Adaptive Learning Explanation to non-Normal Output Growth and Output Gap

Vytautas Valaitis

Sutmitted to

Department of Economics

Central European University

In a partial fulfillment of the requirements for the degree of

Master of Arts

Supervisor: Professor Istvan Konya

Budapest, Hungary

2015

Acknowledgements

I would also like to thank my supervisor Istvan Konya for useful comments and suggestions. also this entire two-year journey would have been much harder without the support from my family and friends. I am mostly thankful to Andreea, Pëllumb and Attila for their continuous support, unselfishness and remarkable teamwork.

Abstract

Empirical studies suggest that the evolution of output across time is described by a distribution with the tails fatter than normal. At the same time standard macroeconomics models are linear and can only replicate this exogenously. I show that replacing the rational expectations with ones formed under statistical learning can deliver the fat tails even with using normally distributed exogenous shocks. Monte Carlo experiments using a small-scale New Keynesian model show that propagation is more reasonable when rational expectations are replaced on the model's final equations rather than when adaptive learning is micro founded. In both cases a model exhibits large and rare exits that become more pronounced under micro founded rule. An even less rational heuristic learning propagates the normal shocks as well, but only together with a very high variance.

Contents

1 Introduction

Most of the time economies are progressing normally, not deviating far from the steady state. Yet short infrequent events that take the economy far away from it happen, leading to a non normal distribution of the output series across time. This phenomenon has been recorded by Fagiolo et. al. (2008),(2009) and Christiano (2007) for the worlds major capitalist economies during the post-war period, in particular for the output growth rate and the output gap. If this is the case, a next question that arises is whether this is due to a large exogenous shocks or whether the normal shocks are endogenously propagated within the economy. The estimation of macroeconomic models using marginal likelihood relies on the assumption of normally distributed exogenous shocks, thus if former is the case, macroeconomic modeling of non normality across time would seriously inhibit the estimation of the models. Therefore in this thesis I provide an explanation how normal exogenous shocks can be endogenously propagated within the New - Keynesian model to give the non normal output dynamics described by fat tails. One of the ways to implement this is through the expectations formation. The same explanation to the non normality has already been given by De Grauwe (2012a) changing the rational expectations with the ones formed under two heuristic rules. In this thesis I implement a two standard versions of a more mainstream statistical learning expectations formation rule into the standard New - Keynesian model and test if it can propagate the normal shocks as well. The results from Monte Carlo simulations show that while a model with rational expectations fails to deliver the desired result, the non normality of output growth rate and gap can be reproduced under learning by varying a constant gain parameter. A model with heuristic learning delivers high kurtosis as well, but at the expense of highly increased second moment of the series analyzed. However, due to either low determinacy of the model or the very strong propagation dynamics or both, the model occasionally exhibits unreasonably sharp and large deviations from the steady state.

The paper is organized as follows: first I provide a brief introduction into the topic, then follow with the description of statistical learning framework. In the section 5 I replicate the results by Fagiolo et al. (2008),(2009) using 10 years longer data sample. In the parts 6-7 I present the model and Monte Carlo simulation results of three different learning rules.

2 Introduction to the topic

It has been documented in a series of papers that the output growth rates across time for a single time series are not normally distributed, but in fact their distribution is is described by fat tails (Fagiolo et al., 2008). In the sequel paper (Fagiolo et al., 2009) the same pattern is observed for the detrended output time series and the finding is robust to the filtering method used. In particular the authors fit the growth rate data with the density that nests Gaussian and Laplace distrubutions:

$$f(x;b,a,m) = \frac{1}{2ab^{\frac{1}{b}}\Gamma(1+\frac{1}{b})}e^{-\frac{1}{b}|\frac{x-m}{a}|^{b}}$$

After observing from the raw data that the kurtosis is above 4, the estimation indicates that the b parameter governing the fatness of the tails is close to 1, whereas the Gaussian distribution would have $b = 2$. A similar pattern emerges in all major economies and is robust to serial correlation, outliers and heteroscedasticity.

The sequel paper Ascari et al. (2013) checks if the business cycle models are able to reproduce this regularity. They show that the basic RBC and medium scale New - Keynessian as in Smets & Wouters 2003 models cannot produce the fat tail distributions of output growth from the normally distributed shocks even with models simulated using second order approximation. This leads the authors to ascribe this to a more established result that the business cycle models do not have endogenous shock propagation mechanism (Cogley and Nason, 1995). Regarding the NK model, a normal technology shock resulted in a slightly thinner tails, which led the authors conclude that, while its frictions and rigidities increase the persistence, they also smooth the series of shocks.

In one of several interrelated papers De Grauwe (2012a) shows that a basic New - Keynessian three equation model with the heuristic forecasting rule described above is able to produce the fat - tail regularity. However, only the distribution plots without any formal testing or density estimation as in Fagiolo et al. (2008) are provided in the paper. The aim of this thesis is to test if the more mainstream approach of adaptive learning is able to produce the fat - tails of output growth rate in the New - Keynessian models.

3 Adaptive Learning

There have been many different attempts to model the beliefs and expectations that may not be rational or formed under an imperfect information. Such attempts include a news - driven business cycles (Beaudry and Portier, 2014), where information friction appears in the expectation about the future technology shock. While in this setup expectation is still formed rationally by the agents using Bayesian updating, the noise of the signal of future technology shock may cause irrational boost of prices and employment in anticipation of the technology boom. Close, but much less explored way to limit the knowledge is to impose the restriction that the agents neither know the inflation target nor are aware of some exogenous processes hitting the economy or to introduce the stickiness of information available to the agents using the Calvo type setting Milani (2010). In all of these frameworks expectations are still formed rationally, but their formation is disturbed or limited by the information available. On the other extreme is the heuristics approach postulated by Paul De Grauwe (2010), (2012a). Here economic agents experience cognitive limitations and keep switching between two biased forecasting rules $E_t(y_{t+1}) = 0$ or $E_t(y_{t+1}) = y_{t-1}$ depending on which performs better in terms of RMSE. Here the only remainder of rational behavior is the willingness of the agents to learn from the past mistakes.

In the middle of the above approaches stands the adaptive learning literature. Under rational expectations the agents are assumed to possess the full knowledge of the model and the exogenous shocks hitting the economy. From that follows that agents know the steady state values of the model's variables and the next periods expectation is formed using the coefficients from the MSV solution of the system. However, according to Honkapohja and Evans (2001):

"In empirical work economists, who postulate rational expectations, themselves do not know the parameter values and must estimate them econometrically. It appears more natural to assume that the agents in the economy face the same limitations on knowledge about the economy"

The basic idea of the method is to impose that the agents do not have a complete knowledge about the the reduced form of the economy. This could happen when the structural parameters of the model are not known, or even if they are, then it is not clear how they map into the reduced form of the model used in forming the expectations. The most common approach is to assume that the exogenous shocks hitting the economy and the structure of the reduced form are known, but the coefficients are re-estimated and updated each period once new data becomes available. A simple example illustrates the basic idea:

Suppose that the model is described by the law of motion

$$y_t = \alpha + \beta_0 E_{t-1}(y_{t+1}) + \beta_1 E_{t-1}(y_t) + \kappa w_t + \epsilon_t \tag{1}$$

The expectation $E(y_{t+1})$ is unknown and is predicted by the agents using their perceived law of motion (PLM), which in this specific case is supposed to take the form

$$y_{t+1} = \hat{a}_t + \hat{b}_t w_{t-1} \tag{2}$$

$$w_t = \rho w_{t-1} + \nu \tag{3}$$

The coefficients $\hat{a}_t, \hat{b}_{1t}, \hat{b}_{2t}$ are updated each period with the least squares, which written recursively takes the form

$$\phi_t = \phi_{t-1} + t^{-1}R_{t-1}^{-1}z_{t-1}(y_t - \phi' z_{t-1})$$
$$R_t = R_{t-1} + t^{-1}(z_{t-1}z_{t-1}' - R_{t-1})$$

Realized values of y_{t_1} are mapped into the model, which this gives the actual law of motion (ALM) of the economy:

$$y_t = T(\phi_{t-1})' z_{t-1} + \kappa\epsilon_t + \nu_t \tag{4}$$

where T() is called the T-mapping. In this particular case

$$T(\phi_{t-1}) = T\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \alpha + (\beta_0 + \beta_1)a \\ (\beta_0 + \rho\beta 1)b + \kappa\rho \end{pmatrix} \tag{5}$$

And R_t is the matrix of the second moments of the variables used in the forecasting equation

Expectations are in turn determined by the evolution of y through equation (1), but the effect of irrational forecast is not conceived by the agents and thus the learning is done by observing the realized values of y

$$\phi_t = \phi_{t-1} + t^{-1}R_{t-1}^{-1}z_{t-1}(T(\phi_{t-1})' z_{t-1} + \kappa\epsilon_t + \nu_t - \phi' z_{t-1})$$
$$R_t = R_{t-1} + t^{-1}(z_{t-1}z_{t-1}' - R_{t-1})$$

This unawareness makes the system self-referential, where a positive shock to y makes the agents revise their forecasts for the next periods value. Those forecasts are subsequently mapped into the actual law of motion of the economy and keep affecting the further forecasts and the evolution of the economy. As shown by Williams (2014) and

Evans and Honkaphoja (2001) a few subsequent exogenous shocks can move the system away from the steady state, to which it may not come back immediately or result in the periods of higher volatility (Slobodyan and Wouters, 2012b).

3.1 Variations

3.1.1 Gain

As mentioned before, every least squares equation can be written in the recursive form

$$a_t = a_{t-1} + t^{-1}(p_t - a_{t-1}) \tag{6}$$

Under some general conditions (Honkapohja and Evans, 2001) this type of learning eventually converges to a rational expectations equilibrium, as when $t \to \infty$, $a_t = a_{t-1}$. This type is called least squares learning. An alternative is to impose some constant value g in place of t^{-1}, which is known as constant gain learning. A higher value of g means that agents forget past information more quickly and are ready to believe that the current changes are due to a structural break in the economy.

Another, computationally even easier way known as stochastic gradient learning is to impose $g = t^{-1}R_{t-1}$. Here the effect of the deviation of a variable on parameter updating does not depend on the variance of the variables used in the forecasting equation.

