Learning, Observability and Time-varying Macroeconomic Volatility

By

Chaozheng Li

M.A., University of Kansas, 2015

M.A., University of Macau, 2013

B.Sc., University of Macau, 2011

Submitted to the graduate degree program in Economics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chair: Shigeru Iwata

Jianbo Zhang

Tarun Sabarwal

Paul Comolli

Xuemin Tu

Date Defended: 30 April 2018

The dissertation committee for Chaozheng Li certifies that this is the approved version of the following dissertation:

Learning, Observability and Time-varying Macroeconomic Volatility

Chair: Shigeru Iwata

Date Approved: 30 April 2018

Abstract

Our paper provides a theoretical explanation for the time-varying macroeconomic volatility by introducing the unobservability of regime switching and learning. With the unobservability of regime switching, agents must endogenously form their expectations using best-performed forecasting models. We find that if the regime switching is observable to agents, agents do not shift their expectation frequently and so will not generate a larger macroeconomic volatility. However, with the unobservability of regime switching where no agents can know which regime is dominant, allowing endogenous expectation formation would give rise to larger macro fluctuations (first-layer amplification mechanism), which is made through agents frequently shifting their expectations. Furthermore, we consider the policy implication under the zero lower bound. Our simulations show that in the unobservable regime switching, the economy is more likely to fall into a deflationary trap. To avoid the deflation risk, the policy maker should set a higher expected inflation based threshold. If the expected inflation is under the threshold, an aggressive policy rate will be implemented; otherwise, the normal Taylor-rule monetary policy will be used. Furthermore, to reduce the deflation risk, the strategy for the policy maker is to raise the threshold, and this will generate larger macroeconomic fluctuations (second-layer amplification mechanism) due to more frequent policy strategy switching. We argue that sometimes only with unobservability, the policy maker faces a dilemma between avoiding deflation risk and maintaining macroeconomic stability, and huge macroeconomic fluctuations do not necessarily result from bad luck or bad policy but from the two-layer amplification mechanism caused by the unobservability.

Acknowledgment

I would like to thank my advisor Professor Shigeru Iwata. I feel lucky to be his student at the University of Kansas. I like to attend his workshop held every week, and I learned a lot from our presentations and our discussions not only on academic thoughts but also on interesting events around the world. He has been supporting me to explore new problems and phenomena since I joined his workshop. It is he that introduced me into the field of macroeconomics, and he spent a lot of time explaining to me the academic values and practical values through his academic experience in the IMF and the Bank of Japan. Also, his rich insights and thoughts open a door to a broader world beyond the academia for me.

I am grateful to my other committee members: Professor Tarun Sabarwal, Professor Jianbo Zhang, Professor Paul Comolli and Professor Xuemin Tu. They are nice professors who are very passionate to help me come up with some valuable comments and suggestions for my dissertation, especially in my oral comprehensive exam and intermediate defense. Professor Sabarwal showed much patience to listen to my research details and offered many valuable suggestions. Professor Zhang asked a lot of aggressive questions about my dissertation that pushed me forward in a fast pace. I am grateful to Professor Comolli who offered many comments that made my paper more interesting. Lastly, I must thank Professor Tu from Math department who has been showing a strong interest on my research that encourages me to improve my idea and details. I also would like to thank Professor Shu Wu. I still remember that when I was doing my third-year paper, he taught me how to do a serious research paper a lot, such as how to do shock identification in VAR systems. He is one of my favorite professors and provides a perfect model of an economist and a professor. I still always remember all those moments he smiled to me and discussed questions with me.

I must thank my friends at the University of Kansas because they are my best audience in my life. Sharing my thoughts and ideas is most enjoyable with them at the University of Kansas, which also provides me a lot of opportunities to practice my spoken English and opens a door to a brand-new cultural world.

Finally, I must be very grateful to my parents and my brother for their substantial encouragement. My parents gave me the best value system when I grew up and let me understand that any great

success is from the hardship instead of from the easy life. During the five years of my PhD study, my brother has been supporting me and talking with me continuously about all kinds of interesting scientific fields, such as bioinformatics, AI and even mathematical philosophy. It makes me go through and experience so many beautiful theories in the human beings' history and makes me understand the nature of doing research.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Time-varying volatility has been an attractive topic in macroeconomics and financial markets in recent years. From the volatility evolution of aggregate macroeconomic variables in US economy in figure 1 and 2, two regularities can be clearly seen. First, the volatility of macroeconomic data including the real GDP growth and the GDP deflator is time-varying through the whole sample. Second, the absolute deviation of real GDP growth and GDP deflator before 1984 is much larger than that after 1984 (before 2008). For the first regularity, many scholars explored the question by using ad hoc setting. Some of them used GARCH approach or stochastic volatility approach to relax the setting of the time-invariant variance. Shephard (2008) argues that the stochastic volatility setting can surprisingly capture some important features of economic data. Justiniano and Primiceri (2006) and Fernandez-Villaverde and Rubio-Ramirez (2007) introduce stochastic volatility approach in the DSGE framework and improve the model's performance in data fitting. For the second regularity, many researchers modeled the scenario in time-varying and structural break setting and explored the interesting phenomenon. Kim and Nelson (1999) used Bayesian Markov-Switching model to find structural changes after 1980 and McConnell and Perez-Quiros (2000) explored the interesting phenomenon that the volatility of US output growth has shown a substantial decline in the early 1980s. Afterwards, Sensier and Dijk (2004) used 214 US macroeconomic time series over the period 1959-1999 to test for a change in the volatility. They find that after 1980 structural breaks occurred more than before 1980, which is supported by 80% of these series. However, all models use the ad hoc setting and improve the data fitting.

Researchers usually model variances of shocks as constant throughout the whole sample, for example, smets and Wouters (2003, 2007), Lubik and Schorfheide (2004) and An and Schorfheide (2007). However, in many areas such as asset pricing, monetary policy and term structures, scholars usually find the empirical regularity that the volatility of macroeconomic data is time-varying and often exogenously assume stochastic process for volatility and explore the effect of this setting. Their objective is to improve the performance in data fitting and the model forecasting.

Unfortunately, those setups for the time-varying volatility were ad hoc and had no sound micro-foundation. Early important literature concentrates on empirical studies that tried to measure the time component in the variance of inflation, for example, Khan (1977) used the absolute value of the first difference of inflation and Klein (1977) used a moving variance around a moving mean as a key measure.

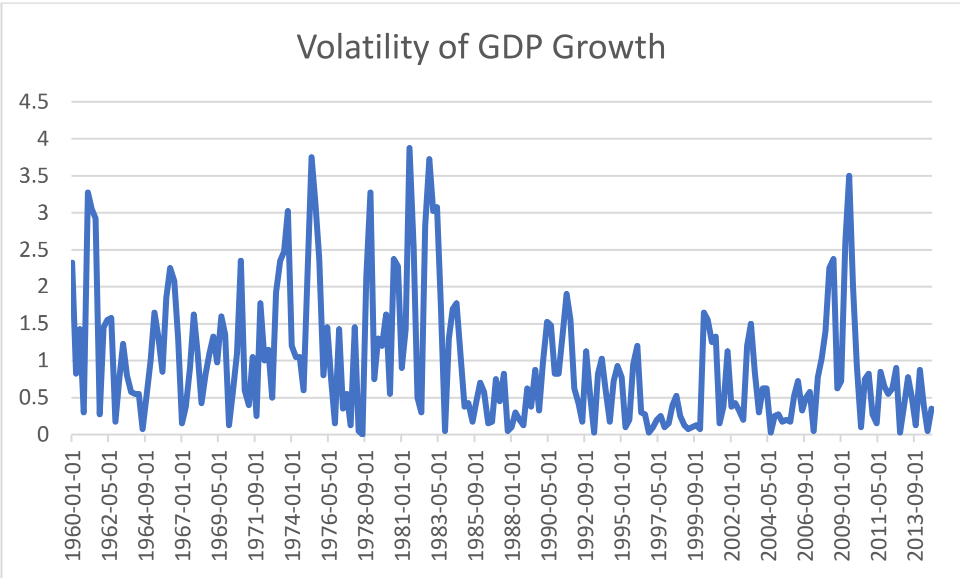

Figure 1. Real GDP growth, absolute deviation from one-year-rolling-window mean

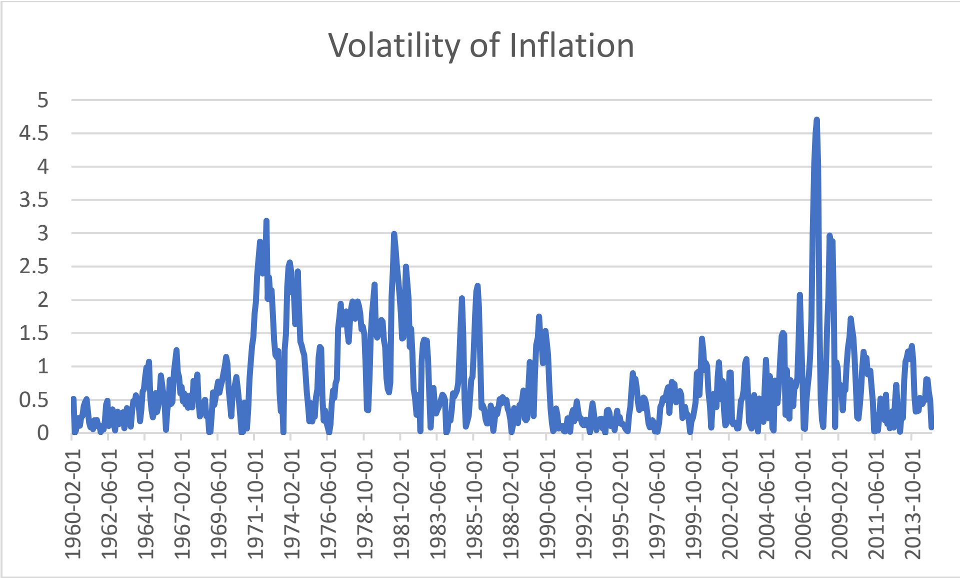

Figure 2. CPI-based inflation, absolute deviation from one-year-rolling-window mean

The main breakthrough was made by Engle (1982) who proposed autoregressive conditional heteroscedasticity (ARCH). In the literature about ARCH, the evolution of variance of time series variables is modeled as autoregressive process. The advantage of ARCH is that people can easily deal with a scoring iterative maximum likelihood method and OLS method. Many studies also used the original idea to obtain the time-varying dynamic features of macroeconomic variables including that Engle (1982) found the British inflation has time-varying behaviors. After Engle (1982), there are 139 variations of Engle (1982), see Bollerslev (2010). For example, Generalized ARCH (GARCH) was created by Bollerslev (1986), Nonlinear GARCH was proposed by Engle and Ng (1993). Nelson (1991) put forward Exponential GARCGH, ZakoÔan (1994) came up with threshold GARCH and Sentana (1995) raised Quadratic GARCH, and so on.

