The Limits of Learning

Elliot Aurissergues

The Limits of Learning

Elliot Aurissergues

DECEMBER 2014

First Draft

ABSTRACT

In this paper, we criticize the current adaptive or statistical learning literature. Instead of emphasizing asymptotical results, we focus on the short run forecasting performance of the different algorithms before convergence to rational expectation solution occurs. First, we suggest that the literature should drop ordinary least squares techniques in favor of the more efficient Bayesian estimation. Second, we cast doubt on the rationality of the behavior implied by the theory. We argue that agents do not use all available information in these models. Past prices carry some information about expectations of others and some algorithms are able to exploit this information. In a very simple case, this algorithm is simply naive expectations. In more complex one, we augment the usual learning with an estimation of past expectation errors using Kalman Filter. Interestingly, we find that some of these algorithms are divergent and may beat convergent ones in the short run. For a large set of parameters, their dominance is too short to be significant. However, when the sensitivity of the actual price to the expected one is close to one, divergent algorithms should be considered.

JEL Classification: D83,D84
Keyword: Adaptive learning, cobweb model, naive expectations, Kalman Filter

Introduction

Rational expectations is the dominant way of thinking about how agents forecast the future in macroeconomic models. Since the Muth's seminal paper, the central idea of rational expectations is that the economic agents of the model and the theorist who builds the model have the same level of knowledge about the model and make similar forecasts if their information set is also similar.
Since its adoption by most of macroeconomics, This hypothesis has attracted a lot of controversies. The biggest issue is it implies thet agents have an enormous level of knowledge. They know the structure of the model, the behavioral parameters, the laws followed by random variables. It seems very unrealistic. the learning literature have been developed to adress these issues. They relax the assumption of parameters knowledge. Instead of behaving like theorist, agents are "econometricians" and guess parameters values using ordinary least squares, maximum likelihood, or bayesian estimation of the reduced form of the Rational Expectation solution.
However, the learning literature have failed to develop a credible alternative to rational expectations. Conversely, Its main finding has become a frequent justification of rational expectations. Indeed, Bray and Savin (1986) have shown that recursive ordinary least squares algorithm usually converges to true RE parameters in a cobweb economic model if the actual price is not too sensitive to the expected price. This result is robust to other model's specifications and other econometric techniques.
Although the results of the learning theory are quite powerful, there are reasons to be unconfortable with some trends of the literarure.
First, The literature focus mostly on asymptotical result. But the asymptotical forecasting performance is not really relevant for an economic agent which wants to make decisions in the short run. The adoption of a forecasting method should be the outcome of "rational" individual choices and a common algorithm for all agents suppose some form of Nash Equilibria. Suppose that we are in an economy dominated by some algorithms, if this algorithm is dominated by another one, it should not be considered. We think that the literature have not explored the performance of the different algorithms enough. For example, it relies mostly on recursive least squares techniques. But this type of algorithm performs poorly when only few datas are available which is the case at the beginning of the process. In some cases, it is outperformed by simple naive forecasting for a long period of time. By contrast, bayesian estimation algorithm performance is much better. We recommand that future works switch to bayesian methods.
More importantly, the exploration of the short run performance of learning leads us to question its rationality. Indeed, in the literature, agents learn the rational expectation or fundamental solution.We label this theory the "fundamental learning" But, the true data generating process is affected by expectations errors and thus by the learning process itself. Learning only the fundamental solution is a misspecification. This is aknowledge by the literature but the answers was not really satisfactory. The main argument is that reflects some form of bounded rationality justified by the difficulty to detect the misspecification (Bary and Savin 1986, Evans and Honkapoja 2001). Our issue is that algorithms or similar complexity or even simpler may take this misspecification into account and outperforms those considered by the literature. When expectations

are formed using learning algorithm, they are slowly updated with recent values of forecasts errors. Current and past expectations are correlated. But, past prices incorporates some information about past expectations. If an agent is able to extract this information, he may guess current forecasts and in principle may beat fundamental learning.
In a nonstochastic, simple framework, an obvious way to extract such information is to use past price as the best forecast for the new one. It is naive expectations. We show that they beat bayesian estimation in many cases.
In a more complex framework with exogenous observables and noise, information extraction is much more challenging. We show that an improved algorithm using a Kalman filter may outperform fundamental learning especially in the case of strategic complementarity. An interesting aspect is that Some variants of this algorithm are divergent but outperforms the convergent one in the short run.
The first section presents framework, a cobweb model derived from Evans and Honkapoja (2001). The second compares the performance of recursive least squares and recursive bayesian estimation. the third examine the reasonability of the fundamental learning in various environments.

1 Cobweb Model

1.1 The learning framework

The classical example of the expectation literature is the cobweb model, introduced by Ezekiel (1938). In order to get tractable results and to compare them easily with the previous literature, we also adopt this framework. Our notations follows closely the first chapter of the Evans and Honkapoja (2001) textbook.

1.1.1 The cobweb model

We study a partial equilibrium problem where production should be decided in advance. Unlike walrasian markets, supply reacts with expected prices and not with actual prices. Demand equation is usual giving system of equations of the type

$$Q_t^d = m_I + m_w w_t - m_p p_t + v_{1t} \tag{1a}$$
$$Q_t^s = r_p p_t^e \tag{1b}$$

p_t is the actual price, p_t^e is the expected one.
w_t is a vector of exogenous variables affecting demand. We assume that w_t follows an AR(1) process defined by

$$w_t = \rho w_{t-1} + \varepsilon_t \tag{2}$$

ε_t follows a white noise of standard deviation σ_ε v_{1t} is a random component of demand. It is not observed by producers. The former is observable by suppliers before they decide their production whereas the latter is not. Without loss of generality, the supply function depends only from expected prices.
Market clearing leads to the following reduced form for the equilibrium price.

$$p_t = \mu + \delta w_t + \alpha p_t^e + \eta_t \tag{3}$$

with $\alpha = -\frac{r_P}{m_P}$, $\delta = \frac{m_w}{m_p}$, $\mu = \frac{m_I}{m_p}$ and $\eta_t = \frac{v_{1t}}{m_I}.\eta_t$ follows a white noise process of standard deviation σ_η
The equation establishes a clear link between equilibrium price and price expectations. The determination of the former requires to make an assumption about how agents forecast. Muth (1956) have proposed the now standard rational expectation assumption $p_t^e = E_t p_t$ where E is the "true" expected value of the price.
p_t^{RE} is the rational expectation solution of the model. It is straightforward to compute it.

$$p_t^{RE} = \frac{\mu}{1-\alpha} + \frac{\delta}{1-\alpha} w_t \tag{4}$$

The rational expectation solution of the model takes the form $p_t^{RE} = \phi^* z_t$ where ϕ^* is the vector $\begin{pmatrix} \frac{\mu}{1-\alpha} \\ \frac{\delta}{1-\alpha} \end{pmatrix}$ and z_t is the vector $\begin{pmatrix} 1 \\ w_t \end{pmatrix}$.