3.1.2 Horizon

There are two main ways used in the literature to incorporate learning into the model. A more traditional way also known as Euler equation learning requires to simply replace a one period ahead expectation with its forecast under the adaptive learning procedure. While easier to implement, the rule is not derived from anywhere and sometimes it becomes not clear what is forecasted by whom. For instance with the Euler equation taking the form

$$\hat{c}_t = E_t(\hat{c}_{t+1}) - \frac{1}{\sigma}(i_t - E_t\pi_{t+1} - \rho) \tag{7}$$

this would mean that the agent is forecasting his own consumption next period. This may be justifiable in the model, where all the income goes to consumption, but most of the models are more complex. Another way introduced by Preston (2005) is to derive the individual consumption rule depending on the variables exogenous to the agent from the intertemporal budget constraint and Euler equation. Then expectation of $(\hat{c}_{t+1})$ is formed by forecasting the exogenous variable using the same least squares procedure. The forecasts are made for infinite future not realizing that the parameters will be updated in the next period. In case of a basic NK model, such a rule would look like

$$\hat{c}_t = (1-\beta)\bar{\omega}_t^i + \hat{E}_t^i \sum_{T=t}^{\infty} \beta^{T-t}[(1-\beta)\hat{Y}_T^i - \beta\sigma(\hat{i}_T - \hat{\pi}_{T+1}) + \beta(g_T - g_{t+1}) \tag{8}$$

where Y^i is individual income and g the subjective discount factor and $\bar{\omega}_t^i$ is individual wealth. Such a rule has to be derived for each variable that is maximized inter temporally in a model

3.1.3 Other

Other possible variations include varying the perceived law of motion - it could be a simple VAR wich variables chosen by discretion or observed by the agent, or an MSV solution of the model Also it matters if the exogenous processes are included in the forecasting equation. Alternatively it has been shown by Slobodyan and Wouters (2012b) and Garceles-Poveda (2007) the specification of initial beliefs about the parameters of the forecasting equation matter, in particular whether they are consistent with rational expectations or not. Lastly, instead of uncertainty about the parameters of the model, it is possible to introduce uncertainty about the structural features of the economy, such as the production function, so that agents optimize given their beliefs Williams (2004).

However due to i.i.d. technology shock assumption this approach thus far could only be implemented in a highly simplified models.

4 Relevant Literature

The first strand of the literature on adaptive learning was mostly preoccupied with the stability and learnability of the equilibria under the least squares learning. The fact that the steady state is learned through time gave a greater support for the rational expectations assumption. However, constant gain and stochastinc gradient learning imply perpetual learning and convergence of beliefs to a stochastic distribution around the steady state (Honkapohja and Evans, 2001). After convergence of adaptive learning mechanism had been established, a news strand started the explore the response of the systems with adaptive learning to exogenous shocks. Using the long-horizon consumption decision rule as in Preston (2005) in the RBC model with the capacity utilization Eusepi & Preston (2011) find that learning mechanism delivers the same volatility of output as RE models using technology shock with 20 % smaller standard deviation. This comes from the amplification of the substitution effect due to changing beliefs in response to the technology shock, which makes the investment and working hours to increase more relative to RE model. The authors also note that the effect comes from the differences from the RE in the forecasted variables at longer horizons, meaning that the Euler equation learning would not be able to reproduce this. Milani (2006), (2007) asserts that the model dynamics to a very large extent depends on the choice of the gain parameter and therefore estimates it jointly with other parameters of the model using Bayesian methods. It appears that the models with learning fits the data better in terms of marginal likelihood and that under learning, the mechanical sources of persistence, such as habit formation and inflation indexation become superfluous. Also the model with Euler equation learning fits the data better than the one with infinite horizon learning (Milani, 2007), as the need to make infinite horizon forecasts also means large forecast errors and thus lower fit. Milani (2007) suggest incorporation of learning into the medium-scale DSGE models as an area for future research. This is done by Slobodyan & Wouters (2012a), (2012b) using the Euler equation type learning. Besides providing evidence that the model with learning gives a better fit the first paper (2012a) also reveals that among the two possible sources of endogenuous propagation - misspecified information set in forecasting equation and the parameter updating - the former is more important. Also it is shown that the existence of constant in the forecasting equation, meaning that the belief about the trend growth rate or the steady state is being updated, matters for the fit. The other paper (2012b) experiments with different specifications of learning - initial beliefs, information set and gain parameter in the Smets & Wouters 2007 model. In most of the cases the behavior of the system exhibits so called rare events with higher volatility and large deviations from the steady state values, similar to those in Williams (2004). They occur with even at small constant gain parameter [1], but become more frequent and large with the large g values and more misspecified forecasting equation. The reason why learning gives a better fit comes from the fact that it can be perceived as a relaxation of restriction of a constant parameters in the forecasting equation, thus the model uses more information from the data. At the same time the forecasting power of the model deteriorates, especially for consumption and the interest rate (Slobodyan and Wouters, 2012b).

Perhaps the most similar although not directly related to this work is the study by Benhabib and Dave(2014), who use a Lucas asset pricing model with stochastic gradient learning to generate the fat-tail distribution of price to dividend ratio. Under some restrictions for the distribution of variables governing the evolution of forecasting coefficient, they analytically prove that the model produces fat tails, more specifically that the distribution follows a power law. Also De Grauwe (2013) tries to explain the movement of the exchange rate using a model with either heuristic forecasting rules as in (Grauwe, 2010) and (DeGrauwe, 2012a) or with the adaptive learning. Reportedly both mechanisms are able to replicate the fat-tails of the exchange rates, but only adaptive learning

[1] small means around 0.01 or 0.02

can account for the volatility clustering while only heuristics can reproduce large and prolonged deviations from the fundamental. [2]

There is also a huge literature analyzing the optimal monetary policies under learning, but since this is not an issue of this study, it will not be reviewed here. However it is worth mentioning that a policy not taking into account the existence of learning in the model is suboptimal and even amplifies variability of economic aggregates. Under this setting the aim of the central bank becomes to tame the forecasted persistence of inflation (Gaspar et al., 2006). To my best knowledge there have been no studies made so far trying to explain the non-normal output growth using adaptive learning.

5 Empirical evidence

Before going into the modeling I replicate the major results from Fagiolo et. al (2008), (2009). The empirical distributions in figures 1-2 from the post-war US data including years 1947-2015 suggest non-normality for both GDP growth rate and gap. [3]. This is especially acute for output growth rate, which most of the time appears to be clustered around the mean, but with a low probability exhibits a values far from the mean in absolute value. The output gap [4] distribution looks more spread, yet still contains value that would be unlikely under normal distribution.

[2]This is referred to as the disconnectedness phenomenon in the original paper

[3] GDP data is taken from research.stlouisfed.org/fred2. The data series used is a chain linked seasonally adjusted quarterly US GDP

[4]series for output gap end in 2013. The estimates of the trend at the end of the sample are unreliable, so it was estimated using all the data available, but only values until 2013 are used as reliable

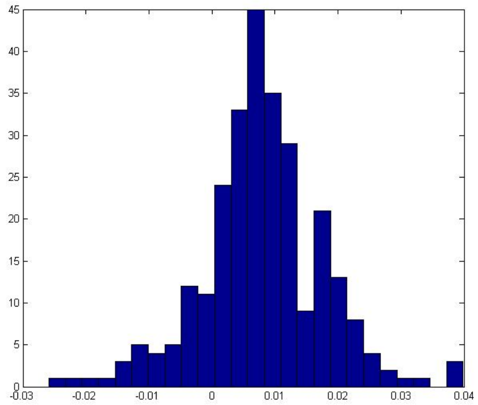

Figure 1: US GDP growth rate, postwar sample

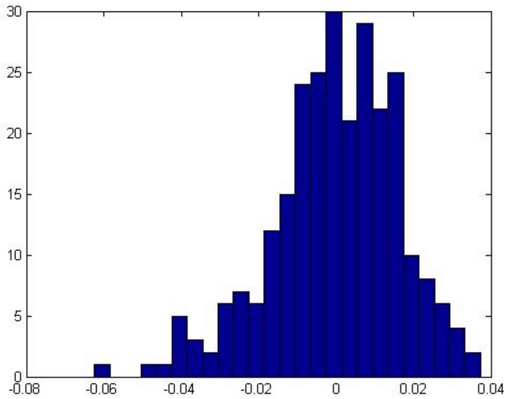

Figure 2: US output gap,postwar sample

Table 1: Empirical moments: US data

	Growth rate	Output gap
Mean	0.0079	-0.0001
Standard deviation	0.0097	0.0166
Skewness	-0.0537	-0.5768
Kurtosis	4.3753	3.5845
J-B p-value	0.0017	0.0027

Table 1 reports the first four moments of the GDP growth rates and the output gaps in a given period [5]. It is apparent that both series are not normal. Output growth rate has kurtosis 4.37, which is way higher that 3 implied by the normal distribution. At the same time its skewness is negligible. This allows to conclude that the Jarque-Berra test rejects the non-normality due to fat tails as indicated by a low p-value. The output gap is non-normal as well according to Jarque-Berra test, yet it is described by both higher than normal kurtosis and sizeable skewness.

Next I fit the data to the subbotin density used in Fagiolo et. al (2008), (2009). Since it is characterized my many local maxima, the usual estimates obtained by minimizing of the likelihood are unreliable (Botazzi and Secchi, 2008). For this purpose I use a "Subbotools" package, which performs the maximization over multiple sub intervals and then picks the highest one [6]. Parameters a,b,m describe the scale, shape and the location. b=2 recovers a normal distribution, while b=1 gives Laplace. The function

$f(x;b,a,m) = \frac{1}{2ab^{\frac{1}{b}}\Gamma(1+\frac{1}{b})}e^{-\frac{1}{b}|\frac{x-m}{a}|^b}$

is symmetric, while the output gap of the data series has skewness different from 0. To check if it is significantly different I repeated the estimation of skewness over 1000 bootstrapped samples and it appeared that it is more than 2 standard deviations away from 0. In this cases I also fit the distribution with the asymmetric Subbotin density available in the Subbotools package as well.[7]

. The result is provided in the appendix 2.

The results from the estimated densities are in line with Fagiolo et. al. (2008),(2009). The estimated parameter $\hat{b}$ for the output growth rate is 1.1 with the 95% confidence interval not including 2, which suggest that it is better described by the Laplace distribution. The estimated $\hat{b}$ for the output gap is 1.6, still lower than normal, however, asymptotic confidence interval includes 2. The results from fitting it with the asymmetric

[5] growth rate is definted here as the first difference of log(Y), output gap is the deviation of log(Y) from the trend obtained using Hodrick - Prescott filter

[6] The package is freely available at http://cafim.sssup.it/ giulio/software/subbotools/. It is a part of General Scientific Library software

[7] Asymmetric density is of the same form, but allows for different b and a parameters for the data above and below 0

Subbotin density presented in the appendix 2 table 12 reveal that it is in indeed skewed with fat left and thinner that normal right tail.

Overall the empirical evidence supports the non normality of the series found in Fagiolo et. al. (2008) (2009) and Chistiano (2007). Contrary to this Frenke (2015) asserts that the non normaliyt is a result of a structural break in the US monetary policy in 1983 that resulted in lower output volatility. Then non normality according to him is a result of pooling two subsamples with different volatilities. However even taking this into account the non-normality cannot be rejected universally.

