

Cognitive Consistency, Signal Extraction, and Macroeconomic Persistence

Hyein Chung*and Wei Xiao†

July 3, 2013

Abstract

A guiding principle for the adaptive learning approach of bounded rationality is “cognitive consistency,” which requires that learning agents be no more knowledgeable than real-life economists. Much of current research falls short of satisfying this principle because agents are often assumed to observe the true state vector of the economy. In this paper we propose a method of learning that better satisfies this requirement. We assume agents observe the same information as econometricians: they cannot observe fundamental shocks and they only observe a finite sample of past endogenous variables. Under this assumption, agents must extract information from historical data using statistical techniques. The learning dynamics are considerably different from those under full information learning. In a new Keynesian model, limited information learning generates empirically plausible levels of information persistence that a standard learning model cannot produce.

JEL Classification: E31, E37, D84, C62

Keywords: Adaptive learning, Misspecification, Inflation persistence, Restricted perceptions equilibrium, Model selection.

*Department of Economics, Binghamton University, Binghamton, NY 13902. Email: hchung3@binghamton.edu.

†Department of Economics, Binghamton University, Binghamton, NY 13902. Email: wxiao@binghamton.edu.

1 Introduction

At the beginning of the rational expectations revolution in the 1970s and 80s, adaptive learning was primarily used as a device to understand how agents learn to possess the knowledge required for rational expectations (RE).¹ Subsequent research, however, demonstrates that not all rational expectations equilibrium (REE) can be learned (see Evans and Honkapohja, 2001, for a survey). Burgeoning research in this area and recent development in the economy have motivated economists to reassess the role of learning in macroeconomic models. More and more researchers are now willing to accept adaptive learning as an independent, competing approach to model agent expectations. The RE assumption is “implausibly strong,”² it is argued, for it fails to recognize the *cognitive* and *information* limitations real life agents face. Rational agents not only know the correct form of the structural model, the values of parameters, and the true distributions of exogenous processes, but also can form accurate expectations based on such information. The promise of the learning approach is to offer a more realistic way to formulate expectations. The guiding principle is “cognitive consistency,” which requires that learning agents be no more knowledgeable than real-life economists. In econometric learning, for example, agents are modeled as econometricians who form expectations by estimating and updating a subjective forecasting model in real time. The approach is appealing as it can potentially capture how economic forecasting works in real life.

Our observation is that much of current research in adaptive learning still falls short of satisfying the cognitive consistency principle. There are at least two reasons why the rationality assumption is deemed “implausibly strong.” One is the behavioral

¹Key papers include Bray (1982), Evans (1985), Lucas (1986), and Marcet and Sargent (1989), among others.

²The quotation is from Evans and Honkapohja (2011). See the same paper for a survey of recent research that treats learning as a competing approach for modeling agent expectations.

assumption that rational agents are aware of the equilibrium dynamics of the economy almost immediately. This issue has been dealt with extensively by the adaptive learning approach. A typical learning agent does not know the equilibrium law of motion of the economy and have to learn about them with econometric techniques. The second reason is that rational agents are able to observe variables that real-life agents cannot observe. This issue has not been studied nearly as much as the first, and as a consequence, the cognitive consistency requirement is often compromised.

One such information assumption is routinely made in the learning literature but is conspicuously at odds with reality: agents can observe the entire state of the economy. By “state of the economy,” we mean the state vector of a dynamic general equilibrium model, which includes not only “state variables” such as capital stocks and monetary aggregates, but also exogenous variables and processes which constitute the fundamental driving force of economic fluctuations. Examples of the latter would include technology shocks, shocks to potential GDP, consumer preference shocks, and shocks to the marginal cost of production. The question is, when we model agents as econometricians, is it plausible to assume that they can observe these variables?

Consider technology shocks. In practice, econometricians do not observe such shocks directly. What they do observe are estimates of input and output data, which are often subject to various degrees of inaccuracy themselves. For example, one crucial input is capital stock, but measures of capital stock are not easily obtained, because their estimation requires compilations of other data as well as sophisticated estimation techniques (Schreyer and Webb, 2006). Estimates of technology shocks are also model-dependent: different modeling assumptions can lead to dramatically different estimates. For instance, King and Rebelo (1999) find that with the assumption of varying capital utilization, estimated technology shocks have a much smaller variance and lower chances of technological regress than conventional measures. Due to these difficulties,

econometricians rarely consider technology shocks as an observable variable. Next consider shocks to the output gap, an important exogenous variable in the popular new Keynesian model. In order to calculate this gap, the potential level of GDP must be estimated first. But its estimates are often inaccurate due to possible structural breaks in employment and production. For this reason, econometricians usually do not treat shocks to output gaps as a known variable. Similar arguments can be made about other commonly used variables in the state vector, such as preference shocks and cost-push shocks.

The consequence of assuming observability for the state vector is that in models of learning, agents can run regressions that are beyond the capability of real-life econometricians. Consider a generic linear model:

$$y_t = \mu + ME_t y_{t+1} + P v_t, \quad (1)$$

$$v_t = \rho v_{t-1} + e_t, \quad (2)$$

where y_t is a vector of endogenous variables, v_t is a vector of serially correlated exogenous processes, e_t represents i.i.d. white noises, and the rest are parameters. A “minimum state variable” (MSV) rational expectations solution is in the form

$$y_t = \alpha + \beta v_t, \quad (3)$$

along with (2), where α and β are vectors of parameters.

If the assumption is that agents can observe the state vector v_t , then they are able to learn (2) accurately as the equation does not involve any expectations. Moreover, the agents’ perceived law of motion (PLM) is often modeled as being consistent with the MSV solution form:

$$y_t = a_{t-1} + b_{t-1} v_t, \quad (4)$$

where a_{t-1} and b_{t-1} are parameter vectors that agents update in real time. Agents will then use this equation to make forecasts of the future.

Thus, agents are able to use the state vector as right-hand-side variables in their regression analysis, and their forecast equation consists only of those state variables. This is where the model and reality diverge. Real-life econometricians rarely have the luxury of an observable state vector, and their forecast equations consist primarily of past values of endogenous variables. In fact in the time series econometrics literature, considerable research has been devoted to designing techniques that can obtain parameter estimates when the state vector is not observable. The Kalman filter, for example, is the canonical device to deal with this situation. When econometricians estimate dynamic general equilibrium models, the variables that they decide to be unobservable and must be inferred with the Kalman filter are precisely those included in the state vector (see, for example, Ireland (2004) and Schorfheide (2013), among many others).³

This “cognitive inconsistency” exists in a large number of research papers of learning, including some that have made milestone contributions in the discipline. Does this inconsistency matter? Is the assumption of full observability a harmless abstraction from reality, or does it have significant impact on learning dynamics? The goal of this paper is to investigate this issue and explore a new approach to abridge the gap between adaptive learning and cognitive consistency. We do this by imposing two realistic *information limitations* on learning agents: one, they do not observe the state vector, and two, they do not possess an infinite series of the past values of endogenous variables. (They can observe past values of *all* endogenous variables. Our only requirement is that the length of their dataset is not infinity.) With these limitations, agents face a signal extraction problem: they must use available data to infer the nature of the

³Our example is a simple one. In richer models of learning, the forecasting equation can include other variables such as lagged endogenous variables. However, the key issue is whether or not the vector of exogenous shocks enters the PLM. As long as this vector enters agents’ forecast equations, the divergence between reality and model assumptions remains.

unobservables.

There are, however, many ways to solve a signal extraction problem. We need a tractable and parsimonious solution method that is consistent with the adaptive learning framework. To this end we draw on the methodology developed by Marcet and Sargent (1989). Marcet and Sargent (1989)'s primary goal is to study the convergence of recursive least square learning to an REE when there is private information. In their setup, there are two types of agents who can only observe a distinct subset of the state vector. The agents are aware that some variables are hidden from them, and they need to extract information from the observables. To do this, they run finite order vector autoregressions (VAR) and use the estimated equations to forecast future endogenous variables. Marcet and Sargent (1989) prove that under certain conditions the learning process converges, and the resulting REE is a "limited information equilibrium." Our objective is quite different from theirs, as we are not interested in heterogeneous agents or private information. Nevertheless, we note that in their model, it is variables from the state vector that are unobservable – a critical feature that we require for our environment. This makes it possible for us to adopt their method to handle our information extraction problem. Indeed, we show that our model can be conveniently formulated as a special case of the Marcet and Sargent model, in which all agents are identical, and all state variables are hidden. Our approach leads to an equilibrium that is similar to Marcet and Sargent's limited information equilibrium, but is conceptually different. Their equilibrium is essentially an REE with private information. Learning helps explaining how agents learn to reach the REE. In our model, the REE is the standard minimum state variable (MSV) solution. Limited information only becomes an issue when learning is considered, as we require the learning agents to possess no more observations than real-life econometricians. In other words, our limited information equilibrium only exists under bounded rationality. We call this equilibrium "limited

information learning equilibrium.”

The learning agents’ beliefs are characterized by stationary processes such as finite order VARs, as in Marcet and Sargent (1989). An immediate question is which belief function should prevail. To answer it, we turn once again to the cognitive consistency principle, and ask what econometricians do in practice. Our conclusion is that agents should use statistical standards, such as the Bayesian information criterion (BIC) or likelihood ratio tests, to “select” a model. These statistical standards are typically constructed to pick the best specification among many to maximize the probability of recovering the true model, thus capturing the problem faced by our learning agents. Our decision to include a model selection device is inspired by recent works of Branch and Evans (2007) and Cho and Kasa (2011), who consider similar situations where agents have multiple models and must select one for learning.