Many studies also use the stochastic volatility approach to discuss the macroeconomic volatility. The first literature that stochastic volatility approach was introduced in the DSGE framework is Justiniano and Primiceri (2006) and then Fernandez-Villaverde and Rubio-Ramirez (2007), which relax the assumption of time-invariant-variance shocks. The introduction of stochastic volatility into DSGE models shows the volatility of shocks has been changed significantly over time and improve the model's performance. Shephard (2008) argues that the stochastic volatility setting can surprisingly capture some important features of economic data. However, those papers only set the volatility of shocks exogenously. There are some debates about advantages and disadvantages of the stochastic volatility approach and the GARCH approach, for example, Villaverde and Ramirez (2010) argue that there are no advantages to using GARCH process instead of SV for four reasons. First, GARCH process has one less degree of freedom. Second, separating level from volatility shocks in GARCH process is significantly difficult. Third, it is hard to incorporate GARCH approach in the DSGE framework, preventing DSGE modelers from combining theoretical and empirical exploration. Fourth, the GARCH models usually have a worse data fitting than the stochastic volatility model.

Another commonly used approach to modeling volatility is Markov regime switching. This approach is a discrete setting compared with the GARCH approach and the stochastic volatility approach. All changes occurring will be discrete jumps from one regime to another regime. In the real world, we are often not able to clearly figure out that the discrete-sampling data behavior results from Markov regime switching or some continuous process, which is first pointed out by Diebold (1986). In finance, Ait-Sahalia, Hansen and Scheinkman (2009) propose a continuous-time Markov process to deal with data. There is consensus that in the real world the volatility in many cases is probably a mix of continuous and discrete events, but still, there are events that are easier to interpret a discrete change. For example, the federal funds rate after 2008 reached zero lower bound (ZLB) and the Fed lost the leverage of the overnight interest rate. Moreover, the approval of Dodd-Frank Act in 2010 and the Large-Scale Asset Purchase (LSAP) announced in late 2008 are all discrete events. Afterwards, the change in operating LSAP can be interpreted as continuous events. Later, to make our economic story simpler and more significant, we will use regime switching to model the case of observability and that of unobservability in exogenous shocks.

Having discussed the exogenous setting for generating the time-varying volatility, we still have a fundamental question still unresolved: where does the time-varying volatility come from? This is a more challenging question than modeling the time-varying volatility behavior. Economists started to explore some endogenous channels to generate more volatile data. In recent studies, Navarro (2014) develops a novel mechanism in which firms' volatility arises endogenously because of financial disruptions. Basu and Bundick (2015) discovered that introduction of fluctuations in uncertainty and zero lower bound can well explain the stochastic volatility in recent macroeconomic data. Gomes (2017) explores the effect of heterogenous wage setting strategies in a macro framework where endogenous fluctuations emerge.

However, the literature uses rational expectation to model agents' behavior. As we know, the information available in the economy is always limited and usually prevents agents from forming rational expectation. Thus, there is a big strand of literature that uses boundedly rational expectation or learning to model the endogenous volatility. Marcet and Nicolini (2003) use the endogenous switching gain to study hyperinflation. Lansing (2006) uses learning in New Keynesian Phillips curve to endogenously generate time-varying volatility. His paper derives the optimal variable gain as the fixed point of a nonlinear map that relates the gain to the autocorrelation of inflation changes.

Besides, the observability of macroeconomic shocks has a significant impact on agents' belief or functional form of the forecasting model. Even though we replace rational expectation models with learning-based models, without the ability to observe important economic data, the assumption for having "correct" functional form of the forecasting model is still unrealistic. For simplicity, we only focus on the discussion about the observability of exogenous shocks. Agents can only use available data to choose feasible forecasting models and condition expectations on the available data and the chosen forecasting model. Lucas (1973) assumed that aggregate price levels were unobserved, and he leveraged this friction to impart real effects of surprise money on output. Mankiw and Reis (2002) and Sims (2003) consider informational frictions on rational inattention. King and Rebelo (1999) assume that certain types of productivity shocks are unobserved. Woodford (2003) argues that expectations are likely to be formed before certain shocks are realized. Levine et al (2012) assumes that shocks are not observed by agents leads to improved empirical performance in the DSGE framework. Cochrane (2009) argues that the monetary policy shock (which is captured by an innovation associated to an instrument rule) should not be taken as observable; he finds that there are multiple learnable equilibria even when the model is determinate.

The main contribution of this paper is that we incorporate learning, unobservability and regime switching mechanism in the New Keynesian model to provide a theoretical explanation for the time-varying macroeconomic volatility. There are several papers that are closely related to our paper. Milani (2014) also used learning to generate macroeconomic time-varying volatility. But in the model setting, agents endogenously update the gain coefficient according to the past forecast errors. The paper does not discuss about the observability of exogenous shocks. Branch and Evans (2007, 2011), presenting a framework in which regime changes in volatility arise, is related with our paper. However, the paper emphasizes that the under-parametrization of forecasting models is important to endogenously generating volatility. But this is unrealistic for third reasons. First, the model setting assumes that agents favor parsimony in their forecasting model and select the best performing model from the set of underparameterized forecasting models. However, since agents have incentives to combine a broader set of variables in improving forecasting performance, the selection's separation of parsimonious forecasting models will not exist. Second, we argue that no agents or professional economists in the market would unrealistically use the supply shock as a proxy of the demand shock or use the demand shock as a proxy of the supply shock. Third, their papers ignore the behavior of agents estimating unobservable shocks according to advanced computational techniques. For example, consider preference shocks in the demand side and productivity shocks in the supply side. However, when the estimation is considered, some estimates are high-quality while some are low-quality. In our paper, we use a regime switching between high-quality and low-quality estimates to unify different scenarios: sometimes, some agents can perfectly observe true regime-switching-based shocks and some agents can only "observe" low-quality shocks which can be interpreted as low-quality-estimates for shocks; sometimes, all agents cannot observe true shocks and they can only accurately estimate the shocks in turn according to "hard-to-estimate" structural changes occurring alternatively, which is modeled as regime switching mechanism.

In our paper, we can view the economy as an "expectation-based" game. We discuss three cases. First, there is no regime switching. Second, there are observable regime switching where some agents (perfect observers) can observe[1] which regime dominates and its realized data. Third, there is an unobservable regime switching that all agents (imperfect observers) can only observe the regime-1 and regime-2 shocks but they cannot know which regime dominates. The second case can be usually modeled as a small-sized structural break and the third case can be treated as a big-sized unprecedented structural break. In small-sized structural breaks, some people can still know which regime the economy is stuck in and estimate the macroeconomic shocks by using computational methods. However, in unprecedented structural breaks (e.g. 2008 financial crisis), people even cannot figure out which regime the economy is in because of the increasing uncertainty.

The three cases will discuss the effect of endogenous expectation mode selection on the macroeconomic volatility. We find that in the first case the macroeconomic volatility does not seem to behave in the time-varying manner even though the endogenous expectation mode selection is allowed. The reason is intuitive. In the structural-break-free economy agents that can observe the exogenous shocks will not deviate from their original best-performing forecasting model because they will form more "correct" belief on the economy after some-period model training. There is no any force to deviate from the stable equilibrium. The volatility of macroeconomic data in the stable equilibrium remains constant. However, when structural breaks are introduced in the second case and the third case, the regime switching will lead to a time-varying feature, which is consistent with a large literature about structural changes, see Kim and Nelson (1999) and Sensier and Dijk (2004). But the difference is that the second case cannot generate a larger volatility when allowing endogenous selection of forecasting models while the third case can generate a larger volatility. The reason is about the observability of structural breaks.

[1] In this paper, "observe" does not necessarily mean directly "observe" in datasets, and it might also mean "estimate" by using advanced computational techniques or algorithms.

In the “easy-to-observe” regime-switching economy (case 2), agents who can observe the true regime are more likely to continue to use their own forecasting model, thus, the agents do not have incentives to deviate from their forecasting model, while other agents without the ability to observe the true regime can also only use their “compromising” forecasting models for the expectation formation. One thing that is worth pointing out is that when a regime is switched, the perfect observer’s forecasting performance is not necessarily best because the regime switching as a shock has different impacts on different agents, but after short periods (still in the same regime) the perfect observer’s forecasting performance will continue to be the best when the economy in that regime converges to the corresponding equilibrium. Hence, the regime switching in the second case can give rise to a time-varying volatility, but the exogenous mechanism of expectation mode shifts cannot contribute to an extra volatility due to their unwillingness to deviate from their forecasting models. In the third case, however, the deviations from their forecasting models will happen very frequently, which causes a larger volatility. There are the direct effect and the indirect effect. When a regime equilibrium is formed, agents will use this regime’s forecasting model to form expectation. Once the regime is switched to the other one, the direct effect is first triggered: the regime switching itself as a shock will cause the economy to be volatile; after the regime switching is turned on, agents will change their forecasting model to be consistent with the new regime, which will result in an extra volatility of macroeconomic data. This is the indirect effect. Therefore, continuously speaking, when regimes are switched back and forth, the volatility of macroeconomic data is time-varying and larger. In this sense, the indirect effect of the second case is very weak.