Table 2: Estimated parameters of the Subbotin density:US data

Growth rate			Output gap		
parameter	value	std. err		value	std. err
b	1.102	0.1268	b	1.603	0.2057
a	0.007312	0.000549	a	0.01517	0.001051
m	0.007759	0.000464	m	0.000388	0.000992
log lik	-3.2657		log lik	-2.6866	

(Fagiolo et al., 2008) also estimate the data purified from autocorrelation nad heteroscedasticity by using the residuals from the best fitting arima model to fit the shocks hitting the economy with the same density. Such problems could in fact have biased the estimates reported here, but according to Frenke (2015),the usage of residuals from the fitted model means to assume that the shock propagation mechanism in the economy is linear. In this work I do not take a stance on which one, the propagation mechanism is nonlinear or the shocks are non normal. Instead, the purpose is to analyze how a model could give rise to a non normally distributed variables given a normal shocks. For this reason I abstain from estimating the density of the fitted residuals.

6 Model

The model used in here is a basic small scale DSGE model including trend growth, Calvo price setting,external habit formation in consumption and price indexation to past inflation. The persistence parameters are included to have the model as comparable as possible with the one used in De Grauwe (2012a), that is to have the lagged variables in the aggregate demand equation and the NK Philips curve in order to have the MSV solution of the form:

$$\begin{aligned} k_t &= \lambda_0 + \lambda_1 k_{t+1} + \lambda_2 k_{t-1} + \lambda_4 z_t \\ z_{t+1} &= \rho z_t + e_{t+1} \end{aligned} \tag{9}$$

where k is a vector of endogenous state variables and z is a vector of exogenous states. The trend growth is included to back out the model implied output growth rate. The model is simulated under both Euler equation learning and the infinite-horizon learning rule as in Preston (2005). For such a study it would be more meaningful to use a richer model that is more widely used in practice and is able to fit the other stylized facts, i.e. Smets and Wouters (2007) or Christiano et. al (2005). However, for the models of this size the derivation of infinite horizon learning is hardly feasible, which would not allow for a comprehensive study of the effect of learning. In the following sections I lay out the model.

6.1 Households

Households maximize the infinite stream of utility.

$$max \sum_{T=t}^{\infty} \beta^{T-t} \big[(ln(C_t - \eta C_{t-1}) - \frac{N_t^{1+\phi}}{1+\phi}\big] \tag{10}$$

Parameter η reflects the degree of habit and C_{t-1} is the aggregate level of consumption that is used as a point of reference and is external to the agent. Parameter ϕ is the inverse of the Frisch elasticity of labor supply and N_t stands for the supply of labor. Since the economy exhibits trend growth, the inter temporal elasticity of substitution parameter σ is imposed to be 1 and the utility is additively separable to give the standard Euler equation without labor. Households maximize utility subject to a budget constraint:

$$C_t + \frac{B_t}{P_t} \leq \frac{W_t}{P_t} N_t + \frac{\Pi_t}{P_t} + i_{t-1} \frac{B_{t-1}}{P_t} \tag{11}$$

where i_{t-1} is defined as the gross interest rate

The implied first order conditions are

$$C_t : \lambda_t = \frac{1}{C_t - \eta C_{t-1}}$$

$$N_t : N_t^{\phi} = \lambda_t W_t$$

$$B_t : \lambda_t = \lambda_{t+1} \beta i_t \pi_{t+1}^{-1}$$

All the variables except N_t π_t and Π_t are growing at the trend rate γ. Before log-linearizing, first order conditions are written in terms of stationary variables

Using $c_t = \gamma^{-t} C_t$, $w_t = \frac{W_t}{P_t \gamma^t}$, $\Xi_t = \lambda_t \gamma^{t-1}$ etc. The labor supply relation and the consumption euler equation become:

$$\text{Labor supply: } N_t^{\phi} = w_t \xi_t \gamma$$

$$\text{Consumption Euler equation: } \Xi_t = \beta \gamma \xi_{t+1} i_t \pi_{t+1}^{-1}$$

6.2 Production side

Intermediate goods are produced with the diminishing returns to scale production function

$$Y_t(j) = A_t N_t^{1-\alpha}(j) \tag{12}$$

The technology shock follows an AR(1) process

$$ln(A_t) = ln(\gamma) + \rho ln(A_{t-1}) + \epsilon_t \tag{13}$$

Intermediate goods are aggregated into the final ones using Dixit-Stiglitz aggregator

$$Y_t = \Big[\int_0^1 Y_t(i)^{\frac{\epsilon-1}{\epsilon}} di\Big]^{\frac{\epsilon}{\epsilon-1}} \tag{14}$$

Given this, a cost minimization problem gives the demand for an intermediate good as a function of aggregate output and prices

$$Y_t(i) = \Big[\frac{P_t(i)}{P_t}\Big]^{-\epsilon} Y_t \tag{15}$$

The labor market is competitive, meaning that the W_t equals the marginal product of labor. Profit maximization under flexible prices yields

$$P_t(i) = \frac{\epsilon}{\epsilon-1} MC_t \tag{16}$$

with nominal marginal cost being $MC_t = \frac{1}{1-\alpha} W_t Y_t^{\frac{\alpha}{1-\alpha}} A_t^{\frac{-1}{1-\alpha}}$

The prices evolve under Calvo type price setting, where ξ is the probability that the price will stay fixed for the next period. When not changed, the price is indexed to the inflation rate in the previous period with the extent of indexation controlled by the parameter φ, 1 meaning complete indexation and 0 - none. Under this scheme the aggregate price evolves according to the following law of motion:

$$P_t = \left[(1-\xi)\big(P_{t-1}\pi_{t-1}^{\varphi}\big)^{1-\epsilon} + \xi(P_t^*)^{1-\epsilon}\right]^{\frac{1}{1-\epsilon}} \tag{17}$$

The inflation rate is defined as $\pi_t \equiv \frac{P_t}{P_{t-1}}$

The firm profit maximization problem then takes the form

$$max\sum_{T=t}^{\infty}(\beta\xi)^{T-t}\frac{U'(C_T)}{U'(C_t)}\left(\pi_{t-1,t_t-1}^{\varphi}P_t^*(\frac{\pi_{t-1,t_t-1}^{\varphi}P_t^*}{P_{t+T}})^{-\epsilon}Y_{t+T} - MC_{t,t+T}(\frac{\pi_{t-1,t_t-1}^{\varphi}P_t^*}{P_{t+T}})^{-\epsilon}Y_{t+T}\right) \tag{18}$$

The stationarized first order condition becomes

$$max\sum_{T=t}^{\infty}(\beta\xi)^{T-t}\frac{U'(c_T)}{U'(c_t)}y_T\left(\frac{P_t^*}{P_t}\pi_{t-1,t_t-1}^{\varphi} - \frac{\epsilon}{1-\epsilon}mc_{t,t+T}\frac{P_{t+T}}{P_t}\right) \tag{19}$$

A more detailed derivation of the model is provided in the appendix.

6.3 Final equations

The aggregate dynamics of the model are given by

$$\gamma\hat{y}_t - \eta\hat{y}_{t-1} = \gamma E_t(\hat{y}_{t+1}) - \eta\hat{y}_t - (\gamma-\eta)(\hat{i}_t - \hat{\pi}_{t+1})$$

$$\hat{\pi}_t - \varphi\hat{\pi}_{t-1} = \beta(\hat{\pi}_{t+1} - \varphi\hat{\pi}_t) + \kappa\left((\frac{\phi+\alpha}{1-\alpha} + \frac{\gamma}{\gamma-\eta})\hat{y}_t - \frac{\eta}{\gamma-\eta}\hat{y}_{t-1} - \frac{\phi+1}{1-\alpha}\hat{a}_t\right)$$

$$\hat{a}_t = \rho\hat{a}_{t-1} + \epsilon_t$$

$\kappa = \frac{1-\xi}{\xi}\frac{1-\beta\xi}{1+\omega\beta}$

The system is closed with the smoothed interest rate rule

$$\hat{i}_t = \rho_i\hat{i}_{t-1} + (1-\rho_i)\psi_\pi\hat{\pi}_t + \nu_t \tag{20}$$

where ν_t is the monetary policy shock.

In is a common practice to write the model equations in terms of output gap and to use the Taylor rule. In the adaptive learning setup this raises some conceptual issues. The main premise for using the adaptive learning framework is the idea that agents do not know the structural parameters of the model or even if they do, it is not known how they are mapped into the model's reduced form. At the same time the agents are aware of the shocks hitting the economy and know their laws of motion. The main idea of this particular study is to test the implication of the framework for the dynamics of output growth rate and the output gap. However, if written in terms of output gap the technology shock appears in the model through the natural rate of interest. In this case given that habit formation natural interest rate is $r_t^n = \gamma(E\hat{y}_{t+1}^f - \hat{y}_t^f) - \eta(\hat{y}_t^f - \hat{y}_{t-1}^f)$ and each $\hat{y}_t^f$ is an AR(1) process, so there is no direct correspondence with the current a_t. Writing the model in this way would violate the above mentioned assumption. An alternative way to overcome this problem, as done by Milani (2006), (2007) is to simply impose that r_t^n follows an AR(1) with persistence parameters estimated separately from other structural parameters. This approach simplifies the model, but is not entirely sound and does not allow to back out the sticky price output. In the end, having the output gap or the Tailor rule in the model requires to have the flexible price output in the aggregate equations. Having it as an exogenous variable violates the major assumption regarding the information contained by the agents. Including flexible price output as a state is technically complicated and conceptually daunting, as this would mean that it is actually observed and forecasted. Instead after simulating the model I additionally calculate the y_t^f to back out the model implied output gap. For the same reasons the interest rate rule does not contain the output gap.

It is also common to use expected value of inflation in the interest rate rule. Such a case would pose a choice whether the central bank forms expectations rationally or the same learning procedure as all the agents in the economy. However if the central bank is aware that expectations in the economy are not formed rationally, its policy problem becomes to tame the expectations instead of stabilize the inflation (Gaspar et al., 2006). Since the analysis of the policy rules is beyond the scope of this study, I keep the agnostic stance on this issue by using the current value of the variable in the interest rate rule.