1.1.2 The learning algorithm

We can see in equation (3) that rational expectations requires to know the vector of parameters. Otherwise, computing the expected value of the equilibrium price is impossible. Obviously, it is a very strong assumption. Agents are not naturally endowed with the knwoledge of these parameters. It seems a decisive flow in the REH. The learning theory adress this problem. The dominant approach have been to relax the assumption of parameter knowledge but instead to assume that agents learn the parmaeter values of the Rational Expectation solution using econometric techniques. Agents do not behave like theorists but like econometricians.
More precisely, at period t, using datas from the beginning of the learning process to period $t-1$, Agents estimate a model of the form

$$p_t = a + b w_t \tag{5}$$

The estimator is denoted p_t^e and is written

$$p_t^e = a_t + b_t w_t \tag{6}$$

Agents learn the reduced form of the rational expectation solution.
The model is reestimated at each period t using the new data provided by the price and observables of the last period. Thus, the learning process take the form of a recursive algorithm.
There are several proposal for the econometric techniques and the recursive algorithm. They have similar asymptotic properties but may widely differ at the early stage of the recursive process. The dominant approach in the literature have been the ordinary least squares estimation and the recursive least squares algorithm. Let's denote R the variance covariance matrix of parameter estimates and ϕ_t the estimation of the parameter vector at period t. The algorithm is given by the two recursive equations

$$R_{t+1} = R_t^{-1} + \frac{1}{t}(z_t z_t' - R_t) \tag{7a}$$

$$\phi_{t+1} = \phi_t + R_{t+1}\frac{1}{t} z_t \left(p_t - z_t' \phi_t\right) \tag{7b}$$

We do not follow that path. Instead, we prefer bayesian estimation. Although both converges to the RE solution, bayesian estimation peforms better with few datas. In other word, the expectatioonal errors generated by the bayesian estimation is inferior to those created by ordinary least squares for a significant aprt of the algorithm. We want precisely to compare the forecasting performance of the different algorithm, especially at the initial stages of the process, so using the more performant algorithm seems logical.
The recursive algorithm for bayesian estimation follows Bullard and Suda (2009).

$$R_{t+1}^{-1} = R_t^{-1} + \nu^{-2} z_t z_t^{'} \tag{8a}$$

$$\phi_{t+1} = \phi_t + R_{t+1} \nu^{-2} z_t \left(p_t - z_t^{'} \phi_t \right) \tag{8b}$$

ν is the standard deviation of the random component η. An issue is that, in principle, bayesian estimation cannot be implemented without knwoing ν and there are no reasons that agents know this parameter. However, the results are not strongly affected if agents take a wrong value for ν. Thus, we prefer interpret it like a gain paramater. The gain is optimal if it corresponds to the true value of the standard deviation, but falls reasonably for other values.
Once you have computed the parameter estimates at date t, you have the perceived law of motion (PLM thereafter) which is the equation (5). Combining this perceived law of motion with the equilibrium price equation (2), you get the actual law of motion

$$p_t = \mu + \alpha a_t + \delta + \alpha b_t w_t + \eta_t \tag{9}$$

1.1.3 Convergence

Bray and Savin (1986) have shown that the recursive least squares algorithm converges to Rational Expectations in a simplified version of the model. this result has been extended by Marcet and Sargent(1989) and Evans and Honkapoja (2001) in more complex environments. Bullard and Suda (2009) provides a similar proof for bayesian estimation algorithm. Modern proofs (Evans and Honkapoja 1989,1992) are based on the stochastic differential equations theory. Indeed, recursive algorithm may be approximated by differential equations. The convergence of the differential equations imply the convergence of the algorithm under some conditions.
The differential equations considered is

$$\frac{d\phi}{d\tau} = T(\phi) - \phi \tag{10}$$

The application T maps the parameter estimates with the "true" value of parameters

$$T \begin{pmatrix} a_\tau \\ b_\tau \end{pmatrix} = \begin{pmatrix} \mu + \alpha a_\tau \\ \delta + \alpha b_\tau \end{pmatrix} \tag{11}$$

Recursive algorithm defines for example by the systems (6) or (7) converge to the true Rational Expectation solution if the differential equation (9) is asymptotically stable. In our model it is the case if $\alpha < 1$.
The intuition of this result is relatively straightforward. Suppose that expectations have no impact on equilibirum outcomes. Agents learn the true model

and the econometric theory tells us that they will converge to true parameter values. Back to our case, expectations affects equilibrium outcomes, so residuals are no longer normally distributed around true values. This is not an issue if the forecast errors remains with the same sign. If $\alpha \geq 1$ the expected price is between the rational expectation price and the equilibrium price, so the forecast error push expectations far from the rational expectation price.

2 Bayesian Estimation vs Recursive Least Squares

Most of the literature have considered the algorithm (6) or some variants of this algorithm. This choice was logical. Ordinary least squares is the simplest econometric technique and remains one of the more popular among macroeconometricians when the learning literature began. The posterior literature have not modified this dominance. The literature focus mainly on asymptotical result, and the unavoidable conclusion was that all the algorithms converges to Rational Expectations for a similar set of parameters, so there was no need to switch to other algorithms.
Although all the algortithm converges under the same criteria, their behavior before convergence differs. in particular, Recursive OLS implies huge initial errors if we initialize the process with a minimal dataset. Indeed, compare the system (6) and (7). There are two differences. First the gain parameter is $\frac{1}{t}$ for OLS and ν^{-2} for bayesian estimation. The first is decreasing whereas the other is constant. [1] But the main difference is the updating of the variance covariance matrix. In the OLS learning, priors on standard deviation of exogenous variables play no role whereas they do in bayesian estimation. Indeed, suppose that $t = 1$. In the OLS algorithm