Furthermore, we consider the zero lower bound environment, and policy makers must consider deflation risk when the economy suffers from a negative shock. Taking unobservability into account, macroeconomic volatility will increase. Some action must be taken for avoiding inflation risk. Our recommended policy solution is setting an expected inflation based threshold under which the policy maker will use a low enough interest rate to boost the economy and above which the policy maker will use Taylor-rule based monetary policy. In this setting, we find two interesting implications. First, the sharp fluctuations of output and expectation switching keep the same pace with policy regime switching, but the inflation does not respond to the policy switching strongly. Second, raising expected inflation based threshold reduces the likelihood of the economy falling into the deflationary trap, but the likelihood of the economy staying in the aggressive policy regime increases. Raising the threshold boosts the economy more strongly and increases the duration of unsustainable boom in output.

Then we explore two important extensions. First, with unobservability, the policy maker faces a dilemma: avoid the deflation risk and maintain the macroeconomic stability. Without the unobservability problem, the policy maker may not face such dilemma. The reason is that when regime switching occurs endogenous expectation formation will generate larger macroeconomic fluctuations (the first-layer amplification mechanism). When the policy maker observes such large fluctuations, she must raise the expected inflation based threshold to avoid the deflation spiral. However, the higher threshold makes the economy more frequently enter and exit the aggressive policy regime, leading to a higher volatility (he second-layer amplification mechanism). Hence, it is impossible that the policy maker maintains the macroeconomic stability and avoid the deflation risk at the same time. Second, large macroeconomic fluctuations are not necessarily from a bad luck or a bad policy. Put it differently, a situation with good-luck shocks and a reasonable policy can also generate unexpected fluctuations with unobservability. Put it simply, with unobservability, the two layers of amplification mechanism will generate substantial macroeconomic fluctuations from a "good-luck" shock. Moreover, the endogenous amplified fluctuations sometimes are not mistakenly made by the policy maker because the policy must take a responsibility of avoiding deflation risk.

Before starting our formal model, we generalize the concept of the rational expectation based on observable data. Following Evans and McGough (2015), the rational expectation is a fixed point of agents' beliefs. Formally, a general model is

$$y_t = f(E_t y_{t+1}, v_t) \tag{1}$$

Let vector spaces of real sequences Y and V be copies of $\mathbb{R}^\infty$. Assume that there is only one agent forming expectation and there is a probability distribution $\mathcal{H}$ over $Y \times V$. Let y^t and v^t be the respective history vectors, the belief $\mathcal{H}$ determines the conditional distribution of y_{t+1} on y^t and v^t. Agents are said to be internally rational if they form expectation as follows

$$E_t y_{t+1} = E^{\mathcal{H}}(y_{t+1} | y^t, v^t) \tag{2}$$

We say that $\mathcal{H}$ tracks the joint distribution over $Y \times V$. New data y_t is generated after the expectation formation $E^{\mathcal{H}}(y_{t+1}|y^t, v^t)$, and the new data y_t has a realized probability distribution $T(\mathcal{H})$ over $Y \times V$ that depends on the agents' belief $\mathcal{H}$. It is said that internal rational agents are externally rational if there is a fixed point of belief $T(\mathcal{H}) = \mathcal{H}$. Thus, we have a formal definition of rational expectation equilibrium.

Definition. A rational expectation equilibrium of a model (1) with expectation formation (2) is a probability distribution $\mathcal{H}$ over $Y \times V$ where $T(\mathcal{H}) = \mathcal{H}$.

However, rational expectation is not a realistic assumption. There are several reasons. First, the rational expectation requires agents to know functional form, but the structure of the economy is usually unobservable to agents and needs agents to use data to update their belief on the structure of the economy. Second, there are some unobservable fundamental variables that determine the evolution of the economy, so agents in this case cannot form rational expectation. Third, agents often have different belief systems to choose in the real world, but rational expectation cannot allow this situation to occur. In this paper, we introduce the unobservability where agents can only forecast using relatively better-performed models which is trained by updated data. In following sections, we will also consider heterogenous beliefs. We assume that the aggregate expectation operator $\tilde{E}_t$ is a linear combination of individual expectation operators $\tilde{E}_t = \sum_i^n n_i \tilde{E}_t^i$ where $\tilde{E}_t^i y_{t+1} = E_i^{\mathcal{H}_i}(y_{t+1}|y^t, v^t)$.

Chapter 2: Model

We start with the hybrid New Keynesian model. Following the hybrid IS curve with the backward-looking term, see Fuhrer (2000), we have

$$x_t = \alpha_1 x_{t-1} + \alpha_2 \tilde{E}_t x_{t+1} - \alpha_3\left(i_t - \tilde{E}_t \pi_{t+1}\right) + e_t \tag{3}$$

$$e_t = \rho_e e_{t-1} + \varepsilon_t^e$$

Where x_t is the output gap, $e_t \sim AR(1)$ is the demand shock and $\varepsilon_t^e \sim iid(0, \sigma_e^2)$. On the other hand, following Galí et. al. (2005), we can write the hybrid New Keynesian Phillips curve, which is caused by nominal price rigidity as follows

$$\pi_t = \lambda_1 \pi_{t-1} + \lambda_2 \tilde{E}_t \pi_{t+1} + \lambda_3 x_t + u_t \tag{4}$$

$$u_t = \rho_u u_{t-1} + \varepsilon_t^u$$

Where π_t is the inflation, $u_t \sim AR(1)$ is the supply shock and $\varepsilon_t^u \sim iid(0, \sigma_u^2)$. We follow the forward-looking Taylor-type monetary policy rule

$$i_t = \chi_x \tilde{E}_t x_{t+1} + \chi_\pi \tilde{E}_t \pi_{t+1} \tag{5}$$

Where χ_x and χ_π are the expectational responses from the output gap and the inflation, respectively. Putting monetary policy rule back to hybrid IS curve and hybrid Phillips curve, a compact form can be written as follow

$$y_t = A_1 y_{t-1} + A_2 \tilde{E}_t y_{t+1} + A_3 s_t \tag{6}$$

$$s_{t+1} = P s_t + \varepsilon_t$$

Where $y_t = (x_t, \pi_t)'$, $s_t = (e_t, u_t)'$, $\varepsilon_t = (\varepsilon_t^e, \varepsilon_t^u)'$, $A_1 = \begin{bmatrix} 1 & 0 \\ -\lambda_3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1 & 0 \\ 0 & \lambda_1 \end{bmatrix}$, $A_2 = \begin{bmatrix} 1 & 0 \\ -\lambda_3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \alpha_2 - \alpha_3 \chi_x & \alpha_3 - \alpha_3 \chi_\pi \\ 0 & \lambda_2 \end{bmatrix}$, $A_3 = \begin{bmatrix} 1 & 0 \\ -\lambda_3 & 1 \end{bmatrix}^{-1}$ and $P = \begin{bmatrix} \rho_e & 0 \\ 0 & \rho_u \end{bmatrix}$.

In the following, we will discuss three cases: no regime switching, observable regime switching and unobservable regime switching.

2.1. Case 1: No Regime Switching

There are two agents in the market, one with the mass n can observe the shock s_t and the other with the mass 1-n cannot observe s_t. The former (s_t agent) is a fundamental leaner who uses fundamental solution to form her expectation and the latter (VAR agent) is VAR learning who uses VAR to form her expectation due to unobservability of shock s_t. In the beginning, we treat n exogenously. We assume that agents form their expectations after observing s_t at time t. Thus, the forecasting model, also called perceived law of motion (PLM), of the two agents are

PLM 1: $y_t = B_t s_t$

PLM 2: $y_t = C_t y_{t-1}$

B_t and C_t are updated by agents after new data is realized. It is worth pointing out that the agent 1 at time t uses the key exogenous data to predict the macroeconomic data at time t, here, only fundamental solution (i.e. no sunspot exists) is considered. while the agent 2 due to a lack of exogenous data can only form expectation about the macroeconomic data by using lagged endogenous data (VAR). The aggregate expectation is $\tilde{E}_t y_t = \text{n}\tilde{E}_t^1 y_t + (1-n)\tilde{E}_t^2 y_t$. The timing of the "expectation-based" economy is as follows

- B_{t-1} and C_{t-1} are determined at the end of time t-1;
- $s_t = Ps_{t-1} + \varepsilon_t$ is realized at the beginning of time t.
- Form Expectation

  PLM 1: $y_t = B_{t-1}s_t$ => $\tilde{E}_t^1 y_{t+1} = \tilde{E}_t^1 B_{t-1}s_{t+1} = B_{t-1}Ps_t$

  PLM 2: $y_t = C_{t-1}y_{t-1}$ => $\tilde{E}_t^2 y_{t+1} = C_{t-1}^2 y_{t-1}$

- Generate time-t data y_t

  ALM: $y_t = \xi_{1t}y_{t-1} + \xi_{2t}s_t$, where $\xi_{1t} = A_1 + (1-n)A_2C_{t-1}^2$ and $\xi_{2t} = nA_2B_{t-1}P + A_3$,

- Update B_t and C_t with moment method

  $B_t = \xi_{1t}r_{1t} + \xi_{2t}$, where $r_{1t} = (\sum_{i=1}^{t} y_{i-1}\, s_i{}')(\sum_{i=1}^{t} s_i\, s_i{}')^{-1}$

$C_t = \xi_{1t} + \xi_{2t} r_{2t}$, where $r_{2t} = (\sum_{i=1}^{t} s_t y'_{i-1})(\sum_{i=1}^{t} y_{i-1} y'_{i-1})^{-1}$