6.4 Model with Infinite Horizon Learning

The final equations for this type of learning are derived from the combination of consumption Euler equation with the intertemporal budget constraint (21) and from the firms optimal price - setting decision (22). A more detailed derivation can be found in the appendix. Law of motion for output

$$y_t\left(\frac{\beta\eta}{\gamma\gamma}+\frac{\gamma}{\gamma-\beta}-1\right)=+y_{t-1}\left(\frac{\gamma\eta}{\gamma(\gamma-\beta)}-\frac{\eta}{\gamma}\right)-\frac{\beta(\gamma-\eta)i_t}{\gamma(\gamma-\beta)}+ \\ +\sum_{T=t}^{\infty}(\frac{\beta}{\gamma})^{T-t}\left(\frac{\beta(\gamma-\eta)p_{t+1}}{\gamma(\gamma-\beta)}-y_{t+1}\left(\frac{\eta}{\gamma}\left(\frac{\beta}{\gamma}\right)^2-\frac{\beta}{\gamma}\right)-\left(\frac{\beta}{\gamma}\right)^2\frac{(\gamma-\eta)i_{t+1}}{\gamma-\beta}\right) \tag{21}$$

Law of motion for inflation

$$\pi_t-\varphi\pi_{t-1}=\frac{1-\xi}{\xi}\kappa\hat{y}_t\big(\omega_2-\frac{\eta}{\gamma-\eta}\beta\xi\big)-\frac{1-\xi}{\xi}\kappa\frac{\eta}{\gamma-\eta}\hat{y}_{t-1}-\frac{1-\xi}{\xi}\beta\xi\varphi\hat{\pi}_t+ \\ +\frac{1-\xi}{\xi}\sum_{T=t}^{\infty}\big(\hat{\pi}_{t+1}(\beta\xi-\varphi(\beta\xi)^2)+ \\ +\frac{1-\xi}{\xi}\sum_{T=t}^{\infty}(\beta\xi)^{T-t}\big(\hat{y}_{T+1}(\kappa\omega_2\beta\xi-\kappa\frac{\eta}{\gamma-\eta}(\beta\xi)^2\big)+\frac{1-\xi}{\xi}\sum_{T=t}^{\infty}(\beta\xi)^{T-t}\frac{\phi+1}{1-\alpha}\hat{a}_T\big) \tag{22}$$

The interest rate rule is the same as before

$$\hat{i}_t=\rho_i\hat{i}_{t-1}+(1-\rho_i)\psi_\pi\hat{\pi}_t+\nu_t$$

6.5 Calibration

Since a model with the infinite-horizon learning cannot be estimated with the regular software, a problem arose getting the values for the parameters that cannot be calibrated from the steady-state. At the same the the purpose of the study is not to fit the model to the data, which makes it sensible to have the reasonable and regular parameter values. The parameters β, γ and α where calibrated from the steady state using the US quarterly data from 1984 to 2015. [8]. Since the 0 steady state inflation is assumed throughout, $\beta = \frac{1}{i}$, i being the sample average of the nominal 3-month T-bill. γ is the average quarterly growth rate of the real US GDP and α is the sample average of the capital share of gross domestic income.

According to Milani (2006), (2007), when adaptive learning is added to the model, the parsistence parameters η and ϑ become superfluous. The model used in his papers is the same model as here, the only difference being the interest rate rule, definition of the shock process and the fact that habit is internal. Anyway I use the estimates for η and ϑ from Milani (2007). The value for the substitutability parameter ϵ is usually assumed and varies significantly in different studies. Simulations showed that this parameter affects the model dynamics very little, so I simply use the value 7.69 from Milani (2007).

For the other parameters I simply assumed reasonable values. The parameter θ is assumed to be 0.5, to give Frisch elasticity of labor supply equal to 2, consistent with the 40 hour work week. Calvo parameter ξ is set to 0.67, implying an average price duration of 9 months. The interest rate smoothing is assigned 0.5 to be the same as in the DeGrauwe (2012a). Ideally a monetary policy shocks should be an i.i.d. process, but to make the matrix defining the evolution of z_t in (9) invertible [9] is assigned to be an AR(1) process with 0.01 coefficient on the lag. $\sigma = 1$ to allow for a trend growth in the

[8]Data was taken from St. Louis FRED database at research.stlouisfed.org/fred2/

[9]This is required by the matlab codes which simulates the model

model.

Table 3: Structural parameters

Parameter		Value	Source
Discount rate	β	0.9629	Calibrated from the data
Trend Growth rate	γ	1.0066	Calibrated from the data
Labor Share	α	0.362	Calibrated from the data
Risk aversion	σ	1	Imposed
External Habit	η	0.117	Taken from Milani 2007
Inverse Frisch elasticity of labor supply	θ	0.5	Assumed
Inflation Indexation	ϑ	0.032	Taken from Milani 2007
Degree of Substitutability	ϵ	7.69	Taken from Milani 2007
Calvo Price Stickiness	ξ	0.67 (0.92 in inf-hor)	Assumed /Milani 2007
Interest Rate Smoothing	ρ	0.5	Assumed
Int. Rate reaction to Inflation	ψ_π	1.5	Assumed
Persistence of Technology	ρ_a	0.9	Assumed
Persistence of mon. pol. Shock	ρ_i	0.01	Assumed
Standard Deviation of techn shock	σ_a	0.0072	Assumed
Standard Deviation of mon. pol shock	σ_i	0.01	Assumed

6.6 Learning

In this exercise I implement to most standard form of statistical learning with the constant gain. Since the model is small it is reasonable to assume that the MSV form is known to the agent, also all the endogenous variables are observed. Therefore the perceived law of motion is assumed to take the same form as the MSV solution, meaning that the agents know the structure of the economy, but not how the structural parameters are mapped into the reduced form. The forecasting equation takes the form:

$$k_t = P^k_{t-1} k_{t-1} + P^z_{t-1} z_{t-1} \tag{23}$$

where $k_t = \begin{pmatrix} \hat{y}_t \\ \hat{\pi}_t \\ \hat{i}_t \end{pmatrix}$ and z = $\begin{pmatrix} \hat{a}_t \\ \hat{\nu}_t \end{pmatrix}$ and $x_t = \begin{pmatrix} k_t \\ z_t \end{pmatrix}$ The forecasting coefficients are updated once the new information becomes available

$$R_t = R_{t-1} + g(x_{t-1}x'_{t-1} - R_{t-1})$$
$$P_t = P_{t-1} + gR^{-1}x_{t-1}(k_t - x'_{t-1}P_{t-1})$$

There are many different ways how to initiate the learning algorithm and the way it is done may lead to different dynamics and the forecast errors. The initial values of P and R can be estimated from the pre-sample data (Slobodyan and Wouters, 2012b), estimated from the model data, simulated under rational expectations, started at ad-hoc values (Garceles-Poveda and Giannitsarou, 2007) or may begin at the rational expectations equilibrium. Since this model is not estimated on the real world data, a natural alternatives would be to start from the rational expectations equilibrium or from using the coefficients obtained from regressing a model simulated data. The latter is conceptually puzzling due to a self-referential nature of the model, as according to Garceles - Poveda (2007), this would require the agents to form the correct expectations and thus know the REE before they start learning about it. Initiating the system at REE suffers from a similar problem, but at least can provide an illustration of how the economy behaves if we start observing it at REE. Since under constant gain learning the forecast parameter values converge to the ergodic distribution around the REE (Honkapohja and Evans, 2001), it being at REE is not unrealistic. For this reason I initiate the simulation at the RE values of P and R.

7 Simulations

Simulating the model with adaptive learning is not a trivial task. The official Dynare does not contain the option for it as it cannot deal with time-varying law of motion parameters. The most popular alternatives available online are the replication code from the Slobodyan and Wouters (2012a), which includes the Dynare modification for Euler equation learning. However, it is a replication code not supposed to work for every model and has a very limited range of learning options. Also it is possible to use a Macroeconomic Model Base Dynare package[10]. Through a user - friendy interface it allows to run a set of the most popular DSGE models estimated under euler equation learning. Still both of these option lack flexibility in choosing a model, initial conditions, PLM and most importantly do not allow for an infinite horizon learning. Bruce Preston‘s matlab code is available online as well, but is very model specific. For my purposes I used a code provided by Garceles-Poveda & Giannitsarou (2007), which runs the model under learning after calculating its MSV solution. Written for an Euler equation learning, the code is relatively simple to modify and works well with a small scale the models like the one used here [11].

Model is written and simulated using the deviations from the stationarized steady states of the variables and the flexible price output was calculated outside the model once the series of technology shocks was known. The the output growth rate was obtained by getting the level of output from its deviation, re-trending it and applying the transformation

$$g(t) = \frac{y(t) - y(t-1)}{y(t-1)}$$

The output gap was backed out using $x_t = \hat{y}_t - \hat{y}f_t$

[10]available at macromodelbase.com

[11]A detailed explanation about the modifications made to the code can be found in the appendix

7.1 Euler equation learning

To get a better sense of how learning affects the behavior of the system I provide graphs of an ordinary paths [12] of the output growth rate and gap simulated for 1000 periods with the normally distributed shocks under the REE and with the gain parameter taking the values 0.01, 0.02 and 0.05 implying the usage of 25, 12.5 and 5 years of data respectively. The effect on the growth rate can be seen more clearly as it is less persistent. Overall as the gain increases series become more volatile, large absolute values become more frequent. Under the gain 0.05 both of the series start to exhibit large jumps known as exits or escapes happening when the model is hit by an unlikely combination of either positive or negative large shocks causing the beliefs to severely propagate the response.

In simulating models with learning, it is common to use a projection facility - to stop updating the coefficients in the PLM once they become explosive - to avoid an unreasonably high escapes. The most usual way to implement it is to simply set the coefficients in the PLM to their values at t-1. Theoretically it should not be invoked very often under the gain as large as 0.05. Also it is normal to eliminate the simulations when it is invoked too often. In a current case quite often the facility was invoked in more than 30 % of the periods, possibly reflective a small determinacy space of the model [13]. Moreover while being able to keep the series in a tight neighborhood of the steady state, it was also giving an even more rare and exponentially larger deviations than those seen in the graphs below, which I suspect be the result resetting the parameters for too many times, while forecast errors continue to accumulate. This led me to abstain from using the facility.