We emphasize that a key assumption of our approach is that agents do not observe an infinite series of past endogenous variables. We make this assumption for two reasons. One, if agents possess infinite data, and if there is a representative agent, then under certain conditions, the signal extraction process can fully recover the true values of the state vector, and the full information REE can be revealed (see Fernandez-Villaverde et al., 2007). This is uninteresting to us, because the assumption of infinite data availability violates the cognitive consistency principle and defeats our purpose of modeling real-life econometricians. What we are interested in is a truly limited information economy, in which a representative agent must use finite data to infer about the fundamentals of the economy. Two, the finite data assumption captures another aspect of reality: econometricians are constantly evaluating the possibility of time-varying system dynamics and structural breaks. Consequently, they may not *want* to use old data even when such data are available. For example, economists typically do not employ data from the Great Depression era when forecasting modern economic performances,

largely because enough structural changes have taken place which render old data not very useful.

After obtaining some theoretical results, we consider an application. Our economy is the standard reduced-form new Keynesian model. We compute the limited information learning equilibrium of this model, and compare the model's implications with those of the full-information model. Our study leads to some interesting conclusions. It is well known that in a full information REE, the standard new Phillips curve does not include any lagged terms of inflation, and the model cannot generate the level of inflation persistence that we observe in the data (Fuhrer, 2009). Empirical researchers often include a lagged term in the Phillips Curve to fit the data better. Considerable amount of theoretical work has also been done to try to derive a Phillips Curve with lagged terms. In our model, the Phillips curve remains a standard one, but the actual law of motion of the economy includes lagged terms of inflation. The reason is that agents use past inflation data to infer about the state vector and form forecasts of future values. The feedback from agents' forecast to the actual law of motion generates inflation persistence. Moreover, the level of inflation persistence is significantly higher than that in the model with full information. We thus uncover an intrinsic mechanism of inflation persistence – persistence which arises from interactions between the information extraction process and the model structure itself.

In our model, agents select the number of VAR lags endogenously. With finite data, the model that survives the BIC test is often the one that strikes a good balance between accuracy and parsimony. For example, when the sample size matches that of real-life macro dataset, agents often select short belief functions such as an AR(1) in a standard new Keynesian model. This leads to model dynamics that are distinct from full information models. The optimal lag becomes a function of sample sizes, which are in turn affected by agents' beliefs about the degree of time variation in the

data. Thus, the model’s dynamics has a self-confirming nature. Our analysis also provides a justification for using simple “misspecified” belief functions for boundedly rational agents. Most researchers argue for simple belief functions on the ground that they are consistent with experimental or survey results (Adam, 2007). Our analysis demonstrates that even agents who are equipped with complex statistical tools may choose simple belief functions, simply because with finite data, these functions perform the best statistically.

Our application demonstrates that if we impose cognitive consistency and require agents not to possess knowledge of the entire state vector, the dynamics of the economy differ considerably from that of full observability. We believe this is an indication that this line of research is valuable, and should be extended to study many macroeconomic issues.

2 Related Literature

Earlier works in the literature consider signal extraction problems in dynamic macro models. In his path-breaking paper that laid the foundation for rational expectations, Muth (1961) notes that in a Cob-web model, it is possible that the exogenous variable is not observed by agents. He states that “if the shock is not observable, it must be estimated from the past history of variables that can be measured.” In Lucas (1975), agents do not observe the state of the economy, and must use observable variables to infer it. Consequently serially uncorrelated shocks propagate through the economy slowly and generates persistence. Townsend (1983) studies a model in which heterogeneous agents possess different information about the state vector, and must use available information to learn about the forecasts of other agents. In Marcet and Sargent (1989), agents use recursive learning techniques to extract information from observable endogenous

variables. If the learning process converges, the economy will converge to a limited information equilibrium. Their work is perhaps the first important paper that considers the signal extraction problem for adaptive learning agents. A common feature of these models is that the source of information – the endogenous variables – is determined itself by the information extraction problem that the agents must solve.

Recent years have seen a revival of research in the area of incomplete information and signal extraction. Gregoir and Weill (2007), Hellwig (2006), Pearlman and Sargent (2005), and Rondina and Walker (2011) are just a few examples of a large number of papers devoted to this area of research. Most of this research focuses on understanding the characteristics of *rational* equilibrium when there are heterogeneous agents and dispersed information. The case of a representative agent is often treated as a simple benchmark to compare with more elaborate scenarios. Levine et al. (2007) do consider a representative agent model with partial information. However, in their paper agents are rational; the case of limited cognitive abilities such as adaptive learning is not studied. The limitation of finite data is also rarely a concern in this literature.

The learning literature made significant advances in recent years, but works in the area of signal extraction are relatively rare. There are some papers that implicitly incorporate a signal extraction problem in their models. For example, Bullard et al. (2008) define an “exuberance equilibrium,” where agents use an ARMA-type PLM to form their forecasts. In Preston (2005)’s infinite horizon learning framework, agents often use VAR-type PLMs. By not formulating their PLMs in the MSV form, these authors implicitly recognize that agents need not possess knowledge of the state vector and/or the MSV solution form. Preston emphasizes that researchers must distinguish between what agents possess in their information set and what not. This is, in spirit, very similar to our argument for cognitive consistency.

Another area of learning research, the “misspecification” approach, also considers

unobservable state variables. In those models, agents typically only observe a subset of state variables. Their PLM therefore only consists of this subset of variables, rendering their model “misspecified,” compared with the full observability MSV solution (Evans and Honkapohja, 2001). When learning converges, these economies reach a “restricted perceptions equilibrium.” The difference between this approach and ours is as follows. In misspecification models, agents still observe a state vector. What they observe is not the entire state – they are not aware that some other state variables exist and should be included. They still run regressions of endogenous variables against the observed state variables, believing they are searching for the MSV solution.⁴ In this respect, the cognitive inconsistency in these models is the same as that in full information models. In our model, agents are aware that there are unobservable state variables, which is precisely why signal extraction is important to them.

We apply our method to the new Keynesian model, and examine the impact of signal extraction learning on inflation persistence. That inflation has inertia is well documented, but its causes are not well understood. Standard rational expectation models do not generate persistent dynamics due to their forward-looking nature. Efforts to explain inflation persistence can generally be divided into two categories: structural causes and intrinsic causes. The structural approach emphasizes that certain features of the economy leads to inflation persistence. Fuhrer and Moore (1995), for example, argue that US inflation dynamics becomes backward-looking if workers care about past wages. Christiano et al. (2005), Smets and Wouters (2003) and Giannoni and Woodford (2003) advocate price indexation as an explanation of inflation persistence. The intrinsic persistence approach emphasizes that persistence may arise even if the underlined economic structure does not contribute to such dynamics. Galí and Gertler (1999), for

⁴If agents are aware of the hidden state variables, they would include past values of endogenous variables in their forecasting functions to extract information and to improve the accuracy of their forecasts.

example, demonstrate that firms' rule of thumb behavior can lead to persistent price dynamics. Mankiw and Reis (2002) argue that persistence is caused by agents' inability to process information swiftly due to its high costs.

Orphanides and Williams (2004) and Milani (2007) study inflation persistence in learning models. In Orphanides and Williams (2004), agents know the MSV form of the model, but use constant gain learning to estimate the model's parameters. Inflation inertia arises from two sources: agents' imperfect information about model parameters and the inability of the model to converge, caused by the constant gain learning algorithm agents adopt. Milani (2007) uses US data and a hybrid new Keynesian model to determine which mechanism, price indexing or adaptive learning, contributes more to inflation persistence. He finds that the latter is more important. In both papers, there are no hidden variables and the PLMs are consistent with the MSV solution. In our model, agents' beliefs are not consistent with the MSV solution due to limited information. This distinguishes our work from theirs.

The rest of the paper is organized as follows. In the next section, we set up a general model, and defines a limited information REE as in Marcet and Sargent (1989). In the following section, we apply the general theory to a new Keynesian model, and discuss the empirical implications of our approach. The last section concludes the paper.

3 Adaptive learning with unobservable state variables

Consider a reduced form model

$$y_t = ME_t^* y_{t+1} + Pv_t, \quad (5a)$$

$$v_t = Fv_{t-1} + e_t, \quad (5b)$$

where y_t is an $n \times 1$ vector of endogenous variables, and v_t is an $m \times 1$ vector of exogenous variables which follow a stationary autoregressive process. e_t is an $m \times 1$ vector of white noise. E_t^* represents expectations of private agents that are not necessarily rational.

This is a generic reduced form model that many contemporary macro models fit in. It can, however, be made even more general. For example, in some macro models the lagged values of endogenous variables play a role in agents' decision making. That would require that we add a vector y_{t-1} to (5a). We choose not to include it for our benchmark setup here, because doing so will make the contrast between the cases of unobservable and observable state vector less distinct. We consider this case in Appendix A.1. We also choose not to add any extra expectations terms (such as $E_{t-1}^* y_t$) to (5a), as the benchmark result can be extended to models with those terms. The exogenous process can also be made more elaborate by adding extra noises and lags. We choose not to do so in order to keep the analysis tractable.

Our information assumptions are as follows. In the benchmark case of full observability, agents can observe both the vector y_{t-1} and v_t . They use this information to forecast future values y_{t+1} , and make decisions about current values y_t . In the case of limited information, agents only observe past values of y and do not observe the state vector v . As a consequence, their forecasting functions cannot contain any elements of v . They need to use statistical techniques to extract information from the observables, and forecast the future values of the endogenous variables. We note that the self-referential nature of the system is evident in (5a): agents' forecasts can clearly

alter the time path of the economy itself. Following the learning literature, we assume that learning agents do not have this knowledge and will not attempt to “game” the system by altering their expectations strategically. They estimate system dynamics based on the premise that the economic system is a stable stationary process.