$$R_1 = z_1 z_1^{'}.$$

R_0 disappear. In bayesian estimation

$$R_1^{-1} = R_0^{-1} + \nu^{-2} z_1 z_1^{'}$$

A R_0 term remains. Parameters are treated like random variables and R_0 is a prior on their standard deviation. The matrix is less sensitive to the first data point because it incorporates prior information. With one data point, the estimation is much more reliable with bayesian method. To iluustrate that, we compare in figure 1 the forecasts performance of OLS and Bayesian estimation in an economy where the majority of agents use OLS (more precisely a negligible measure of agents use bayesian methods). The calibration is similar to the first chapter of Evans and Honkapoja (2001). The two algorithm are initialized with the same prior values. After two data points, the error of the bayesian algorithm is three times inferior to those of the OLS algorithm. The latter is equivalent to sixty percent of the average rational expectation price. The bayesian algorithm outperforms OLS for fifty periods. Next, the OLS algorithm holds a slight advantage but both are near the true solution.
A related problem of ordinary least squares is that you need as much data point as exogenous variables to initialize the process. For example, if you have ten

[1] A rich literature have studied the OLS algorithm with a constant gain parameter. Unlike Bayesian estimation, they generally do not converge

exogenous variables, you need at least ten data points to form an estimation of parameters. But the ten value of prices depend from expectations which are not defined! A solution is to use the prior value for parameters for ten periods and start the process after. When you have less exogenous variables, it can also be useful to reduce the error at the beginning of the algorithm. The problem is that these datas are contaminated by inaccurate expectations and slowdown convergence reducing the overall performance of the algorithm. Let's compare it with naive forecasting in figure 2. Naive expectations dominates statistical learning for 40 periods! By contrast look at figuer five which compares naive expectations and bayesian estimation in a similar economy but where teh majority of agents use bayesian estimation. Naive expectations dominates for only four or five periods. The incentive to deviate from the fundamental learning is much less important.

3 Learning fundamentals or learning expectations?

3.1 Intuition

We can rewrite the reduced form market clearing condition on our market.

$$p_t = \underbrace{P_t^{RE}}_{\text{Fundamental part}} + \underbrace{\alpha\left(p_t^e - P_t^{RE}\right)}_{\text{Expectationnal errors}} + \underbrace{\eta_t}_{\text{Noise}} \tag{12}$$

The equilibrium price have three components. A fundamental part corresponding to the rational expectation price, expectationnal errors and the random noise. The problem is the feedback effect of expectationnal errors on the equilibrium prices. This feedback effect is important to understand. In our cobweb model, an overestimation of the equilibrium price leads suppliers to increase their production above its optimal level lowering the equilibrium price if the demand function is standard. A similar mechanism occurs in most macroeconomic models. Take the Ramsey optimal growth model. Suppose that agents overestimate the path of future wages. They overestimate their wealth leading to higher consumption and lower capital accumulation and lower future wages. In a new keynesian model, strategic complementarity may appear. Forecasts of low future output gap leads to low present consumption and investment and so high current output gap.

The learning strategy emphasized by the learning literature is to learn the fundamental part of the price. But, they estimate a misspecified model. they omit expectationnal errors. The learning literature aknwoledges this misspecification but gives a different interpretation. The dominant interpretation is that agents treats parameters as if they were constant whereas they are actually time varying. It has lead a fraction of the literature to develop a "rational learning" theory treating parameters as time varying following some random law. It is the path of Bullard (1992) and Mcgough (2003) for example. We think that this strategy is wrong. First, It is true that if everyone use this type of learning strategy, parameters are time varying. Recall that in the Actual Law of motion, the parameter vector is $T(\phi_t) = \begin{pmatrix} \mu + \alpha a_t \\ \delta + \alpha b_t \end{pmatrix}$. But they are not really random. They have two components a fundamental one and a expectationnal one, the latter corresponding to the belief about the fundamental parameters and the impact of these beliefs on the equilibrium price. Second, parameters are time varying only if everyone use fundamental learning. If they follow other forecasting schemes, for example if they believe irrationnaly in sunspots, parameters are no longer time varying because only the fundamental part remains. Even if it is more convenient for the theorician to assume that everyone uses statistical learning, an indiviudal agent which wants to make accurate forecasts should adopt more agnostic assumptions

We prefer adopting a very different strategy concerning this misspecification. The optimal strategy should not be to treat parameters as time varying but to learn both the fundamental part and the expectationnal part. The problem is that the expectationnal errors are in principle not observable. But, suppose that other agents use a recursive learning algorithm. The estimation of the new price will depend from past expectaionnal errors. Thus, there is a correlation between past and current expectation errors.Consider the square expected differences (without the contemporaneous noise term η_t) between a fundamental forecasts using ecoometric learning and the actual price. Using the definition of expected price, equation (12), and the formula for bayesian learning, we can show the relation between current and past forecasts errors.

$$
\begin{aligned}
E_t(p_t^e - p_t)^2 &= E_t[(p_t^e - p_t^{RE}) + (p_t^{RE} - p_t)]^2 \\
&= E_t[(z_t'\phi_t - z_t'\phi^*) - \alpha\left(p_t^e - P_t^{RE}\right)]^2 \\
&= E_t[z_t'(\phi_t - \phi^*)(1-\alpha)]^2 \\
&= E_t[(z_t' - z_{t-1}')(\phi_t - \phi^*)(1-\alpha) + (1-\alpha)z_{t-1}^{'}(\phi_t - \phi^*)]^2 \\
&= E_t[(1-\alpha)\left((z_t' - z_{t-1}')(\phi_t - \phi^*) + z_{t-1}^{'}R_t\nu^{-2}z_{t-1}\left(p_{t-1} - z_{t-1}^{'}\phi_{t-1}\right)\right) \\
&+(1-\alpha)z_{t-1}^{'}(\phi_{t-1} - \phi^*)]^2 \\
&= E_t[(1-\alpha)\left((z_t' - z_{t-1}')(\phi_t - \phi^*) + z_{t-1}^{'}R_t\nu^{-2}z_{t-1}E_{t-1}\left(p_{t-1} - p_{t-1}^e\right)\right) + E_{t-1}(p_{t-1}^e - p_{t-1})]^2
\end{aligned}
$$

We see clearly that past forecasts error terms are included in present one. If the gain and innovations in observables are not too important, the correlation is important and should be exploited by a rational agent. Past prices are affected by past expectations errors. Thus, they carry some information on them. If agents succed to extract this information, they may outperform "rational" learners. Thus, it is not obvious that the fundamental learning is the best strategy at the individual level when other agents use this type of learning. If it is not, there are some incentives to deviate from that strategy and it is not reasonable to expect that agents use it. In other words, we try to determine if the fundamental learning is some sort Nash Equilibria. Obviously, in our context, there is no Nash equilibria in a usual sense because the information is not complet by definition. However, the forecasting performance of different methods may be compared. In next sections, we adopt the following strategy. We simulate an economy where agents form a bayesian estimation of the fundamental solution. It is the model of the section defined by equations (3), (6) and (8). We get a certain path for equilibrium prices. Next, we compute price forecasts with other methods. These forecasts does not affect equilibrium price. they are forecasts of an individual agent without impact on aggregate variables. We compare their accuracy with those of the bayesian fundamental forecasts. In that context, p_t^e

always refer to the expected price using bayesian estimation of the RE solution. The price expectation using alternative strategy is denoted p_t^a.