[12] simulations were done using the same seed

[13] I also tried to run the model with high values of η and ϑ, but could not find a parameter combination that would not violate Blanchard - Kahn conditions

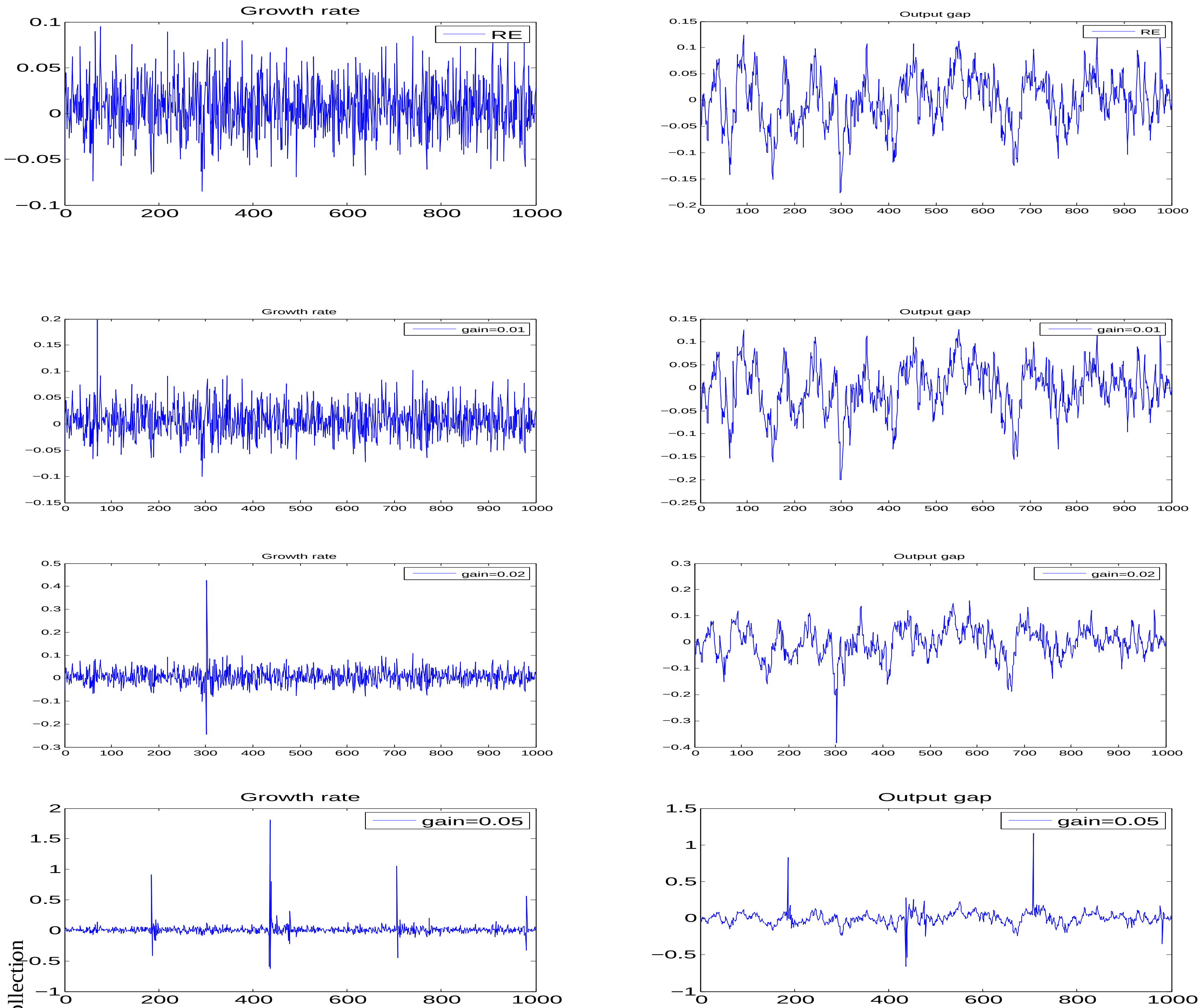

Figure 3: typical bahavior of simulated output growth rate and gap series with different constant gain parameters

For the statistical inference I performed a 700 Monte Carlo simulations for each value of the gain with the length of 270 quarters, approximately equal to the length of the longest available post-war data series. The numbers reported in the tables 4-7 are the means over the given statistic or estimate of each Monte Carlo simulation. Occasional huge exits were contaminating all the estimates. To tackle this I removed the series containing a large deviation, which I defined as an observation that is more than seven times larger in absolute value than the standard deviation of the series that would prevail under the rational expectations[14].

As seen in the tables 4 and 5, increasing the gain leads to consistently higher standard deviations and the kurtosis, even after rare events are excluded. Also the average p-values of the Jarque-Berra test decline. Even though they are still above 10% on average with the gain being 0.05, it is apparent that the normality is being rejected in the increasing share of the simulated series. Under the 0.05 gain for both growth rate and the gap it is rejected in around 2/3 of the sample, suggesting the relatively high average p-value is affected by a small number of outliers. The means and the skewness are not affected, indicating that both of the series remain symmetric. Also as the gain goes up, an increasing share of the simulated series have to be discarded due to containing an unlikely event.

Table 4: growth rate statistics

gain	Mean	St. dev	Skewness	Kurtosis	JB p-value	# of p-values < 0.05	sample
0	0.007003	0.028786	0.080713	3.047506	0.347497	64	700
0.01	0.00704	0.029964	0.099807	3.607333	0.245692	200	500
0.02	0.007051	0.03037	0.089993	3.837327	0.220389	157	404
0.05	0.007081	0.031382	0.161483	4.79983	0.105751	128	183

The estimation results in the tables 5, 6 show that learning mechanism is generally able to produce the fact tails of the growth rate and gap series. As in Ascari et. al (2013) the sources of stickiness and persistence limit the propagation of the shocks and give the

[14]A very similar approach was taken in (Slobodyan and Wouters, 2009)

Table 5: output gap statistics

gain	Mean	St. dev	Skewness	Kurtosis	JB p-value	# of p-values < 0.05	sample
0	3.33E-05	0.048205	0.015464	2.876194	0.197917	202	700
0.01	-9.2E-05	0.048336	-0.00017	3.286533	0.166645	252	581
0.02	0.000419	0.049318	0.025963	3.607906	0.15998	225	505
0.05	0.00016	0.049242	-0.02933	4.930768	0.111643	217	343

tails thinner than normal, especially in case of output gap. As the gain increases, the parameter b stably declines towards 1. However at the gain 0.05 it is still not as low as seen in the data (see table 2 pooled growth rate sample). It could be possible to keep increasing the gain, but this posits a trade off that huge share of the simulations would have to be discarded. At the same time larger gains imply a very short memory and are not consistent with the empirical estimates around 0.01 and 0.02 (Orphanides and Williams, 2005)[15]. Another limitation is that the learning method does not differentiate between going up or down and thus cannot on average generate a skewed distribution, similar to that of the output gap series for instance. A single simulated series can be highly positively or negatively skewed with the same probability however.

Table 6: growth rate estimations

	b	a	m	likelihood
RE	2.007817	0.028608	0.006991	-2.133884
	-0.299913	-0.001932	-0.000383	
0.01	1.852834	0.028767	0.006951	-2.098553
	0.309770	0.002471	0.000447	
0.02	1.800000	0.028700	0.006930	-2.090000
	0.339112	0.002544	0.000516	
0.05	1.537931	0.027801	0.006908	-2.065006
	0.277716	0.003492	0.000723	

[15] gain estimated by fitting the model to match the expectations in the professional forecasters survey

Table 7: output gap estimations

	b	a	m	likelihood
RE	2.288053	0.049529	-0.000012	-1.629459
	0.667595	0.008004	0.012994	
0.01	2.163203	0.048556	-0.000049	-1.634029
	0.707790	0.010210	0.012346	
0.02	2.006480	0.048128	0.000377	-1.618417
	0.595835	0.009980	0.012683	
0.05	1.722022	0.045352	0.000212	-1.641723
	0.594743	0.013876	0.012146	

7.2 Infinite Horizon Learning

Next I repeat the same exercise for he model with the infinite horizon forecasting rule. The first two terms in the consumption rule (see appendix) reflect the expected total lifetime income and make the rule in line with the permanent income hypothesis (Preston, 2005). Since there is a direct correspondence between consumption and output in the model, this provides some intuition why the output growth and the gap are less volatile than under the previous rule. At the same time the agents know the persistence of the technology process and use this in calculating its infinite discounted sum, which means that this system is much more responsive to the current shocks. To limit this sensitivity I increased the Calvo parameter to 0.92, which is a value estimated in Milani (2007) under the same learning rule. However even then the simulated series exhibit unreasonably large deviations both with and without projection at the gain parameter as low as 0.01 as can be seen in the last row of simulated paths. In the least squares constant gain learning deviations may arise either due to large gain or low variance-covariance matrix. In a current case, this is in a large part due to the latter (see appendix for the REE coefficients and covariance matrices), as the exogenous process have low standard deviations meaning that the new information is trustable and taken into account with a large weight in the learning process. In contrast, stochastic gradient learning [16] provides a much more reasonable dynamics of

[16]in stochastic gradient learning parameters are uptated with $P_t = P_{t-1} + gx_{t-1}(k_t - x'_{t-1}P_{t-1})$ not taking R into account

both variables. From the first look at the system dynamics larger gain tends to increase the variability of both output growth rate and the gap relative to the REE. At the same time escapes begin to occur at a relatively high constant gain value of 0.2.

The statistics and estimates from the Monte Carlo simulations show a similar dynamics to the previous model under REE. Since least squares constant gain learning gives too much variability and too many samples have to be discarded, it is not analyzed here. The stochastic gradient learning (tables 8-11) on the other hand does not generate that many exits so that even at the gain equal 0.4 a half and 1/3 of the samples for the growth rate and the gap respectively could be included in the estimation. However being able to provide more stability, the method does not create enough propagation. In case of output growth rate it can produce kurtosis only slightly higher than 3 with the constant gain as large as 0.2 and 0.4 implying memory of 5 and 2.5 quarters respectively. The same pattern can be seen in the estimation results in the table 10 the average of the estimate for $\hat{b}$ parameter being 1.7218 at the largest gain used.

Table 8: Statistics, growth rate

gain	Mean	St. dev	Skewness	Kurtosis	JB p-value	# of p-values < 0.05	sample
RE	0.006828	0.021393	0.063751	3.018645	0.350522	62	1000
0.01	0.006861	0.022651	0.065879	3.032529	0.353346	81	1000
0.05	0.006857	0.023234	0.066964	3.049591	0.345828	95	1000
0.1	0.006875	0.024142	0.078734	3.080547	0.330432	114	999
0.2	0.006934	0.026391	0.079428	3.239235	0.293288	191	975
0.4	0.007027	0.030313	0.083129	3.528523	0.201091	202	507

As in the previous case the model smooths and gives even thinner tails than normal for the output gap under RE. The presence of learning increases the moments to the desired direction, yet the extent of it is too small. The small values of constant gain increase the variability, but the kurtosis remains the same, meaning that in this case the shocks of all sizes are propagated evenly. Even though there is less propagation than for the growth rate, Jarque-Bera p-values are on average lower. This could possibly come

