplete in the sense that there are no swaps, only markets for capital and labour. Since the model is standard, only the linearised constraints and optimality conditions are presented here[7]. More workings are in Appendix A.1.

2.1 Households

A typical household on island s consumes (c_t^s) and rents capital (k_t^s) and labour (h_t^s) to firms. Household labour on each island has idiosyncratic productivity (z_t^s) whereas capital is homogenous, so households earn the aggregate gross return (r_{kt}) on capital but an idiosyncratic wage (w_t^s) on their labour. Households on different islands are unconditionally identical.

The Euler equation for a typical household s is

$$\tilde{E}_t^s \Delta c_{t+1}^s = \sigma \tilde{E}_t^s r_{t+1} \tag{1}$$

where r_t is the net return to capital (related to the gross return by $r_{kt} = r_t + (1-\delta)$) and σ the coefficient of relative risk aversion. The expectations operator for household s, with a tilde since in the general case individuals will have non-model-consistent expectations,

[6] Examples of papers which use similar models include Krusell and Smith (1998), Graham and Wright (2010) and Lorenzoni (2010).

[7] The linear model presented here can be thought of as a first-order approximation to a non-linear model (for details of the linearisation see Graham and Wright, 2010). However since the focus of this paper is on learning and as is conventional only linear learning rules are investigated, it may be better simply to think of the model as linear *per se*.

is defined as the expectation given the household's information set Ω_t^s, i.e. for some variable a_t $\tilde{E}_t^s a_t = \tilde{E} a_t | \Omega_t^s$.

The first-order condition for labour is

$$n_t^s = \frac{1-n}{n}\eta \left(w_t^s - c_t^s \right) \tag{2}$$

where n is steady-state labour and $\frac{1}{\eta}$ the intertemporal elasticity of labour supply. Household capital evolves according to

$$k_{t+1}^s = (1-\delta)\, k_t^s + \delta x_t^s \tag{3}$$

where δ is the rate of depreciation and x investment and the budget constraint is

$$\frac{c}{y} c_t^s + \left(1 - \frac{c}{y}\right) x_t^s = \alpha \left(w_t^s + n_t^s \right) + (1-\alpha) \left(r_{kt} + k_t^s \right) \tag{4}$$

where α is the labour share and $\frac{c}{y}$ is the steady-state consumption share of output. The left-hand side shows spending on consumption and investment weighted by their steady-state shares of output. The right-hand side shows income from renting labour and capital to firms weighted by their factor shares.

2.2 Firms

A typical firm on island s faces a production function

$$y_t^s = \alpha \left(a_t + z_t^s \right) + \alpha n_t^s + (1-\alpha)\, j_t^s \tag{5}$$

where a_t is an aggregate productivity shock, z_t^s island-specific productivity and j_t^s is the capital rented by the firm: in general, $j_t^s \neq k_t^s$, since capital will flow to more productive islands. The firm chooses capital and labour to satisfy first-order conditions

$$r_{kt} = (1-\alpha) \frac{k}{y} \left(y_t^s - j_t^s \right) \tag{6}$$

$$w_t^s = y_t^s - n_t^s \tag{7}$$

2.3 Aggregates

Aggregate quantities are sums over household or firm quantities, calculated as quantities per household. For example aggregate consumption is given by

$$c_t = \frac{1}{S} \sum_{s=1}^{S} c_t^s. \tag{8}$$

2.4 Markets

Markets are incomplete in the sense that the only markets are those for output, labour and capital - there are no swaps which allow idiosyncratic risk to be transferred between households.

The labour market is segmented in that firms on island s only rent labour from households on island s, and the wage on island s, w_t^s, adjusts to set labour supply (2) equal to labour demand (7). Appendix A.1.3 shows the relation between the idiosyncratic wage and the aggregate wage is:

$$w_t^s = w_t + z_t^s \tag{9}$$

Recall that in the general case no household will be able to observe w_t or z_t^s.

In contrast, capital is homogenous and tradeable between islands, so flows to islands with more productive labour. The gross return, r_t, adjusts to make the aggregate demand for capital from firms consistent with each household's Euler equation (1) and the aggregate budget constraint.

Market clearing prices (in the general case no household in the model could calculate these) are given by

$$w_t = \lambda_{wk} k_t + \lambda_{wa} a_t + \lambda_{wc} c_t \tag{10}$$

$$r_t = \lambda_{rk} k_t + \lambda_{ra} a_t + \lambda_{rc} c_t \tag{11}$$

where expressions for the coefficients are given in Appendix A.2. Note in the case of fixed labour supply ($\eta \to \infty$), $\lambda_{wc} = \lambda_{rc} = 0$.

2.5 Shocks

For both the aggregate and idiosyncratic productivity shocks, assume autoregressive processes

$$a_t = \phi_a a_{t-1} + \varepsilon_t \tag{12}$$

$$z_t^s = \phi_z z_{t-1}^s + \varepsilon_t^s \tag{13}$$

where ε_t and ε_t^s are iid mean-zero errors, and $E\varepsilon_t^2 = \sigma_a^2$; $E\left(\varepsilon_t^s\right)^2 = \sigma_z^2$. The innovation to the idiosyncratic process satisfies an adding up constraint, $\sum_{s=1}^{S} \varepsilon_t^s = 0$ which implies

$$\sum_{s=1}^{S} z_t^s = 0. \tag{14}$$

2.6 Information

Definition 1. ***(Full information)*** *Full information for a typical household on island* s*, denoted by an information set* Ω_t^{s*}*, is knowledge of the aggregate states and the relevant idiosyncratic states*

$$\Omega_t^{s*} = [k_t, a_t, k_t^s, z_t^s]$$

Graham and Wright (2010) argues that in a decentralized equilibrium the states will not in general be known by agents, so the assumption of full information is a strong one. Instead that paper proposes the following definition of an information set consistent with a decentralized equilibrium, reproduced here:

Definition 2. ***(Market-consistent information)*** *Households' information sets consists of the prices in the markets in which they participate.*

With only capital and labour markets the market-consistent information set of a household on island s at time t is[8]

$$\Omega_t^s = \left[\{r_i\}_{i=0}^t, \{w_i^s\}_{i=0}^t\right]$$

Define the innovation to this information set as

$$i_t^s = \begin{bmatrix} r_t & w_t^s \end{bmatrix}' \tag{15}$$

such that the information set evolves according to $\Omega_{t+1}^s = \Omega_t^s \cup i_{t+1}^s$.

2.7 Equilibrium

Definition 3. ***Equilibrium with market-consistent information:*** *a competitive equilibrium in which the law of motion of the economy is consistent with each agent solving a decentralized optimisation problem. A sequence of plans for allocations of households* $\left\{c_t^s, n_t^s, k_{t+1}^s\right\}_{t=1:\infty}^{s=1:S}$ *and prices* $\{r_t, w_t^s\}_{t=1:\infty}^{s=1:S}$

1. *Given prices and informational restrictions, the allocations solve the utility maximization problem for each household*

2. $\{r_t, w_t^s\}_{t=1:\infty}^{s=1:S}$ *are the marginal products of aggregate capital and island-specific labour.*

3. *All markets clear*

[8] Households also have knowledge of the history of their own decisions, $\{c_i^s\}_{i=0}^t, \{n_i^s\}_{i=0}^{t-1}, \{k_i^s\}_{i=0}^t$ however, since each of these histories embodies the household's own responses to the evolution of Ω_t^s, it contains no information not already in Ω_t^s.

2.8 Benchmark cases

The above model nests four familiar cases. With complete markets, the market-consistent information set is invertible (in the sense of Baxter et al, 2011) and full information is revealed (the result of Radner, 1979). All idiosyncratic risk is diversified away and the model is identical to the representative agent real business cycle model.

With incomplete markets and assumed full information, the path of the aggregate economy is identical to the complete markets case. This is related to Krusell and Smith's (1998) result that an economy with incomplete markets can closely resemble one with complete markets - the resemblance is exact in the model because it is linear. However the economy differs markedly at a household level since household wealth follows a unit root process.

With complete markets and learning, the model is the real business cycle model with learning, though it differs from most standard treatments (e.g. Carceles-Poveda and Giannitsarou) in that households are assumed to have infinite horizons (see Preston, 2005; Honkapohja et al, 2011; and Graham, 2011 for further discussion of this issue).

With market-consistent information and model-consistent expectations, the model is that studied in Graham and Wright (2010). Since this represents the limit to which a model with learning might converge, it is worth reviewing its properties. Market-consistent information implies heterogeneity of information across households, so to form model-consistent expectations households need to estimate an infinite hierarchy of expectations. Numerically, this leads to the properties of the model looking quite different from under full information, notably the sign of the impact response of aggregate consumption to an aggregate technology shock reverses. This is discussed further in section 3.6.

2.9 Optimal consumption

To solve for optimal consumption, substitute the budget constraint (4) into the capital evolution equation (3), solve forward and use the transversality condition on capital to give an expression relating the path of future consumption to current capital, current prices and expected future prices

$$\gamma_2 \tilde{E}_t^s \sum_{j=0}^{\infty} \beta^j c_{t+j}^s = \gamma_1 k_t^s + \gamma_3 w_t^s + \gamma_5 r_t + \tilde{E}_t^s \sum_{j=1}^{\infty} \beta^j \left(\gamma_3 w_{t+j}^s + \gamma_5 r_{t+j} \right) \tag{16}$$

where the constants are defined (along with a full derivation) in Appendix A.3. Iterate the Euler equation (1) forward to give

$$E_t c_{t+j}^s = c_t^s + \tilde{E}_t^s \sum_{i=1}^{j} r_{t+i} \tag{17}$$

Combining these give

$$c_t^s = \frac{r}{1+r}\frac{1}{\gamma_2}\left(\gamma_1 k_t^s + \gamma_3 w_t^s + \gamma_5 r_t\right) + \gamma_{cw}\tilde{E}_t^s \sum_{j=1}^{\infty} \beta^j \left(\gamma_{cw} w_{t+j}^s + \gamma_{cr} r_{t+j}\right) \tag{18}$$

where $1 + r = \frac{1}{\beta}$ is the steady state interest rate and

$$\gamma_{cw} = \frac{r}{1+r}\frac{\gamma_3}{\gamma_2} \tag{19}$$

$$\gamma_{cr} = \left[\frac{r}{1+r}\frac{\gamma_5}{\gamma_2} - 1\right] \tag{20}$$

The first term shows how consumption depends on current wealth consisting of capital, and income from labour and capital (the constants pick up the fact that quantities are substituted out). The second term shows how consumption depends on expected future prices. In the case of fixed labour supply the term on capital becomes $\frac{k}{c}\frac{r}{1+r}$ which is the familiar propensity to consume out of wealth - the constant scales linearised capital to consumption.

2.10 Calibration

The benchmark calibration follows Graham and Wright (2010). Values for most of the parameters are chosen following Campbell (1994): $\sigma = 1$, $\delta = 0.025$, $\alpha = 0.6$, $\beta = 0.99$, $n = 0.2$. The intertemporal elasticity of labour supply $\frac{1}{\gamma}$ is chosen to be 5. The aggregate productivity shock is given the benchmark RBC values, $\phi_a = 0.9$, $\sigma_a = 0.7\%$ per quarter. Graham and Wright (2010) uses empirical estimates of labour income process to calibrate the idiosyncratic shock $\phi_z = 0.9$, $\sigma_z = 5\sigma_a$.

3 Market-consistent information and learning

This section studies the case of learning from a market-consistent information set, which, with incomplete markets, will consist of the aggregate return on capital and the island-specific wage. A perceived law of motion (PLM) is first defined, then, conditional on the PLM, an expression is derived for the actual law of motion (ALM). It is shown that in general the state space of the ALM will expand to the history of the economy. Finally, a condition for e-stability is given

3.1 The perceived law of motion

Assume households estimate a VAR in the prices in their market-consistent information set, then use this estimated process to forecast future prices. The perceived law of motion

(PLM) for a household on island s at time t is

$$i_t^s = \phi_{t-1}^s (L) \, i_{t-1}^s + \varpi_{it} \tag{21}$$

where ϕ_t^s, a polynomial of order l in the lag operator L, i_t^s is the measurement vector defined in (15) and ϖ_{it} is the estimation error.[9]

Define matrices $T_r = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and $T_w = \begin{bmatrix} 0 & 1 \end{bmatrix}$ which pick out the return and the wage respectively from the measurement vector, then

$$\tilde{E}_t^s r_{t+i} = T_r \left(\phi^s\right)^i i_t^s \tag{22}$$

$$\tilde{E}_t^s w_{t+i} = T_w \left(\phi^s\right)^i i_t^s \tag{23}$$

and using this consumption (18) can be written

$$c_t^s = \frac{r}{\gamma_2 (1+r)} \left[\gamma_1 k_t^s + \gamma_3 w_t^s + \gamma_5 r_t + \theta_{ci} \left(\phi_{t-1}^s (L) \right) i_t^s \right] \tag{24}$$

where

$$\theta_{ci} = \left(\gamma_3 T_w + \gamma_5 T_r \right) \beta \phi^s \left(I - \beta \phi^s \right)^{-1} \tag{25}$$

3.2 The actual law of motion

To derive the actual law of motion for the economy individual consumption and labour supply must be aggregated. It is important to note that no household in the economy has sufficient knowledge, either in terms of the structure of the economy or information about other households, to do this (this is the same as saying households are unable to form model-consistent expectations). What follows is from the modeler's perspective.

Firstly, following Honkapohja and Mitra (2006), stack the PLMs for all households in the economy to give

$$I_t = \Phi_{t-1} (L) \, I_{t-1} \tag{26}$$

where $I_t = \begin{bmatrix} i_t^1 & i_t^2 & .... & i_t^S \end{bmatrix}'$ and $trace \left(\Phi_t \right) = \begin{bmatrix} \phi_t^1 & \phi_t^2 & .... & \phi_t^S \end{bmatrix}'$.