3.2 A simple nonstochastic framework

We illustrate our argument in a very simple nonstochatsic framework where we can easily derive analyctical results. Consider the model of the previous section with a simplified demand function

$$Q_t^d = m_I - m_p p_t \tag{13}$$

Exogenous variables and random noise disappear. The Rational expectation solution becomes just a constant $\frac{\mu}{1-\alpha}$. The reduced form of the equilibrium price can be written in a more direct way

$$p_t = \mu + \alpha p_t^e \tag{14a}$$

$$p_t = \frac{\mu}{1-\alpha} + \alpha \left(p_t^e - \frac{\mu}{1-\alpha} \right) \tag{14b}$$

The learning consists in guessing the value of the constant. With ordinary least squares, the estimation is simply the average value of the observed price. The bayesian estimation algorithm is a little bit more complex but collapse to one equation. [2]

$$p_t^e = p_{t-1}^e + \frac{\gamma}{R_0^{-1} + \gamma t}(p_{t-1} - p_{t-1}^e) \tag{15}$$

In this section,we consider naive expectations as the alternative forecasting strategy $p_t^a = p_{t-1}$.

In the previous section, we express the fundamental current forecasts errors in function of the past one. Here, the relation is considerably simplified. Indeed, observables z_t are simply equal to 1. A consequence is that the expected price is simply equal to the parameter estimate $p_t^e = \phi_t$

$$E_t[p_t - p_t^e]^2 = E_t[(p_{t-1} - p_{t-1}^e) \left(1 - \frac{\gamma}{R_0^{-1} + \gamma t} + \alpha \frac{\gamma}{R_0^{-1} + \gamma t} \right)]^2 \tag{16}$$

The current forecast error is a linear function of the previous one We compare the square forecasting errors of the bayesian estimation strategy and the alm naive forecasting strategy. The forecasts errors of naive expectations can be expressed in avery simple way here

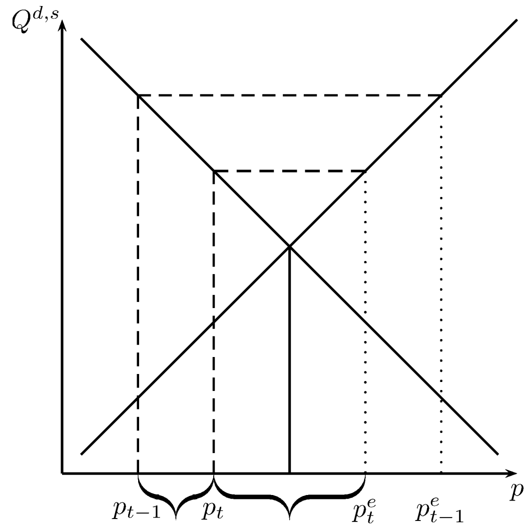
$$E_t[p_t - p_t^a]^2 = [p_t - p_{t-1}]^2 = [\alpha(p_t^e - p_{t-1}^e)]^2 \tag{17}$$

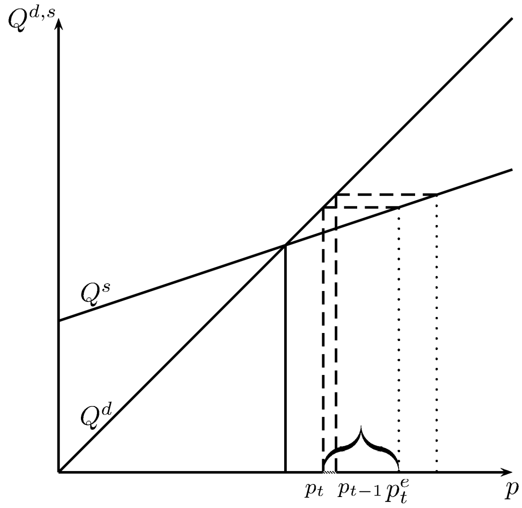
The error of the bayesian estimation of the RE solution is proportionnal to the differences between p_t^e and the rational expectation solution. By contrast, the naive forecasting error depends from the difference between current and past expected prices. Using the updating rule of the bayesian algorithm.

$$E_t[p_t - p_t^a]^2 = [\alpha \frac{\gamma}{R_0^{-1} + \gamma t}(p_{t-1} - p_{t-1}^e)]^2 \tag{18}$$

[2] the model is no more stochastic so using bayesian algorithm is a bit weird but similar results can be obtained with recursive least squares for example

So, forecasts errors of both method can be expressed with respect to past forecasts errors of the fundamental learning. The two following graph illustrate the intuition for the case of strategic complementarity and strategic substituability. We represent demand and supply curve. The demand function is traidtionnal in the first figure whereas it is increasing in the second one. Forecasts errors may be represented graphically and correspond to braces. In both cases, the previous price is much nearer to actual price than p_t^e and it depends from the difference between current expected price and past one. The speed of the learning determining by the gain parameter is critical for our result. It is also the reason why we compare to bayesian algorithm because the gain is higher at early periods. The gain cannot be represented on the graphs. However, the two previous equations integrates it and may be easily compared. We get two propositions. This second effect dominates the first. We show that naive strategy always dominates the rational one in the case of strategic complementarity and dominates asymptotically if instead subsituability is predominant.