- Go back to the first step and repeat the same process, t = t+1

It is assumed that s_t is realized before agents' expectation formation. It is easy to obtain the evolution for ξ_{1t} and ξ_{2t} below

$$A_1 + (1-n)A_2(\xi_{1t} + \xi_{2t} r_{2t})^2 \rightarrow \xi_{1t+1}$$

$$nA_2(\xi_{1t} r_{1t} + \xi_{2t})P + A_3 \rightarrow \xi_{2t+1}$$

Here, there is a fixed point $(\bar{\xi}_1, \bar{\xi}_2)$, and Restricted Perception Equilibrium (RPE) is written as $y_t = \bar{\xi}_1 y_{t-1} + \bar{\xi}_2 s_t$. The difference between RPE and REE is that REE makes PLM and ALM consistent while RPE cannot let ALM consistent with PLM. Before obtaining the local stability condition, we write the evolution system in a compact way

$$\xi_{t+1} = T_n(\xi_t)$$

Where

- $\xi_t = (\xi_{1t}, \xi_{2t})'$
- $T_n = (T_{n1}, T_{n2})'$
- $T_{n1}(\xi_t) = A_1 + (1-n)A_2(\xi_{1t} + \xi_{2t} r_{2t})^2$
- $T_{n2}(\xi_t) = nA_2(\xi_{1t} r_{1t} + \xi_{2t})P + A_3$

Now the local stability condition can be given: $DT_{\xi_{it}} = \frac{dT_{ni}(\bar{\xi}_t)}{d\xi_{it}}$ has all eigenvalues within the unit circle for i=1, 2

- $\bar{\xi}_t$ is RPE value of ξ_t
- $DT_{\xi_{1t}} = \frac{dT_{n1}(\bar{\xi}_t)}{d\xi_{1t}} = (1-n)\left(\bar{\xi}_{1t} + \bar{\xi}_{2t} r_{2t}\right)' \otimes A_2 + (1-n)I \otimes A_2\left(\bar{\xi}_{1t} + \bar{\xi}_{2t} r_{2t}\right)$
- $DT_{\xi_{2t}} = \frac{dT_{n2}(\bar{\xi}_t)}{d\xi_{2t}} = nP'_1 \otimes A_2$
- $\otimes$ denotes the Kronecker product

Now consider that n is endogenously determined every period. The mechanism is that s_t agent can change her own forecasting model to VAR forecasting model once VAR forecasting model performs better than s_t forecasting model in terms of mean square errors (MSE). It is important to mention that VAR agent cannot use s_t forecasting model due to her unobservability of s_t and she can only use VAR forecasting model to form her expectation. Formally, at time t, the performance is measured by (negative) mean square errors (MSE) $U_t^j = -\left(y_{t-1} - \tilde{E}_{t-1}^j y_{t-1}\right)'\left(y_{t-1} - \tilde{E}_{t-1}^j y_{t-1}\right), j = 1, 2$. Here, 1 stands for s_t agent and 2 stands for VAR agent. When $U_t^1 \geq U_t^2$, the s_t agent still uses s_t forecasting model for her expectation formation. When $U_t^1 < U_t^2$, the s_t agent uses VAR forecasting model to form expectation. However, when the pure forward looking New Keynesian (NK) model is used and $U_t^1 < U_t^2$ occurs, the mass of agents using VAR forecasting model will be 1, which will lead to an explosion in macroeconomic system. Thus, we set a threshold $\bar{n}$ such that the mass of agents using VAR forecasting model cannot exceed $\bar{n}$. Thus, formally put, evolution system can be written as

$$\xi_{t+1} = \begin{cases} T_n(\xi_t), \text{when } U_t^1 \geq U_t^2 \\ T_{1-\bar{n}}(\xi_t), \text{when } U_t^1 < U_t^2 \end{cases}$$

We can interpret that T_n and $T_{1-\bar{n}}$ define two different paths for the evolution of ξ_t, where ξ_t is switched between the two paths based on the relative performance about s_t forecasting model and VAR forecasting model.

According to figure 3, the graph of model performance shows that s_t agents always choose s_t model instead of VAR model. It means that s_t agents always informative signals to forecast and VAR model does not contain fundamental information to explain what is happening today, let alone forecast for the future. The graphs of output gap and inflation around the steady state show that endogenous expectation formation and exogenous expectation formation do not have any effects on their fluctuations, because s_t agents do not change s_t model to the uninformative VAR model. From the graph of volatility, both output volatility and inflation volatility are very stable through the whole window. It means that allowing the endogenous expectation formation does not help build up the volatility. The figure 6 gives the reason. We see that after the 450th period agents in most periods still endogenously choose s_t forecasting model for expectation formation, which is almost the same as the pre-450th-period case.

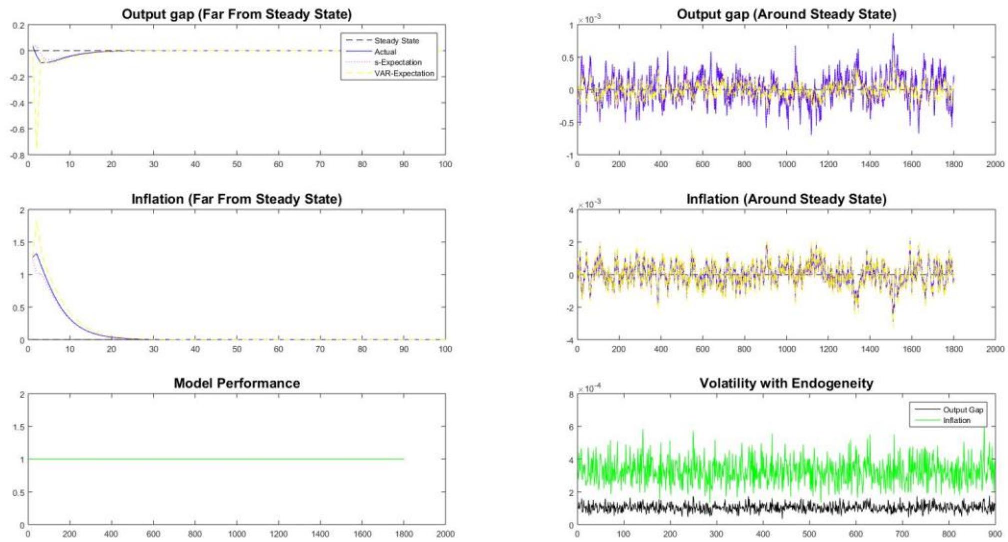

Figure 3. This is the case without regime switching where s_t agents and VAR agents are in the market. In the model performance, 1 stands for s_t forecasting model performing best, 0 stands for VAR forecasting model performing best. In volatility and endogeneity, only exogenous expectation formation is allowed before 900 periods and endogenous expectation formation is allowed after 900 periods. Mass of s_t agents n=0.95 and $\bar{n} = 0.2$.

2.2. Case 2: Observable Regime Switching

Having discussed a basic model only with a single exogenous shock, we transfer our attention to more than one exogenous shocks. The reason is that in the real world the shock hitting the macroeconomy is different from shocks used to form agents' expectation. In this paper, we introduce regime switching mechanism to formulate our economy.

Assume that the economy has a shock set $\mathrm{S} = \{s_t, s_t^0, s_t^1\}$. There is regime switching mechanism $s_t = 1_{[z_t=0]} s_t^0 + 1_{[z_t=1]} s_t^1$ and the state $z_t = 0, 1$ can be determined by $z_t = \begin{cases} 0, & t = 4k+1, \ldots, 4k+3 \\ 1, & t = 4k \end{cases} (k \in \mathbb{N})$, and the two shocks have the following evolution $s_{t+1}^0 = P_0 s_t^0 + \sigma_0 \varepsilon_t$ and $s_{t+1}^1 = P_1 s_t^1 + \sigma_1 \varepsilon_t$, where $\varepsilon_t \sim N(0,1)$. Thus, the evolution of the exogenous shock is $s_{t+1} = 1_{[z_t=0]} P_0 s_t^0 + 1_{[z_t=1]} P_1 s_t^1 + (1_{[z_t=0]} \sigma_0 + 1_{[z_t=1]} \sigma_1) \varepsilon_t$. There are three agents, s_t agent who can observe s_t, s_t^0 agent who can only observe s_t^0 and the VAR agent who can observe nothing in the shock set S. The mass distribution of the three agents is n, n_0, $1-n-n_0$, respectively. Intuitively, if $z_t = 0$, the shock s_t^0 dominates; if $z_t = 1$, the shock s_t^1 dominates. We can simply interpret the three agents as perfect observer, imperfect observer and blind observer. They use signals of different quality to form their own expectations. The quality of signals is ordered as $s_t > s_t^0 > VAR$. In terms of shock formation, this is a discrete setup. The reason why such discrete instead of continuous setup is introduced in our model is that there are many structural changes when some events occur in the economy, such as the zero-lower bound for the monetary policy is reached or the Dodd-Frank Act was approved. However, it is worth pointing out that the real-world shocks must be a mix of a discrete case and a continuous case but we want to use discrete model for some reasons. First, we want to emphasize the effect of a discrete event on the economy. A lot of macroeconomic volatility is closely related to some specific events, for example, the economy after 2008 hit zero lower bound. Second, functional forms in continuous models are difficult to determine in a micro foundation, and there are many variations in ad hoc settings. Originally, we have initial mass distribution of agents n, n_0 and $1-n-n_0$. However, over time, when agents release their forecasting performance based on mean square error every period, observe each other's performance and determine their own forecasting models for a period ahead. Due to the limitation in the observability of shocks, different agents have different feasible set of forecasting models. VAR agent cannot choose other forecasting models no matter how good other models perform. s_t^0 agent can use VAR forecasting model if her s_t^0 forecasting model is weaker than VAR model, but she cannot choose s_t forecasting model due to her unobservability of shock s_t. s_t agent can shift either to VAR forecasting model or to s_t^0 forecasting model if needed. There is one point needed to emphasize. When the VAR forecasting model has best performance and all agents will choose VAR to form expectation, but in the standard New Keynesian model (pure forward looking), if VAR's expectation formation is made, the economic system may explode. To avoid an explosion of macroeconomic system, we set a threshold $\bar{n} = 0.2$ such that there is only a mass $\bar{n}$ of VAR agents. The three agents' forecasting models (i.e., PLM) are as follows,