Then sum (24) across households and substitute for market clearing prices from (10) and (11) to give an expression for aggregate consumption

$$c_t = \Theta_{cY} \left(\Phi_{t-1} \right) Y_t + \Theta_{cI} (L) \, I_{t-1} \tag{27}$$

where $Y_t = \begin{bmatrix} k_t & a_t & z_t^1 & .... & z_t^S \end{bmatrix}'$ is the current vector of states and Φ_t, defined in (26) stacks the PLMs for all households. Note that aggregate consumption is independent of the wealth distribution - this is related to Krusell and Smith's (1998) finding that the

[9] As in full-information learning (Carceles-Poveda and Giannitsarou, 2008) circularity is avoided by assuming that to form estimates at time t the agents use only information from $t-1$ and earlier.

wealth distribution only has a small effects on the dynamics of the model. There is no effect at all here because of the linearity of the model.

Given states, lagged prices, last period's perceived law of motion Φ_{t-1} and a knowledge of the current state vector (27) is sufficient to solve for aggregate consumption and hence all other aggregate and idiosyncratic quantities.

3.3 A state space representation

From (27) aggregate consumption depends on lagged prices; and from (10) and (11), lagged prices depend on lagged aggregate consumption. So repeatedly substituting shows that current consumption depends on the full history of consumption. In other words, the state space expands to the full history of the economy. Writing the full state vector as a stack of the current state vectors Y_t

$$X_t = \begin{bmatrix} Y_t & Y_{t-1} & \dots & Y_0 \end{bmatrix}' \tag{28}$$

and substituting lagged prices into (27) gives

$$c_t = \Theta_{cX}\left(\{\Phi_i\}_{i=0}^{t-1}\right) X_t \tag{29}$$

The law of motion for the full state vector is then found by substituting into the law of motion for the non-expectational states and stacking this on top of the exogenous processes for aggregate and idiosyncratic technology to give:

$$X_t = \Theta_{XX}\left(\{\Phi_i\}_{i=0}^{t-1}\right) X_{t-1} + \Theta_{XW} W_{t-1} \tag{30}$$

where $W_t = \begin{bmatrix} \varepsilon_t & \varepsilon_t^1 & \dots & \varepsilon_t^S \end{bmatrix}'$ is a vector of innovations. This assumes there is an initial period with Φ_0, X_0 exogenously given. For a detailed derivation see Appendix A.4.

The expansion of the state space is a normal feature of models with heterogeneous information across agents. With model-consistent expectations, it is usually formulated as a hierarchy of average expectations of the current state vector (Townsend, 1983, Woodford, 2003, Nimark, 2007). However there is always an equivalent representation in terms of lags of the non-expectational state vector (Lorenzoni, 2010, Mackowiak and Wiederholt, 2009), analogous to the representation derived here.

Proposition 1. *Special cases (a) If labour supply is constant (*$\lim \eta \to \infty$*) the state vector comprises* l *lags of the current state vector state* Y_t*. (b) If there is a single lag in the PLM (*$l = 1$*) the state vector comprises the current state vector* Y_t*.*

Proof. See Appendix A.4. □

Part (b) of the proposition is related to the result of Graham and Wright (2010) that in an economy with model-consistent expectations the hierarchy of expectations collapses in the limiting cases of perfect heterogeneity ($\sigma_z \to \infty$) and perfect homogeneity ($\sigma_z = 0$) and in this case the economy evolves according to a first-order autoregressive process.

Expressions for the observable prices in terms of the state vector can be found by substituting (29) into (10) and (11), and stacking them on top of each other to give

$$I_t = \Phi_{iX}\left(\{\Phi_i\}_{i=0}^{t-1}\right) X_t \tag{31}$$

This is the actual law of motion (ALM) for the observables.

3.4 Learning rules

A standard learning rule is

$$\begin{aligned} \phi_{t+1}^{s} &= \phi_t^s + \gamma_t R_t^{-1} i_{t-1}^s \left(i_t^{s\prime} - i_{t-1}^{s\prime}\phi_t^{s\prime}\right) && (32)\\ R_{t+1}^{s} &= R_t^s + \gamma_t \left(i_{t-1}^s i_{t-1}^{s\prime} - R_t^s\right) && (33)\end{aligned}$$

where $\{\gamma_t\}_{t=0}^{\infty}$ is the gain sequence which needs to satisfy standard conditions. The paper studies two gain sequences, ordinary least squares learning, with $\gamma_t = t^{-1}$ and constant gain learning, with $\gamma_t = \gamma$. Such rules for each household can be stacked on top of each other to give a learning rule for Φ of the form

$$\Phi_{t+1} = \Phi_t + \gamma_t \mathcal{H}\left(\Phi_t, I_t\right) \tag{34}$$

3.5 E-stability and learnability

The standard analysis of the stability of economies under learning is given in Evans and Honkapohja (2001). Honkapohja and Mitra (2006) extend this to a model with both structural and learning heterogeneity. This section draws on these techniques to an economy with heterogeneous and incomplete information sets. Conditions for convergence of Φ_t to an equilibrium Φ are found by defining an associated ordinary differential equation (ODE)

$$\frac{d\Phi}{d\tau} = h\left(\Phi\right), \text{where } h\left(\Phi\right) = \lim_{t\to\infty} E\mathcal{H}\left(\Phi, X_t\right) \tag{35}$$

The economy with learning will converge to Φ only if Φ is a locally stable fixed point of the associated ODE.

The state-space representation allows an expression to be derived for $\mathcal{H}$. First express $\mathcal{H}$ in terms of lagged states and innovations by substituting for prices from (31) then for current states from (30) so (34) can be rewritten as

$$\Phi_{t+1} = \Phi_t + \gamma_t \mathcal{H}\left(\{\Phi_i\}_{i=0}^{t-1}, X_{t-1}, W_{t-1}\right) \tag{36}$$

Then to obtain an expression for $h(\Phi)$ take expectations, picking a particular Φ and take the limit:

$$h(\Phi) = \lim_{t\to\infty} E\mathcal{H}(\Phi, X_t) = \left(\Theta_{IX} M_X \Theta'_{IX}\right)^{-1} \Phi_{IX} M_X \left(\Theta_{XX}\Theta'_{IX} - \Theta'_{IX}\Phi'\right) \tag{37}$$

using $EX_tX_t' = M_X$, $EX_tW_t' = 0$. For a detailed derivation see Appendix A.5.

An equilibrium is a zero of $h(\Phi)$. It is e-stable and learnable if the eigenvalues of the Jacobian of $h(\Phi)$ have real parts which are negative.

3.6 The nature of the equilibrium

Proposition 2. *Except in the cases of proposition 1 an equilibrium under learning must be a restricted-perceptions equilibrium in the sense of Evans and Honkapohja (2001, p320)*[10].

Proof. In general, the PLM (21) depends on some limited history of the observables, whereas the ALM (31) depends on the full history. □

For $l = 1$, the PLM (21) is of the form

$$w_t^s = \phi_{wwt}^s w_{t-1}^s + \phi_{wrt}^s r_{t-1} + \varpi_{wt}^s \tag{38}$$

$$r_t = \phi_{rwt}^s w_{t-1}^s + \phi_{rrt}^s r_{t-1} + \varpi_{rt}^s \tag{39}$$

Table 1 shows the coefficients of the PLM for different lag lengths at the restricted perceptions equilibrium, i.e. the elements of the ϕ which is a zero of (37), along with the PLM under model-consistent expectations. PLMs at all lag lengths are characterized by strong first-order autoregressive components. As the number of lags increases, the PLM approaches the true law of motion at the MCEE.

[TABLE 1 HERE]

How different are the properties of the economy at the restricted-perceptions equilibrium from the model-consistent equilibrium[11]? To study the equilibrium, first "switch off" learning and fix the beliefs of all households at their value at the restricted-perceptions equilibrium (this can be thought of as the non-stochastic steady state of the economy, in

[10] See also Branch (2004).

[11] The answer to this question is complicated by the fact that the MCEE can only be solved approximately by truncating the hierarchy of expectations. However Graham and Wright (2010) shows that in practice the weight on orders of the hierarchy declines quickly so an solution to machine precision can be found, at least for the calibration used here.

contrast to the stochastic steady state in which there is a non-degenerate distribution of beliefs, discussed in section 5.2).

As with all RBC-type models, the main driver of the response of the economy is the behaviour of consumption, so begin by considering that.

[FIGURE 1 HERE]

Figure 1 shows the impulse response of consumption to an aggregate productivity shock at the MCEE and at the RPE with 1 lag in the PLM. First note that under model-consistent expectations the impact effect of the positive technology shock on consumption is *negative*. This contrast with the full information response is one of the results of Graham and Wright (2010). A full discussion can be found there but brief intuition is as follows. With market-consistent information sets, households do not observe the aggregate technology shock directly, but instead see its effect as a positive innovation to both the wage and the return to capital. Consider the response to the latter signal. An unexpected increase in the return could either be caused by a positive productivity shock, or because the household overestimated aggregate capital in the previous period. The certainty equivalent response to the first is to increase consumption, to the second to reduce consumption. Graham and Wright (2010) shows under all plausible calibrations the latter effect dominates so the impact response of consumption is negative.

Another way of putting this is, even with model-consistent expectations, households' limited information sets mean they make predictable (from the modeler's point of view) forecast errors in response to the technology shocks (though the forecast errors are white noise conditional on households' information sets). Turning to the restricted perceptions equilibrium, households make larger forecast errors (since their PLM is misspecified) and so consumption responds by more on impact. These bigger forecast errors mean bigger positive income surprises in subsequent periods so consumption rises above its value at the MCEE after a few periods then falls back to the steady state. It is possible to show numerically that welfare is unambiguously lower at the RPE.

Table 2 shows a number of statistics comparing the economy with households' PLMs fixed at the RPE with various lag lengths to the economy with model-consistent expectations. First note the modest magnitude of the misspecification - aggregate consumption is 2% and output 3.5% more volatile in the case with 1 lag in the PLM than in the economy with model-consistent expectations. Also note that as the number of lags increases the volatility falls towards its value at the MCEE.

[TABLE 2 HERE]

One further point: the RPE will depend on the choice of the number of households in the economy. To see this, consider a shock to a single household's idiosyncratic productivity. With many households, the impact on the aggregate economy will be small and its main affect on the learning rule will be in updating the elements of ϕ_t^s which relate to the idiosyncratic wage w_t^s. With fewer households, the impact on the aggregate economy will increase so will also lead to the updating of elements of ϕ_t^s which relate to the aggregate return r. This can have a significant effect on the properties of the economy so the number of households needs to be chosen to be sufficiently high. Another, more interesting, way in which the equilibria with learning will depend on the number of households is discussed in section 5.5.

3.7 Projection

The consumption function (24) is well defined as long as $I - \beta\phi_t^s$ is invertible. Since this term comes from computing the discounted sum of the expected future path of prices, the invertibility condition is the same as requiring the sum to be bounded. This is summarised in the following definition

Definition 4. *(stable PLM).A given ϕ^s is stable if it results in consumption being bounded. This will be the case if the eigenvalues of ϕ_t^s are less than $\beta^{-1} > 1$ in absolute value.*

Theorem 4 of Ljung (1977, p. 557), which forms the basis of many convergence results in the learning literature employs a "projection facility" constraining estimates to remain in a region around the REE. This has been widely criticized (e.g. Grandmont and Laroque, 1991 and Grandmont, 1998) since it involves endowing households with knowledge of what they are supposed to be learning. Even though a projection facility has been shown not to be necessary to proofs of convergence and stability in models with a unique REE (Bray and Savin, 1986) or more generally (Evans and Honkapohja, 1998), it is crucial for any numerical implementation of learning. To see this note that with a non-zero gain there is always a finite probability that particular sequence of shocks will lead to a household estimating a PLM that is unstable in the sense of definition 4, leading forecasts to grow without limit and consumption to be undefined.

The form of the consumption function (24) gives a natural way to define a projection algorithm which escapes the critiques of Grandmont and Laroque.

Definition 5. *(projection facility).After estimating the PLM households check the eigenvalues of ϕ_t^s. If they are greater than q the household discards the estimated ϕ_t^s and chooses a different one.*

If the projection facility is used there are many ways to pick a ϕ_t^s which do not involve endowing households with knowledge of the RPE. The simplest way is to use the value from the previous period[12].

[12] Other possibilities are to pick one from a random household; to use the average across households etc. As long as the number of households is sufficiently large, the choice makes no difference to the properties of the economy.

In the remainder of the paper, q is taken to be unity which can be interpreted as endowing households with the knowledge that the macroeconomy is stationary. There are two justifications for this. Firstly, estimating a VAR of the form (21) is problematic with non-stationary variables. Secondly, the consumption function is strongly non-linear for PLMs with eigenvalues greater than unity (recall that as $eig(\phi^s) \to \beta^{-1}, c^s \to +\infty$) and allowing beliefs to enter this range means arbitrary amounts of volatility can be generated in the macroeconomy.

Projection is rarely discussed in the context of numerical analysis. Williams (2003) and Eusepi and Preston (2011) both mention they discard explosive values though it is not clear if this includes rational bubble paths, and in the latter paper at least the extremely small gains used means that such paths will be very rare events. With "Euler equation learning" (Preston, 2005; Honkapohja et al, 2011), there is no infinite forward sum in the consumption function so the issue does not arise although Carceles-Poveda and Giannitsarou (2007, p2673) explicitly exclude non-stationary paths.

4 Ordinary least squares learning

This section investigates the convergence properties of the model under ordinary least squares (OLS)[13]. Why does convergence matter? When studying the properties of models with learning they are usually initialized with PLMs at the MCEE (or RPE, if appropriate). This avoids transitional dynamics, governed by an arbitrary choice of prior, affecting the results. However without fast convergence this is informationally implausible - households are being endowed with what the nature of learning models assumes they are unable to calculate.