A more general setup would encompass the possibility that a subset of v is observable, and a subset of y is unobservable to agents. Unobservable endogenous variables are possible: in a multi-agent model, for example, it is possible that some variables are private to certain agents, and each agent only observes a subset of y . However, creating the most general environment is not the optimal strategy for us, as the technical complications may obscure the main point that we attempt to make. So we decide to make the assumption that the entire v vector is unobservable, and all past values of y_t are observable to agents. This way, agents’ main task will be to extract from past endogenous variables information of the fundamental shocks v_t . We believe this setup captures the essence of the argument that we made earlier, that fundamental shocks should not be assumed observable, and should not be used as right-hand-side variables in agents’ learning process. This setup will also make a sharp distinction between our limited information approach and the conventional approach, as we show below.

3.1 REE and learning with full observability

With full observability, both the REE and the adaptive learning solutions of this model are well known. We briefly review the results below.

To derive the REE, assume agents have the perceived law of motion (PLM) that corresponds to the minimum state variable (MSV) REE of the model:

$$y_t = av_t. \tag{6}$$

To obtain the T-mapping for the PLM to the actual law of motion (ALM), compute the one-step forecast as

$$E_t^* y_{t+1} = aFv_t. \quad (7)$$

Plugging (7) into (5a) and (5b), we obtain the ALM as

$$y_t = (MaF + P)v_t. \quad (8)$$

The REE is simply defined as the fixed point of

$$a = T(a) = MaF + P. \quad (9)$$

We define more formally the REE solution below.

Proposition 1. *The full information rational expectations equilibrium (REE) of the system (5a) and (5b) is a matrix a that satisfies*

$$a = T(a) = MaF + P.$$

In the REE, the law of motion of the economy is

$$y_t = T(a)v_t, \quad (10)$$

$$v_t = Fv_{t-1} + e_t. \quad (11)$$

When learning is considered, agents are assumed to be running regressions of the form

$$y_{t-1} = c_{t-1} + a_{t-1}v_{t-1}, \quad (12)$$

where c is a constant term for the regression. The learning process can also be expressed more generally in recursive form (Evans and Honkapohja, 2001). Note that (11) does not include any expectation terms. With full observability, agents should be able to learn the parameters easily. Hence it is typically assumed that (11) is known to agents.

Learnability or “E-stability” (Evans and Honkapohja, 2001) of the REE is governed by the differential equation

$$\frac{d}{dt}(c, a) = T(c, a) - (c, a). \quad (13)$$

The REE is learnable if the above equation converges in time. Note that throughout the analysis, a critical assumption is that at time t , agents not only observe all the endogenous variables y_{t-1} , but can also observe all the variables in the vector v_t .

3.2 Limited information equilibrium with unobserved state variables

Now we consider the case in which the shocks v_t are not observable. Agents must use past values of endogenous variables $(y_{t-1}, y_{t-2}, \dots)'$ to extract information about the exogenous variables and forecast future endogenous variables. A foremost issue we need to address is what “models” our agents should use to forecast future variables with limited information.

3.2.1 Forecasting

After the publication of Sims’s seminar work *Macroeconomics and Reality (1980)*, VARMA (vector autoregressive moving average) methods have become the most widely used tool for forecasting macroeconomic variables. There are at least two reasons why this method is useful for our context. One, equilibrium solutions of structural macro

models, such as the model we lay out above, can often be reduced to VARMA forms, especially VARs. In a recent paper, Fernandez-Villaverde et al. (2007) derive conditions under which a wide class of linearized DSGE models with hidden state variables can be completely identified by running an infinite order VAR. Ravenna (2007) points out that under certain conditions, the same class of models can also be represented by finite order VARs. Both papers report that if the conditions for a VAR are not met, the models can often be represented by more general VARMA forms. A second reason to consider the VARMA approach is that it is also a relatively “model-free” method of forecasting. VAR forecasting, for example, does not have to be based on any structural models. It is sometimes regarded as an advantage that economists can rely on them to make reasonable forecasts without committing to any specific model structures.

Much like real-life economists, agents in our model must forecast future variables with limited knowledge. If they do understand the general structure of the economic environment they live in, they will quite naturally use VARs to forecast the future. As the next proposition shows, the REE solution of our model has a simple VAR representation.

Proposition 2. *Let the rational expectations equilibrium solution with full observability be represented by (10) and (11). Suppose the matrix $T(a)$ is non-singular, then the solution has a VAR(1) representation*

$$y_t = T(a) \cdot F \cdot T(a)^{-1} y_{t-1} + T(a) \cdot e_t,$$

Proof: solve for v_t as $v_t = T(a)^{-1} y_t$ from (10) and plug into (11), and rearrange.

Note that the non-singularity condition requires that there are equal number of fundamental shocks in the vector v_t as the number of endogenous variables in y_t . This condition is not as restrictive as it seems. As Ireland (2004) points out, measurement

errors can always be added to an equation like (10) to ensure that a DSGE model can be properly identified.

What if the agents do not understand the structure of the model? As we argue above, it is still reasonable to assume that they use VARMA models to make forecasts, as this is precisely what many real-life economists do. In this case, agents do not necessarily have to use VAR(1) forecasting, as Proposition 2 would suggest. The problem facing them is one that requires them to make a forecast as accurate as possible without any definite knowledge of the economic structure. What they need is a general but parsimonious forecasting function that finite data can support.

In principle, they can use any stationary VARMA models. Without loss of generality, let us assume the forecasting model is a p th order VAR of the form

$$y_t = \sum_{j=1}^p b_j y_{t-j} + \epsilon_t. \quad (14)$$

This assumption is an extension of Marcet and Sargent (1989)'s VAR(1) belief function and is almost the same as Bullard et al. (2008)'s assumption of VAR forecasting in an exuberance equilibrium.⁵ It nests the rational expectations solution in Proposition 2.

We note that since agents know that there are unobserved fundamental shocks, the above equation is not strictly their “perceived law of motion (PLM)” of the economy. It should be more accurately called a “forecasting model.” To be consistent with the literature, however, we shall still use the terminology “PLM.” We will use it and the terms “forecasting model” and “belief function” interchangeably.

⁵As an example of other possible forecasting models, in Sargent (1991), learning agents run ARMA(1,1) models to forecast future variables.

Defining $Y_{t-1} = (y'_{t-1}, y'_{t-2}, \dots, y'_{t-p})'$, we can rewrite the above equation as

$$y_t = bY_{t-1} + \epsilon_t, \quad (15)$$

where $b = (b_1, b_2, \dots, b_p)$. The agents form their expectations

$$E_t^* y_{t+1} = \sum_{j=1}^p \tilde{b}_j y_{t-j} \quad (16)$$

where

$$\tilde{b}_j = \begin{cases} b_1 b_j + b_{j+1}, & j = 1, \dots, p-1 \\ b_1 b_j, & j = p. \end{cases}$$

Given the PLM (14), the ALM of the economy can be derived as

$$y_t = M \sum_{j=1}^p \tilde{b}_j y_{t-j} + PFv_{t-1} + Pe_t, \quad (17)$$

which can also be written as

$$z_t = \sum_{j=1}^p T_j(b) z_{t-j} + Ve_t, \quad (18)$$

where $z_t = (y'_t, v'_t)'$, $T_j(b) = \begin{pmatrix} M\tilde{b}_1 & PF \\ 0 & F \end{pmatrix}$ if $j = 1$, $T_j(b) = \begin{pmatrix} M\tilde{b}_j & 0 \\ 0 & 0 \end{pmatrix}$ otherwise,

and $V = \begin{pmatrix} P \\ I \end{pmatrix}$.

An even more compact expression for the ALM is

$$z_t = T(b)Z_{t-1} + Ve_t, \quad (19)$$

where $T(b) = (T_1(b), T_2(b), \dots, T_p(b))$ and $Z_{t-1} = (z'_{t-1}, z'_{t-2}, \dots, z'_{t-p})'$.

With the above assumptions and derivations, our model becomes a special and extended case of Marcet and Sargent (1989). It is a special case because our model does not have heterogeneous agents and private information; it is an extension because in our model agents run p th order VARs, while in Marcet and Sargent (1989) agents only run first-order VARs. In Appendix A.2, we show mathematically how our model fits into Marcet and Sargent (1989)'s framework. This setup allows us to define an equilibrium without proof in the subsequent discussion, as the theoretical foundations have already been provided by Marcet and Sargent (1989).

3.2.2 Equilibrium and stability under learning

With the PLM and ALM given as above, the linear least-squares projection of y_t on Y_{t-1} can be calculated as

$$P[y_t|Y_{t-1}] = s(b)Y_{t-1}, \quad (20)$$

where $s(b)$ is the projection operator defined as

$$s(b) = i_n T(b) (E_t Z_{t-1} Y'_{t-1}) (E_t Y_{t-1} Y'_{t-1})^{-1}. \quad (21)$$

The selection matrix $i_n = [I \ 0]$ picks out the vector y_t from the projected z_t . Intuitively, the coefficients of the population regression of Z_{t-1} on Y_{t-1} are computed first. The result is then fed into the ALM to yield the optimal projection coefficient matrix $s(b)$. Note that the theoretical population regression is not one that the learning agents can really run, as they do not observe Z_{t-1} . If the agents' forecast (16) converges to this (optimal) least-squares projection, the system has reached an equilibrium. We are going to call this a “limited information learning equilibrium.”

Proposition 3. *A limited information learning equilibrium (LILE) is a matrix $b =$*

(b_1, b_2, \dots, b_p) that satisfies

$$b = s(b), \tag{22}$$

where b is defined by (21).

Hence, this equilibrium is a fixed point of the mapping s .