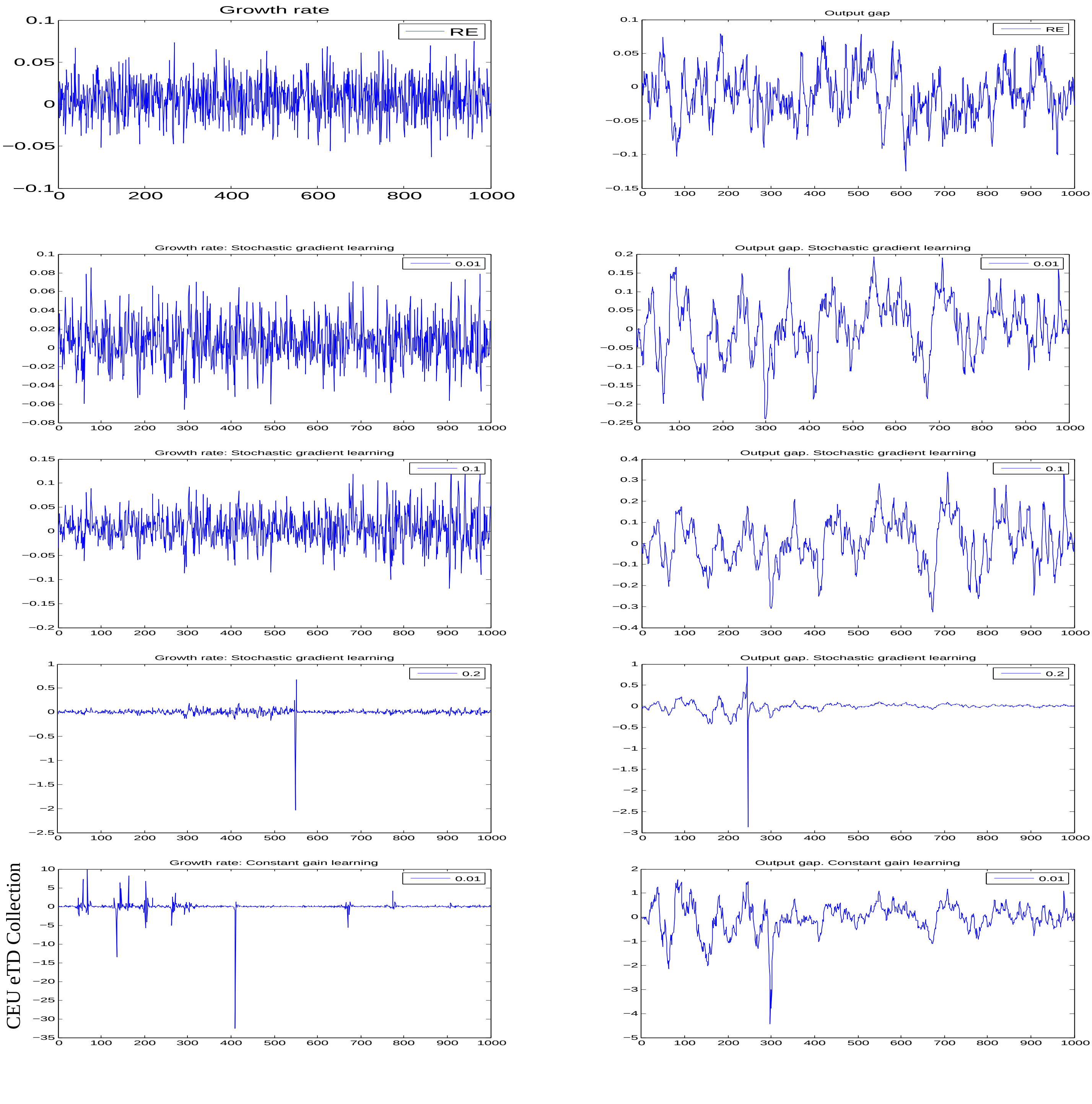

Figure 4: typical bahavior of simulated output growth rate and gap series with different constant gain parameters

Table 9: Estimates, growth rate

	b	a	m	lik
RE	2.0296	0.0214	0.0068	-2.4306
	0.2949	0.0014	0.0003	
0.01	2.0138	0.0225	0.0069	-2.3739
	0.3023	0.0017	0.0004	
0.05	2.0043	0.0231	0.0069	-2.3486
	0.2975	0.0017	0.0004	
0.1	1.9825	0.0239	0.0069	-2.3110
	0.3008	0.0019	0.0004	
0.2	1.8800	0.0255	0.0069	-2.2300
	0.3067	0.0023	0.0005	
0.4	1.7218	0.0282	0.0069	-2.0942
	0.3058	0.0033	0.0007	

from the fact that here some of the samples had to be discarded, meaning that there are some containing the outliers close to the cutoff affecting the whole statistic. A more sizable increase in the kurtosis and the $\hat{b}$ estimate can only be achieved be increasing the constant gain to 0.4, but even then the learning only outweighs the internal smoothing of the normally distributed shocks.

Table 10: Statistics, output gap

gain	Mean	St. dev	Skewness	Kurtosis	JB p-value	# of p-values < 0.05	sample
RE	0.000701	0.035115	0.002244	2.890697	0.178955	357	1000
0.01	-8.7E-05	0.054456	-0.00791	2.860067	0.146643	469	997
0.05	0.000926	0.056424	-0.00741	2.855983	0.142875	467	992
0.1	-0.00039	0.059587	-0.01046	2.846652	0.143037	451	969
0.2	9.69E-05	0.063664	-0.02662	2.866667	0.156779	363	826
0.4	0.001524	0.066339	-0.00725	2.996784	0.153416	136	311

One of the most straightforward ways to make the system more stable under least-squares constant gain learning would be to decrease the persistence of technology process. Yet $\rho = 0.9$ is a value consistent with the data and trying to decrease it to 0.6 still gives very similar dynamics. It is as well possible to increase the Calvo parameter ξ. However 0.92 is already an unrealistic value implying expected price duration of 12.5 quarters. At the same time a value of 0.67 is giving little stability under both least-squares

Table 11: Estimates, output gap

	b	a	m	lik
RE	2.2739	0.0360	0.0007	-1.9464
	0.6920	0.0059	0.0093	
0.01	2.4642	0.0566	-0.0001	-1.5130
	1.0956	0.0111	0.0159	
0.05	2.4845	0.0589	0.0010	-1.4794
	1.0457	0.0124	0.0166	
0.1	2.4794	0.0621	-0.0005	-1.4262
	1.2359	0.0141	0.0172	
0.2	2.4242	0.0659	0.0001	-1.3619
	1.0593	0.0153	0.0184	
0.4	2.1635	0.0664	0.0014	-1.3196
	0.7690	0.0144	0.0173	

and stochastic gradient frameworks. Alternatively taming the inflation with higher λ_π increases the variability of output and makes the deviations larger. However decreasing it would not give strong enough monetary policy response to endogenous variables and would lead to indeterminacy.

7.3 Heuristics

For comparison I simulate the model with the same structural parameters under heuristic learning rule as in De Grauwe (2012a) using the code published in De Grauwe (2012b). Heuristic learning method like the Euler equation learning replaces one period ahead forecasts with the ones formed subject to cognitive biases. In short, the agents are now switching between two forecasting rules: fundamentalist expecting the steady state and extrapolative - predicting the next period's value to be equal to that in the previous period:

$$E(\hat{y}_{t+1}) = \hat{y}_{t-1} \text{ and } E(\hat{y}_{t+1}) = 0 \text{ for output} \quad (24)$$

$$E(\hat{\pi}_{t+1}) = \hat{\pi}_{t-1} \text{ and } E(\hat{\pi}_{t+1}) = 0 \text{ for inflation}$$

The ratio of population using one rule versus the other is referred to as 'animal spirit' by the author and is determined by

$$\alpha_f = \frac{exp(\gamma U_{f,t})}{exp(\gamma U_{f,t}) + exp(\gamma U_{e,t})} = 1 - \alpha_e$$

and

$$U_{f,t} = \rho U_{f,t-1} - (1-\rho)(y_{t-1} - \tilde{E}_{f,t-2} y_{t-1})^2 \tag{25}$$

Parameter γ determines the intensity of learning. When $\gamma = 0$, agents keep switching between the two rules randomly and as $\gamma \to \infty$ there is no noise in the utility and all the agents switch to the same rule when they do so.

Within this framework I carry out the estimations by varying σ and the parameter ρ determining how much weight an agent put to the new information relative to the entire previous history, which can be considered as an analogy of a gain parameter in the previous setup. $\rho = 1$ means perfect memory and $\rho = 0$ - no memory. The information from the Monte Carlo simulations is summarized in tables 13-16 in the appendix 2. First I simulate the model with the values used in the book (Grauwe, 2012b), $\gamma = 1$,$\rho = 0.5$, meaning that agents assign equal weights to new information relative to entire history and that the learning is relatively noisy. The evidence from this benchmark model is mixed. It is indeed able to give the fat tails in the output growth rate, but not for the output gap. In fact the propagation is too high than necessary. Also it should be noted that the model implied variance is much higher than in the data or in the model with statistical learning.

A benchmark ρ parameter implies unreasonably short memory, something comparable to a constant gain value of 0.5 so it would natural to increase it. However at its higher values keeping the γ constant makes the system too dependent on the first few shocks in the simulation that determine to which forecasting rule all the agents converge to and are very unlikely to switch in the future. In this case outputs deviation from the steady either gets stuck at the level where the system is taken by initial sequence of shocks and exhibits very little variability (when everyone using fundamentalist rule), or has an upward/downward trend (when everyone is extrapolating). For this reason I perform the estimates by increasing both ρ and γ. I increase ρ to 0.9 and 0.98, which could be compared to the constant gain equal 0.(1) and 0.204. The γ parameter is increased to 1000 implying very little noise in the information observed and giving an almost perfectly correlated choices of all the agents [17]. At these new values, the propagation in output growth rate becomes enormous, giving the kurtosis above 100. This could happen due to exits described previously, but in this framework the learning rule do not nest the REE therefore it is impossible to come up with a cutoff value to discard the samples or to induce a projection facility. A higher variance here is partly caused by these exits. However while under statistical learning the system is evolving similarly as under RE with some rare and brief deviations, here the system fluctuates between between two moderately high values on both sides of the steady state (figure 5 in Appendix 2 gives a visual illustration of this). Increased learning efficiency does not propagate the shocks to the output gap although it does increase the overall variability as well. The means of the estimated $\hat{b}$ parameters are sizeably larger than two. This happens due to several huge outlier estimates in the sample as the parameter is bound by 0 from below, but is unbounded from above. In fact the median over the samples of estimates appears to be between 2.05 and 2.07 at $\gamma = 1000$.[18]

[17] Experiments with different values of γ revealed that the problems associated with $\gamma = 1$ disappear only at such a high values

[18] the result is not reported here for compactness.