First, consider the benchmark case of full information. Convergence with ordinary least squares learning is typically found to be slow. To illustrate this take the representative household RBC model of Evans and Honkapohja (2001) or Carceles-Poveda and Giannitsarou (2007). In such a model, the perceived law of motion is

$$k_{t+1} = \phi_{kk} k_t + \phi_{aa} a_t \tag{40}$$

and figure 2 shows the convergence of ϕ_{kk} starting from a prior of 0.5 of its value at the MCEE. Even after 10,000 periods, the parameter is a long way from its value at the MCEE.

[FIGURE 2 HERE]

Turning to the model of this paper, take the benchmark calibration with one lag in

[13] Since under OLS the gain tends to zero as time passes, it is rarely used to study business cycle dynamics, but remains an important benchmark case.

the perceived law of motion ($l = 1$)[14]. Choose a prior to be very different from the RPE, for example one drawn across households from $N(0.3, 0.1)$. This choice is of course arbitrary, but a choice needs to be made if figures are to be shown.

[FIGURE 3 HERE]

Figure 3 shows the convergence of this economy in terms of deviation along the convergence path of the elements of the average PLM from their value at the RPE. The lines are the average values across 10,000 runs of the model; 95% of these runs lie in the shaded areas. Convergence is remarkably fast when compared with the standard case of figure 2 (note the figures have different scales) with all elements of the PLM close to their value at the RPE within a few hundred periods. Graham (2011) shows (in a full-information model) that this is a consequence of modelling households as having infinite forecast horizons in contrast to the "Euler equation learning" of the standard model. The intuition for this is that individual rationality leads to behaviour away from the RPE being much closer to that at the RPE, and hence convergence is much faster.[15]

4.1 $\sqrt{t}$ convergence

Theorem 3 of Benveniste et al (1990, p110)[16] studies a system of the form of (26) and (30) under OLS learning ($\gamma_t = t^{-1}$). It states that if the derivative of $h(\Phi) = E\mathcal{H}(\Phi, X_t)$ has all eigenvalues with real parts less than -0.5 then

$$\sqrt{t}\left(\Phi_t - \Phi^*\right) \xrightarrow{\mathcal{D}} N(0, P) \tag{41}$$

where the matrix P satisfies the Lyapunov equation

$$\left[I/2 + h_\Phi(\Phi^*)\right] P + P\left[I/2 + h_\Phi(\Phi^*)\right]' + E\mathcal{H}(\Phi^*, X_t)\,\mathcal{H}(\Phi^*, X_t)' = 0 \tag{42}$$

As pointed out by Marcet and Sargent (1995), this means that if the conditions are satisfied, there is root - t convergence, although the formula for the variance of the estimators is modified from the classical case. As the eigenvalues become larger, convergence is slower in the sense that the variance covariance matrix of the limiting distribution P is larger.

In the RBC case discussed in the previous section, the eigenvalues are -0.074 and -0.042, too large to apply the theorem. In the model of this paper, no analytical

[14] Using more lags makes no significant difference to the results.

[15] Another interesting feature is the "notch" in the confidence interval for ϕ_{ww}. This shows that when the PLM is far from the RPE its properties is dominated by the (mostly aggregate) transitional dynamics so the distribution across agents remains narrow. When the PLM is close to the RPE, this strong aggregate component fades and the distribution is dominated by idiosyncratic variation across households.

[16] Also used by Marcet and Sargent (1995) and Ferrero (2007).

expression is available for the eigenvalues so they were calculated numerically. For the baseline calibration, the eigenvalues lie in the range $[-1.26, -1.00]$ so the condition of Benveniste et al (1990) is satisfied and convergence is at the rate root-t or faster. The eigenvalues were then calculated for around 30,000 calibrations[17] and across all of these the upper bound of the eigenvalues was found to be -1. So root-t convergence appears to be a robust property of this model.

5 Constant gain learning

Constant gain learning is often used to study business cycle dynamics since it captures the idea that learning is perpetual and allows households to respond to changes in the structure of the economy. The gain parameter can be chosen in various ways. Milani (2007, 2009) estimates it along with the other parameters of the model. Eusepi and Preston (2011) use survey data. Evans and Ramey (2006) allow households to choose it optimally. This paper will study gain parameters in the range [0.001 0.05] which encompasses all the values commonly used. A baseline value of 0.01 is chosen.

A simple way to interpret the gain is by noting that the weight on the forecast error from τ periods ago relative to the weight from the most recent forecast error is given by $(1-\gamma)^s$. So a gain of 0.02 (as estimated in Milani, 2007) implies data from around 34 quarters ago is given approximately half the weight of current data. On the other hand, a gain of 0.002 (the baseline value of Eusepi and Preston, 2011) means households put half as much weight on data from 84 years ago as they do on current data.

5.1 Convergence

The economy with constant gain converges to a stationary distribution of beliefs (see Evans and Honkapohja, 2001, p162, for conditions under which such convergence will occur), the properties of the distribution depending on the size of the gain and the stochastic properties of the model. As the gain increases, convergence will generally be faster and figure 4 shows this for a single component of the PLM, ϕ_{ww}. There are a number of interesting aspects to this figure. Firstly, with low gains the economy takes an extremely long time to converge to the RPE. Secondly, as the gain increases the economy seems to converge to a PLM with a mean lower than at the RPE. Thirdly, the economy converges to a limiting distribution, and the variance of this distribution increases with the gain. The properties of the distribution is investigated in more detail in section 5.2.

[FIGURE 4 HERE]

[17] The ranges were chosen to encompass values commonly used in the literature. The grid was not particularly fine, but experimentation showed no evidence of any non-linear effects. $\delta \in [0.01, \mathbf{0.025}, 0.10]$; $\alpha \in [0.4, \mathbf{0.6}, 0.8]$; ; $\beta \in [0.96, \mathbf{0.99}, 0.999]$; $\eta \in [0, \mathbf{0.2}, 1, \infty]$; $\rho_a \in 0.7, \mathbf{0.9}, 0.95$,0.99]; $\sigma_a \in [0.5, \mathbf{0.7}, 1]$; $\rho_z \in [0.7, \mathbf{0.9}, 0.95, 0.99]$; $\sigma_z \in [0, 1, \mathbf{3.5}, 5, 7, 10]$. The bold figure represents the baseline calibration.

5.2 The steady state distribution of beliefs

In a representative household model, the intuition for how constant gain learning affects the economy seems straightforward. Compared to the model-consistent expectations equilibrium, learning makes the representative household's expectations more volatile and this volatility of expectations translates into higher volatility of aggregates. However, Graham (2011) shows that this intuition only goes through in special cases and the distribution of beliefs across time must be taken into account. In a heterogeneous agent model, things are more complicated still since the stochastic steady state of the economy is now characterized by a distribution across households of their beliefs. In other words, while in the representative agent case the distribution of beliefs is a time-series, with heterogeneous households it is also a cross section. The next section describes this distribution.

Figure 5 shows the steady state distribution of beliefs for various values of the gain parameter. With one lag in the PLM (and for the rest of the section I shall use this specification) the PLM is represented by the 4 elements of ϕ. It is important to remember that the distributions in the figure are for each element of ϕ taken alone, whereas in fact they are jointly distributed.

The figure shows three interesting features. As the gain parameter increases the means of the distributions (particularly of the AR coefficients ϕ_{ww} and ϕ_{rr}) fall; the standard deviations of the distributions increase and they become more asymmetrical with a long leftward tail and a short rightward tail. This is a consequence of projection (see section 3.7). Realizations of ϕ on the right-hand side of the distributions ,which correspond to non-stationary paths of expected prices, will be discarded, so the distributions are truncated.

[FIGURE 5 HERE]

Table 3 shows the moments of a distribution fitted to the steady state distribution of each element of the PLM. This confirms the impression from figure 5: as the gain increases the mean of the distribution falls; and both its standard deviation and skewness increase. Again remember that the elements of ϕ are in fact jointly distributed

[TABLE 3 HERE]

A further feature is shown by table 3. The mean of the distribution is lower than the MCEE even for very small gains in which the projection facility is not invoked. To understand this and to clarify the impact of the distribution on aggregates, consider a simple example.

5.3 A simple example

To understand the effect of a stationary distribution of beliefs on the macroeconomy, it is helpful to consider a simple univariate example[18] in which capital and labour are fixed and income follow an exogenous $AR(1)$ process:

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{43}$$

Beliefs are parametrized by a scalar ϕ such that

$$\tilde{E}_t y_{t+i} = \phi^i y_t \tag{44}$$

then the consumption function is

$$c_t = \frac{r}{1+r}\left[(1+r)\,b_t + \frac{1}{1-\phi_t\,(1+r)^{-1}} y_t\right] \tag{45}$$

where b_t is current wealth and the second term represents expectations about future income. Note the first and second derivatives to ϕ of the second term are positive capturing the positive and increasing effect of income persistence on consumption. Although only the infinite forecast horizon case will be considered here, the second derivative of f is positive as long as $T > 0$.

When beliefs are model-consistent, i.e. $\phi_t = \rho\ \forall t$ consumption will be a random walk and the standard deviation of the first difference of consumption is

$$\sigma^*_{\Delta c} = \frac{r}{1+r}\frac{1}{1-\rho\,(1+r)^{-1}}\sigma_\varepsilon \tag{46}$$

Beliefs are updated by a simplified constant gain learning algorithm

$$\phi_{t+1} = \phi_t + \gamma\,(y_t - \phi_t y_{t-1}) \tag{47}$$

How does the stationary distribution of ϕ affect the economy? Firstly, assume that the distribution has a mean of ρ (the value of beliefs at the MCEE); non-zero standard deviation and is symmetric. To understand the impact of this distribution on the unconditional properties of consumption consider the response of consumption to a positive innovation to income. Taking $\rho = 0.9$, figure 6 shows the response in the three cases of $\phi_0 = \phi^* = \rho$; $\phi_0 = 0.95 > \rho$ and $\phi_0 = 0.85 < \rho$.

[FIGURE 6 HERE]

[18] Full details are in Appendix B.

5.3.1 Case 1: $\phi = \rho$

If households' beliefs are correct, then the impact response of consumption will be exactly that at the MCEE. In the second period, beliefs will be revised upwards. This will mean consumption in period 2 is higher than it would be in at MCEE since households believe income will be more persistent than it actually is. In the third period, there are two effects. Firstly, beliefs will be revised downward towards the MCEE. Secondly, household wealth will be lower than expected. Both of these tend to reduce consumption. As time passes, these two effects continue, and at some point consumption will fall below its value at the MCEE and remain there for the rest of history (as is required to satisfy the intertemporal budget constraint).

To summarise, learning has no impact effect but causes consumption to rise above its value at the MCEE for a number of periods after the initial one, then fall below this value for the rest of time.

Proposition 3. *If beliefs are initialized at the MCEE, the impulse response function with learning implies a higher volatility of consumption growth than without learning*

Proof. See Appendix B.1 □

5.3.2 Case 2: $\phi > \rho$

In this case households believe that income is more persistent that it is at the MCEE so on impact increase their consumption by more than with correct beliefs. In subsequent periods there are two effects. Firstly, households wealth will be lower than expected which will tend to reduce consumption. Secondly, beliefs will be revised, in the second period upward and in subsequent periods downward back towards the MCEE. In the second period the second effect dominates so consumption increases further, in subsequent periods both effects go in the same direction and as time passes, consumption will fall below its value at the MCEE and stay there for the rest of time. So the overall effect is higher consumption than at the MCEE for some initial periods, then consumption lower than at the MCEE for the rest of time.

5.3.3 Case 3: $\phi < \rho$

The intuition for this case is simply the mirror image of that with $\phi > \rho$. However note the difference in magnitude. Since the derivative of the consumption function is increasing in ϕ, the response is much smaller to a lower value of ϕ than to the higher one of the previous section.

Given these three cases, the unconditional properties of consumption will be the average of the three cases weighted by the stationary distribution of ϕ. Since the distribution is assumed to be symmetric, the larger impact of case 2 will dominate the smaller one of case 3 and the volatility of consumption will increase.

So the distribution of beliefs will unambiguously increase the volatility of consumption. The higher the gain, the higher will be the standard deviation of beliefs so the higher will be the standard deviation of consumption.

There is a further effect. The theorem of Evans and Honkapohja (2001) that states the mean of the distribution will be at the MCEE only holds for small values of the gain. In practice, the mean will often be different from the MCEE. Since the distribution of beliefs causes the mean response of consumption to be different from that at the MCEE, the response of capital will also be different (if consumption responds by more capital would be expected to be less persistent) and hence the mean ALM will be different from the MCEE. So the mean of the distribution will be different from the MCEE, in this case lower.

How do the properties of consumption change if the mean of the distribution is lower than at the MCEE (either for the reason given in the previous paragraph or due to the projection facility, as will be discussed in the next section)? If the mean is lower, draws of ϕ from case 3 are more likely than those from case 2, and if the it is sufficiently low this will result in the standard deviation of consumption falling below its value at the MCEE. Similarly, if the distribution is sufficiently skewed to the left this will result in the standard deviation of consumption falling.

To summarise, this simple example suggests that the stationary distribution of beliefs will have the following effects:

1. If it is symmetrical, the non-linearity of consumption to beliefs will mean consumption responds by more on impact and be more volatile. This will imply the mean of the distribution is slightly lower than at the MCEE.

2. If the mean of the distribution is lower, this will offset the effects in (1) and make consumption respond by less on impact and be less volatile

3. If the distribution is skewed to the left, this will further offset the effects.

5.4 The aggregate economy

To analyse the properties of the aggregate economy, first take the gain to be $\gamma = 0.01$. Sensitivities to different gains will be considered later. Figure 7 shows the impulse responses of aggregates to a 1% positive innovation in the process for aggregate technology[19]. For each variable three lines are plotted. The heavy line is the response of the model with learning, starting from the steady state distribution of beliefs. The dashed

[19] For clarity, the figures omit to show the distribution of responses of the variables across households.

line is the response of the model at the RPE (i.e. with all household beliefs fixed at the RPE). The light line is the response of the model at the mean of the steady state distribution (i.e. with all household beliefs fixed at the mean).