Proposition 1 *If* $\alpha \geq 0$, $\forall t$ $E_t[p_t - p_t^a]^2 < E_t[p_t - p_t^e]^2$
square of forecasts errors of a naive agent are always inferior to the forecast errors of a rational agent

Proposition 2 *If* $\alpha < 0$, *There exist a certain period* $T \in N$ *such that forecasts errors of a naive agent are inferior to those of bayesian agents after* T
$\exists T \;\; such \;\; that \;\; for \;\; t > T \; E_t[p_t - p_t^a]^2 < E_t[p_t - p_t^e]^2$

Indeed, we prove (see appendix) that the naive strategy beats the rational one if

$$\alpha \geq \frac{\gamma - R_0^{-1} - \gamma t}{2\gamma} \tag{19}$$

It is always verified for strategic complementarity $\alpha > 0$ and verified for big values of t in case of strategic substituability.
If the estimation of the RE solution is not accurate at the beginning and is corrected only slowly, the error of the Rational strategy will be important whereas the error of the naive strategy will be low. Only, fast learning algorithm may outperform the naive strategy in that context. In case of strategic substituability $\alpha < 0$, prices expectations have a negative impact on equilibrium prices, magnifying expectationnal errors, but because expectationnal errors are important, the learning algorithm is faster. Thus, the parameter α have an ambiguous effect on our result, affecting negatively the accuracy of contemporaneous forecasts but positively the learning gain. Our simualtions and theoretical results clearly shows that the latter dominates the former.

Figure (3) and (4) displays some simulation results which supports our analytical findings. The first is performed under strategic complementarity assumption. The dominance of the blue line (naive forecasts) is obvious. The second figure displays results for strategic substituability. Only a very limited set of periods is represented, both forecasts errors are negligible after. The dominance of the blue line intervenes after the second period.

3.3 Extensions to the stochastic case

This result is interesting but it does not hold in a more sophisticated framework. Consider expectationnal errors of naive forecasting in the complete model

$$E_{t-1}p_t - p_{t-1} = \alpha\left(p_t^e - p_{t-1}^e\right) + \underbrace{(\alpha b_t + \delta)(w_t - w_{t-1})}_{\text{Innovations in observables}} + \underbrace{\eta_{t-1}}_{\text{Past noise}} \tag{20}$$

We can see two new error terms. The first correspond to innovations in exogenous variables. Past prices are affected only by past values of w and do not include any information about current innovations. The second term is the random noise of the previous period. Because these residual is iid, it does not affect current prices. Thus, past prices incorporates an irrelevant information. Not only these two inefficiencies affects the contemporaneous forecasting performance but they alter the asymptotic properties. The two error terms remains once the expectationnal error of the bayesian estimation is zero. Thus, if naive forecasting may continue to beat bayesian estimation, its dominance is a very short one.

Simulations suggest that it depends mainly from four parameters: the level strategic complementarity/substituability α, the persistence of innovations in observablesρ, the standard deviation of the random noise and from the spread between priors and true values, especially for the constant term. A low persistence in observable innovations and a high standard deviation for them means that the contemporaneous information missed by naive expectations is huge. Using this information even with inaccurate parameter estimates is better. With high persistence and strategic substituability, the convergence of the learning process is too fast leading to an unambiguous dominance of bayesian estimation. The case of strategic complementarity is more contrasted. Figure 6 displays forecasts errors of naive expectations and of bayesian estimation in an economy where the latter is dominant. We consider the case of strategic complementarity with $\alpha = 0.5$. We calibrate $\frac{\mu}{1-\alpha} = 4$, $\frac{\mu}{1-\alpha} = 2$. The prior are $a_0 = 1$ for the constant term and $b_0 = 1$. The persistence parameter is $\rho = 0.9$ and $\sigma_\varepsilon = 0.2$. Standard deviation of the noise process is $\sigma_\eta = 0.5$. Naive expectations are more accurate for five periods but bayesian estimation is dominant for the rest of the simulation. It seems unlikely that this short period constitutes a sufficent incentive to deviate from fundamental learning. However, if we consider a lower volatility of noise like in figure 5, with $\sigma_\eta = 0.1$, and a lower volatility for the innovations of observables, we can restore a good performace of naive forecasting. This result is quite logical as we tend to the case where both are equal to zero which is the case considered in the previous section. At this stage, there are several strategies. The first is to look at macroeconomic models and at what case they are likely to correspond. This is the section 5. The second is to correct naive forecasting to alleviate the bias introduced by past noise and innovations in observables. Our solution is to augment the bayesian estimation of the fundamental solution with an estimation of expectationnal errors.

Exogenous variables First, we study the problem when there is only exogenous observables but no random component. The solution is straightforward. You can always rewrite the equilibrium price condition like

$$p_t = p_t^{RE} + \alpha \left(p_t^e - p_t^{RE}\right)$$
$$p_t^{RE} = \frac{\mu}{1-\alpha} + \frac{\delta}{1-\alpha} w_t$$

The price is the sum of a fundamental component and a expectationnal error. Naive forecasting incorporates information about the second but fails to exploit all available information about the fundamental component. Next, the agent may implement the following algorithm. In a first step, he makes a bayesian recursive estimation of the vector of parameters of the rational expectations solution. He compares its forecast of the past fundamental price with the actual price of the previous period. The residual should be the expectationnal errors. For the next period, its foecast for the price will be composed of two parts. The first is expectationnal errors of the previous period. The second is the estimation of the fundamental price. The expected price will be the following one

$$p_t^a = p_{t-1} + (z_t' - z_{t-1}')\phi_t \tag{21}$$

Conjecture 1 *If $\alpha > 0$, the square errors of the augmented algorithm is inferior to the square error of the fundamental learning.*

$\forall t \quad E_t(p_t^a - p_t)^2 \leq E_t(p_t^e - p_t)^2$

Conjecture 2 *If $\alpha \leq 0$, the square errors of the augmented algorithm and square errors of the fundamental learning are similar.*
$\forall t E_t(p_t^a - p_t)^2 \sim E_t(p_t^e - p_t)^2 \Rightarrow \frac{E_t(p_t^a - p_t)^2}{E_t(p_t^e - p_t)^2} \simeq 1$

The figure 7 illustrates the first conjecture. The calibration is similar to previous one with a relatively low prior error about the constant term (around 20 percent) and by definition a zero standard deviation for random noise. Optimal forecasts clearly dominates fundamental ones.
It is interesting to note that this case is quite relevant for macroeconomics. Indeed, modern macroeconomic models, especially DSGE models like Smets and Wouters (2007) assume that every shock is perfectly observable by economic agents, so there is no noisy term.