Evans and Honkapohja (2001) formalize the idea of an equilibrium with restricted perceptions, which requires that given agents' misspecified perceptions, the resulting economic dynamics is such that agents are not able to detect any systematic errors. The limited information equilibrium satisfies this requirement. The coefficient matrix of the OLS regression (15) is unbiased and converges in the long run to the theoretical population regression coefficient matrix

$$b^* = (E_t y_t Y'_{t-1})(E_t Y_{t-1} Y'_{t-1})^{-1}.$$

One requirement of the population regression is the orthogonality condition

$$E Y'_{t-1} [y_t - b^* Y_{t-1}] = 0.$$

It follows that in equilibrium,

$$E Y'_{t-1} [y_t - s(b) Y_{t-1}] = 0.$$

That is, agents' forecasting errors are orthogonal to the information set Y_{t-1} . It implies that for any specific forecasting model, agents can no longer detect any systematic errors and improve upon their existing forecasts.⁶ The difference between a limited information equilibrium and a restricted perceptions equilibrium is that in the latter case, agents are not aware that there are hidden variables and do not use past values

⁶Agents can still improve their forecasts by selecting an optimal model among a set of models.

of endogenous variables to extract information.

While the mathematical foundations of Marcet and Sargent (1989) allow us to define the LILE, our equilibrium is conceptually different from theirs. Their limited information equilibrium is an REE with private information, and learning serves to explain how agents learn to reach it. In our model, the limited information equilibrium only occurs when agents are not rational. It arises because under bounded rationality, agents are assumed unable to observe the exogenous processes. That’s why we call it a “learning equilibrium.”

Under learning, agents update the parameters of the VAR (equation 14) every period as new data become available. The coefficients of the VARs can be computed using the recursive least squares formulas

$$\begin{aligned} b_t &= b_{t-1} + \gamma_t R_t^{-1} Y_{t-1} (y_t - b'_{t-1} Y_{t-1}), \\ R_t &= R_{t-1} + \gamma_t (Y_{t-1} Y'_{t-1} - R_{t-1}), \end{aligned}$$

where γ_t represents a “gain” parameter, and R_t is a moment matrix. The convergence of least squares learning is governed by the “E-stability principle,” established by Evans and Honkapohja (2001), which we now turn to.⁷ We state the following proposition:

Proposition 4. The limited information learning equilibrium is stable under adaptive learning if the following differential equation is stable:

$$\frac{db}{dt} = s(b) - b. \tag{23}$$

Thus, the LILE is reachable under adaptive learning if the E-stability equation is stable. Note that this is a local result that governs the stability of learning within a small neighborhood of the equilibrium values of b .

⁷Marcet and Sargent (1989) also establish the same stability conditions.

Unobservability and the limited information equilibrium lead to two immediate consequences. First, a comparison of (9) and (21) reveals that the LILE generally differs from the full information REE in functional forms. This suggests that model dynamics can be significantly different, not only between equilibrium behavior of the REE and the LILE, but also between learning dynamics of two models. We believe this is worth further exploration. In the next section of the paper, we examine the implied levels of inflation persistence in the LILE and the REE.

The second consequence is the possibility of multiple equilibria. There is no restriction as to what order of VAR agents should run, and even what variables agents should include in the PLM. In principle, every selection of PLM should lead to a new limited information REE, as long as the fixed point of the mapping from the PLM to the ALM exists. Thus, multiple equilibria become the norm in such models. The implication is that if unobservability is a prevailing phenomenon, then multiple equilibria should also be equally common in macroeconomic models. The unique REE solution in many models is merely a full-information special case of a more complex reality.⁸

3.2.3 Determination of the VAR order p

A remaining issue to address is how agents determine the order p when they run the VAR(p) model.

In the literature there are at least two methods to handle similar issues. One approach is the “misspecification equilibrium” approach suggested by Branch and Evans (2006). In their setup, there are multiple possible underparameterized PLMs (predictors) and equilibria. Their solution is to allow heterogeneous agents to consider all possible predictors. In equilibrium, agents find their own best performing predictor

⁸In full information models, multiple equilibria are also possible, but criteria such as the MSV requirement or the determinacy requirement can often help find a unique solution. The same criteria cannot be applied to the limited information REE, as the source of multiplicity is not the same.

given their choices, and it is possible that more than one predictor is utilized (by different agents). This approach requires the assumption of heterogeneous agents. The second approach is advocated by Cho and Kasa (2011) and others. The idea is to let agents use a statistical criterion to select the best performing model among all possible models.

Since we do not consider heterogeneous agents, we will follow the second line of thought. When facing model uncertainty in real life, econometricians use statistical criteria to decide which model is their optimal choice. In time series models, information criteria such as the Bayesian information criterion (BIC) are often used to determine the order of VARs. This is what agents will do in our model.

With infinite data, the more lags a PLM has, the more accurate it is, in the sense that it reduces the gap between the incomplete information ALM and the full information ALM (recall that most DSGE models can be reduced to infinite order VARs). But with finite data, the accuracy of parameter estimates deteriorates as the number of lags increases. The agents therefore face a trade-off between a more precise model and more precise parameters. The law of parsimony or Occam's razor provides the remedy to this dilemma.⁹ The idea is to maintain a good balance between the goodness-of-fit and model complexity. In addition to the usual statistics that measures a model's standard error, this approach adds a penalty to each additional parameter adopted. The BIC, proposed by Schwarz (1978), is one of the popularly used criteria that implement this principle.¹⁰ Specifically, the BIC weighs the error and the number of parameters in the

⁹The principle has a long history. A 14th-century English logician and theologian, William of Ockham, wrote "entities must not be multiplied beyond necessity." This is interpreted that the simplest explanation is the correct one. The principle has played a crucial role in the model selection literature, and also in many other disciplines including statistics, physics, chemistry, and biology.

¹⁰Besides BIC, Akaike information criterion (Akaike, 1974), minimum description length (Rissanen, 1978), and minimum message length (Wallace and Boulton, 1968) are also widely used as model selection criteria.

following way (e.g., see Lütkepohl, 1991):

$$BIC = \log \det(\Sigma_{err}) + \frac{k \log N}{N}, \quad (24)$$

where $\det(\Sigma_{err})$ refers to the determinant of the variance matrix of the error, N and k denote the number of data and the number of parameters in the model, respectively. The first term corresponds to the goodness-of-fit and the second term represents the overparameterization penalty. A model with a smaller BIC value is considered to be more preferable than a model with a larger BIC value.

Agents compute the BIC and determine the number of VAR lags endogenously during real-time learning. We describe the details in the next section, where we apply the methodology to a new Keynesian model.

4 Application: a new Keynesian model

In this section, we apply our learning approach to a standard new Keynesian model. Our purpose is twofolds. First, we want to provide an example to illustrate how our learning and signal extraction approach works. Second, as we argue in the previous section, one consequence of a limited information equilibrium is that model dynamics may differ significantly from the full information case. We examine such differences for the new Keynesian model.

On the empirical front, one issue that we are particularly interested in is whether the new framework implies higher levels of inflation persistence than conventional rational expectations models. We conjecture that there would be more inflation persistence for the following reason: first we note that the original structural model (5a) is purely forward looking. There is no built-in “intrinsic persistence,” except for the AR(1) process of the shock in (5b). Consequently, in the full information REE (8), the only

persistence comes from the variable v_t . This is essentially why in many models, the level of macroeconomic persistence fails to match the data. With limited information, system dynamics have changed. In (17), current endogenous variables are not only functions of v_{t-1} , but also functions of past levels of endogenous variables. There should be more persistence in model dynamics.

4.1 A New Keynesian Model

The baseline framework for analysis is the standard New Keynesian model studied by Woodford (2003), Clarida et al. (1999) and many others. The model economy consists of a representative household, a continuum of monopolistically competitive firms, and a central bank. The economy can be described by the following set of linearized equations, each derived from the optimization behavior of households and firms.

The first equation in the reduced form system is the New Keynesian Phillips curve (NKPC)

$$\pi_t = \beta E_t^* \pi_{t+1} + \delta x_t + u_t, \quad (25)$$

and the second equation is the IS curve

$$x_t = -\sigma(i_t - E_t^* \pi_{t+1}) + E_t^* x_{t+1} + g_t. \quad (26)$$

π_t is inflation rate, x_t is the output gap and i_t is the nominal interest rate, all representing the percentage deviations of variables from their steady state values. The terms $E_t^* \pi_{t+1}$ and $E_t^* x_{t+1}$ represent the private sector's expectations of inflation and output, respectively. The parameters β and σ denote the discount factor and the elasticity of intertemporal substitution of a representative household. δ is related to the degree of price stickiness. The disturbance terms u_t and g_t represent a cost push shock and a

shock to aggregate demand. They are assumed to follow AR(1) processes

$$u_t = \rho u_{t-1} + \tilde{u}_t, \quad (27a)$$

$$g_t = \mu g_{t-1} + \tilde{g}_t, \quad (27b)$$

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

The monetary policy of the central bank is represented by the Taylor rule (Taylor, 1993)

$$i_t = \pi_t + \psi_\pi(\pi_t - \pi_t^*) + \psi_x(x_t - x_t^*), \quad (28)$$

where π_t^* and x_t^* are the target levels of inflation rate and the output gap. Without loss of generality, we assume $\pi_t^* = x_t^* = 0$. Then, the rule is reduced to

$$i_t = \psi_\pi \pi_t + \psi_x x_t, \quad (29)$$

where $\psi_\pi > 0$ and $\psi_x > 0$.