PLM 1 for type 1 agent: $y_t = B_{t-1} s_t$

PLM 2 for type 2 agent: $y_t = B_{t-1}^0 s_t^0$

PLM 3 for type 3 agent: $y_t = C_t y_{t-1}$

The expectation formations of s_t agent, s_t^0 agent and VAR agent are $\tilde{E}_t^1 y_{t+1} = B_{t-1}(P_1 s_t + 1_{[z_t=0]}(P_0 - P_1)s_t^0)$, $\tilde{E}_t^2 y_t = B_{t-1}^0 P_0 s_t^0$ and $\tilde{E}_t^3 y_t = C_{t-1}^2 y_{t-1}$, respectively. It is easy to obtain the aggregate expectation $\tilde{E}_t y_t = n\tilde{E}_t^1 y_t + n_0 \tilde{E}_t^2 y_t + (1 - n_0 - n)\tilde{E}_t^3 y_t$. Now we write the macroeconomic system in a compact way

$$y_t = A_1 y_{t-1} + A_2 \tilde{E}_t y_{t+1} + A_3 s_t$$

$$s_t = 1_{[z_t=0]} s_t^0 + 1_{[z_t=1]} s_t^1$$

$$s_{t+1}^0 = P_0 s_t^0 + \sigma_0 \varepsilon_t$$

$$s_{t+1}^1 = P_1 s_t^1 + \sigma_1 \varepsilon_t$$

$$\varepsilon_t \sim N(0,1)$$

$$z_t = \begin{cases} 0, & t = 4k+1, \ldots, 4k+3 \\ 1, & t = 4k \end{cases} (k \in \mathbb{N})$$

At the time t, the timeline is as follows

- B_{t-1} , B_{t-1}^0 and C_{t-1} are updated at the end of time t-1;
- Based on Q and z_{t-1}, z_t is realized at time t.
- $s_t = P_1 s_{t-1} + 1_{[z_t=0]}(P_0 - P_1)s_{t-1}^0 + (1_{[z_t=0]}\sigma_0 + 1_{[z_t=1]}\sigma_1)\varepsilon_t$ is realized at the beginning of time t.
- Form Expectation

PLM 1: $y_t = B_{t-1} s_t$ => $\tilde{E}_t^1 y_{t+1} = \tilde{E}_t^1 B_{t-1} s_{t+1} = B_{t-1}(P_1 s_t + 1_{[z_t=0]}(P_0 - P_1)s_t^0)$

PLM 2: $y_t = B_{t-1}^0 s_t^0$=> $\tilde{E}_t^1 y_{t+1} = \tilde{E}_t^1 B_{t-1}^0 s_{t+1}^0 = B_{t-1}^0 P_0 s_t^0$

PLM 3: $y_t = C_{t-1} y_{t-1}$ => $\tilde{E}_t^2 y_{t+1} = C_{t-1}^2 y_{t-1}$

- Endogenize n and n_0: $U_t^j = -(y_{t-1} - \tilde{E}_{t-1}^j y_{t-1})'(y_{t-1} - \tilde{E}_{t-1}^j y_{t-1})$, j=1, 2, 3 for s_t, s_t^0 and VAR
  - $U_t^1 \geq \max\{U_t^2, U_t^3\}$: the mass of s_t agent is n.
    - $U_t^2 \geq U_t^3$: the mass of s_t^0 agent is n_0 and the mass of VAR agent is $1 - n_0 - n$
    - $U_t^2 < U_t^3$: the mass of s_t^0 agent is 0 and the mass of VAR agent is $1 - n$
  - $U_t^2 > \max\{U_t^1, U_t^3\}$: the mass of s_t^0 agent is $n_0 + n$, the mass of s_t agent is 0 and the mass of VAR agent is $1 - n_0 - n$
  - $U_t^3 > \max\{U_t^1, U_t^2\}$: the mass of VAR agent is $\bar{n}$, the mass of s_t agent is $n + \frac{1-\bar{n}-n-n_0}{2}$ and the mass of s_t^0 agent is $n_0 + \frac{1-\bar{n}-n-n_0}{2}$
- Generate time-t data y_t

  ALM: $y_t = \xi_{1t} y_{t-1} + \xi_{2t} s_t + \xi_{3t} s_t^0$

  - $\xi_{1t} = A_1 + (1 - n - n_0) A_2 C_{t-1}^2$,
  - $\xi_{2t} = nA_2 B_{t-1} P_1 + A_3$
  - $\xi_{3t} = nA_2 B_{t-1} 1_{[z_t=0]}(P_0 - P_1) + n_0 A_2 B_{t-1}^0 P_0$
- Update B_t, B_t^0 and C_t with moment method (see Appendix)

  $B_t = \xi_{1t} r_{yst} + \xi_{2t} + \xi_{3t} r_{sst}$, where $r_{yst} = (\sum_{i=1}^t y_{i-1} s_i')(\sum_{i=1}^t s_i s_i')^{-1}$ and $r_{ss0t} = (\sum_{i=1}^t s_i^0 s_i')(\sum_{i=1}^t s_i s_i')^{-1}$

  $B_t^0 = \xi_{1t} r_{ys0t} + \xi_{2t} r_{ss0t} + \xi_{3t}$ where $r_{ys0t} = (\sum_{i=1}^t y_{i-1} s_i^{0\prime})(\sum_{i=1}^t s_i^0 s_i^{0\prime})^{-1}$ and $r_{ss0t} = (\sum_{i=1}^t s_i s_i^{0\prime})(\sum_{i=1}^t s_i^0 s_i^{0\prime})^{-1}$

  $C_t = \xi_{1t} + \xi_{2t} r_{syt} + \xi_{3t} r_{sy0t}$, where $r_{sy0t} = (\sum_{i=1}^t s_i y_{i-1}')(\sum_{i=1}^t y_{i-1} y_{i-1}')^{-1}$ and $r_{sy0t} = (\sum_{i=1}^t s_i^0 y_{i-1}')(\sum_{i=1}^t y_{i-1} y_{i-1}')^{-1}$

- Update t = t+1, repeat this process

Now fixing z_t, n and n_0, the evolution system for ξ_{1t}, ξ_{2t} and ξ_{3t} is

$$A_1 + (1 - n - n_0) A_2 (\xi_{1t} + \xi_{2t} r_{syt} + \xi_{3t} r_{sy0t})^2 \rightarrow \xi_{1t+1}$$

$$nA_2\big(\xi_{1t}r_{yst} + \xi_{2t} + \xi_{3t}r_{sst}\big)P_1 + A_3 \to \xi_{2t+1}$$

$$nA_2(\xi_{1t}r_{yst} + \xi_{2t} + \xi_{3t}r_{sst})1_{[z_t=0]}(P_0 - P_1) + n_0A_2(\xi_{1t}r_{ys0t} + \xi_{2t}r_{ss0t} + \xi_{3t})P_0 \to \xi_{3t+1}$$

Compactly put, we have

$$\xi_{t+1} = T_{n,n_0,z_t}(\xi_t)$$

- $\xi_t = (\xi_{1t}, \xi_{2t}, \xi_{3t})'$
- $T_{n,n_0,z_t} = (T^1_{n,n_0,z_t}, T^2_{n,n_0,z_t}, T^3_{n,n_0,z_t})'$
- $T^1_{n,n_0,z_t}(\xi_t) = A_1 + (1 - n - n_0)A_2(\xi_{1t} + \xi_{2t}r_{sy0t} + \xi_{3t}r_{sy1t})^2$
- $T^2_{n,n_0,z_t}(\xi_t) = nA_2\big(\xi_{1t}r_{yst} + \xi_{2t} + \xi_{3t}r_{ss0t}\big)P_1 + A_3$
- $T^3_{n,n_0,z_t}(\xi_t) = nA_2(\xi_{1t}r_{yst} + \xi_{2t} + \xi_{3t}r_{ss0t})1_{[z_t=0]}(P_0 - P_1) + n_0A_2(\xi_{1t}r_{ys1t} + \xi_{2t}r_{ss1t} + \xi_{3t})P_0$

The local stability condition is that $DT_{\xi_{it}} = \frac{dT_{n,n_0,z_t}(\bar{\xi}_t)}{d\xi_{it}}$ has all eigenvalues within the unit circle for i=1, 2, 3

- $\bar{\xi}_t$ is RPE value of ξ_t
- $DT_{\xi_{1t}} = \frac{dT^1_{n,n_0,z_t}(\bar{\xi}_t)}{d\xi_{1t}} = (1 - n - n_0)\big(\bar{\xi}_{1t} + \bar{\xi}_{2t}r_{sy0t} + \bar{\xi}_{3t}r_{sy1t}\big)' \otimes A_2 + (1 - n - n_0)I \otimes A_2\big(\bar{\xi}_{1t} + \bar{\xi}_{2t}r_{sy0t} + \bar{\xi}_{3t}r_{sy1t}\big)$
- $DT_{\xi_{2t}} = \frac{dT^2_{n,n_0,z_t}(\bar{\xi}_t)}{d\xi_{2t}} = nP_1' \otimes A_2$
- $DT_{\xi_{3t}} = \frac{dT^3_{n,n_0,z_t}(\bar{\xi}_t)}{d\xi_{3t}} = n\left(r_{ss0t}1_{[z_t=0]}(P_0 - P_1)\right)' \otimes A_2 + n_0{P_0}' \otimes A_2$
- $\otimes$ denotes the Kronecker product

Now consider that n and n_0 are endogenously determined. This endogenous evolution path can be split into eight paths, four paths for $z_t = 0$ and four paths for $z_t = 1$. Formally,

$$\xi_{t+1} = \begin{cases} T_{n,n_0,z_t}(\xi_t), \text{when } U_t^1 \geq U_t^2 \geq U_t^3 \\ T_{n,0,z_t}(\xi_t), \text{when } U_t^1 \geq U_t^3 \geq U_t^2 \\ T_{0,n+n_0,z_t}(\xi_t), \text{when } U_t^2 \geq \max\{U_t^1, U_t^3\} \\ T_{n+\frac{1-\bar{n}-n-n_0}{2}, n_0+\frac{1-\bar{n}-n-n_0}{2}, z_t}(\xi_t), \text{when } U_t^3 \geq \max\{U_t^1, U_t^2\} \end{cases}$$

Now, we assign $n = 0.85$ and $n_0 = 0.1$, and $\bar{n} = 0.2$, and simulate the economy 50 times.