[FIGURE 7 HERE]

Start by considering the first panel of the figure, the response of aggregate consumption. First compare the heavy line with the dashed line to see that the impact response of consumption is smaller in magnitude with learning than at the RPE, in other words learning mutes the impact of the shock. To understand this, recall the three effects of beliefs on consumption described in the previous section. Since the mean of the distribution is lower than at the RPE (or in other words households expect the shock to have less persistent effects on prices than it actually does), households respond as if the shock were less persistent and this reduces its impact. The combination of the second and third effects could go in either direction - increasing the impact if the nonlinearity dominates, or reducing it if skewness dominates. To show the magnitude of these effects, the light line on the figure shows the response if all beliefs were fixed at the mean of the distribution. The gap between the light line and the heavy line shows that the combined effect of skewness and non-linearity works to offset the effect of the lower mean.

In subsequent periods, two things happen. Firstly households get unexpected factor income (since prices are more persistent than they on average expected); secondly households update their PLMs so the entire distribution of beliefs shift. In practice, this last effect is too small to see on the figure, in the second period the difference between the path of consumption with a gain of 0.01 and that with no learning (a gain of 0) is of the order of 10^{-3}. The effect is so small because the idiosyncratic volatility is so much more volatile than the aggregate so aggregate shocks get a small weight in the updating rule (32). This is in contrast to representative agent models (Graham, 2011 or Eusepi and Preston 2011) in which impulse response show a pronounced kink in the period after the shock when beliefs are updated.

This is an important sense in which heterogeneity changes the effect of learning on the economy. Since the volatility of aggregate shocks is small compared to that of the idiosyncratic shocks, an innovation in prices due to an aggregate shock only has a small effect on households' beliefs.

The combination of all these effects means the magnitude of the response of consumption is smaller than at the RPE for the first 60 periods or so from the impact of the shock. After this (just off the right-hand side of the figure) consumption with learning stays above that at the RPE as both adjust back to the steady state.

To summarise, learning mutes the response of consumption on impact but makes the response more persistent. Other variables show similar qualitative patterns.

[TABLE 4 HERE]

Table 4 shows how the standard deviations of aggregate variables to their values at the RPE change with the gain. Looking first down the columns of the table there is evidence of non-linearity with respect to the gain - the standard deviations of variables first increases then, for gains above 0.01 starts to decrease. As the gain increases, the distribution becomes more skewed with less mass at very persistent values of beliefs, so the skewness effect starts to dominate the non-linearity effect, reducing the impact of the shock still further. Across all gains, the standard deviation of never more than 1% higher than at the RPE, and falls to a 2% lower at higher gains. Again, therefore, the simple intuition that learning increases volatility does not go through to this model, but more important than the sign of the changes is how modest they are, particularly given how little information households are using to form forecasts. In practice the economy with learning would be indistinguishable from one with model-consistent expectations, at least to an econometrician subject to the typical limits on macroeconomic data.

However, recall that the model-consistent expectations equilibrium which the economy with learning resembles is that with market-consistent information of Graham and Wright (2010) which differs in significant ways from the equilibrium with full information. In other words, assumptions about information have a large effect on the properties of aggregates; assumptions on whether households can form model-consistent expectations seem much less important.

5.5 The idiosyncratic economy

The previous section discussed the response of aggregates. What about household variables? First recall that household variables are non-stationary since idiosyncratic shocks are pure permanent income and have a permanent effect on household wealth and consumption (as is the case at the equilibrium with model-consistent expectations). So one appropriate measure is the standard deviation of consumption growth (an alternative would be to use any of the wide range of filters available). Table 5 shows this statistic, averaged across households, for different values of the gain parameter. For low values of the gain, the volatility of household consumption is very close to that at the RPE. As the gain increases, the standard deviation increases to a maximum (at to a gain of 0.01) of four times that at the RPE.

[TABLE 5 HERE]

This is an appealing feature of the model. At the restricted perceptions equilibrium, the standard deviation of household consumption growth is 0.51%, much lower than the $2-3\%$ found in the data (e.g. Attanasio et al, 2002). With a gain of 0.01, this becomes 2.4% per quarter, within the range of observed values. As the gain increases above 0.01, the volatility of consumption growth falls.

For a particular household, the steady state distribution of beliefs discussed in section 5.2 is a time series distribution: in a period when the PLM represents a persistent path for expected prices households change their consumption by a large amount. However some households with such beliefs will receive an idiosyncratic shock which lead them to increase their consumption, and some will receive an idiosyncratic shock which lead them to decrease their consumption. Both shocks result in higher volatility of household consumption, but in aggregate their effect cancels out to leave the much small aggregate effects described in the previous section. This gives a further mechanism in addition to that described in section 3.6 by which the number of households affects the dynamics of the model. If the number of households is small, idiosyncratic volatility will contaminate the aggregate economy. As shown in table 6, with a small number of households effectively arbitrary amounts of aggregate volatility can be generated[20].

[TABLE 6 HERE]

5.6 Sensitivities

While the structural parameters of the model $(\beta, \sigma, \delta, \alpha, \eta)$ change the equilibrium, they do not change the informational problem and so have little effect on the properties of the economy with learning relative to the economy with model-consistent expectations. As discussed in Graham and Wright (2010), it is the properties of the shocks which change the informational problem in interesting ways, and the same is true of the model with learning.

[TABLE 7 HERE]

Table 7 shows how the ratio of the standard deviation of aggregate consumption in the model with learning to its value at the RPE changes with the persistence of the aggregate and the idiosyncratic shocks. To understand these results, remember that the return to capital is an aggregate object and so its persistence is largely determined by the persistence of the aggregate shock. In contrast, because the idiosyncratic shock is much more volatile than the aggregate, the persistence of the household wage is largely determined by the persistence of the idiosyncratic shock. Thus increasing the persistence of the aggregate shock is like shifting the distribution of ϕ_{rr} to the right; and increasing the persistence of the idiosyncratic shock does the same for ϕ_{ww}. As the distributions shift to the right, projection is more likely to happen so the mean of the distribution falls further below its value at the RPE and the effects described in section 5.4 become stronger. On the other hand, for lower values of persistence, projection is less likely to

[20] This suggests a simple rule of thumb for picking the number of agents to use for simulations. Increase the number of agents until doubling this number has no effect on the statistics of interest at the desired level of accuracy.

happen so the distribution becomes more symmetrical with its mean closer to the value at the RPE. Both these effect can be seen in table 7. The left-hand column sees a smaller effect of learning, but one more linear in the gain. The right-hand column sees a larger effect, but more non-linear in the gain.

5.7 An intercept in the learning rule

A number of recent papers (Milani, 2011, Eusepi and Preston, 2011) include an intercept in the learning rule, interpreted as capturing households' uncertainty about the steady state. It is straightforward to augment the model of this paper with an intercept by changing the measurement vector (15) to $i_t^s = \begin{bmatrix} 1 & r_t & w_t^s \end{bmatrix}'$, (see the Appendix for details). When this is done the restricted-perceptions equilibrium (which is unchanged) remains e-stable and the convergence properties of the model are very similar.

However the addition of an intercept has significant effects on the properties of the model - table 8 shows the moments for different gains. Now at a gain of 0.001 consumption and output are around 10% more volatile than at the RPE, in contrast to the model without an intercept in which there was almost no amplification.

[TABLE 8 HERE]

To see why this happens, firstly consider the steady state distribution of beliefs. Figure 8 shows the distribution of beliefs about the intercept of the wage equation (the distributions of the other components of beliefs are broadly similar to those shown in figure 5). Note the high volatility of beliefs: this is a consequence, from (9), of the high volatility of the idiosyncratic shock.

[FIGURE 8 HERE]

Why should uncertainty about the intercept translate into high volatility? To answer this, modify the simple example of section 5.3 to include an intercept. Consumption is then

$$c_t = \frac{r}{1+r}\left[(1+r)\,b_t + \frac{1}{1-(1+r)^{-1}}\phi_1 + \frac{1}{1-\phi_t\,(1+r)^{-1}}y_t\right] \tag{48}$$

where the second term picks up the effect of the intercept, a discounted forward sum of a constant. For the discount factor of the baseline calibration, $\frac{1}{1-(1+r)^{-1}} \approx 100$ which, if the persistence of income is 0.9 is around 10 times higher than the coefficient on income. So variations in ϕ_1 are greatly amplified[21]. This happens to such an extent that for values of the gain greater than 0.005 , the economy becomes unstable so no values are reported.

[21] This also explains the strong effect of updating beliefs present in Eusepi and Preston (2011) but absent from the impulse response of this paper - a small change in the element of ϕ relating to the intercept has a massively amplified effect on consumption.

Why does this lead to instability? Imagine a household learning a large positive value for the constant. This means their consumption increases dramatically, which will increase aggregate consumption. Other things equal, higher aggregate consumption means lower aggregate capital and lower aggregate labour supply, so the return to capital and the wage increase and in the next period which will increase ϕ_1 further; and hence consumption still further leading to instability.[22] Such instability is more of a problem in a heterogeneous agent economy than in one with identical agents since the higher volatility of idiosyncratic shocks is translated into wider distributions of beliefs so a higher probability of a draw leading to instability. It also arises only if households are very forward looking - in models that take the "Euler equation learning" approach (for example Milani, 2011, which uses an intercept) it is not an issue.

While using a low gain, as in Eusepi and Preston (2011) avoids this problem, it seems informational implausible for two reasons. Firstly, a gain of 0.001 implies that households place around half as much weight on data two centuries old as they do on new information[23]. Secondly, as shown in section 5.1, with a gain of 0.001 the economy takes many thousands of periods to converge to the RPE - so starting the economy from the steady state distribution is equivalent to endowing households with the very knowledge they are supposed to be learning.

6 Discussion

This paper has taken a model in which agents have limited information, both about the structure of the economy and the variables relevant to their decisions. Despite this, the economy is shown to converge quickly to an equilibrium which is similar to the equilibrium with model-consistent expectations. Learning does not introduce strong dynamics at the aggregate level, though more interesting things happen at a household level. Another way of putting this is, at least in the context of this model, assumptions about information are important for aggregates; assumptions about the ability of households to form model-consistent expectations less so.

One strength of the approach taken in this paper is that the informational assumptions are clear. Households' information sets are constrained by the markets in which they trade and they use all the information at their disposal to make optimal decisions. They have no other knowledge either of the structure of the aggregate economy or of the values of aggregate variables.

On the one hand, as pointed out in Graham and Wright (2010) the assumption of

[22] Clearly this will only lead to instability of the aggregate economy if households are sufficiently large, however even with $20,000$ households the economy shows instability with gains much above 0.005 and computational constraints prevent more households being used.

[23] The weight on information τ periods old is $\gamma(1-\gamma)^{\tau}$.

market-consistent information is itself a strong one - households clearly have many other sources of information than factor prices - but adding such information (for example a noisy signal of output) would only further reduce the impact of learning. On the other hand, section 5.7 showed that if an intercept is added to the learning rule, learning can have a bigger effect on the properties of the economy. However this comes at the cost of instability unless the gain is small, and although plausible, the addition of an intercept seems arbitrary. This is related to a point made by Grandmont (1998) on the specification of perceived laws of motion. What variables should be included in them? What econometric specifications should be used? Such choices would be far more complicated if the model included features such as non-linearity, structural breaks or non-ergodic shocks.

References

Attansio, Orazio P., James Banks, and Sarah Tanner, 2002, Asset holding and consumption volatility, Journal of Political Economy 110, 771–792.

Baxter, B., Graham, L., Wright, S., 2010. Invertible and non-invertible information sets in linear rational expectations models. Journal of Economic Dynamics and Control, 35,3 pp. 295-311.

Benveniste, A., Metivier, M., Priouret, P., 1990. Adaptive Algorithms and Stochastic Approximations. Springer, Berlin.

Berardi, M., 2007. Heterogeneity and misspecifications in learning. Journal of Economic Dynamics and Control, 31, 10, pp 3203-3227.

Branch, W.A., 2004. Restricted Perceptions Equilibria and Learning in Macroeconomics. In Post Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model. ed. David Colander, Cambridge University Press. March 2004

Branch, W.A., Evans, G.W., 2005. Intrinsic heterogeneity in expectation formation. Journal of Economic Theory 127, 264–295.

Branch, W.A., McGough, B., 2004. Multiple equilibria in heterogeneous expectations models. Contributions in Macroeconomics 4.

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

Brav, A, G.M. Constantinides, and C.C. Geczy, 2002. Asset pricing with heterogeneous consumers and limited participation: Empirical evidence, Journal of Political Economy

110, 793–824.

Campbell, John Y, 1994. Inspecting the mechanism: an analytical approach to the stochastic growth model. Journal of Monetary Economics 33, pp 463-506.

Carceles-Poveda, E., Giannitsarou, C., 2007. Adaptive learning in practice. Journal of Economic Dynamics and Control, 31, 8, pp. 2659-2697.

Eusepi, S., Preston, B., 2011. Expectations, Learning and Business Cycle Fluctuations. American Economic Review, forthcoming.

Evans, G.W., Honkapohja, S., 1998. Economic Dynamics with Learning: New Stability Results. Review of Economic Studies, 65(1), pp. 23-44.

Evans, G.W., Honkapohja, S., 2001. Learning and Expectations in Macroeconomics. Princeton University Press, Princeton.

Evans, G.W., Honkapohja, S & Mitra, K, 2009. Anticipated fiscal policy and adaptive learning. Journal of Monetary Economics, Elsevier, 56(7), pp 930-953.

Evans, G.W., Ramey, G., 2006. Adaptive expectations, underparameterization and the Lucas critique. Journal of Monetary Economics, 53(2), pp. 249-264.