These equations can be combined into a reduced form as (5a) and (5b), which we reproduce here:

$$y_t = M E_t^* y_{t+1} + P v_t,$$

$$v_t = F v_{t-1} + e_t,$$

where $y_t = (\pi_t, x_t)'$, $v_t = (u_t, g_t)'$, and $e_t = (\tilde{u}_t, \tilde{g}_t)'$. In (5b), $F = \text{diag}(\rho, \mu)$ and the disturbance term e_t follows $\text{iid}(0, R)$ where $R = \text{diag}(\sigma_u^2, \sigma_g^2)$. The matrices M and P

Table 1: Parameter values

β	δ	σ	ρ	μ	ψ_π	ψ_x
0.99	0.024	1/0.157	0.35	0.35	1.05	0.05

depend on the choice of the monetary policy. They are given by

$$M = \frac{1}{1 + \sigma\psi_x + \delta\sigma\psi_\pi} \begin{pmatrix} (1 + \sigma\psi_x)\beta + \delta\sigma & \delta \\ -\beta\sigma\psi_\pi + \sigma & 1 \end{pmatrix}, \quad (30a)$$

$$P = \frac{1}{1 + \sigma\psi_x + \delta\sigma\psi_\pi} \begin{pmatrix} 1 + \sigma\psi_x & \delta \\ -\sigma\psi_\pi & 1 \end{pmatrix}. \quad (30b)$$

We follow Woodford (2003) and Bullard et al. (2008) to calibrate the parameters of the model. Table 1 lists the values of parameters we choose to use. Since most parameter values are quite standard, we do not elaborate on the calibration process.

4.2 Limited information learning equilibrium

The full information REE of this model is well known. With standard parameter values, the system has a unique REE as long as the central bank's policy follows the long-run Taylor principle (Bullard and Mitra, 2002).

With full information, the learning agents' PLM is given by (6), which corresponds to the MSV solution under rational expectations. Agents estimate the parameters of (6) and use the results to make one-step forecast (7). Plugging (7) into (5a), we obtain the ALM as (8). It is also well known that the REE is E-stable if the Taylor principle is satisfied.

The equation that corresponds to inflation dynamics is

$$\pi_t = b_\pi v_t,$$

where b_π is a parameter. The inflation rate is a function of the shock variable alone. If the shock is not persistent, then inflation will “jump” to equilibrium on impact of any shock, and there will be zero persistence. With persistent shocks, the sole source of inflation persistence is the AR(1) shock process. As documented in the literature, the exogenous process is not persistent enough to match the level of inflation persistence we observe in the data.

With limited information, the exogenous processes, (27a) and (27b), are not observable. Agents will need to run VARs to form their forecasts.¹¹ In the following exercise, we put aside the issue of the endogenous determination of VAR lags, and consider six variants of VARs as agents’ forecasting models (the case of lag selection will be dealt with in the last subsection). In addition to the full VAR(p) type of forecasting functions, we will allow our agents to consider AR(p) models. These models have the same number of lags as VAR(p) but have fewer parameters (no off-diagonal elements). We take an interest in them because various studies, such as Adam (2007), find that in experiments, real-life agents’ expectations are often better characterized by simple ARs than more sophisticated models. Hommes and Zhu (2012) define a “behavioral learning equilibrium” for economies resided by agents who do AR(1) forecasting.

Without loss of generality, we let $p = 1, 2, 3$. That is, our agents have a set of six models to choose from. Table 2 lists the forecasting functions agents use.

The complexity of the model does not allow us to derive the limited information equilibrium analytically. We compute the equilibrium numerically as follows.

¹¹We do not consider other belief functions such as ARMA or VARMA. The results of our paper should extend to those specifications.

Table 2: Forecasting functions

VAR(p)	$y_t = \sum_{j=1}^p b_j y_{t-j} + \epsilon_t, \quad p = 1, 2, 3.$
AR(p)	$\pi_t = \sum_{j=1}^p b_{\pi,j} \pi_{t-j} + \epsilon_{\pi,t}, \quad p = 1, 2, 3.$ $x_t = \sum_{j=1}^p b_{x,j} x_{t-j} + \epsilon_{x,t}, \quad p = 1, 2, 3.$

The key is to compute the fixed point of $s(b) = b$. We adopt the numerical approach of Bullard et al. (2008), which they call the “*E-stability algorithm*.” It is a fixed point iteration

$$b_t = b_{t-1} + \kappa(s(b_{t-1}) - b_{t-1}), \quad (31)$$

where κ is a small positive number. An advantage of this approach is that it not only computes the equilibrium, but also eliminates unstable solutions. In other words, any equilibrium found by this algorithm is guaranteed to satisfy the E-stability condition.

Using this approach, we attempted to compute the limited information equilibrium for every forecasting function we defined in Table 2. It turns out that for each PLM, there exists a limited information learning equilibrium that is stable under learning. The convergence is robust given different initial values. We list the computed equilibrium matrix b for each case in Table 3.

Table 3: The equilibrium coefficients in various PLMs

VAR(1)	$\begin{pmatrix} 0.5819 & -0.0175 \\ 0.1390 & 0.5067 \end{pmatrix} y_{t-1} + \epsilon_t$
VAR(2)	$\begin{pmatrix} 0.5102 & 0.0028 \\ -0.0291 & 0.4567 \end{pmatrix} y_{t-1} + \begin{pmatrix} -0.0959 & -0.0024 \\ 0.0242 & -0.0503 \end{pmatrix} y_{t-2} + \epsilon_t$
VAR(3)	$\begin{pmatrix} 0.5390 & 0.0042 \\ -0.0443 & 0.4582 \end{pmatrix} y_{t-1} + \begin{pmatrix} -0.0983 & -0.0026 \\ 0.0269 & -0.0493 \end{pmatrix} y_{t-2} + \begin{pmatrix} 0.0198 & 0.0007 \\ -0.0078 & 0.0053 \end{pmatrix} y_{t-3} + \epsilon_t$
AR(1)	$\begin{pmatrix} 0.7478 & 0 \\ 0 & 0.4978 \end{pmatrix} y_{t-1} + \epsilon_t$
AR(2)	$\begin{pmatrix} 0.4980 & 0 \\ 0 & 0.4595 \end{pmatrix} y_{t-1} + \begin{pmatrix} -0.0850 & 0 \\ 0 & -0.0527 \end{pmatrix} y_{t-2} + \epsilon_t$
AR(3)	$\begin{pmatrix} 0.5198 & 0 \\ 0 & 0.4627 \end{pmatrix} y_{t-1} + \begin{pmatrix} -0.0859 & 0 \\ 0 & -0.0519 \end{pmatrix} y_{t-2} + \begin{pmatrix} 0.0156 & 0 \\ 0 & 0.0061 \end{pmatrix} y_{t-3} + \epsilon_t$

4.3 Inflation persistence

First, we compute the level of persistence for the economy that is in the limited information learning equilibrium. That is, we assume that agent learning has already led the economy to converge to the equilibrium, and we compute inflation persistence using those equilibrium dynamic equations. We do this to make a direct comparison of system dynamics in the two distinct equilibria: the limited information equilibrium and the full information REE. Dynamics of learning is considered in the next section.

Following Fuhrer (2009), we employ six measures of persistence for inflation. The first measure (AR1) is simply the autoregressive coefficient obtained by estimating an AR(1) process for inflation data. The second measure is the largest root of the univariate time series process (LR): if the autoregressive representation of inflation is of lag of length k

$$\pi_t = c_1\pi_{t-1} + c_2\pi_{t-2} + \dots + c_k\pi_{t-k} + e_t,$$

then the companion matrix for the state space representation is

$$C = \begin{bmatrix} c_1 & \dots & c_k \\ I & 0 & \end{bmatrix}.$$

LR is the largest root of the matrix C . The next two measures are closely related. $\text{sum}(\text{AR})$ is the sum of the autoregressive coefficients $c_1 \dots c_k$, and $\text{sum}(|\text{AR}|)$ is the sum of the absolute values of these coefficients. The fifth measure is originally suggested by Cogley et al. (2007). Persistence is measured by the R^2 of the j -step-ahead forecast of an VAR of inflation. The rationale is that higher persistence makes future inflation more predictable. Our sixth measure is the autocorrelation function (ACF). It is a vector of the autocorrelations of current inflation with a series of its own lags. For

Table 4: Inflation persistence in different equilibria (full and limit information)

Models	Measures of persistence				
<i>Limited Information</i>	AR1	LR	sum(AR)	sum(AR)	predictability
VAR(1)	0.6506	0.3500	0.5990	0.9111	0.4368
VAR(2)	0.4584	0.3500	0.4205	0.6042	0.2160
VAR(3)	0.4814	0.3500	0.4486	0.6258	0.2366
AR(1)	0.7478	0.5503	0.7016	1.0649	0.5737
AR(2)	0.4590	0.3500	0.4211	0.6045	0.2165
AR(3)	0.4811	0.3500	0.4481	0.6255	0.2363
<i>Full information</i>					
REE	0.3500	0.3500	0.3500	0.3500	0.1225

example, the i th autocorrelation coefficient in the vector is defined as

$$\rho_i = \frac{E(\pi_t \pi_{t-i})}{var(\pi_t)},$$

where the denominator is the variance of inflation. Given the ALM and parameter values, the ACF can be analytically computed for each LILE. Five scalar measures can then be calculated from the computed ACFs.