Noise We consider the second problem of random component. The true model is

$$p_t = \frac{\mu}{1-\alpha} + \alpha \left(p_t^e - \frac{\mu}{1-\alpha} \right) + \eta_t \tag{22}$$

The problem is more complex than the previous one. Indeed, the residual of the bayesian estimation of the fundamental price has two components, expectationnal errors and random noise. Both are unobserved. If you equalize expectationnal errors with the whole residual, you get the previous and irrelevant error term in your forecast of the new price. An alternative is to use Kalman filter to distinguish the two terms. The model can be written in a space state representation

$$p_t = \frac{\mu}{1-\alpha} + f_t + \eta_t \tag{23a}$$

$$f_t = f_{t-1} \tag{23b}$$

The challenge compare to the traditionnal Kalman Filter problem is that the measurement equation is unknown. So, we should perform simultaneously the bayesian estimation of the fundamental solution and the bayesian inference of the expecationnal error part. We propose the following algorithm

- **Step 1**: Using your forecast $f_{t-1|t-1}$, you get an estimation of the "true" residual and you update your forecast for ϕ_t.

$$R_t^{-1} = R_{t-1}^{-1} + \nu^{-2} z_{t-1} z_{t-1}^{'} \tag{24a}$$

$$\theta_t = \theta_{t-1} + R_t \nu^{-2} z_{t-1} \left(p_{t-1} - f_{t-1|t-1} - z_{t-1}^{'} \theta_{t-1} \right) \tag{24b}$$

- **Step 2** : Using your new estimation of ϕ_t you reestimate your total residual. Using kalman filter, you make a guess for $f_{t-1|t}$. You get $f_{t|t}$ using a random walk hypothesis $f_{t|t} = f_{t-1|t}$

$$f_{t|t-1} = f_{t-1|t-1} \tag{25a}$$
$$\Omega_{t|t-1} = \Omega_{t-1|t-1} \tag{25b}$$
$$\tilde{y}_t = p_{t-1} - \theta_t - f_{t|t-1} \tag{25c}$$
$$S_t = \Omega_{t|t-1} + \nu^2 \tag{25d}$$
$$K_t = \Omega_{t|t-1} S_t^{-1} \tag{25e}$$
$$f_{t|t} = f_{t|t-1} + K_t \tilde{y}_t \tag{25f}$$
$$\Omega_{t|t} = (I - K_t)\Omega_{t|t-1} \tag{25g}$$

- **Step 3** Your forecast for the price in t is the sum of the estimation of the random component and the estimation of the expectationnal errors

$$p_t^e = z_t'\theta_t + f_{t|t} \tag{26}$$

It is harder to form a conjecture about the respective performance of this algorithm and fundamental learning. Simulations suggests that fundamental and optimal learning provides very similar forecasts in case of strategic substituability. There is no strong incentive to deviate from fundamental learning. In case of strategic complementarity, results are more ambiguous. For low prior errors on paramters and low level of α, fundamental forecasts are more accurate. Otherwise, the algorithm augmented with Kalman filter dominates. Figure 8 and 9 compares forecasts errors for high and low prior errors and shows clearly this last result.

3.4 A random walk problem?

An interesting case occurs when the level of strategic complementarity is close to one. For example, we calibrate $\alpha = 0.9$. This value may seem extreme but could be relevant in asset pricing literature. Model are not similar because asset price model links p_t and p_{t+1}^e. But, similar issues appears in both. In asset prices literature, the α parameter corresponds to the interest rate and may be close to one.

We consider not only the algorithm above but variants using random walk. The model estimated becomes

$$p_t = \frac{\mu}{1-\alpha} + f_t + \eta_t \tag{27a}$$
$$f_t = f_{t-1} + \omega_t \tag{27b}$$

Where ω_t is simply white noise. The main difference is the equation $\Omega_{t|t-1} = \Omega_{t-1|t-1}$ becomes $\Omega_{t|t-1} = \Omega_{t-1|t-1} + var(\omega_t)$ Both algorithms dominates fundamental learning at early stage. Random walk algorithm are not convergent. But, figure 10 shows a clear domination of divergent algorithms in the short run over our converging algorithm of the previous section.

Conclusion (provisionnal)

In this paper, our contribution is twofold. First, we try to compare learning algorithms on the ground of their contemporaneous forecasting performance

and not of their asymptotical properties. We conclude that bayesian methods are largely superior to ordinary least squares. Second, we argue that agents should learn the fundamental solution but also expectationnal errors if past and present forecasts errors are correlated which is the case if agents learn. In that respect, the literature is incomplete. We show how we can learn expectationnal errors in various environments, the caveats mainly the identification between expectationnal errors and noise, and the possible solutions including the use of Kalman filter.
The next step is the application to general equilibrium macroeconomic models. the asset pricing literature could be especially relevant.

References

1. **Adam, K. and A. Marcet.** 2011. "Internal rationality, imperfect market knowledge and asset prices." *Journal of Economic Theory* 146:1224-1252

2. **Bray, M., and N. Savin.** 1986. "Rational expectations equilibria, learning, and model specification." *Econometrica* 54:1129-1160

3. **Bullard, J.** 1992. "Time varyning parameters and nonconvergence to rational expectations equilibirum under least squares learning" *Economic Ltters* 40:159-166

4. **Bullard, J. and J. Suda.** 2011. "The stability of Macroeconomic systems with Bayesian learners" *Banque de France working papers*

5. **Evans, G., and S. Honkapoja.** 2001. *Learning and expectations in Macroeconomics* Princeton

6. **Marcet, A., and J.P. Nicolini.** 2003. "Recurrent Hyperinflation and Learning." *American Economic Review* 93(5):1476-1498

7. **Marcet, A., and T. Sargent.** 1989. "Convergence of least squares learning mechanisms in self referential linear stochastic models." *Journal of economic Theory* 48:337-368

8. **McGough, B.** 2003. "Statistical learning with time varying parameters." *Macroeconomic Dynamics* 7:119-139

Appendix

Proof of Proposition 1 and 2

We write the expectationnal error in case of naive forecasting

$$\begin{aligned} p_t - p_{t-1} &= \alpha(p_t^e - p_{t-1}^e) \\ p_t - p_{t-1} &= \alpha \frac{\gamma}{R_0^{-1} + \gamma t}(p_{t-1} - p_{t-1}^e) \end{aligned}$$