Ferrero, G., 2007. Monetary policy, learning and the speed of convergence. Journal of Economic Dynamics and Control, 31(9), pp. 3006-3041.

Fout, H.B., Francis, N.R., 2011. Information-consistent learning and shifts in long-run productivity. Economics Letters,111(1), pp 91-94.

Giannitsarou, C., 2003. Heterogeneous learning. Review of Economic Dynamics 6, 885–906

Graham, L., 2011. Individual rationality, model-consistent expectations and learning. Working paper.

Graham, L., Wright S., 2010. Information, market incompleteness and heterogeneity. Journal of Monetary Economics, 57, 2, pp. 164-174.

Grandmont, J.-M., 1998. Expectations formation and stability of large socioeconomic systems. Econometrica, 66(4), pp. 741-781.

Grandmont, J.-M., Laroque G., 1991. Economic dynamics with learning: some instability examples, In: Barnett, W.A. et al. Eds.., Equilibrium Theory and Applications, Proceedings of the Sixth International Symposium in Economic Theory and Econometrics. Cambridge University Press,Cambridge, pp. 247–273.

Hellwig, C., 2006. Monetary Business Cycle Models: Imperfect Information. UCLA Working Paper

Honkapohja, S., Mitra, K., 2006. Learning Stability in Economies with Heterogeneous Agents. Review of Economic Dynamics, 9(2), pp. 284-309.

Honkapohja, S., Mitra, K., Evans, G.W., 2011. Notes on Agents' Behavioral Rules Under Adaptive Learning and Studies of Monetary Policy. CDMA Working Paper 11/02.

Krusell, P., Smith, A., 1998. Income and Wealth Heterogeneity in the Macroeconomy. Journal of Political Economy 106(6), pp. 867-96.

Ljung, L., 1977. Analysis of recursive stochastic algorithms. IEEE Trans. Auto. Control AC 22, pp. 551–575.

Lorenzoni, G., 2009. A Theory of Demand Shocks. American Economic Review, 99(5), pp. 2050-84.

Mackowiak, B., Wiederholt, M.„ 2009. Optimal Sticky Prices under Rational Inattention. American Economic Review, 99(3), pp 769-803.

Marcet, A., Sargent, T.J., 1995. Speed of convergence of recursive least squares: learning with autoregressive moving-average perceptions. In: Kirman, A., Salmon, M. (Eds.), Learning and Rationality in Economics. Basil Blackwell, Oxford, pp. 179–215.

Milani, F., 2007. Expectations, learning and macroeconomic persistence. Journal of Monetary Economics, 54, pp. 2065–2082.

Milani, F., 2011. Expectation Shocks and Learning as Drivers of the Business Cycle. The Economic Journal, 121 (552), pp 379 - 401

Nimark, K., 2007. Dynamic Higher Order Expectations. Working Paper

Preston, B., 2005. Learning About Monetary Policy Rules when Long-Horizon Expectations Matter. International Journal of Central Banking, 1(2), 81—126.

Radner, Roy, 1979. Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices. Econometrica, 47, 3, pp. 655-678

Townsend, R.M., 1983. Forecasting the Forecasts of Others. Journal of Political Economy 91 (4), pp 546-588.

Williams, N., 2003. Adaptive Learning and Business Cycles. Working paper.

Woodford, M, 2003. Imperfect Common Knowledge and the Effects of Monetary Policy. In P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, eds. Knowledge, Information,and Expectations in Modern Macroeconomics: In Honour of Edmund S. Phelps. Princeton:Princeton University Press.

Figures and Tables

Figure 1: Response of consumption to a positive aggregate technology shock at the RPE

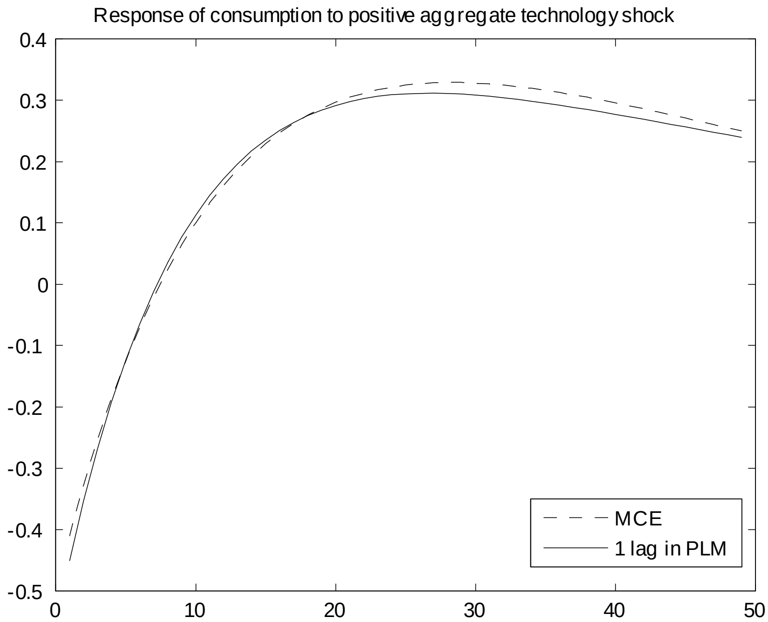
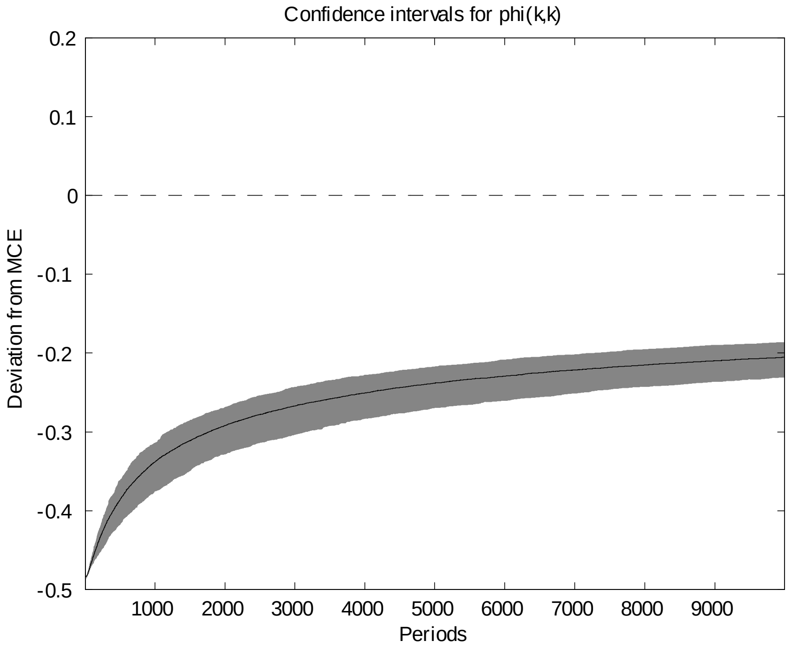


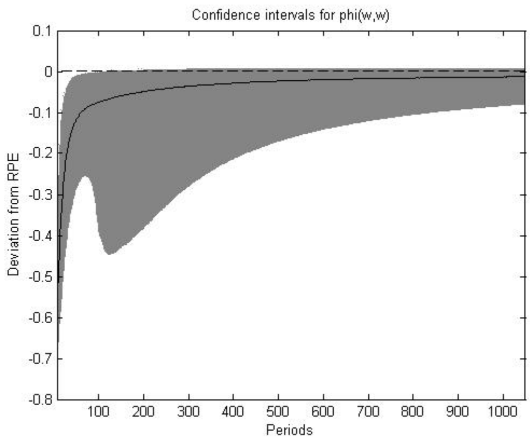
Figure 2: Convergence in the RBC model with learning over 10,000 periods



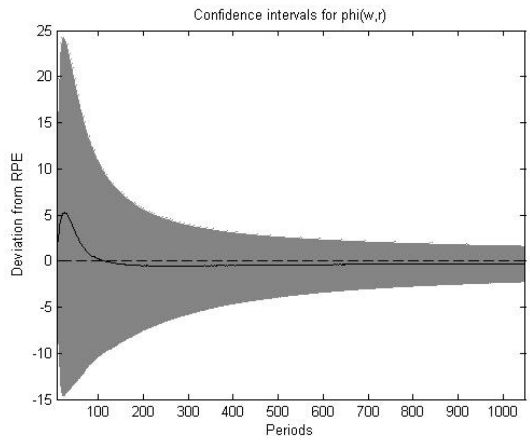
x-axis shows number of periods; y-axis the deviation of the autoregressive component of the PLM from its value at the MCEE. Line is mean value. 95% of values lie within the shaded area.

Figure 3: Convergence of the economy with OLS learning - elements of the PLM

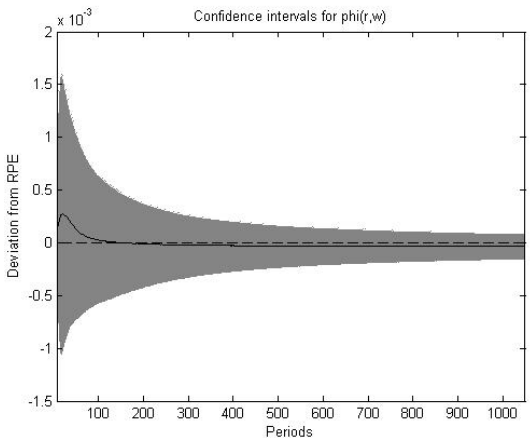
ϕ_{ww}



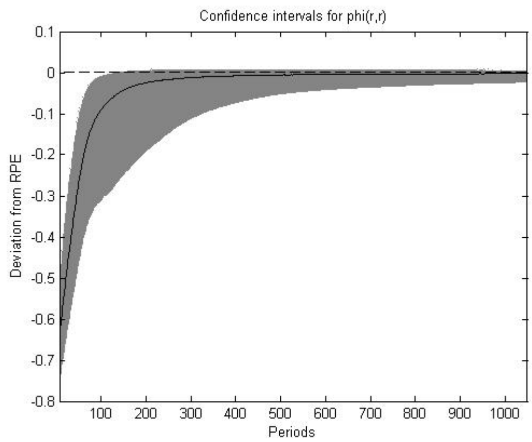
ϕ_{wr}



ϕ_{rw}



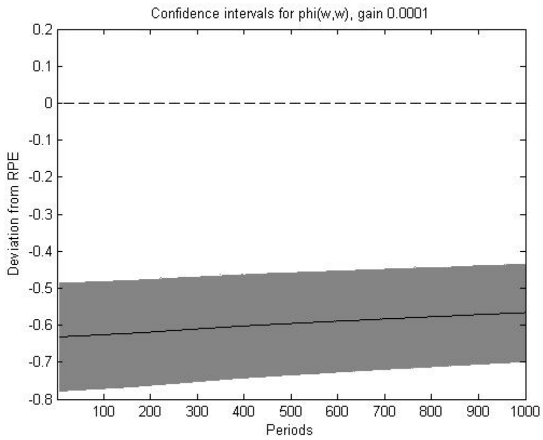
ϕ_{rr}



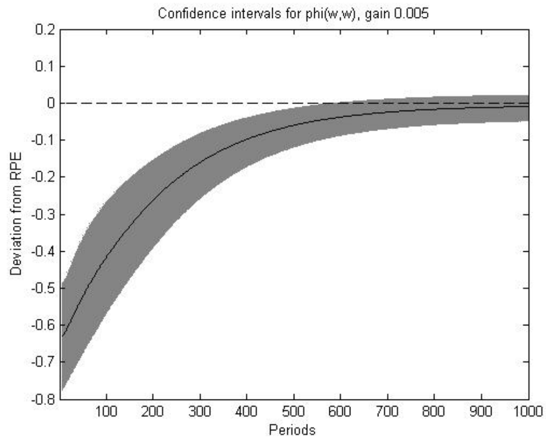
x-axis shows number of periods; y-axis the deviation of the element of the PLM from its value at the RPE. Line is mean value of element. 95% of values lie within the shaded area.

Figure 4: Convergence of the economy with constant gain learning, ϕ_{ww}

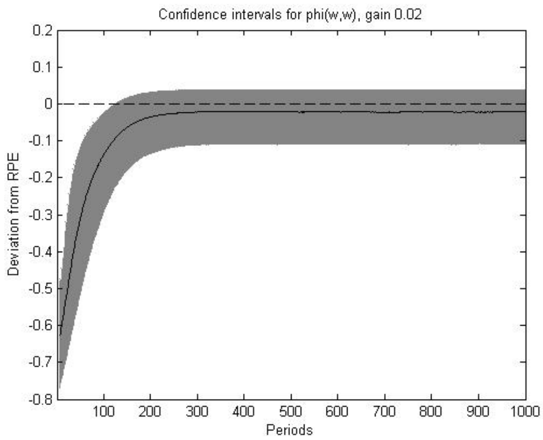
Gain $= 0.001$



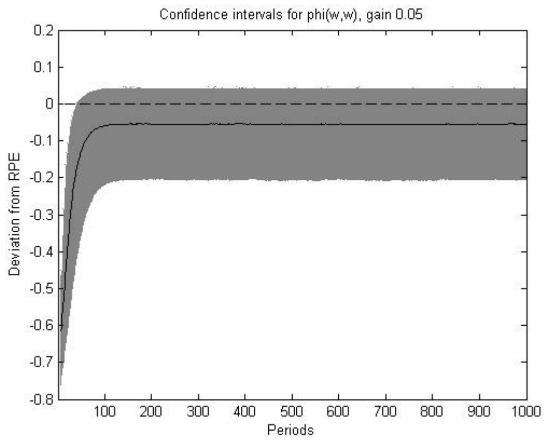
Gain $= 0.005$



Gain $= 0.01$



Gain $= 0.05$



x-axis shows number of periods; y-axis the deviation of the element of the PLM from its value at the RPE. Line is mean value of element. 95% of values lie within the shaded area.