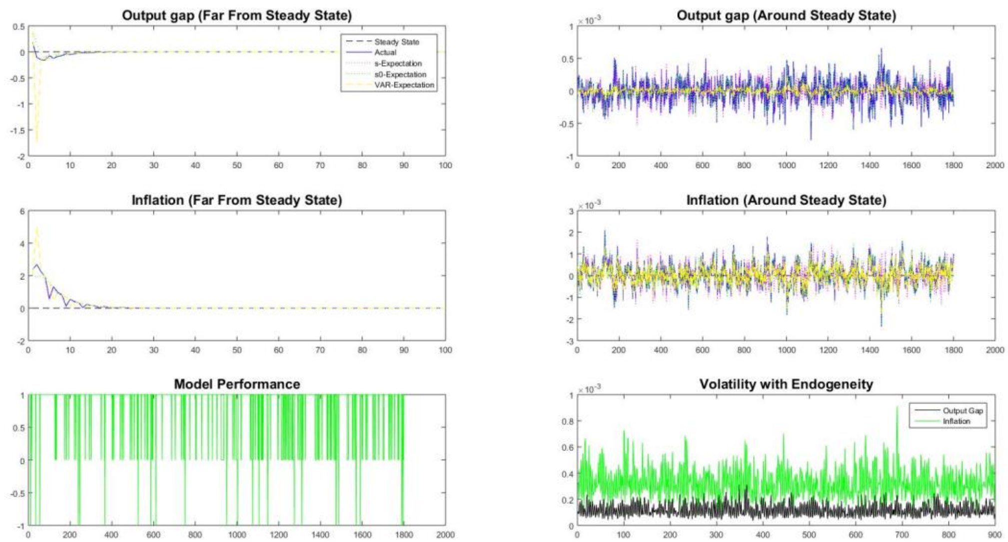

Figure 4. This is the case of regime switching without unobservability. green line is s_t^0-model-based expectation and yellow line is VAR-based expectation. In the model performance, 1 stands for s_t forecasting model performing best, 0 stands for s_t^0 forecasting model performing best and -1 stands for VAR forecasting model performing best.

According to figure 4, there are two interesting phenomena.

1. There is no significant difference after allowing endogenous expectation formation, and the expectation is usually switched between s_t^0 model and s_t model.
2. Endogenous expectation formation does not generate more volatility in output and inflation

For the first phenomenon, after allowing endogenous expectation formation (after 450th period), the frequency of agents choosing s_t^0 forecasting model compared to that before 100th period does not significantly change, which implies that the endogenous expectation formation mechanism does not lead to a huge change in the way of agents choosing the best-performed models. Another interesting thing is that the model choice is always between s_t^0 model and s_t model, not VAR model. It means that VAR model does not give a better fitting and explain what is going on today and let alone forecast for the future. When s_t^0 is dominant, s_t^0 model might be selected by s_t agent; when s_t^1 is dominant, s_t model might be chosen by s_t agent. Even though the best-performed models are often switched between s_t model and s_t^0 model, the aggregate expectation does not necessarily shift that much. The reason is that when s_t^0 model overperforms other models, the dominant regime is very likely to be s_t^0 model, and s_t agent's expectation should be the same no matter whether she use s_t^0 model or s_t model. Hence, the mass distribution change does not mean that the aggregate expectation changes. For the second phenomenon, since the aggregate expectation does not change that much between exogenous expectation formation and endogenous expectation formation, it is clear that the evolution path of output gap and inflation should not be different.

2.3. Case 3: Unobservable Regime Switching

Now we consider a further case where all agents cannot observe s_t which has a direct impact on the economy. Assume that the economy has a shock set $\mathrm{S} = \{s_t, s_t^0, s_t^1\}$. There is regime switching mechanism $s_t = 1_{[z_t=0]} s_t^0 + 1_{[z_t=1]} s_t^1$ and the state $z_t = 0, 1$ can be determined by $z_t = \begin{cases} 0, & t = 4k+1, .., 4k+3 \\ 1, & t = 4k \end{cases} (k \in \mathbb{N})$, and the two shocks have the following evolution $s_{t+1}^0 = P_0 s_t^0 + \sigma_0 \varepsilon_t$ and $s_{t+1}^1 = P_1 s_t^1 + \sigma_1 \varepsilon_t$, where $\varepsilon_t \sim N(0,1)$. Thus, the evolution of the exogenous shock is $s_{t+1} = 1_{[z_t=0]} P_0 s_t^0 + 1_{[z_t=1]} P_1 s_t^1 + (1_{[z_t=0]} \sigma_0 + 1_{[z_t=1]} \sigma_1) \varepsilon_t$. There are two agents, s_t^{10} agent who can observe s_t^1 and s_t^0 and the VAR agent who can observe nothing in the shock set S. The mass distributions of the two agents is n_{10} and $1 - n_{10}$, respectively. For convenience, we will use s_t^0 agent and s_t^1 agent to represent the s_t^{10} agent in different cases, one for s_t^{10} agent using s_t^0 forecasting model and one for using s_t^1 forecasting model. The mass distribution of s_t^0 agent and s_t^1 agent and VAR agent is n_0, n_1 and $1 - n_0 - n_1$. Here, $n_{10} = n_0 + n_1$. When s_t^0 forecasting model overperforms other models, $n_1 = 0$; symmetrically, when s_t^1 forecasting model overperforms other models, $n_0 = 0$. We can simply interpret this case as one where all agents do not have perfect information about which exogenous shock hits the economy. In some sense, in some parameter setting, s_t^1 and s_t^0 are good proxies for s_t. When $z_t = 0$, s_t^0 is exactly s_t and s_t^1 is a "bad-quality" proxy of s_t. When $z_t = 1$, s_t^1 is exactly s_t and s_t^0 is a "bad-quality" proxy of s_t. Different from the previous model, s_t^{10} agent can observe two exogenous shocks but is not sure about the true exogenous shock s_t. She can base her expectation formation on the relative performance among all forecasting models. She would use VAR forecasting model if both of s_t^0 and s_t^1 forecasting model are weaker than VAR model and in many cases she would be more likely to use one of s_t^1 and s_t^0 forecasting models due to the fact that some information is better than no information, but since the shock s_t is unobservable to s_t^{10} agent, she cannot use s_t forecasting model for her expectation formation; On the other hand, VAR agent cannot choose s_t^1 forecasting model or s_t^0 forecasting model due to her unobservability of shocks in the shock set S. Similarly, when the VAR forecasting model is best in performance we set the threshold $\bar{n} < 1$ such that there is only a mass $\bar{n}$ of VAR agents in order for avoiding the explosion in the pure forward-looking NK model. The three agents' forecasting models (i.e., PLM) are as follows,

PLM 1: $y_t = B_{t-1}^1 s_t^1$

PLM 2: $y_t = B_{t-1}^0 s_t^0$

PLM 3: $y_t = C_t y_{t-1}$

The expectation formations of s_t^1 agent and s_t^0 agent and VAR agent are $\tilde{E}_t^1 y_{t+1} = B_{t-1}^1 P_1 s_t^1$, $\tilde{E}_t^2 y_t = B_{t-1}^0 P_0 s_t^0$ and $\tilde{E}_t^3 y_t = C_{t-1}^2 y_{t-1}$, respectively. It is easy to obtain the aggregate expectation $\tilde{E}_t y_t = n_1 \tilde{E}_t^1 y_t + n_0 \tilde{E}_t^2 y_t + (1 - n_0 - n_1)\tilde{E}_t^3 y_t$. Now we write the macroeconomic system in a compact way

$$y_t = A_1 y_{t-1} + A_2 \tilde{E}_t y_{t+1} + A_3 s_t$$

$$s_t = 1_{[z_t=0]} s_t^0 + 1_{[z_t=1]} s_t^1$$

$$s_{t+1}^0 = P_0 s_t^0 + \sigma_0 \varepsilon_t$$

$$s_{t+1}^1 = P_1 s_t^1 + \sigma_1 \varepsilon_t$$

$$\varepsilon_t \sim N(0,1)$$

$$z_t = \begin{cases} 0, & t = 4k+1, .., 4k+3 \\ 1, & t = 4k \end{cases} (k \in \mathbb{N})$$

Expectation-based economy is evolving as follows

- B_{t-1}^1 , B_{t-1}^0 and C_{t-1} are determined at the end of time t-1;
- Based on Q and z_{t-1}, z_t is realized at time t.
- $s_t = P_1 s_{t-1} + 1_{[z_t=0]}(P_0 - P_1)s_{t-1}^0 + (1_{[z_t=0]}\sigma_0 + 1_{[z_t=1]}\sigma_1)\varepsilon_t$ is realized at the beginning of time t.
    - Derivation: $s_t = 1_{[z_t=0]}P_0 s_{t-1}^0 + P_1(s_t - 1_{[z_t=0]}s_{t-1}^0) + (1_{[z_t=0]}\sigma_0 + 1_{[z_t=1]}\sigma_1)\varepsilon_t$
- Form Expectation

PLM 1: $y_t = B_{t-1}^1 s_t^1$ => $\tilde{E}_t^1 y_{t+1} = \tilde{E}_t^1 B_{t-1}^1 s_{t+1}^1 = B_{t-1}^1 P_1 s_t^1$

PLM 1': $y_t = B_{t-1}^0 s_t^0$=> $\tilde{E}_t^1 y_{t+1} = \tilde{E}_t^1 B_{t-1}^0 s_{t+1}^0 = B_{t-1}^0 P_0 s_t^0$