Now, we can write the expectationnal error in case of bayesian estimation of the RE solution

$$\begin{aligned} p_t - p_t^e &= p_t - p_{t-1}^e - \frac{\gamma}{R_0^{-1} + \gamma t}(p_{t-1} - p_{t-1}^e) \\ p_t - p_t^e &= p_t - p_{t-1} + (p_{t-1} - p_{t-1}^e)[1 - \frac{\gamma}{R_0^{-1} + \gamma t}] \\ p_t - p_t^e &= (p_{t-1} - p_{t-1}^e)[1 - \frac{\gamma}{R_0^{-1} + \gamma t} + \alpha \frac{\gamma}{R_0^{-1} + \gamma t}] \end{aligned}$$

We study conditions under which

$$\begin{aligned} &(p_t - p_t^e)^2 > (p_t - p_{t-1})^2 \\ &[1 - \frac{\gamma}{R_0^{-1} + \gamma t} + \alpha \frac{\gamma}{R_0^{-1} + \gamma t}]^2 > [\alpha \frac{\gamma}{R_0^{-1} + \gamma t}]^2 \\ &[1 - \frac{\gamma}{R_0^{-1} + \gamma t}]^2 + 2\alpha[1 - \frac{\gamma}{R_0^{-1} + \gamma t}] > 0 \\ &\alpha > \frac{\gamma - R_0^{-1} - \gamma t}{2\gamma} \end{aligned}$$

For $t > 0$, the right hand side of the equation is always neagtive. So, the inequality is always verified if $\alpha > 0$. Moreover, the right hand side of the equation is decreasing with respect to t. As α is constant even if it is negative there is a certain value T of t such that the inequality is verified.