Figure 5: Steady state distribution of beliefs

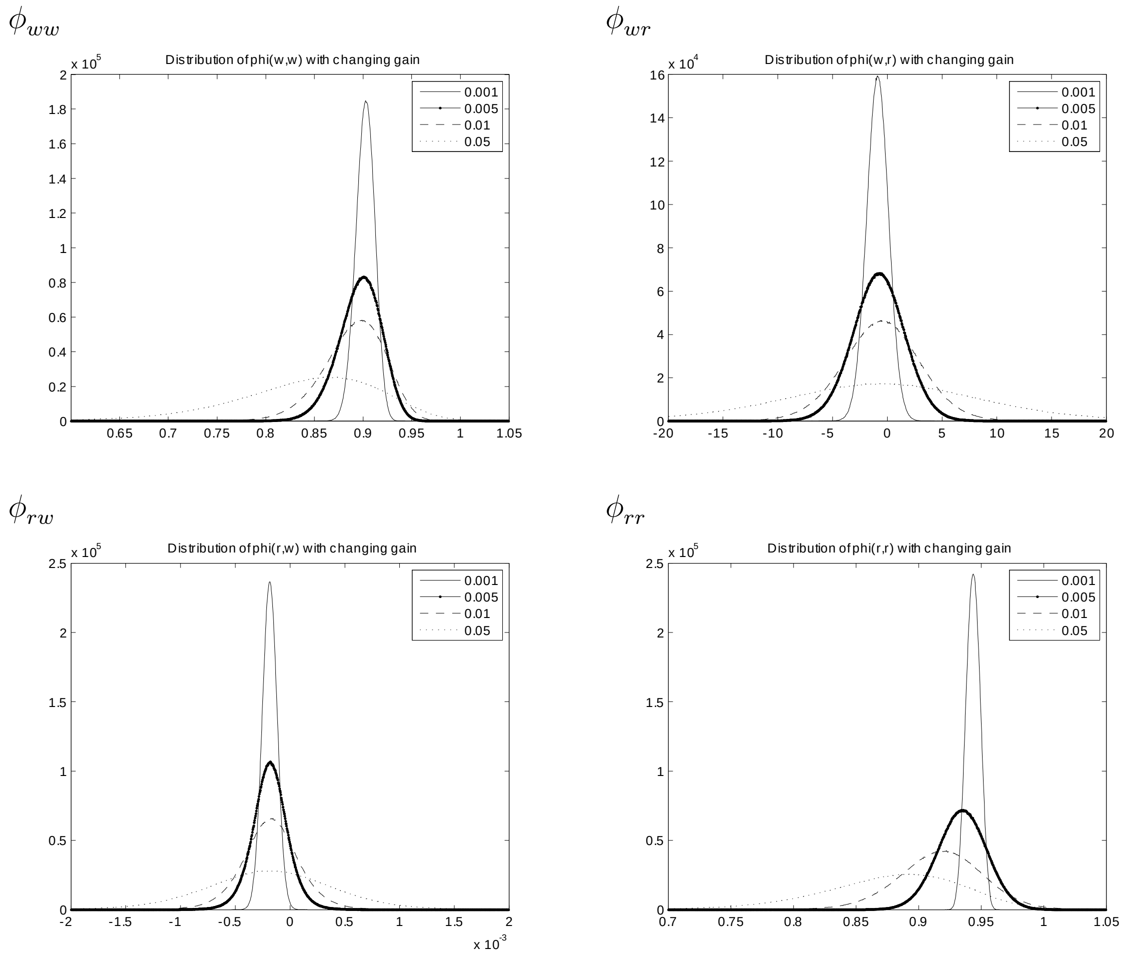


Figure 6: A simple example: impulse responses of consumption with different beliefs

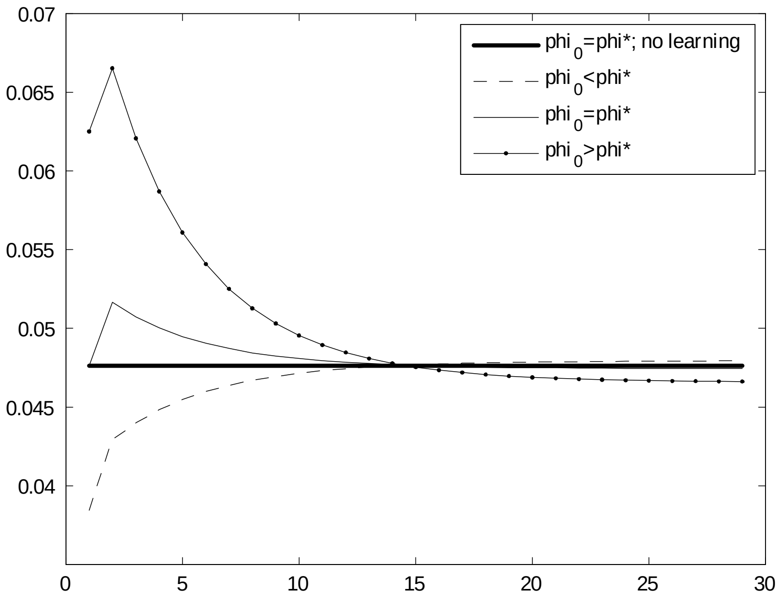
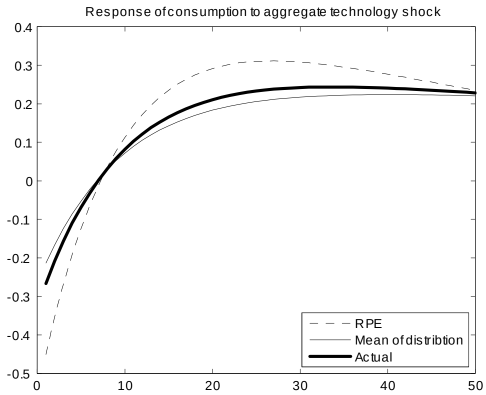
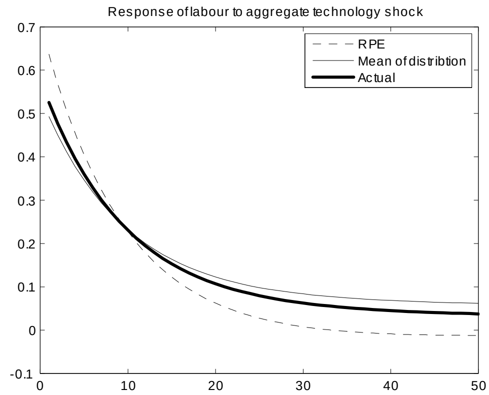


Figure 7: Impulse response functions to a unit innovation in the process for aggregate technology

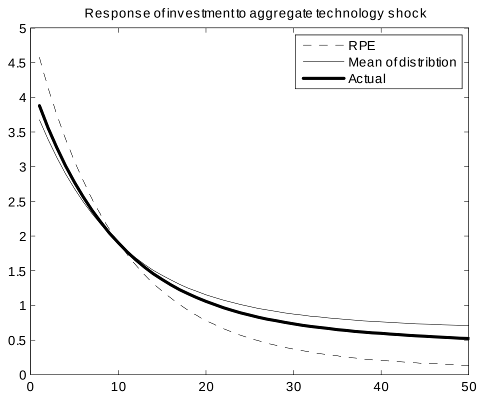
Consumption



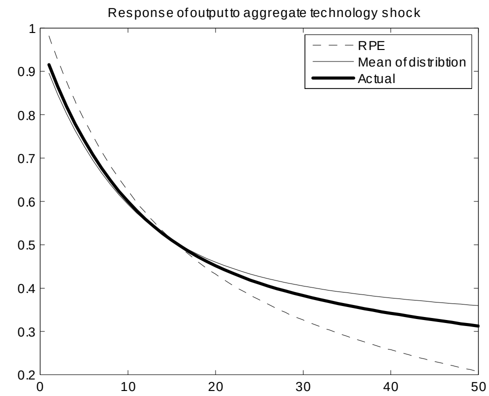
Labour



Investment



Output



Capital

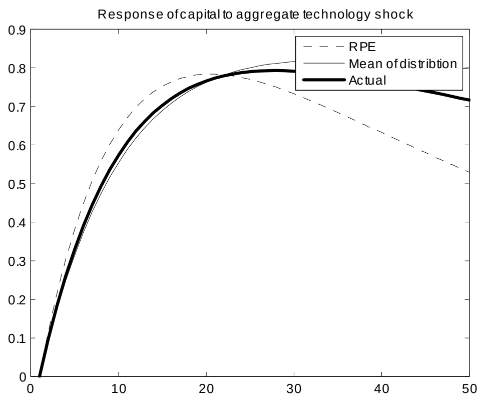


Figure 8: Steady state distribution of beliefs with intercept in learning rule

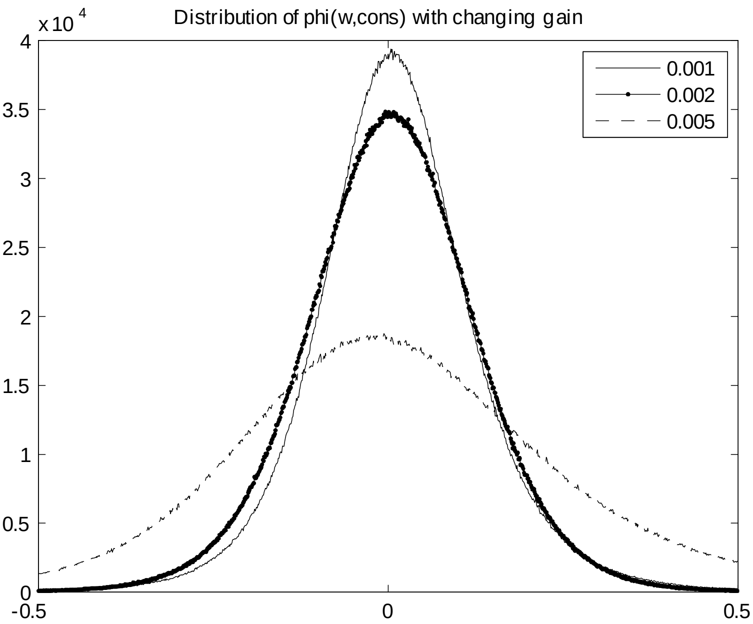


Table 1: The perceived law of motion at the RPE

Wage equation

l	w_{t-1}	r_{t-1}	w_{t-2}	r_{t-2}	w_{t-3}	r_{t-3}	w_{t-4}	r_{t-4}	w_{t-5}	r_{t-5}
1	0.9034	−0.8506								
2	0.9026	−0.6324	0.0009	−0.2171						
3	0.9026	−0.6270	0.0001	−0.0201	0.0008	−0.2012				
4	0.9025	−0.6223	0.0001	−0.0197	0.0001	−0.0186	0.0008	−0.1865		
5	0.9025	−0.6181	0.0001	−0.0193	0.0001	−0.0182	0.0001	−0.0173	0.0008	−0.1730
∞	0.9025	-0.5846	0.0001	−0.0171	0.0001	−0.0158	0.0001	−0.0147	0.0001	−0.0137

Return equation

l	w_{t-1}	r_{t-1}	w_{t-2}	r_{t-2}	w_{t-3}	r_{t-3}	w_{t-4}	r_{t-4}	w_{t-5}	r_{t-5}
1	−0.0002	0.9443								
2	−0.0001	0.9330	−0.0000	0.0113						
3	−0.0001	0.9327	−0.0000	0.0010	−0.0000	0.0105				
4	−0.0001	0.9324	−0.0000	0.0010	−0.0000	0.0010	−0.0000	0.0097		
5	−0.0001	0.9322	−0.0000	0.0010	−0.0000	0.0010	−0.0000	0.0009	−0.0000	0.0090
∞	−0.0001	0.9305	−0.0000	0.0009	−0.0000	0.0008	−0.0000	0.0008	−0.0000	0.0007

Table 2: Volatility of economy at the restricted-perceptions equilibrium

l	y	c	n	x	a
1	2.491	1.575	0.998	7.702	1.605
2	2.483	1.571	0.995	7.683	1.605
3	2.476	1.566	0.992	7.667	1.605
4	2.469	1.562	0.990	7.657	1.605
5	2.463	1.558	0.988	7.648	1.605
$MCEE$	2.406	1.547	0.985	7.5900	1.605

Table 3: The stationary distribution of the autoregressive component of beliefs

ϕ_{ww}

$Gain$	$Mean$	$StdDev$	$Skewness$
RPE	0.903	0	0
0.001	0.902	0.010	−0.154
0.002	0.901	0.014	−0.244
0.005	0.900	0.022	−0.389
0.01	0.890	0.032	−0.523
0.02	0.877	0.047	−0.651
0.05	0.839	0.077	−0.805

ϕ_{rr}

$Gain$	$Mean$	$StdDev$	$Skewness$
RPE	0.945	0	0
0.001	0.944	0.006	−0.011
0.002	0.942	0.008	−0.025
0.005	0.935	0.020	−0.062
0.01	0.918	0.034	−0.156
0.02	0.902	0.037	−0.312
0.05	0.877	0.058	−0.645

Table 4: Ratio of standard deviation of aggregates with constant gain learning to that at the RPE

$Gain$	y	c	n	x	$Projection$
0.001	1.001	1.000	1.000	1.001	0.00%
0.002	1.002	1.001	1.001	1.002	0.00%
0.005	1.003	1.003	1.003	1.003	0.02%
0.01	1.003	1.005	1.004	1.004	0.16%
0.02	0.984	0.989	0.990	0.986	0.62%
0.05	0.975	0.984	0.986	0.976	3.04%

Table 5: Ratio of standard deviation of household consumption growth rates with constant gain learning to that at the RPE

Gain	$\sigma_{\Delta c}$
0.001	0.57%
0.002	0.63%
0.005	1.34%
0.01	2.41%
0.02	2.31%
0.05	1.78%
$MCEE$	0.51%

Table 6: Ratio of standard deviation of consumption with constant gain learning, $\gamma = 0.01$ to that at the RPE, σ_c/σ_c^{RPE}, sensitivity to number of households

N	50	100	500	$1,000$	$10,000$
σ_c/σ_c^{RPE}	2.341	1.874	1.123	1.005	1.005

Table 7: Ratio of standard deviation of aggregate consumption with constant gain learning to that at the RPE

Sensitivity to persistence of aggregate shock

$Gain/\rho_a$	0.8	0.9	0.95
0.001	1.000	1.000	1.000
0.002	1.000	1.001	1.004
0.005	1.001	1.003	1.006
0.01	1.002	1.005	0.992
0.02	0.998	0.989	0.971
0.05	0.991	0.984	0.951

Sensitivity to persistence of idiosyncratic shock

$Gain/\rho_z$	0.8	0.9	0.95
0.001	1.000	1.000	1.000
0.002	1.001	1.001	1.008
0.005	1.003	1.003	1.008
0.01	1.004	1.005	0.994
0.02	1.000	0.989	0.978
0.05	0.997	0.984	0.961

Table 8: An intercept in the learning rule, ratio of standard deviation of aggregates with constant gain learning to that at the RPE

$Gain$	y	c	n	x
0.001	1.102	1.081	1.104	1.091
0.002	1.231	1.149	1.221	1.183
0.005	1.841	1.493	1.739	1.621

A Derivations

In what follows, all of the section except A.3 are from the modeler's perspective i.e. no agent in the economy would be able to perform the calculations involved. This is equivalent to the assumption that households are unable to form model-consistent expectations.