PLM 2: $y_t = C_{t-1} y_{t-1}$ => $\tilde{E}_t^2 y_{t+1} = C_{t-1}^2 y_{t-1}$

- Endogenize n and n_0: $U_t^j = -(y_{t-1} - \tilde{E}_{t-1}^j y_{t-1})'(y_{t-1} - \tilde{E}_{t-1}^j y_{t-1})$, j=1, 2, 3 for s_t^1, s_t^0 and VAR
    - $U_t^1 > \max\{U_t^2, U_t^3\}$: the mass of s_t^1 agent is n_{10}, the mass of s_t^0 agent is 0 and the mass of VAR agent is $1 - n_{10}$
    - $U_t^2 > \max\{U_t^1, U_t^3\}$: the mass of s_t^1 agent is 0, the mass of s_t^0 agent is n_{10} and the mass of VAR agent is $1 - n_{10}$
    - $U_t^3 > U_t^1 > U_t^2$: the mass of s_t^1 agent is $1 - \bar{n}$, the mass of s_t^0 agent is 0 and the mass of VAR agent is $\bar{n}$
    - $U_t^3 > U_t^2 > U_t^1$: the mass of s_t^1 agent is 0, the mass of s_t^0 agent is $1 - \bar{n}$ and the mass of VAR agent is $\bar{n}$
- Generate time-t data y_t

ALM: $y_t = \xi_{1t} y_{t-1} + \xi_{2t} s_t^1 + \xi_{3t} s_t^0$

- $\xi_{1t} = A_1 + (1-n_{10})A_2 C_{t-1}^2$,
- $\xi_{2t} = n_1 A_2 B_{t-1}^1 P_1 + A_3 1_{[z_t=1]}$
- $\xi_{3t} = n_0 A_2 B_{t-1}^0 P_0 + A_3 1_{[z_t=0]}$

- Update B_t^1, B_t^0 and C_t with moment method

$B_t^1 = \xi_{1t} r_{ys1t} + \xi_{2t} + \xi_{3t} r_{s0s1t}$, where $r_{ys1t} = (\sum_{i=1}^t y_{i-1}\, s_t^{1\prime})(\sum_{i=1}^t s_t^1\, s_t^{1\prime})^{-1}$ and $r_{s0s1t} = (\sum_{i=1}^t s_t^0\, s_t^{1\prime})(\sum_{i=1}^t s_t^1\, s_t^{1\prime})^{-1}$

$B_t^0 = \xi_{1t} r_{ys0t} + \xi_{2t} r_{s1s0t} + \xi_{3t}$ where $r_{ys0t} = (\sum_{i=1}^t y_{i-1}\, s_t^{0\prime})(\sum_{i=1}^t s_t^0\, s_t^{0\prime})^{-1}$ and $r_{s1s0t} = (\sum_{i=1}^t s_t^1\, s_t^{0\prime})(\sum_{i=1}^t s_t^0\, s_t^{0\prime})^{-1}$

$C_t = \xi_{1t} + \xi_{2t} r_{s1yt} + \xi_{3t} r_{s0yt}$, where $r_{s1yt} = (\sum_{i=1}^t s_t^1 y_{i-1}')(\sum_{i=1}^t y_{i-1} y_{i-1}')^{-1}$ and $r_{s0yt} = (\sum_{i=1}^t s_t^0 y_{i-1}')(\sum_{i=1}^t y_{i-1} y_{i-1}')^{-1}$

- Update t = t+1, repeat this process

Fixing z_t, n_1 and n_0, the evolution system for ξ_{1t}, ξ_{2t} and ξ_{3t} is written as

$$A_1 + (1-n_1-n_0)A_2(\xi_{1t} + \xi_{2t} r_{s1yt} + \xi_{3t} r_{s0yt})^2 \to \xi_{1t+1}$$

$$n_1 A_2(\xi_{1t} r_{ys1t} + \xi_{2t} + \xi_{3t} r_{s0s1t})P_1 + A_3 1_{[z_t=1]} \to \xi_{2t+1}$$

$$n_0 A_2\big(\xi_{1t} r_{ys0t} + \xi_{2t} r_{s1s0t} + \xi_{3t}\big)P_0 + A_3 1_{[z_t=0]} \to \xi_{3t+1}$$

Writing the system into a compact form, we have

$$\xi_{t+1} = T_{n,n_0,z_t}(\xi_t)$$

- $\xi_t = (\xi_{1t}, \xi_{2t}, \xi_{3t})'$
- $T_{n,n_0,z_t} = (T_{n,n_0,z_t}^1, T_{n,n_0,z_t}^2, T_{n,n_0,z_t}^3)'$
- $T_{n,n_0,z_t}^1(\xi_t) = A_1 + (1-n_1-n_0)A_2(\xi_{1t} + \xi_{2t} r_{s1yt} + \xi_{3t} r_{s0yt})^2$
- $T_{n,n_0,z_t}^2(\xi_t) = n_1 A_2(\xi_{1t} r_{ys1t} + \xi_{2t} + \xi_{3t} r_{s0s1t})P_1 + A_3 1_{[z_t=1]}$
- $T_{n,n_0,z_t}^3(\xi_t) = n_0 A_2\big(\xi_{1t} r_{ys0t} + \xi_{2t} r_{s1s0t} + \xi_{3t}\big)P_0 + A_3 1_{[z_t=0]}$

The local stability condition is obtained: $DT_{\xi_{it}} = \frac{dT_i(\bar{\xi}_t)}{d\xi_{it}}$ has all eigenvalues within the unit circle for i=1, 2, 3

- $\bar{\xi}_t$ is RPE value of ξ_t
- $DT_{\xi_{1t}} = \frac{dT_1(\bar{\xi}_t)}{d\xi_{1t}} = (1 - n_1 - n_0)\left(\bar{\xi}_{1t} + \bar{\xi}_{2t} r_{s1yt} + \bar{\xi}_{3t} r_{s0yt}\right)' \otimes A_2 + (1 - n_1 - n_0) I \otimes A_2 \left(\bar{\xi}_{1t} + \bar{\xi}_{2t} r_{s1yt} + \bar{\xi}_{3t} r_{s0yt}\right)$
- $DT_{\xi_{2t}} = \frac{dT_2(\bar{\xi}_t)}{d\xi_{2t}} = n_1 P_1' \otimes A_2$
- $DT_{\xi_{3t}} = \frac{dT_3(\bar{\xi}_t)}{d\xi_{3t}} = n_0 P_0{}' \otimes A_2$
- $\otimes$ denotes the Kronecker product

Now consider that n_1 and n_0 are endogenized. This endogenous evolution path can be split into eight paths, four paths for $z_t = 0$ and four paths for $z_t = 1$. Formally,

$$\xi_{t+1} = \begin{cases} T_{n_{10},0,z_t}(\xi_t), \text{when } U_t^1 > \max\{U_t^2, U_t^3\} \\ T_{0,n_{10},z_t}(\xi_t), \text{when } U_t^2 > \max\{U_t^1, U_t^3\} \\ T_{1-\bar{n},0,z_t}(\xi_t), \text{when } U_t^3 > U_t^1 > U_t^2 \\ T_{0,1-\bar{n},z_t}(\xi_t), \text{when } U_t^3 > U_t^2 > U_t^1 \end{cases}$$

Now, we assign $n_1 = 0.5$ and $n_0 = 0.5$, and $\bar{n} = 0.2$, and simulate the economy 50 times. Here, we do not let VAR agents exist in the model, but the agents can use VAR model for expectation formation.

According to figure 5, there are two interesting features. After allowing endogenous expectation formation,

1. Endogenous expectation formation makes output and inflation more volatile.
2. Agents' s_t^1 expectation and s_t^0 expectation is more volatile than those in the case with exogenous expectation formation.

From the graph of volatility, it is clear to see that the endogenous expectation formation will make output and inflation volatile. There is an amplification mechanism. When we fix the mass distribution of agents, the aggregate expectation is always from a half-and-half combination of s_t^0 expectation and s_t^1 expectation; however, once the mass distribution of agents can freely shift based on their performance, the aggregate expectation will come from either s_t^0 expectation or s_t^1 expectation. When regime is switching from s_t^0 to s_t^1, the model performance will give agents a signal of expectation switching from s_t^0 expectation to s_t^1 expectation, a structural break will show up in aggregate expectation, which gives rise to a large volatility in output and inflation. When regime switching occurs very frequently, the endogenous expectation switching will happen frequently as well. The second feature is very interesting. First, why are s_t^0 expectation and s_t^1 expectation more volatile than those in the case with exogenous expectation formation? The reason is that the actual output and inflation data is more volatile than the case of exogenous expectation formation, and the volatile data will return a volatile coefficient system for s_t^0 model and s_t^1 model, leading to a volatile expectation.

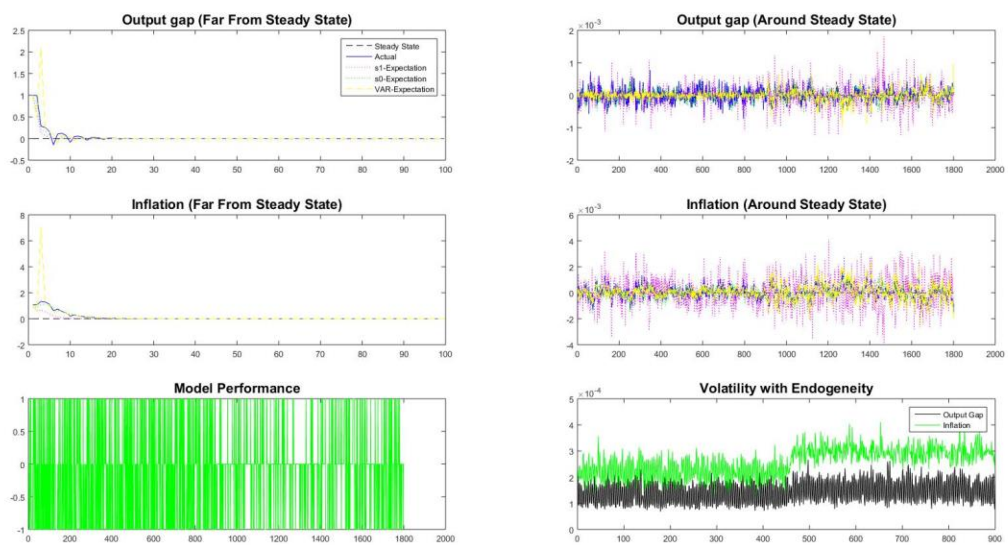

Figure 5. This is the case of regime switching with unobservability. In the model performance, 1 stands for s_t^1 forecasting model performing best, 0 stands for s_t^0 forecasting model performing best and -1 stands for VAR forecasting model performing best.