8 Discussion and Conclusion

It cannot be concluded that any of the frameworks presented here is able to match the fourth moment of the data perfectly, each posing different problems. Euler equation learning is able to deliver reasonable levels of internal propagation, while keeping the standard deviation of the series close to the data as well. Yet when unreasonable exits are discarded the propagation is still not big enough under the reasonable values of a constant gain parameter. Infinite horizon learning on the other hand gives too much propagation as there the current value of output depends on the infinite expectation of the variables exogenous to the agent, which makes the system highly responsive to the current exogenous shocks. Also the need to make infinite horizon forecasts increases the forecast errors and thus the coefficients in the PLM are being updated more drastically. In this case recursive least squares constant gain learning makes the system too unstable, while stochastic gradient counterpart is not able to provide enough internal propagation under a reasonable values of a constant gain. Heuristic learning poses slightly different problems. While the benchmark specification used in the book delivers the kurtosis high enough, the overall variance of both the gap and the growth rate become unrealistically high and the estimated $\hat{b}$ too low for the former. Moreover, the so called 'animal spirits' in this framework do not happen when the agents are equipped with long enough memory comparable to the values of constant gain used before and the imperfect learning capacity. When on the other hand it is increased, the system becomes highly unstable and described by large exits similar to those in the constant gain learning. Also none of these standard frameworks can produce skewed distribution on average.

The model used here is small featuring only one of the many rigidities commonly used in the DSGE models, also persistence parameters are set to low values. A richer model with a non trivial persistence parameters and capital may provide more stability and less exits under learning. At the same time a pattern observed here is expected to remain the same - infinite horizon learning should bring higher propagation. However due to computational and technical issues associated with the infinite horizon learning, it still has not been implemented in medium - scale DSGE models.

Another caveat of the model comes from low persistence parameters making the model highly dependent on the expectations and no response of monetary policy to output, letting it vary more than in the usual models. The combination of these two possibly create the explosiveness of the system observed in all the setups. However, the response to output could not be included due to conceptual issues associated with learning, which caused in the indeterminacy of the model under relatively high persistence parameters. Still the results support the original conjecture that learning is a source of endogenous propagation and helps to match the fourth moments of the series of output growth and gap.

9 References

References

Guido Ascari, Giorgio Fagiolo, and Andrea Roventini. Fat-tail distributions and business cycle models. *Macroeconomic dynamics Vol. 17, 1-12*, 2013.

Paul Beaudry and Franck Portier. News-driven business cycles: Insights and challenges. *Journal of Economic Literature, Vol. 52(4), 993-1074*, 2014.

Jess Benhabib and Chetan Dave. Learning, large deviations and rare events. *Review of Economic Dynamics, Vol.17, issue 3, 367-382*, 2014.

Giolio Botazzi and Angelo Secchi. Maximum likelihood estimation of the symmetric and asymmetric exponential power distribution. *LEM Working Paper Series 2006/19*, 2008.

Lawrence Christiano. On the fit of new keynesian models: Comment. *Journal of Business and Economic Statistics, Vol. 25, No. 2, p. 143-151*, 2007.

Timothy Cogley and James M. Nason. Output dynamics in real-business-cycle models. *The American Economic Review; Vol. 85, 492 - 511*, 1995.

Paul DeGrauwe. Booms and usts in conomic activity: A behavioral explanation. *Journal of Economic Behavior and Organization; Vol. 83, 484-501*, 2012a.

Stefano Eusepi and Bruce Preston. Expectations, learning, and business cycle fluctuations. *American Economic Review, Vol. 101, 2844 - 2872*, 2011.

Giorgio Fagiolo, Mauro Napoletano, and Andrea Roventini. Are output growth rate ddistributions fat-tailed? some evidence from oecd countries. *Journal of Applied Econometrics, Vol 23, 639 - 669*, 2008.

Giorgio Fagiolo, Mauro Napoletano, Marco Piazza, Andrea Roventini, and Marco Piazza. Detrending and the distributional properties of the u.s. output time series. *Economics Bulletin; Vol.29, issue 4, 3155-3161*, 2009.

Reiner Franke. How fat - tailed is us output growth? *Metroeconomica, Vol.66, issue 2, 213-242*, 2015.

Eva Garceles-Poveda and Chryssi Giannitsarou. Adaptive learning in practice. *Journal of Economics Dynamics and Control, Vol.31, 2659-2697*, 2007.

Vitor Gaspar, Frank Smets, and David Vestin. Adaptive learning, persistence and optimal monetary policy. *ECB Working Paper Series, No. 644*, 2006.

Paul De Grauwe. Top - down versus bottom - up macroeconomics. *CESifo Economic Studies, Vol. 56, 465 - 497*, 2010.

Paul De Grauwe. *Lectures on Behavioral Macroeconomics*. Princeton University Press, 2012b.

Seppo Honkapohja and Rober Evans. *Learning and Expectations in Macroeconomics*. Princeton University Press, 2001.

Charles L. Evans Lawrence J. Christiano, Martin Eichenbaum. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy; Vol.113, issue 1, 1-45*, 2005.

Fabio Milani. A bayesian dsge model with infinite-horizon learning: Do "mechanical" sources of persistence become superfluous? *International Journal of Central Banking, Vol.2, issue 3, 87-206*, 2006.

Fabio Milani. Expectations, learning and macroeconomic persistence. *Journal of Monetary Economics, Vol.54, 2065- 2082*, 2007.

Fabio Milani. The modelling of expectations in empirical dsge models: A survey. *Advances in Econometrics; Vol. 28, 3-38*, 2010.

Athanasios Orphanides and John C. Williams. The decline of activist stabilization policy: Natural rate misperceptions, learning, and expectations. *Journal of Economic Dynamics and Control, Vol. 29 p.1927-1950*, 2005.

Agnieszka Markiewicz Paul De Grauwe. Learning to forecast the exchange rate: Two competing approaches. *Journal of International Money and Finance, Vol. 32. 42-72*, 2013.

Bruce Preston. Learning about monetary policy when long-horizon expectations matter. *International Journal of Central Banking, Vol. 1, issue 2, 81-126*, 2005.

Sergey Slobodyan and Raf Wouters. Learning in an estimated medium-scale dsge model. *CERGE-ER Working Paper Series wp396, The Center for Economics Research and Graduate Education - Economic Institute, Prague*, 2009.

Sergey Slobodyan and Raf Wouters. Learning in a medium-scale dsge model with expectations based on small forecasting models. *American Economic Journal:Macroeconomics, Vol.4 issue 2, 65-101*, 2012a.

Sergey Slobodyan and Raf Wouters. Learning in an estimated dsge model. *Journal of Economic Dynamics and Control, Vol. 36, 26-46*, 2012b.

Frank Smets and Rafael Wouters. Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review; Vol.97, issue 3, 586-606*, 2007.

Noah Williams. Adaptive learning and the business cycle. *mimeo, University of Wisconsin - Madison*, 2004.

Noah Willimas. Escape dynamics in learning models. *University of Wisconsin - Madison*, 2014.

10 Appendix 1

10.1 linearization and derivation of the model

Households

Loglinearized household conditions:

Consumption Euler equation: $\hat{\Xi}_t = \hat{\Xi}_{t+1} + \hat{i}_t - \hat{\pi}_{t+1}$

Labor Supply: $\phi \hat{N}_t = \hat{w}_t + \hat{\Xi}_t$

and $\hat{\Xi}_t = \dfrac{-\gamma \hat{c}_t - \eta \hat{c}_{t-1}}{\gamma - \eta}$

Getting the IS relation

Plugging the expession for $\hat{\Xi}_t$ into the Euler equation in terms of consumption. There is no capital or government in the model and the output is perishable implying $C_t(i) = Y_t(i)$. Using this market clearing relation we obtain the IS relation

$$\gamma \hat{y}_t - \eta \hat{y}_{t-1} = \gamma E_t(\hat{y}_{t+1}) - \eta \hat{y}_t - (\gamma - \eta)(\hat{i}_t - \hat{\pi}_{t+1}) \tag{26}$$

Firm side

Loglinearized first order condition to the price setting problem under Calvo

$$log(P_t^*) = \sum_{T=t}^{\infty} (\beta \xi)^{T-t} \big(m\hat{c_{t,t+T}} + log(P_{t+T}) - log(\pi^{\varphi}_{t-1,t+T-1}) \big) \tag{27}$$

real marginal cost depends on the marginal costs at t and the price differential in the subsequent periods before the price is reset again

$$\hat{mc}_{t+T} = \hat{mc}_t - \frac{\alpha\epsilon}{1-\alpha}(log(\pi^{\varphi}_{T-1,t-1}P^*_t) - log(P_{t+T}) \quad (28)$$

using this fact after some manipulation the first order condition can be rewritten in the form

$$log(P^*_t) - log(P_t) = \beta(log(p^*_{t_1}) - log(P_{t+1}) + \frac{1-\beta\xi}{1+\omega\beta}\hat{mc}_t \quad (29)$$

The loglinearized price law of motion

$$\hat{P}_t = \xi\hat{P}_{t-1} + \hat{P}^*_t(1-\xi) + \xi\gamma\hat{\pi}_{t-1} \quad (30)$$

Subtracting $\xi\hat{P}_t$ and rearranging gives the relation between reset price and the inflation rate

$$\hat{P}^*_{t+1} - \hat{P}_{t+1} = \frac{\xi}{1-\xi}(\hat{\pi}_{t+1} - \varphi\hat{\pi}_t) \quad (31)$$

Using the definition of the marginal cost and the labor supply relation the expression for real marginal cost in terms of the output and the level of technology is given by

$$\hat{mc}_t = (\frac{\phi+\alpha}{1-\alpha} + \frac{\gamma}{\gamma-\eta})\hat{y}_t - \frac{\eta}{\gamma-\eta}\hat{y}_{t-1} - \frac{\phi+1}{1-\alpha}\hat{a}_t \quad (32)$$

Using this gives the Philips curve

$$\hat{\pi}_t - \varphi\hat{\pi}_{t-1} = \beta(\hat{\pi}_{t+1} - \varphi\hat{\pi}_t) + \frac{1-\xi}{\xi}\frac{1-\beta\xi}{1+\omega\beta}\left[(\frac{\phi+\alpha}{1-\alpha} + \frac{\gamma}{\gamma-\eta})\hat{y}_t - \frac{\eta}{\gamma-\eta}\hat{y}_{t-1} - \frac{\phi+1}{1-\alpha}\hat{a}_t\right] \quad (33)$$

$\omega = \frac{\alpha}{1-\alpha}$

10.2 Derivation of the infinite horizon decision rule

To derive the infinite horizon consumption decision rule first iterate forwards the linearized household budget constraint

$$i_{t-1}\frac{B_{t-1}}{P_t\gamma^t} \geq \frac{B_t}{P_t\gamma^t} + \frac{C_t}{\gamma^t} - \frac{Y_t}{\gamma^t} \tag{34}$$

to get

$$i_{t-1}\frac{B_{t-1}}{P_t\gamma^t} \geq \sum_{T=t}^{\infty} R_{t,t+T}(c_T - y_T) \tag{35}$$