A.1 Useful formulations

A.1.1 Aggregation

Using the definition of an aggregate quantity (8) along with the adding up constraint (14) allows the household-level relations to be easily aggregated.

The labour supply relation (2) aggregates to

$$n_t = \varsigma \left(w_t - c_t \right) \tag{A.1}$$

where $\varsigma = \frac{1-N}{N}\eta$. The capital evolution equation (3)

$$k_{t+1} = \left(1 - \delta\right) k_t + \delta x_t \tag{A.2}$$

The budget constraint (4)

$$\frac{c}{y} c_t + \left(1 - \frac{c}{y}\right) x_t = \alpha \left(w_t + n_t\right) + \left(1 - \alpha\right) \left(r_{kt} + k_t\right) \tag{A.3}$$

The production function (5) - using the capital market clearing condition that total household capital equals total firm capital

$$y_t = a_t + \alpha n_t + \left(1 - \alpha\right) k_t \tag{A.4}$$

and the factor demand conditions (6) and (7)

$$r_{kt} = \left(1 - \alpha\right) \frac{k}{y} \left(y_t - k_t\right) \tag{A.5}$$

$$w_t = y_t - n_t \tag{A.6}$$

A.1.2 The capital evolution equation

This follows Campbell (1994). Substituting (A.2) into (A.3) gives

$$k_{t+1} = \tilde{\lambda}_1 k_t + \tilde{\lambda}_2 \left[a_t + n_t\right] + \tilde{\lambda}_4 c_t \tag{A.7}$$

where

$$\tilde{\lambda}_1 = (1-\delta) + \frac{y}{k}(1-\alpha) \tag{A.8}$$
$$\tilde{\lambda}_2 = \frac{y}{k}\alpha\left[a_t + n_t\right] \tag{A.9}$$
$$\tilde{\lambda}_4 = -\frac{c}{k} \tag{A.10}$$

Then substituting for labour from (A.1) and for the wage from (A.24) gives

$$k_{t+1} = \lambda_1 k_t + \lambda_2 a_t + \lambda_4 c_t \tag{A.11}$$

where

$$\lambda_1 = \lambda_1 + \frac{\lambda_2 \varsigma (1-\alpha)}{1+(1-\alpha)\varsigma} \tag{A.12}$$
$$\lambda_2 = \lambda_2\left(1 + \frac{\varsigma\alpha}{1+(1-\alpha)\varsigma}\right) \tag{A.13}$$
$$\lambda_4 = \left(\lambda_4 - \lambda_2\varsigma + \frac{\lambda_2\varsigma^2(1-\alpha)}{1+(1-\alpha)\varsigma}\right) \tag{A.14}$$

A.1.3 An expression for the wage

Subtracting (5), and (6) from their aggregate equivalents (A.4) and (A.5) gives

$$y_t - y_t^s = -\alpha z_t^s + \alpha\left(n_t - n_t^s\right) + (1-\alpha)\left(k_t - j_t^s\right) \tag{A.15}$$
$$0 = (1-\alpha)\frac{k}{y}\left(\left(y_t - y_t^s\right) - \left(k_t - j_t^s\right)\right) \tag{A.16}$$

and combining these gives

$$y_t - y_t^s = -z_t^s + n_t - n_t^s \tag{A.17}$$

Subtracting (7) from its aggregate equivalent (A.6) gives

$$w_t^s - w_t = \left(y_t - y_t^s\right) - \left(n_t - n_t^s\right) \tag{A.18}$$

so

$$w_t^s = w_t + z_t^s \tag{A.19}$$

A.2 Market clearing prices

Combining the aggregate production function (A.4) with the aggregate labour demand relation (A.6) gives

$$w_t = \alpha a_t + (\alpha - 1)\, n_t + (1-\alpha)\, k_t \tag{A.20}$$

Substituting the aggregate labour supply relation (A.1) gives

$$w_t = \frac{\alpha a_t + (1-\alpha)\varsigma c_t + (1-\alpha) k_t}{1 + (1-\alpha)\varsigma} \tag{A.21}$$

Rearranging

$$\begin{aligned} n_t &= \frac{1}{1+\varsigma(1-\alpha)} \left[\varsigma\alpha a_t + \varsigma(1-\alpha) k_t - \varsigma c_t\right] &\text{(A.22)} \\ &= \nu\left(\alpha a_t + (1-\alpha) k_t - c_t\right) &\text{(A.23)} \end{aligned}$$

Then substituting this into the aggregate labour demand relation (A.6) gives

$$w_t = \frac{\alpha a_t + (1-\alpha)\varsigma c_t + (1-\alpha) k_t}{1 + (1-\alpha)\varsigma} \tag{A.24}$$

Substitute for labour from (A.22) into the aggregate capital demand relation (A.5) to give

$$r_t^k = \alpha a_t - \alpha k_t + \alpha n_t \tag{A.25}$$

and using (A.22) gives

$$r_t^k = \alpha\left(\nu(1-\alpha) - 1\right) k_t + \alpha(1+\nu\alpha) a_t - \nu\alpha c_t \tag{A.26}$$

Finally note the relation between the gross and net returns to capital

$$r_t = \kappa_2 r_t^k \tag{A.27}$$

where $\kappa_2 = \frac{r^k}{r}$.

A.3 Optimal household consumption

Substitute the capital evolution equation (3) into the budget constraint (4) to give

$$\frac{c}{y} c_t^s + \frac{k}{y} k_{t+1}^s - \frac{k}{y}(1-\delta) k_t^s = \alpha\left(w_t^s + n_t^s\right) + (1-\alpha)\left(r_t^k + k_t^s\right)$$

Rearranging this and substituting for labour using the household's FOC (2) gives

$$k_t^s = \beta k_{t+1}^s - \frac{1}{\gamma_1}\left(\gamma_3 w_t^s + \gamma_5 r_t - \gamma_2 c_t^s\right) \tag{A.28}$$

where

$$\begin{aligned}\gamma_1 &= \frac{k}{y}(1-\delta)+(1-\alpha)\\ \gamma_2 &= \frac{c}{y}+\alpha\varsigma\\ \gamma_3 &= \alpha(1+\varsigma)\\ \gamma_5 &= \frac{1-\alpha}{\kappa_2}\end{aligned}$$

Solving this forward and using the transversality condition on capital gives

$$\gamma_1 k_t^s = -E_t^s \sum_{j=0}^{\infty} \beta^j \left(\gamma_3 w_{t+j}^s + \gamma_5 r_{t+j} - \gamma_2 c_{t+j}^s\right) \tag{A.29}$$

Rewrite this to separate out the expectational part.

$$\gamma_2 E_t^s \sum_{j=0}^{\infty} \beta^j c_{t+j}^s = \gamma_1 k_t^s + \gamma_3 w_t^s + \gamma_5 r_t + E_t^s \sum_{j=1}^{\infty} \beta^j \left(\gamma_3 w_{t+j}^s + \gamma_5 r_{t+j}\right) \tag{A.30}$$

Solving the Euler equation (1) forward gives

$$\gamma_2 E_t^s \sum_{j=0}^{\infty} \beta^j c_{t+j}^s = \frac{\gamma_2}{1-\beta} c_t^s + \frac{\gamma_2}{1-\beta} E_t^s \sum_{j=1}^{\infty} \beta^j r_{t+j} \tag{A.31}$$

and combining these two

$$\begin{aligned}c_t^s &= \frac{1-\beta}{\gamma_2}\left(\gamma_1 k_t^s + \gamma_3 w_t^s + \gamma_5 r_t\right) + (1-\beta)\frac{\gamma_3}{\gamma_2} E_t^s \sum_{j=1}^{\infty} \beta^j w_{t+j}^s + \left[(1-\beta)\frac{\gamma_5}{\gamma_2} - 1\right] E_t^s \sum_{j=1}^{\infty} \beta^j r_{t+j}\\ &= \frac{1-\beta}{\gamma_2}\left(\gamma_1 k_t^s + \gamma_3 w_t^s + \gamma_5 r_t\right) + \left(\gamma_{cw} T_w + \gamma_{cr} T_r\right)\beta\phi^s\left(I - \beta\phi^s\right)^{-1} I_t^s \end{aligned} \tag{A.32}$$

where

$$\gamma_{cw} = (1-\beta)\frac{\gamma_5}{\gamma_2} - 1 \tag{A.33}$$

$$\gamma_{cr} = (1-\beta)\frac{\gamma_3}{\gamma_2} \tag{A.34}$$

Given the PLM (21) and defining matrices T_w and T_r to pick the respective prices out of the measurement vector i_t^s

$$E_t^s w_{t+j}^s = T_w \left(\phi^s\right)^j i_t^s \tag{A.35}$$

$$E_t^s r_{t+j} = T_r \left(\phi^s\right)^j i_t^s \tag{A.36}$$

so

$$c_t^s = \frac{1-\beta}{\gamma_2}\left(\gamma_1 k_t^s + \gamma_3 w_t^s + \gamma_5 r_t\right) + \left(\gamma_{cw} T_w + \gamma_{cr} T_r\right) \beta \phi^s \left(I - \beta \phi^s\right)^{-1} I_t^s \tag{A.37}$$

A.4 State space representation

Given the structure of the measurement vector

$$I_t^s = \left[\begin{array}{cccc} i_t^s & i_{t-1}^s & ... & i_{t-l}^s \end{array} \right]' \tag{A.38}$$

it is helpful to write the term multiplying it in (A.37) as $\left[\begin{array}{cccc} \vartheta_t^s & \vartheta_{t-1}^s & ... & \vartheta_{i-l}^s \end{array} \right]$ and write the term on current prices as $\vartheta_t^s = \left[\begin{array}{ccc} \vartheta_1^s & \vartheta_w^s & \vartheta_r^s \end{array} \right]$ (note this is written to allow an intercept in the PLM as in section 5.7). Then (A.37) becomes

$$c_t^s = \vartheta_1^s + \frac{1-\beta}{\gamma_2}\left(\gamma_1 k_t^s + \left(\gamma_3 + \vartheta_w^s\right) w_t^s + \left(\gamma_5 + \vartheta_r^s\right) r_t\right) + \sum_{l=1}^{L-1} \vartheta_{i-l}^s i_{t-l}^s$$

and aggregating across households using (9) gives

$$c_t = \frac{1}{S}\sum_s \vartheta_1^s + \frac{1-\beta}{\gamma_2}\left(\gamma_1 k_t + \gamma_{3\phi} w_t + \gamma_{5\phi} r_t + \frac{1}{S}\sum_s \vartheta_w^s z_t^s\right) + \frac{1}{S}\sum_s \sum_{l=1}^{L-1} \vartheta_{i-l}^s i_{t-l}^s$$

Using the expression for the market clearing prices (10) and (11) gives

$$c_t = \gamma_{cc}\frac{1}{S}\sum_s \vartheta_1^s + \gamma_{ck} k_t + \gamma_{ca} a_t + \gamma_{cz}\frac{1}{S}\sum_s \vartheta_w^s z_t^s + \frac{\gamma_{cc}}{S}\sum_s \sum_{l=1}^{L-1} \vartheta_{i-l}^s i_{t-l}^s \tag{A.39}$$

where

$$\gamma_{cc} = \frac{1}{1 - \frac{1-\beta}{\gamma_2}\left[\gamma_{3\phi}\lambda_{wc} + \gamma_5 \lambda_{rc}\right]} \tag{A.40}$$

$$\gamma_{ck} = \gamma_{cc}\frac{1-\beta}{\gamma_2}\left(\gamma_1 + \gamma_{3\phi}\lambda_{wk} + \gamma_{5\phi}\lambda_{rk}\right) \tag{A.41}$$

$$\gamma_{ca} = \gamma_{cc}\frac{1-\beta}{\gamma_2}\left[\gamma_{3\phi}\lambda_{wa} + \lambda_{ra}\gamma_{5\phi}\right] \tag{A.42}$$

$$\gamma_{cz} = \gamma_{cc}\frac{1-\beta}{\gamma_2} \tag{A.43}$$

Define a current state vector

$$Y_t = \left[\begin{array}{ccccc} k_t & a_t & z_t^1 & .... & z_t^S \end{array} \right]' \tag{A.44}$$