In Table 4, we compare the first five measures of persistence across six limited-information models and the full-information REE model. The full information benchmark case is listed in the last row of the table. The first column of the table lists all the forecasting functions we consider for learning agents. From the table, we can see that the level of persistence of the REE model is entirely governed by the calibrated persistence of the exogenous process ($\rho = 0.35$). There is no intrinsic persistence, as we argued above. In comparison, the six limited-information models display various degrees of intrinsic persistence. The levels of persistence in the six limited information models are almost all higher than the REE case. The VAR(1) model, for example, implies a first order autoregression coefficient of 0.65, almost twice the level of the full

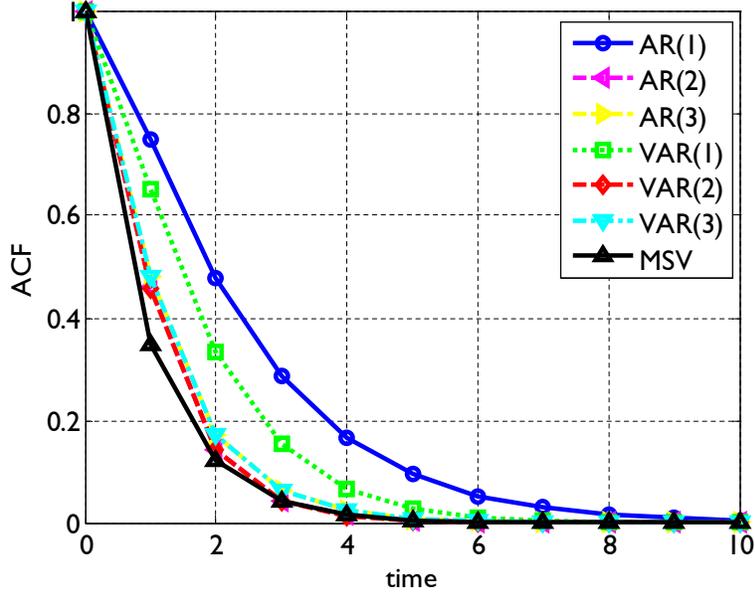


Figure 1: Inflation persistence in perfect information REE (denoted MSV) and LILEs information model. In terms of predictability, the VAR and AR models score at least twice as good as the REE model. For example, the VAR(3) model’s predictability is almost exactly twice that of the REE model. In Figure 1, we plot the ACFs of all seven models together. The autocorrelation functions of the REE model lies to the far left, indicating that inflation correlations over time is the lowest among all models. Among the six belief functions, VAR(1) and AR(1) perform the strongest in terms of generating higher ACFs. Out of the six measures of persistence, only one measure, the largest root (LR) does not produce significant differences between the full information and limited information cases. Only the AR(1) performs better than the REE models when measured with LR. The other five models’ LRs are identical to that of the REE model.

Thus, higher persistence arises in a standard model when agents have bounded rationality and limited information. This type of persistence is intrinsic, as it does not rely on any structural element of the model. Persistence is a direct result of expectation

formation and the self-referential nature of the model.

4.4 Endogenous selection of VAR orders

Our analysis in the previous subsection focuses on equilibrium dynamics. Agents' learning behavior was not explicitly modeled. We take up this task now. Agents will now estimate their VARs in each period to update their parameters and forecasts. They must decide the number of lags to use in the VAR and whether or not the AR models are better. They not only need to run regressions to estimate the parameters of their models, but also have to periodically re-evaluate their forecasting models with the BIC, and switch to the best-performing model accordingly.

One concern is that agents may choose higher and higher number of lags for their VARs as their sample size approaches infinity. This is possible because as the sample size increases, the cost of using more data gradually diminishes. As we state earlier, the case of infinite data availability is uninteresting as it does not realistically reflect the trade-off between model complexity and forecasting accuracy that real-life economists must face. It also does not let agents consider the possibility of varying system dynamics and structural breaks. We therefore implement a learning method that relies on finite data. Specifically, our agents learn with rolling window regressions. They abandon old data and only use the most recent N number of observations to estimate their model. Rolling window regressions are similar to constant gain learning (Friedman, 1979, Orphanides and Williams, 2004), and are frequently employed in empirical studies (e.g., see Swanson, 1998, Thoma, 2007, and Stark, 2010 among many others).

What is critical is the choice of N , the number of observations that the agent uses for model evaluation. Its choice directly affects which model will be chosen by the agent, because in the BIC formula (24), the relative weight between the error term and the penalty term depends on sample size N . In similar studies in the literature, a sample

size of 30–100 is considered to be empirically plausible. For example, Pesaran and Timmermann (2004) let $N = 60$, which corresponds to 15 years of quarterly observations. In constant gain learning, the sample size is approximately equal to one or two times the inverse of the learning gain. Orphanides and Williams (2005)’s estimate of this gain parameter using the Survey of Professional Forecasters is between 0.01 and 0.04. Milani (2006)’s estimate is 0.015. Many other papers report similar results, with the majority of the estimates being around 0.02. This is translated into a sample size of 50–100 in the rolling window regression. In the following analysis, we experiment with three different sample sizes: 100, 2,000 and 200,000. The last number is quite generous. We do this purposely to highlight what happens when the sample size approaches infinity.

We must also decide how often agents re-evaluate their model choices. In principle, agents could perform these evaluations in every period. They can also do so in every finite number of periods. We experiment with a few different values. It turns out that the core results are quite similar across different cases. We report the results for our benchmark case in which agents perform a BIC test every 100 periods.¹²

In summary, our simulation procedures are as follows. In period 0, agents randomly select a forecasting model among the six available choices. Using this model, they do least-square learning to estimate the model’s parameters for a fixed period of time (100 periods). After every 100 periods, they compute the BIC to evaluate which forecasting model is optimal. If the currently used model is optimal, they continue to use the model. Otherwise, agents will switch to the model that performs the best in the statistical test, and continue with parameter learning for another 100 periods.

To highlight the value-added of limited information learning, we also simulate agents’ learning process for the full information case, in which agents’ belief function is $y_t = a + bv_t$ and observe v_t .

¹²We will gladly provide results for other specifications upon requests.

Table 5: Level of inflation persistence with learning and endogenous model selection

Models	Measures of persistence				
	AR1	LR	sum(AR)	sum(AR)	predictability
Data	0.6202	0.7989	0.6784	0.9980	0.4208
Full information learning	0.3557	0.3767	0.3931	0.4532	0.1273
Limited information learning	0.6498	0.4594	0.6622	0.9922	0.4338

We measure the levels of persistence for this simulation and compare with them with the real data. Inflation data is obtained from the St. Louis Fed’s FRED database and spans 1959:Q1 and 2013:Q1. Inflation rates are calculated by taking log-differences of GDP deflators and annualize them. We use a band pass filter to detrend the data. This procedure follows that of Adam (2005).

The result is reported in Table 5 (see also Figure 2 for the ACF). With all six measures, our comparison with the full information case shows that limited information has led to a significant rise in the level of inflation persistence and is much more comparable to the level of persistence in the real data. Persistence comes from two sources: one is the intrinsic persistence embedded in the *equilibrium* dynamics. This is the same mechanism that leads to higher persistence in the previous subsection. The second source of persistence comes from the learning behavior of agents itself.

Two other findings emerge from this analysis. The first is that as the sample size increases, agents tend to select a more complex VAR as their forecasting model. This is not surprising. What is unexpected is that it takes a very large sample size for them to make such decisions. Table 6 reports the estimated probability that each forecasting model will be chosen by agents, based on simulations of the model for 1 million periods.¹³ When the sample size is 100 or 2000, agents tend to select AR(1) or AR(2) models. It is only in the third case, when the sample size is 200,000, that agents

¹³The simulation had to be run for a million periods because in the third case, agents’ sample size is 200,000.

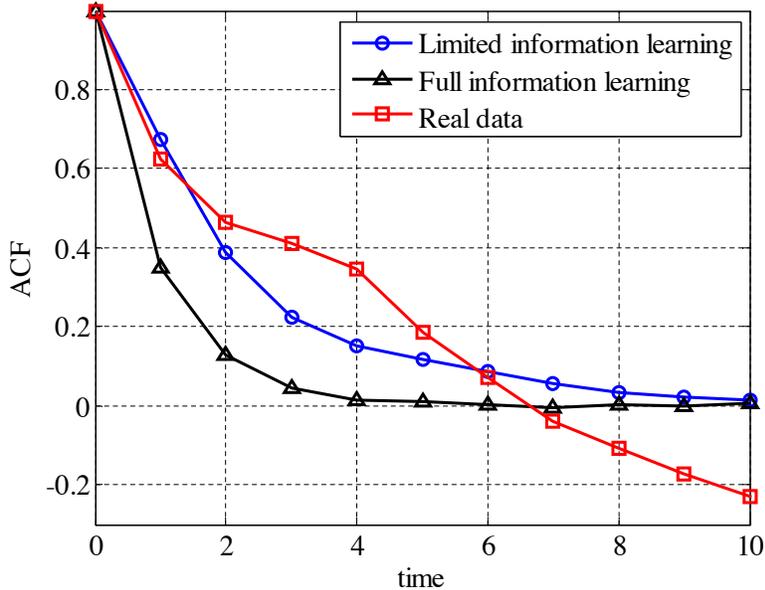


Figure 2: Autocorrelation function for full and limited information.

select the VAR(3) model with a high probability. Since large sample sizes are typically not available to real-life agents, it is reasonable to conclude that simpler belief functions are more likely to prevail in reality. This result provides a justification for using simple “misspecified” belief functions for boundedly rational agents. Most researchers argue for simple belief functions on the ground that they are consistent with experimental or survey results (Adam, 2007). Our analysis demonstrates that even agents who are equipped with complex statistical tools may choose simple belief functions, because when there are limited observations, these functions perform the best statistically.¹⁴

Our second finding is closely related to the first one. Notice that in Table 6, no forecasting model has a convergence probability of 1. This implies that even agents tend to select a particular model most of the time, there will be times when they decide to switch to a different model for forecasting purposes. This phenomenon is sometimes

¹⁴The results in the three tables depend on the model selection criterion we use, the BIC. We also experimented with another criterion, the AIC, and yielded similar results. We conjecture that with other statistical criteria, the quantitative relationship between N and the optimal model may change slightly, but our qualitative evaluation of the relationship should remain true.