$R_{t,t+T} = \prod_t^T i_T$

Bonds are in 0 net supply and cancel out in aggregate. Using this, the steady state relation $i = \frac{\beta}{\gamma}$ and loglinearization gives

$$0 = \sum_{T=t}^{\infty} \left(\frac{\beta}{\gamma}\right)^{T-t} (\hat{c}_T - \hat{y}_T) \tag{36}$$

To get the expression for $\hat{c}_T$ iterate the linearized Euler equation to get

$$\gamma\hat{c}_t - \eta\hat{c}_{t-1} = \gamma\hat{c}_{t+T} - \eta\hat{c}_{t+T-1} - \sum_{T=1}^{T=T-1} (\gamma - \eta)(\hat{i}_{t+T-1} - \hat{\pi}_{t+T}) \tag{37}$$

Inserting the expression for $\hat{c}_T$ into the individual budget constraint gives the infinite horizon consumption decision rule

$$\frac{\gamma}{\gamma - \eta}\hat{c}_t^i = i_{t-1}\frac{B_{t-1}}{P_t\gamma^t} + \sum_{T=t}^{\infty}(\hat{y}_T^i + \frac{\eta}{\gamma}\hat{c}_{t-1} - \frac{\eta}{\gamma}\bar{c}_{t+T-1} - \frac{\gamma - \eta}{\gamma - \beta}\frac{\beta}{\gamma}(i_T - \pi_{T+1})) \tag{38}$$

Aggregating, rearranging and using market clearing condition gives the aggregate law of motion for output under infinite horizon learning

Price decision rule

Starting with (27) and plugging (28) in and subtracting $log(P_t)$ from both sides yields

$$log(P_t^*) - log(P_t) = \sum_{T=0}^{\infty} (\beta\xi)^{T-t} \left(\frac{1-\beta}{1+\omega\epsilon} \hat{mc}_T + \underbrace{(1-\beta\xi)(log(P_{t+T}) - log(P_t))}_{1} - \underbrace{(1-\beta\xi)log(\pi_{t-1,t+T-1})}_{2} \right)$$

Summation of (1) and (2) gives $\sum_{T=t}^{\infty}(\beta\xi)^{T-t}\beta\xi\pi_{T+1}$ and $\sum_{T=t}^{\infty}(\beta\xi)^{T-t}\beta\xi\pi_T^{\phi}$ respectively

This allows to write the reset price decision rule in terms of the infinite forecast of inflation and marginal cost.

$$log(P_t^*) - log(P_t) = \sum_{T=0}^{\infty} (\beta\xi)^{T-t} \left(\frac{1-\beta}{1+\omega\epsilon} \hat{mc}_T + \beta\xi(log(\pi_{T+1} - \phi log(\pi_T)) \right) \tag{39}$$

and the marginal cost is a function of aggregate output and the technology shock. Substituting price with inflation using (31) gives the law of motion for inflation based on this rule.

Law of motion of flexible price output is derived by inserting the linearized production function into the linearized labor supply relation, substituting real wage from the definition of real marginal cost and using the fact that the real marginal cost is constant under flexible prices.

$$\mathrm{yf}_t = \frac{a_t \left(\alpha^2\gamma\epsilon - \alpha^2\gamma - \alpha^2\epsilon + \alpha^2 - \alpha\gamma\epsilon + \alpha\gamma + 2\alpha\epsilon - 2\alpha + \eta\epsilon\phi\mathrm{n}^{\phi+1} - \epsilon\phi\mathrm{n}^{\phi+1} - \epsilon + 1\right)}{\alpha^2\gamma\epsilon - \alpha^2\gamma - \alpha\gamma\epsilon + \alpha\gamma + \alpha\gamma\epsilon\mathrm{n}^{\phi+1} - \gamma\epsilon\mathrm{n}^{\phi+1} + \eta\epsilon\phi\mathrm{n}^{\phi+1} - \epsilon\phi\mathrm{n}^{\phi+1}} + \frac{\mathrm{yf}_{t-1}\left(\alpha\eta\epsilon\mathrm{n}^{\phi+1} - \eta\epsilon\mathrm{n}^{\phi+1}\right)}{\alpha^2\gamma\epsilon - \alpha^2\gamma - \alpha\gamma\epsilon + \alpha\gamma + \alpha\gamma\epsilon\mathrm{n}^{\phi+1} - \gamma\epsilon\mathrm{n}^{\phi+1} + \eta\epsilon\phi\mathrm{n}^{\phi+1} - \epsilon\phi\mathrm{n}^{\phi+1}}$$

and n is the steady state labor supply

$$n = \left(\frac{\epsilon - 1}{\epsilon} \frac{(1-\alpha)}{\gamma - \eta} \right)^{\frac{1}{\phi+1}} \tag{40}$$

11 Appendix 2

Table 12: Output gap: estimates from fitting data with the asymmetric density

Output gap						
	b_l	b_r	a_l	a_r	m	log lik
value	1.259	2.615	0.0149	0.0166	-0.00112	-2.7089
std. err	0.294	0.7763	0.002778	0.004161	0.00518	

11.1 REE

Euler equation learning

$$P^k = \begin{pmatrix} 0.1216 & -0.0425 & -0.8769 \\ -0.0038 & 0.0317 & -0.0058 \\ -0.0029 & 0.0238 & 0.4957 \end{pmatrix} P^z \begin{pmatrix} 2.436 & -0.0177 \\ -0.4393 & -0.0001 \\ -0.3295 & 0.0099 \end{pmatrix}$$

$$var(P) = \begin{pmatrix} 0.0002 & 0 & 0.0009 & -0.0001 & -0.0002 \\ 0 & 0.0001 & -0.0002 & 0 & 0.0001 \\ 0.0009 & -0.0002 & 0.0037 & -0.0004 & -0.0008 \\ -0.0001 & 0 & -0.0004 & 0.0001 & 0.0001 \\ -0.0002 & 0.0001 & -0.0008 & 0.0001 & 0.0002 \end{pmatrix}$$

Infinite horizon learning

$$P^k = \begin{pmatrix} 0.1052 & -0.0193 & -0.4168 \\ 0.0003 & 0.031 & -0.003 \\ 0.0001 & 0.0161 & 0.3449 \end{pmatrix} P^z = \begin{pmatrix} 0.0835 & -0.0083 \\ -0.1905 & -0.0001 \\ -0.099 & 0.0068 \end{pmatrix}$$

$$var(P) = \begin{pmatrix} 0.00030 & 0.00000 & 0.00110 & -0.00060 & -0.00080 \\ 0.00000 & 0.00010 & -0.00010 & 0.00000 & 0.00010 \\ 0.00110 & -0.00010 & 0.00490 & -0.00230 & -0.00360 \\ -0.00060 & 0.00000 & -0.00230 & 0.00130 & 0.00180 \\ -0.00080 & 0.00010 & -0.00360 & 0.00180 & 0.00270 \end{pmatrix}$$

11.2 Estimates under Heuristic learning

Table 13: statistics: output growth rate

parameters	Mean	St. dev	Skewness	Kurtosis	JB p-value	p-values < 0.05	sample
0.5 1	0.047414	0.360803	1.904198	20.94853	0.001465	998	1000
0.5 1000	0.293757	12.52844	0.973507	127.5429	0.012295	971	1000
0.9 1000	0.282395	17.92991	0.990456	119.4801	0.033916	910	1000
0.98 1000	0.243251	9.658872	0.479719	106.4538	0.062998	831	1000

Table 14: Statistics: output gap

parameters	Mean	St. dev	Skewness	Kurtosis	JB p-value	p-values < 0.05	sample
0.5 1	0.000834	0.264112	-0.00315	2.963278	0.334933	85	1000
0.5 1000	6.49E-05	0.436077	-0.00059	2.960202	0.358957	87	1000
0.9 1000	-0.00207	0.411039	0.007794	2.911743	0.33537	134	1000
0.98 1000	-0.00189	0.377176	-0.00809	2.865251	0.302472	214	1000

Table 15: Estimates: output growth rate

	b	a	m	lik
0.5 1	1.1842	0.2356	0.0129	0.1864
	0.2611	0.0181	0.0139	
0.5 1000	0.5173	0.5140	-0.0810	1.2263
	0.3466	0.1194	0.0659	
0.9 1000	0.5900	0.4850	-0.0789	0.9940
	0.4707	0.1601	0.0695	
0.98 1000	0.7639	0.4276	-0.0680	0.4652
	0.6494	0.2181	0.0680	

Table 16: Statistics: output gap

	b	a	m	lik
0.5 1	2.10E	0.266	-0.00181	0.0774
	0.35514	0.0220	0.0325	
0.5 1000	2.2500	0.4350	0.0015	0.5320
	1.5758	0.0813	0.0773	
0.9 1000	2.64	0.4172	-0.0001	0.446
	3.729701	0.104844	0.105305	
0.98 1000	6.39	0.3831	0.0062	0.270
	34.96271	0.136401	0.145634	

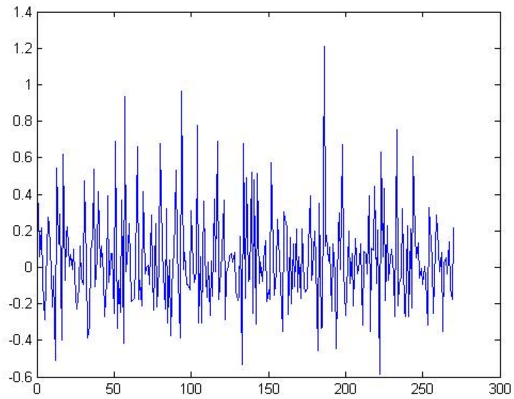

Figure 5: typical bahavior of output growth rate with $\gamma = 1$, $\rho = 0.5$

11.3 Modifications to the Original code used in this thesis

- I extended the original code to run the model with the infinite-horizon learning rule. This required adjustment in the files solving the model and doing the T-mapping. To get the model in the MSV form I substituted the infinite sum of expected values of a variable with the infinite sum of the forecasts. $\sum_{T=t}^{\infty}(\beta\xi)^{T-t}\hat{y}_{T+1}$ with $\underbrace{\frac{\varphi_y}{1-\varphi_y\beta\xi}}_{A} y_t + \underbrace{\frac{\varphi_z}{1-\varphi_z\beta\xi}\frac{1}{1-\rho_z}}_{B} z_t$

  The file solving the model solved for 'A' and 'B', from them I backed out the φ_y and φ_z. In the T-mapping to the Actual Law of Motion I calculated 'A' and 'B' using φ_y and φ_z from the previous period and did the mapping using 'A' and 'B'. In this way to solution, mapping and the model simulation codes where affected minimally.

- I added an option to do the Impulse response functions for the model. I simply created an additional file that creates a shock vector and calls the code, which simulates the model. They are not reported here, but where used to evaluate if the model behaves in line with the macroeconomic theory
- After solving for the functional form of yf, I added a command to calculate it from the technology shocks after the model is simulated.