Then (A.39) can be written

$$c_t = \Theta_{cY} Y_t + \frac{1}{S\Theta_{cc}} \sum_{l=1}^{L-1} \sum_s \vartheta_i^s i_{t-l}^s ... \tag{A.45}$$

where

$$\Theta_{cY} = \frac{1}{\Theta_{cc}} \begin{bmatrix} \Theta_{ck} & \Theta_{ca} & \frac{1}{S}\vartheta_w^1 & .... & \frac{1}{S}\vartheta_w^S \end{bmatrix} \tag{A.46}$$

and

$$\Theta_{cc} = \frac{1}{\gamma_{cc}} \tag{A.47}$$

$$\Theta_{ck} = \frac{\gamma_{ck}}{\gamma_{cc}} \tag{A.48}$$

$$\Theta_{ca} = \frac{\gamma_{ca}}{\gamma_{cc}} \tag{A.49}$$

Then using (10) and (11) to substitute for lagged prices gives

$$c_t = \Theta_{cY} Y_t + \frac{1}{\Theta_{cc}} \frac{1}{S} \sum_{l=1}^{L-1} \sum_s \vartheta_{wl}^s z_t^s + \frac{1}{S} \sum_{l=1}^{L-1} \left[\Theta_{ckl} k_{t-l} + \Theta_{cal} a_{t-l} + \Theta_{ccl} c_{t-l} \right] \tag{A.50}$$

where

$$\Theta_{ckl} = \frac{1}{S\Theta_{cc}} \left(\vartheta_{wl} \lambda_{wk} + \vartheta_{rl} \lambda_{rk} \right) \tag{A.51}$$

$$\Theta_{cal} = \frac{1}{S\Theta_{cc}} \left(\vartheta_{wl} \lambda_{wa} + \vartheta_{rl} \lambda_{ra} \right) \tag{A.52}$$

$$\Theta_{ccl} = \frac{1}{S\Theta_{cc}} \left(\vartheta_{wl} \lambda_{wc} + \vartheta_{rl} \lambda_{rc} \right) \tag{A.53}$$

Since consumption depends on lagged states and on lagged consumption, the true state vector will contain the full history of the current state vector Y_t. However there are two special cases

1. If $l = 1$ (A.50) does not depend on lagged consumption.

2. If labour supply is fixed $\eta = \infty$, $\varsigma = 0$ so aggregate prices (10) and (11) do not depend on consumption and hence (A.50) does not depend on lagged consumption.

This proves Proposition 1

Write

$$c_t = \Theta_{cx} \left(\Phi_t \right) X_t \tag{A.54}$$

where the state vector is given by $X_t = \begin{bmatrix} Y_t & Y_{t-1} & Y_{t-2} & .... & Y_0 \end{bmatrix}'$ and Φ_t is the stack of a households' PLMs ϕ_t^s

Using (A.54), (A.50) can be written

$$c_t = \Theta_{cY} Y_t + \frac{1}{\Theta_{cc}} \frac{1}{S} \sum_{l=1}^{L-1} \sum_s \vartheta_{wl}^s z_t^s + \sum_{l=1}^{L-1} \left[\Theta_{ckl} k_{t-l} + \Theta_{cal} a_{t-l} \right] + \sum_{l=1}^{L-1} \Theta_{ccl} \Theta_{cx} \left(\Phi_{t-l} \right) X_{t-l} \tag{A.55}$$

Then introduce the dynamics of capital by writing the (A.11) in terms of X_t

$$\begin{aligned} k_{t+1} &= \begin{bmatrix} \lambda_1 & \lambda_2 & 0_{1x\infty} \end{bmatrix} X_t + \lambda_4 c_t & \text{(A.56)} \\ &= \left(\begin{bmatrix} \lambda_1 & \lambda_2 & 0 \end{bmatrix} + \tilde{\lambda}_4 \Theta_{cX} \right) X_t & \text{(A.57)} \\ &= \Theta_{kX} \left(\Phi_t \right) X_t & \text{(A.58)} \end{aligned}$$

and stack this on top of the processes for the shocks to get the law of motion for X_t

$$\begin{aligned} X_t &= \begin{bmatrix} \Theta_{kX} \left(\Phi_{t-1} \right) & 0 \\ \begin{matrix} \rho^a & 0 & 0 & 0 \\ 0 & \rho^z & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \rho^z \end{matrix} & 0 \\ 0 & I \end{bmatrix} X_{t-1} + \begin{bmatrix} 0 \\ I \end{bmatrix} W_{t-1} & \text{(A.59)} \\ &= \Theta_{XX} \left(\Phi_{t-1} \right) X_{t-1} + \Theta_{XW} W_{t-1} & \text{(A.60)} \end{aligned}$$

where the innovations to the exogenous technology processes (12) and (13) are stacked as $W_{t-1} = \begin{bmatrix} \varepsilon_t & \varepsilon_t^1 & \dots & \varepsilon_t^s \end{bmatrix}$

Finally write the observables in terms of the states using (11), (10) and (A.54)

$$\begin{aligned} w_t &= \frac{1}{1 + (1-\alpha)\varsigma} \left\{ \begin{bmatrix} (1-\alpha) & \alpha & 0 \end{bmatrix} Y_t + (1-\alpha)\varsigma c_t \right\} & \text{(A.61)} \\ &= \Theta_{wX} \left(\Phi_t \right) X_t & \text{(A.62)} \end{aligned}$$

and into (11)

$$\begin{aligned} r_t &= \frac{\kappa_2 \alpha}{1 + (1-\alpha)\varsigma} \left\{ \begin{bmatrix} -1 & (\varsigma+1) & 0 \end{bmatrix} Y_{t-1} - \varsigma c_t \right\} & \text{(A.63)} \\ &= \Theta_{rX} \left(\Phi_t \right) X_t & \text{(A.64)} \end{aligned}$$

Then using (9) can relate the observables I_t to current prices and exogenous shocks by

$$I_t = \begin{bmatrix} i_t^1 \\ i_t^2 \\ .. \\ ... \\ i_t^S \end{bmatrix} = FF \begin{bmatrix} r_t \\ w_t \\ z_t^1 \\ ... \\ z_t^s.. \end{bmatrix} \tag{A.65}$$

where

$$FF = \begin{bmatrix} 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 1 & ... & 0 \\ 1 & 0 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \end{bmatrix} \tag{A.66}$$

Then using (A.62) and (A.64) can write

$$\begin{bmatrix} r_t \\ w_t \\ z_t^1 \\ ... \\ z_t^s.. \end{bmatrix} = \begin{bmatrix} \Theta_{rX} \\ \Theta_{wX} \\ I \end{bmatrix} X_t \tag{A.67}$$

then combine these to give

$$I_t = FF \begin{bmatrix} \Theta_{rX} \\ \Theta_{wX} \\ I(S)\rho_z \end{bmatrix} X_t \tag{A.68}$$

$$= \Theta_{IX}(\Phi_t) X_t \tag{A.69}$$

So the system in state space form is given by this relation and the state evolution equation (A.60).

A.5 Learning

This section follows Honkapohja and Mitra (2006) - their Appendix, pp302 - 303 is particularly relevant. Conditions for convergence of Φ_t to an equilibrium Φ are found be defining an associated ordinary differential equation (ODE)

$$\frac{d\Phi}{d\tau} = h(\Phi), \text{where } h(\Phi) = \lim_{t\to\infty} E\mathcal{H}(\Phi, X_t) \tag{A.70}$$

The economy with learning will converge to Φ only if Φ is a locally stable fixed point of the associated ODE.

An updating rule for the PLM can be written generally as

$$\Phi_{t+1} = \Phi_t + \gamma_t \mathcal{H}\left(\Phi_t, I_t\right) \tag{A.71}$$

Note that this is no more than a stack of the individual updating rules. At the restricted perceptions equilibrium, all households are identical so simplify things by taking a single household, so $\Phi_t = \phi_t$, the stacked aggregate PLM is the same as the PLM of the single household in the economy

The learning rule is given by (32) and (33), reproduced here

$$\phi^s_{t+1} = \phi^s_t + \gamma_t R_t^{-1} i^s_{t-1}\left(i^{s\prime}_t - i^{s\prime}_{t-1}\phi^{s\prime}_t\right) \tag{A.72}$$

$$R^s_{t+1} = R^s_t + \gamma_t\left(i^s_{t-1} i^{s\prime}_{t-1} - R^s_t\right) \tag{A.73}$$

Substitute for prices in the expression for R using (A.68) to obtain

$$R_{t+1} = R_t + \gamma_t\left(\Theta_{IX}\left(\phi_{t-1}\right) X_{t-1} X'_{t-1}\Theta_{IX}\left(\phi_{t-1}\right)' - R^s_t\right) \tag{A.74}$$

In the expression for R, fist substitute for prices using (A.68) to obtain

$$\phi_{t+1} = \phi_t + \gamma_t R_t^{-1}\Theta_{IX}\left(\phi_{t-1}\right) X_{t-1}\left(X'_t\Theta_{IX}\left(\phi_t\right)' - X'_{t-1}\Theta_{IX}\left(\phi_{t-1}\right)'\phi'_t\right) \tag{A.75}$$

then use the state evolution equation (A.60) to substitute for X_t

$$\phi_{t+1} = \phi_t + \gamma_t R_t^{-1}\Theta_{IX}\left(\phi_{t-1}\right) X_{t-1}\left(\left[X'_{t-1}\Theta_{XX}\left(\phi_{t-1}\right)' + W'_{t-1}\Theta'_{XW}\right]\Theta_{IX}\left(\phi_t\right)' - X'_{t-1}\Theta_{IX}\left(\phi_{t-1}\right)'\phi'_t\right)$$

So

$$\mathcal{H}\left(\phi_{t-1}, I_t\right) = R_t^{-1}\Theta_{IX}\left(\phi_{t-1}\right) X_{t-1}\left(\left[X'_{t-1}\Theta_{XX}\left(\phi_{t-1}\right)' + W'_{t-1}\Theta'_{XW}\right]\Theta_{IX}\left(\phi_t\right)' - X'_{t-1}\Theta_{IX}\left(\phi_{t-1}\right)'\phi'_t\right)$$

and

$$\lim_{t\to\infty} E\mathcal{H}\left(\phi\right) = R^{-1}\Theta_{IX}\left(\phi\right) M_X\left(\Theta_{XX}\left(\phi\right)'\Theta_{IX}\left(\phi\right)' - \Theta_{IX}\left(\phi\right)'\phi'\right) \tag{A.76}$$

From (A.74)

$$\lim_{t\to\infty} ER = \Theta_{IX}\left(\phi\right) M_X\Theta_{IX}\left(\phi\right)' \tag{A.77}$$

where $M_X = EX'X$ is the variance covariance matrix of the states, and $EW_t = 0$

$$\lim_{t\to\infty} E\mathcal{H}\left(\phi\right) = \left[\Theta_{IX}\left(\phi\right)' M_X\Theta_{IX}\left(\phi\right)\right]^{-1}\Theta_{IX}\left(\Phi\right) M_X\left[\Theta_{XX}\left(\phi\right)'\Theta_{IX}\left(\phi\right)' - \Theta_{IX}\left(\phi\right)'\phi'\right]$$

B Simple example

Take a representative household maximizing expected discounted lifetime utility

$$\max_{\{c_{t+i}\}_{i=0}^{\infty}} \tilde{E}_t \sum_{i=0}^{\infty} \frac{1}{(1+r)^i} \ln c_{t+i} \tag{B.1}$$

subject to a budget constraint

$$b_{t+1} = (1+r)\, b_t + y_t - c_t \tag{B.2}$$

where y_t is an exogenous process

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{B.3}$$

and the innovation is drawn from $N(0, \sigma)$.

Let the household have belief ϕ_t about the persistence of the income process, so

$$\tilde{E}_t y_{t+i} = \phi_t^i\, y_t \tag{B.4}$$

Note in the case of model-consistent expectations $\phi_t = \rho$

The first-order condition for consumption is

$$c_t = E_t c_{t+1} \tag{B.5}$$

and, using the transversality condition $\lim_{t\to\infty} \frac{1}{(1+r)t} b_t = 0$, optimal consumption is

$$c_t = \frac{r}{1+r} \left[(1+r)\, b_t + \frac{1}{1-\phi_t (1+r)^{-1}} y_t \right] \tag{B.6}$$

Let beliefs be updated according to a simple constant gain algorithm

$$\phi_{t+1} = \phi_t + \gamma \left(y_t - \phi_t y_{t-1} \right) \tag{B.7}$$

Note that at the MCEE consumption is a random walk (taking initial wealth to be zero)

$$\Delta c_t^* = \frac{r}{1+r} \left[\frac{1}{1-\rho(1+r)^{-1}} \right] \varepsilon_t \tag{B.8}$$

and

$$\sigma_{\Delta c}^* = \frac{r}{1+r} \left[\frac{1}{1-\rho(1+r)^{-1}} \right] \sigma \tag{B.9}$$

B.1 Proof of proposition 1

Starting from $\phi_0 = \rho$, in response to an innovation ε_0 beliefs in period 1 are $\phi_1 = \rho + \gamma\varepsilon_0$ and then are revised back towards the MCEE i.e. $\phi_t > \rho \ \forall t > 2$. Since $f'(\phi) > 0$ this implies $c > c^*$ for a number of initial periods then $c < c^*$ for the rest of history. If we define the impulse response of consumption as a function IRF then the standard deviation of the first difference of consumption is given by

$$\sigma_{\Delta c} = \sigma \sum_{0}^{\infty} \Delta IRF_t^2 \tag{B.10}$$

Since $\phi_0 = \rho = \phi^*$, $\Delta IRF_0 = \Delta IRF_0^*$

$$\sigma_{\Delta c} = \sigma_{\Delta c}^* + \sigma \sum_{1}^{\infty} \Delta IRF_t^2 > \sigma_{\Delta c}^* \tag{B.11}$$