Table 6: Estimated probability that each model will be selected

Sample Size	Models: VAR(1), VAR(2), VAR(3), AR(1), AR(2), AR(3)
100	0.0615 0.0105 0.0037 0.7048 0.1923 0.0272
1,000	0.0046 0.0033 0 0.2232 0.7443 0.0246
200,000	0 0.198 0.9332 0 0 0.0470

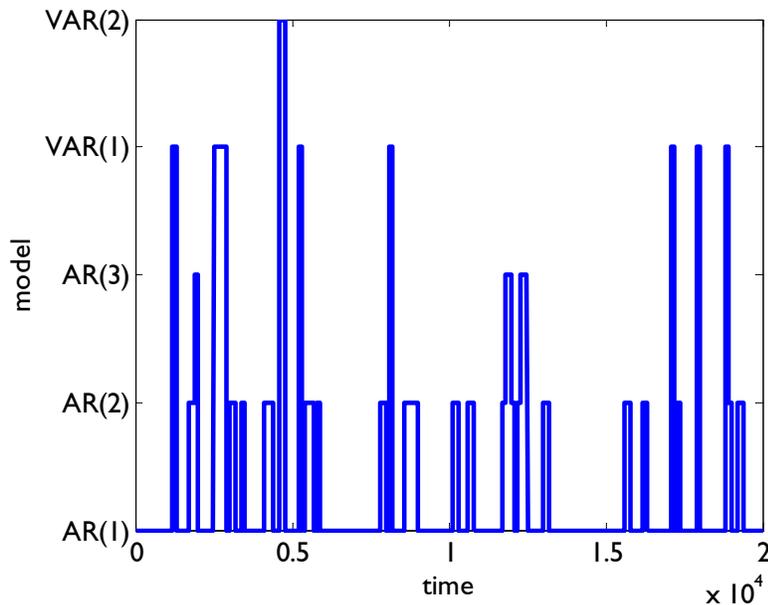


Figure 3: Model switching with limited information learning

referred to as “escape dynamics,” as in Williams (2001), Williams (2002), Cho et al. (2002) and Cho and Kasa (2011).

Figure 3 shows the model switching pattern in the simulation in which the sample size is 100. The AR(1) model has the highest rate of occupancy, and the AR(2) model ranks second. Most of the time, agents’ decisions fluctuate between these two models. But occasionally, agents also switch to other models. Higher order models such as the VAR(3) model is rarely used. This is no doubt due to the limited sample size we impose on the agents. Agents switch their models quite frequently. This is because they are not able to estimate the parameters of the true model robustly, given the

limited number of data available to them. Our underlying assumption is that agents discount old data when estimating their forecast models. This makes their parameter estimates fluctuate around the true equilibrium parameter values but never converge to them. If the estimated parameters deviate too far away from the true values, the “true” forecasting model will predict poorly and be outperformed by a model which is less accurate but have more robust parameters. As we pointed out in Section 3, agents face a dilemma of choice between precise models and precise parameters. Selection criteria such as the BIC help them make such a choice, but also makes it more likely for them to switch away from their initial models.

4.5 Discussion

Early inflation models can reproduce the level of inflation persistence in the data because they incorporate lags of inflation in the Phillips curve (see, for example, Gordon (1982), among others). Standard rational expectations models fail to capture the same level of persistence primarily because these models are forward-looking in the core, which implies prices should jump in response to shocks. To remedy this, there must be a reconciliation between the “backward correlation” in the data and the forward-looking nature of the model. Perhaps because of this, efforts to resolve this puzzle often focus on deriving a new Phillips curve that has lagged terms for inflation. For example, Mankiw and Reis (2002)’s sticky information approach, Christiano et al. (2005)’s price indexation, and Galí and Gertler (1999)’s rule of thumb consumers all produce new Phillips curves that have lagged terms.

Our approach is innovative in that it does not require the new Phillips curve to have any lagged terms. In the system (5a), the structural form of the Phillips curve is purely forward-looking. Yet due to the unobservability of shocks, agents must reply on past values of endogenous variables to extra information about the unobservables.

Consequently, the solution of the model (the actual law of motion) necessarily includes lagged terms of these variables. Consider our version of the inflation equation in the equilibrium:

$$\pi_t = \sum_{j=1}^p c_j \pi_{t-j} + \sum_{j=1}^p d_j x_{t-j} + \epsilon_t.$$

Inflation is a function of lagged inflation in equilibrium. This equation is reminiscent of early models of the accelerationist Phillips curve. Indeed, even the rationale for including the lagged terms is suggestive of the early literature, which argues that agents' expectations should depend on past values of inflation.

Thus, without resorting to model complications such as price indexing, we have found a way to make the Phillips curve more conformable with empirical estimates. The key element that makes this possible is bounded rationality and realistic forecasting behavior of learning agents.

5 Conclusion

In this paper we propose an approach of adaptive learning that can better satisfy the cognitive consistency principle. We impose two information limitations on learning agents: they cannot observe the fundamental shocks, and they only possess finite data. To forecast future endogenous variables, agents solve signal extraction problems by running VARs. We show that under this assumption, agents no longer have perceived laws of motion that are consistent with the rational expectations equilibrium (REE). This leads to limited information learning equilibria that are generally different from full information REEs. Our application of the approach to the new Keynesian model demonstrates that the dynamics of model are significantly different from those of the standard learning models. The level of inflation persistence is amplified considerably in the model with limited information.

Given the result of this paper, chances now exist for us to study other topics using the limited information approach. Potentially a large number of issues can be examined using this new approach. One area that we are particularly interested in is the stability under learning of such equilibria under different monetary policy regimes. This will be our next project.

References

- Adam, Klaus**, “Learning to Forecast and Cyclical Behavior of Output and Inflation,” *Macroeconomic Dynamics*, 2005, *9*, 1–27.
- , “Experimental Evidence on the Persistence of Output and Inflation,” *The Economic Journal*, 2007, *117*, 603–636.
- Akaike, Hirotugu**, “A New Look at the Statistical Model Identification,” *IEEE Transactions on Automatic Control*, 1974, *19*, 716–723.
- Branch, William A. and George W. Evans**, “Intrinsic Heterogeneity in Expectation Formation,” *Journal of Economic Theory*, 2006, *127*, 264–295.
- **and** – , “Model Uncertainty and Endogenous Volatility,” *Review of Economic Dynamics*, 2007, *10*, 207–37.
- Bray, Margaret M.**, “Learning, Estimation, and the Stability of Rational Expectations,” *Journal of Economic Theory*, 1982, *26*, 318–339.
- Bullard, James B. and Kaushik Mitra**, “Learning about Monetary Policy Rules,” *Journal of Monetary Economics*, 2002, *49*, 1105–1129.
- , **George W. Evans**, **and Seppo Honkapohja**, “Monetary Policy, Judgment and Near-Rational Exuberance,” *American Economic Review*, 2008, *98*, 1163–1177.

Cho, In-Koo and Kenneth Kasa, “Learning and Model Validation,” Working paper 2011.

– , **Noah Williams, and Thomas J. Sargent**, “Escaping Nash Equilibrium,” *Review of Economic Studies*, 2002, 69, 1–40.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy,” *Journal of Political Economy*, 2005, 113, 1–45.

Clarida, Richard, Jordi Galí, and Mark Gertler, “The Science of Monetary Policy: A New Keynesian Perspective,” *Journal of Economic Literature*, 1999, 37, 1661–1707.

Cogley, Timothy, Giorgio E. Primiceri, and Thomas J. Sargent, “Inflation-Gap Persistence in the U.S,” working paper 2007.

Evans, George W., “Expectational Stability and the Multiple Equilibria Problem in Linear Rational Expectations Models,” *Quarterly Journal of Economics*, 1985, 100, 1217–1233.

– **and Seppo Honkapohja**, *Learning and Expectations in Macroeconomics*, Princeton University press, 2001.

– **and** – , “Learning as a Rational Foundation for Macroeconomics and Finance,” Bank of Finland Working paper 2011.

Fernandez-Villaverde, J., J. Rubio-Ramirez, T. Sargent, and M. Watson, “ABCs (and Ds) of Understanding VARs,” *American Economic Review*, 2007.

- Friedman, Benjamin M.**, “Optimal Expectations and the Extreme Information Assumptions of ‘Rational Expectations’ Macromodels,” *Journal of Monetary Economics*, 1979, 5, 23–41.
- Fuhrer, Jeffery C.**, “Inflation Persistence,” Federal Reserve Bank of Boston Working Paper 2009.
- **and George R. Moore**, “Inflation Persistence,” *Quarterly Journal of Economics*, 1995, 110, 127–159.
- Galí, Jordi and Mark Gertler**, “Inflation Dynamics: A Structural Econometric Analysis,” *Journal of Monetary Economics*, 1999, 44, 195–222.
- Giannoni, Marc P. and Michael D. Woodford**, “How Forward Looking is Optimal Monetary Policy?,” *Journal of Money, Credit and Banking*, 2003, 35, 1425–69.
- Gordon, Robert J.**, “Price Inertia and Policy Ineffectiveness in the United States, 1890–1980,” *Journal of Political Economy*, 1982, 90, 1087–1117.
- Gregoir, Stephane and Pierre-Olivier Weill**, “Restricted Perception Equilibria and Rational Expectation Equilibrium,” *Journal of Economic Dynamics and Control*, 2007, 31, 81–109.
- Hellwig, Christian**, “Monetary Business Cycle Models: Imperfect Information,” New Palgrave Dictionary of Economics 2006.
- Hommes, C. and M. Zhu**, “Behavioral Learning Equilibria,” 2012. CeNDEF Working Paper.
- Ireland, Peter**, “A Method for Taking Models to the Data,” *Journal of Economic Dynamics and Control*, 2004, 28, 1205–1226.