Figures

Figure 1: Bayes vs OLS

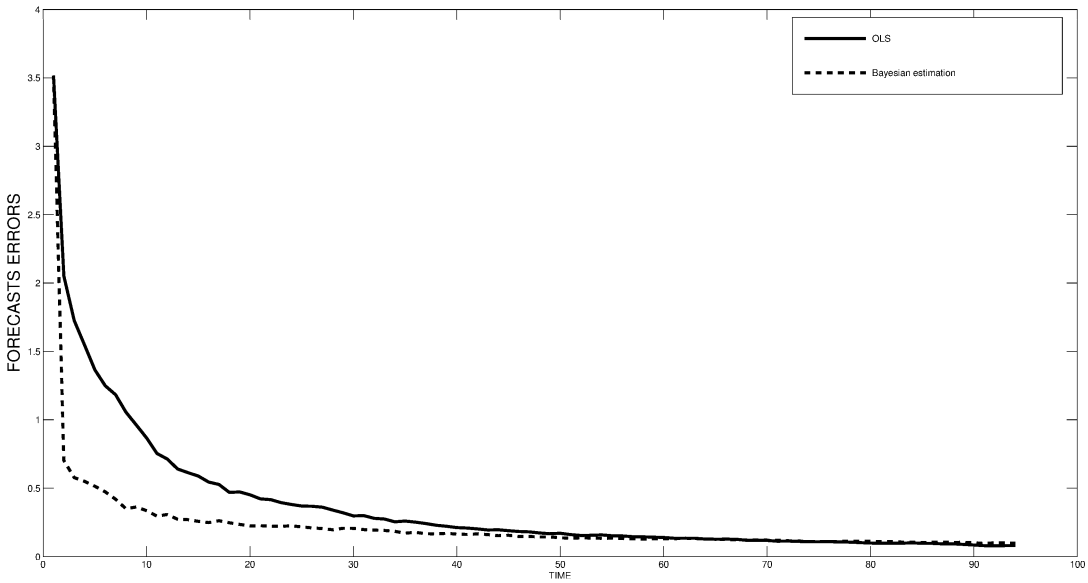


Figure 2: OLS and naive expectations

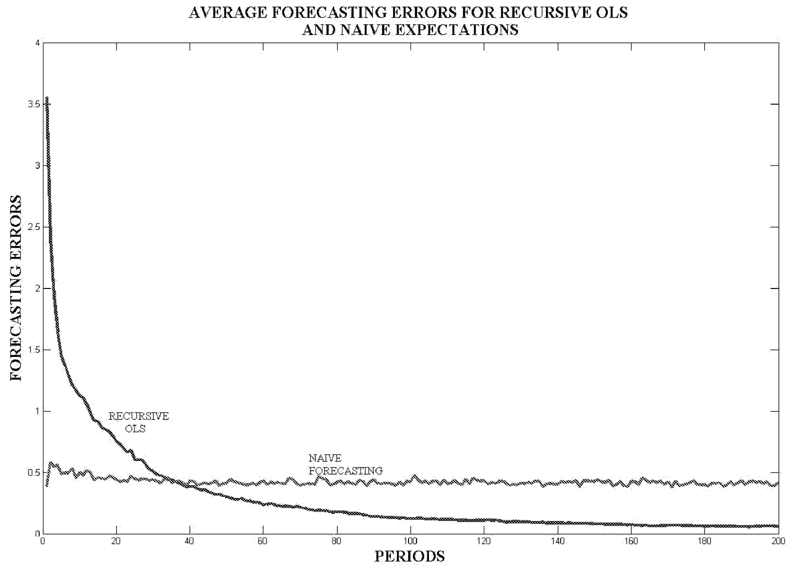


Figure 3: Simple case with strategic complementarity

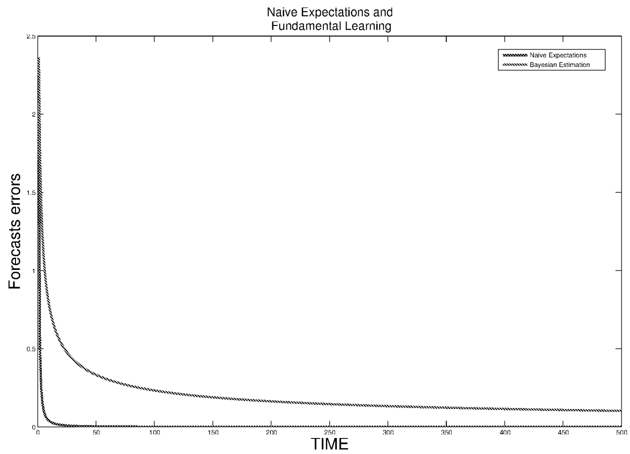


Figure 4: Simple case with strategic substituability

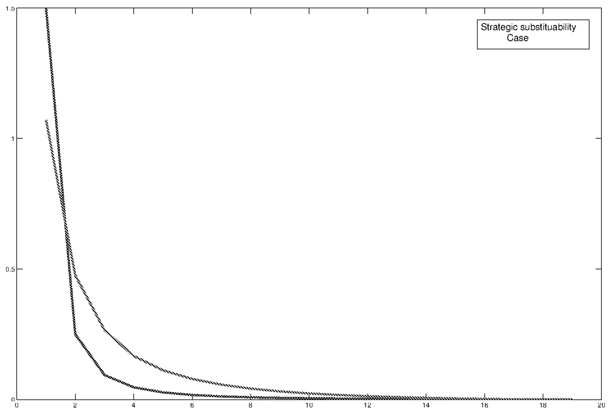


Figure 5: complete case: naive vs fundamental forecasts

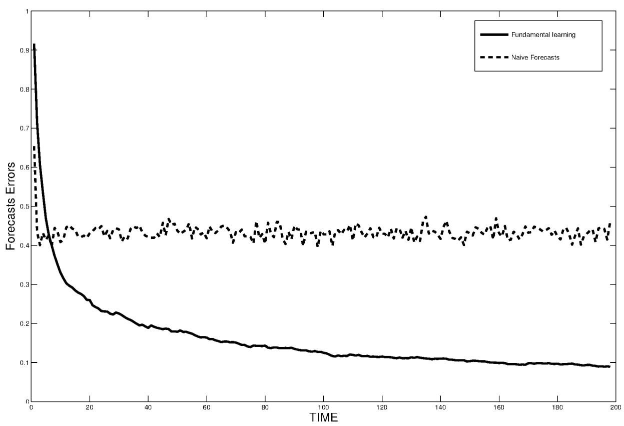


Figure 6: complete case with low volatility: naive vs fundamental forecasts

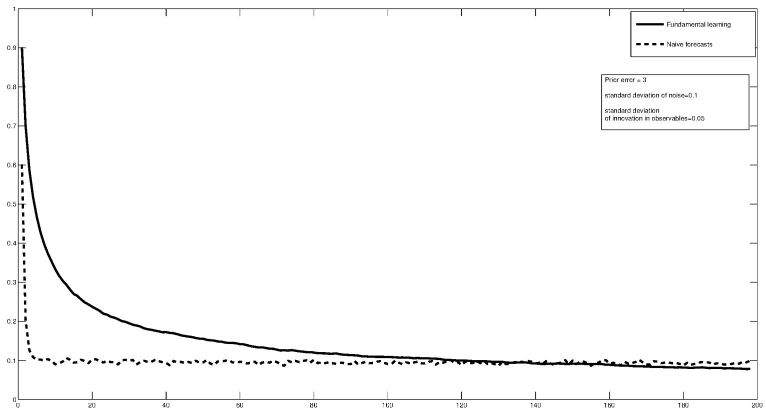


Figure 7: case with exogenous observables: optimal vs fundamental forecasts

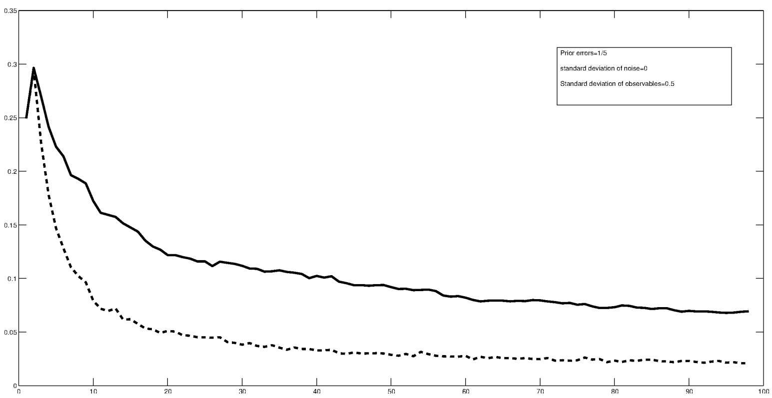


Figure 8: complete case: optimal vs fundamental forecasts

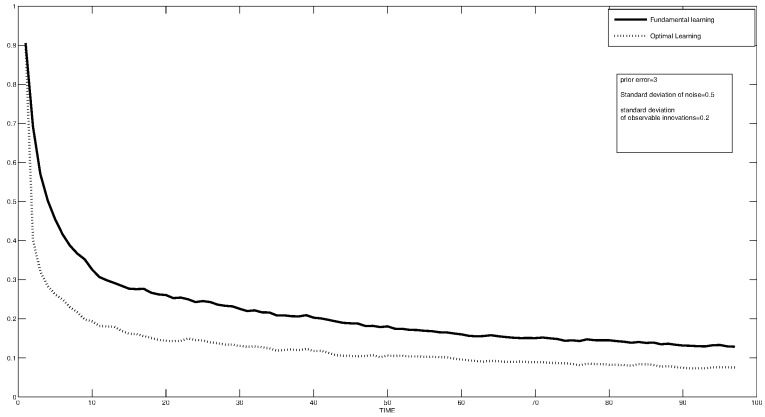


Figure 9: complete case with low prior errors: optimal vs fundamental forecasts

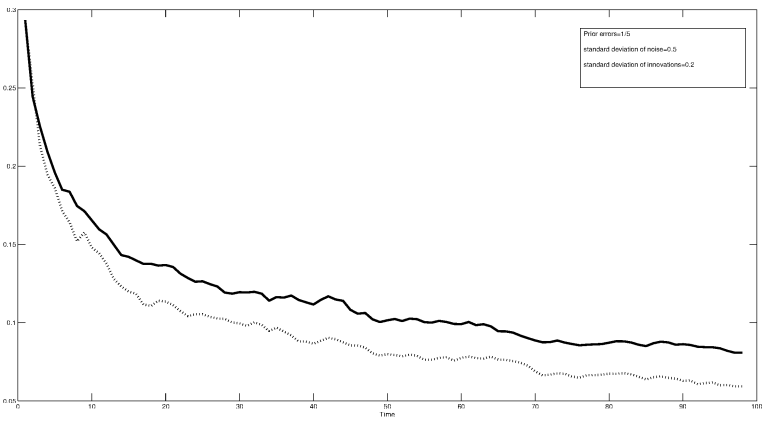


Figure 10: Random Walk performance