- King, Robert G. and Sergio T. Rebelo**, *Resuscitating Real Business Cycles*, *Handbook of Macroeconomics*, Elsevier, 1999.
- Levine, P., Pearlman, and Perendia**, “Estimating DSGE Models Under Partial Information,” 2007. CDMA working paper 07/22.
- Lucas, Robert E., Jr.**, “An Equilibrium Model of the Business Cycle,” *Journal of Political Economy*, 1975, *83*, 1113–1144.
- , “Adaptive Behavior and Economic Theory,” *The Journal of Business*, 1986, *59*, S401–S426.
- Lütkepohl, Helmut**, *Introduction to Multiple Time Series Analysis*, Springer-Verlag, 1991.
- Mankiw, N. Gregory and Ricardo Reis**, “Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve,” *Quarterly Journal of Economics*, 2002, *117*, 1295–1328.
- Marcet, Albert and Thomas J. Sargent**, “Convergence of Least-Squares Learning in Environments with Hidden State Variables and Private Information,” *Journal of Political Economy*, 1989, *97*, 1306–1322.
- Milani, Fabio**, “A Bayesian DSGE Model with Infinite-Horizon Learning: Do ‘Mechanical’ Sources of Persistence Become Superfluous?,” *International Journal of Central Banking*, 2006, *2*, 87–106.
- , “Expectations, Learning and Macroeconomic Persistence,” *Journal of Monetary Economics*, 2007, *54*, 2065–2082.
- Muth, John F.**, “Rational Expectations and the Theory of Price Movements,” *Econometrica*, 1961, *29*, 315–335.

- Orphanides, Athanasios and John C. Williams**, “Imperfect Knowledge, Inflation Expectations, and Monetary Policy,” NBER Chapters, in: *The Inflation-Targeting* 2004.
- **and** –, “The Decline of Activist Stabilization Policy: Natural Rate Misperceptions, Learning and Expectations,” *Journal of Economic Dynamics and Control*, 2005, *29*, 1927–1950.
- Pearlman, Joseph G. and Thomas J. Sargent**, “Knowing the Forecasts of Others,” *Review of Economic Dynamics*, 2005, *8*, 480–497.
- Pesaran, M. Hashem and Allan Timmermann**, “How costly is it to ignore breaks when forecasting the direction of a time series?,” *Journal of International Forecasting*, 2004, *20*, 411–425.
- Preston, Bruce**, “Learning about Monetary Policy Rules when Long-Horizon Expectations Matter,” *International Journal of Central Banking*, 2005, *1*, 81–126.
- Ravenna, Federico**, “Vector autoregressions and reduced form representations of DSGE models,” *Journal of Monetary Economics*, 2007, *54* (7), 2048–2064.
- Rissanen, Jorma**, “Modeling by Shortest Data Description,” *Automatica*, 1978, *14*, 465–471.
- Rondina, Giacomo and Todd B. Walker**, “Information Equilibria in Dynamic Economies,” Working Paper 2011.
- Sargent, Thomas J.**, “Equilibrium with signal extraction with endogenous variables,” *Journal of Economic Dynamics and Control*, 1991, *15*, 245–273.

- Schorfheide, Frank**, *Estimation and Evaluation of DSGE Models: Progress and Challenges*, Vol. Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress, Cambridge University Press, 2013.
- Schreyer, Paul and Colin Webb**, “Capital Stock Data at the OECD – Status and Outlook,” OECD Report 2006.
- Schwarz, Gideon**, “Estimating the Dimension of a Model,” *Annals of Statistics*, 1978, 6, 461–464.
- Smets, Frank and Rafael Wouters**, “Monetary Policy in an Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area,” *Journal of the European Economic Association*, 2003, 1, 1123–75.
- Stark, Tom**, “Realistic Evaluation of Real-Time Forecasts in the Survey of Professional Forecasters,” Special Report, Federal Reserve Bank of Philadelphia 2010.
- Swanson, Norman R.**, “Money and Output Viewed Through a Rolling Window,” *Journal of Monetary Economics*, 1998, 41, 455–474.
- Taylor, John B.**, “Discretion versus Policy Rules in Practice,” *Carnegie-Rochester Conference Series in Public Policy*, 1993, 39, 195–214.
- Thoma, Mark**, “Structural Change and Lag Length in VAR Models,” *Journal of Macroeconomics*, 2007, 30, 965–976.
- Townsend, Robert M.**, “Forecasting the Forecasts of Others,” *Journal of Political Economy*, 1983, 91, 546–588.
- Wallace, Christopher Stewart and David M. Boulton**, “An Information Measure for Classification,” *Computer Journal*, 1968, 11, 185–194.

Williams, Noah, “Escape Dynamics in Learning Models.” PhD dissertation, University of Chicago 2001.

– , “Stability and Long Run Equilibrium in Stochastic Fictitious Play,” Working Paper, Princeton University 2002.

Woodford, Michael, *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, 2003.

A Appendix

A.1 A more general reduced form model

We consider more general form of economic model with a lagged value of endogenous variable y_t

$$y_t = ME_t^* y_{t+1} + Ny_{t-1} + Pv_t,$$

$$v_t = Fv_{t-1} + e_t,$$

where y_t is an $n \times 1$ vector of endogenous variables, and v_t is an $m \times 1$ vector of exogenous variables which follow a stationary VAR. e_t is an $m \times 1$ vector of white noise. E_t^* represents expectations of private agents that are not necessarily rational.

A.1.1 Perfect Information REE

MSV form of PLM is

$$y_t = by_{t-1} + cv_t.$$

Agents form expectations

$$E_t y_{t+1} = b^2 y_{t-1} + (bc + cF)v_t.$$

ALM becomes

$$y_t = (Mb^2 + N)y_{t-1} + (M(bc + cF) + P)v_t.$$

The REE is defined as the fixed point of

$$b = T(b) = Mb^2 + N$$

$$c = T(c) = M(bc + cF) + P.$$

A.1.2 Limited information learning equilibrium (LILE)

In the REE, the actual law of motion can be expressed as VAR(∞) form

$$y_t = \sum_{j=1}^{\infty} D_j y_{t-j} + \epsilon_t$$

where

$$D_j = \begin{cases} b + cP^{-1}, & j = 1 \\ -(cFc - cP^{-1})^{j-2} cP^{-1} (cFc^{-1} + b + cP^{-1} + 2PF P^{-1}), & j \geq 2. \end{cases}$$

Agents may have VAR(p) form of PLM

$$y_t = \sum_{j=1}^p b_j y_{t-j} + \epsilon_t.$$

Defining $Y_{t-1} = (y'_{t-1}, y'_{t-2}, \dots, y'_{t-p})'$, it is also written as

$$y_t = bY_{t-1} + \epsilon_t.$$

Agents form expectations

$$E_t^* y_{t+1} = \sum_{j=1}^p \tilde{b}_j y_{t-j}$$

where

$$\tilde{b}_j = \begin{cases} b_1 b_j + b_{j+1}, & j = 1, \dots, p-1 \\ b_1 b_j. & j = p. \end{cases}$$

The ALM becomes

$$y_t = (M\tilde{b}_1 + N)y_{t-1} + M \sum_{j=2}^p \tilde{b}_j y_{t-j} + PFv_{t-1} + Pe_t.$$

It can also be written as

$$z_t = \sum_{j=1}^p T_j(b) z_{t-j} + Ve_t,$$

where $z_t = (y'_t, v'_t)'$, $T_1(b) = \begin{pmatrix} M\tilde{b}_1 + N & PF \\ 0 & F \end{pmatrix}$, $T_j(b) = \begin{pmatrix} M\tilde{b}_j & 0 \\ 0 & 0 \end{pmatrix}$ when $j \geq 2$,

and $V = \begin{pmatrix} P \\ I \end{pmatrix}$.

The ALM is further simplified to

$$z_t = T(b)Z_{t-1} + Ve_t,$$

where $T(b) = (T_1(b), T_2(b), \dots, T_p(b))$ and $Z_{t-1} = (z'_{t-1}, z'_{t-2}, \dots, z'_{t-p})'$.

With the PLM and ALM given as above, the projection of y_t on Y_{t-1} can be calculated as

$$P[y_t|Y_{t-1}] = s(b)Y_{t-1},$$

where $s(b)$ is the projection matrix defined as

$$s(b) = i_n T(b) (E_t Z_{t-1} Y'_{t-1}) (E_t Y_{t-1} Y'_{t-1})^{-1}.$$

The selection matrix $i_n = [I \ 0]$ picks out the vector y_t from the projected z_t .

A limited information learning equilibrium (LILE) is a matrix $b = (b_1, b_2, \dots, b_p)$ that satisfies

$$b = s(b).$$

A.2 Marcet and Sargent (1989)'s framework

We show how our approach fits into Marcet and Sargent's framework as a special case.

Marcet and Sargent define their model as follows. There is a state vector z_t . Let $z_{it} = e_i z_t$, where e_i is a selector matrix and z_{it} is a subvector of z_t . The index i represents the types of agents. There are two types of agents, a and b . Type a agents only observe z_{at} , and type b agents only observe z_{bt} . Agents of type j want to predict future values of subvectors $z_{k(j)} = e_{k(j)} z_t$, where $k(a) = c$ and $k(b) = d$. z_{ct} and z_{dt} are possibly distinct subvectors from z_{at} and z_{bt} . The beliefs of agents are given by

$$E(z_{ct}|z_{at-1}) = \beta_a z_{at-1},$$

$$E(z_{dt}|z_{bt-1}) = \beta_b z_{bt-1}.$$

Given these beliefs, the actual law of motion is

$$z_t = T(\beta)z_{t-1} + V(\beta)\epsilon_t,$$

where ϵ_t is a vector of white noise.

Our model is

$$y_t = ME_t^* y_{t+1} + Pv_t, \tag{32}$$

$$v_t = Fv_{t-1} + e_t, \tag{33}$$

If we define

$$z_t = \begin{pmatrix} y_t \\ v_t \end{pmatrix},$$

$$z_{at} = z_{bt} = y_t,$$

$$\epsilon_t = e_t,$$

and let $\beta_a = \beta_b$, and $z_{at} = z_{bt} = z_{ct} = z_{dt}$, then our model becomes a special case of Marcet and Sargent's framework.