Combining the first and second feature, we can find a more interesting point, that is, the increased volatility of macroeconomic data and the increased volatility of expectation can be strengthened by each other. The entangled forces drive the economy more volatile.

2.4. Visualization of the Effects: Regime Switching Shocks and Endogenous Expectation Switching Shocks

We use the absolute value of first order difference of macroeconomic paths as measurement for volatility. The reason is that we would like to measure the length of every step the economy moves forward, where this is only dependent on the current state and not on the historical path. In this sense, this measure is better than the traditional variance in a specific sample or rolling-window samples.

First, we look at the effect of regime switching shock. In doing so, we use two cases to compare: one case is that the endogenous expectation switching shock is shut down and the regime switching shock is opened; the other case is that both shocks are shut down. Practically, setting the expectation formation exogenously is the shutdown of endogenous expectation switching shock and setting a very infrequent regime switching $z_t = \begin{cases} 0, & t = 100k+1, \dots, 100k+99 \\ 1, & t = 100k \end{cases}$ $(k \in \mathbb{N})$ represents a shutdown of regime switching shocks and setting a very frequent regime switching $z_t = \begin{cases} 0, & t = 4k+1, \dots, 4k+3 \\ 1, & t = 4k \end{cases}$ $(k \in \mathbb{N})$ represents cyclical regime switching shocks. To avoid the effect of fundamental shocks, we set AR(1) s_t^0 process and s_t^1 process having the same variance by balancing convergence speed and the variance of white noise, see the figure below. In the following figure, we see two features. First, for both output and inflation, the step length of the economy moving forward as volatility with regime switching shocks is much larger than that without regime switching shocks, which is indicated by the fact that the volatility of output increases from less than 10^{-4} to more than $3 * 10^{-4}$ in the following graph. In this sense, the effect of regime switching shocks is more important than the effect of fundamental shocks, especially on output. Second, the output is more sensitive to regime switching shocks than inflation. It is clear that the left graph indicates that frequent regime switching shocks make the volatility of output surpass the volatility of inflation. Moreover, from the right graph, it also shows that the sharp impulse responses of output to regime switching shocks is stronger than that of inflation.

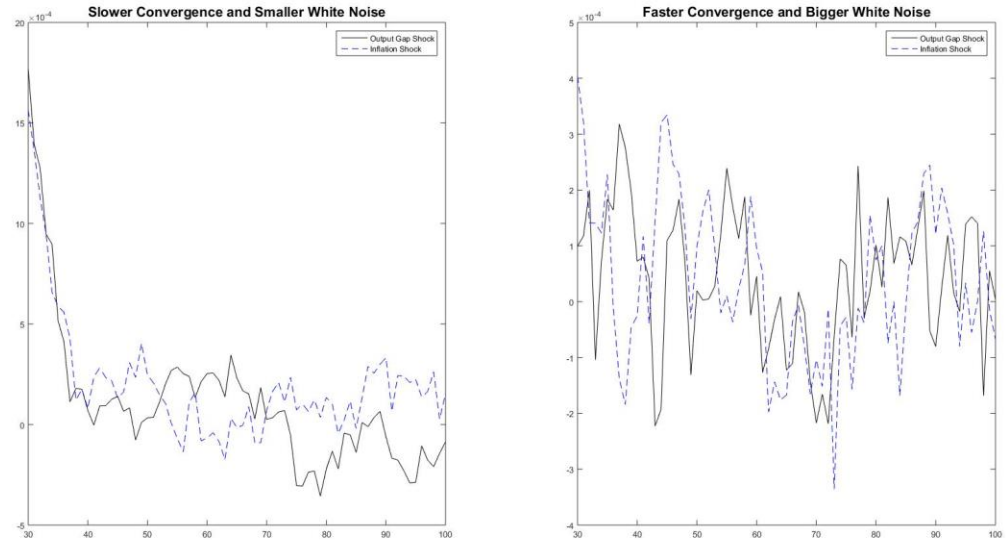

Figure 6. The left graph is s_t^0 process and the right graph is s_t^1 process

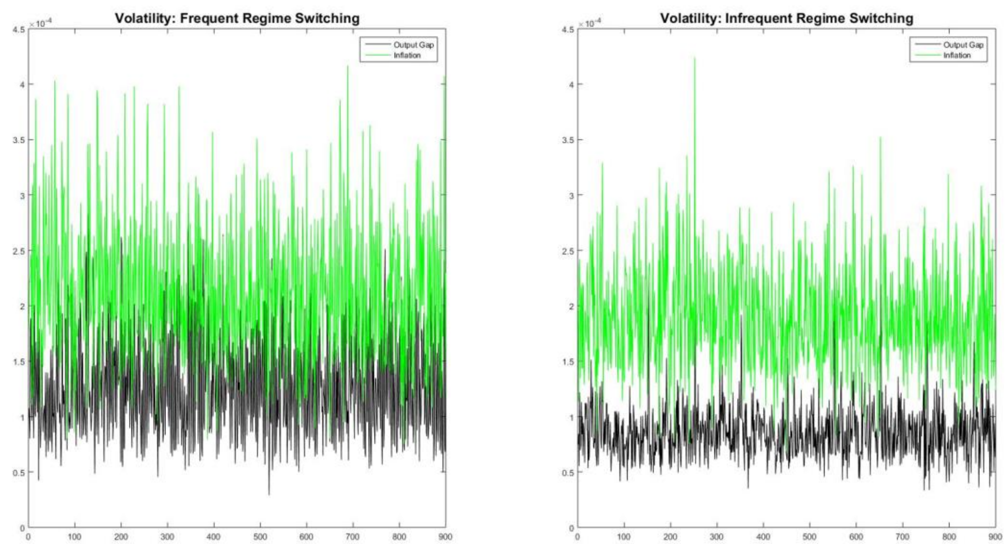

Figure 7. Regime switching shock with exogenous expectation formation

Second, we look at the effect of endogenous switching shock. In similar logic, we set a frequent regime switching as $z_t = \begin{cases} 0, & t = 4k+1, \ldots, 4k+3 \\ 1, & t = 4k \end{cases}$ $(k \in \mathbb{N})$ and allow endogenous expectation formation (the left graph) relative to exogenous expectation formation (the right graph).

There are two interesting points. First, endogenous expectation switching shocks drive more fluctuations in both output and inflation and the output is more sensitive to expectation switching shocks. Second, there are many impulsive peaks in the left-hand graph. In the right-hand graph, without expectation switching shocks, the volatility of output and inflation is roughly between $5 * 10^{-5}$ and $3 * 10^{-4}$. However, in the left-hand graph, the volatility of inflation lies between $5 * 10^{-5}$ and $4 * 10^{-4}$, and the volatility of output gap can reach 10^{-3}, meaning that the effect of endogenous expectation switching shocks is much larger than that of regime switching shocks. There is an interesting amplification mechanism. An endogenous expectation switching shock is usually driven by a regime switching shock and it cannot exist without the regime switching. Therefore, an endogenous expectation switching shock can be viewed as an amplifier based on regime switching shocks.

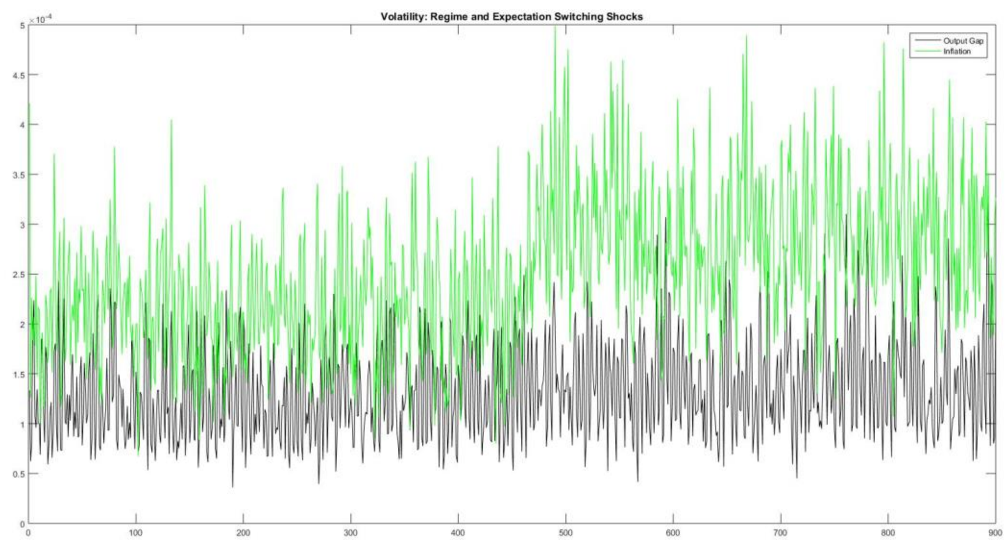

Figure 8. Endogenous expectation switching shocks

In sum, it is seen that the effect of the regime switching shocks is much larger than that of the fundamental shocks on the volatility of output and inflation and the endogenous expectation switching shocks as an amplifier are much larger than that of regime switching shocks. Moreover, the volatility of output is sensitive to both shocks, especially to endogenous expectation switching shock.

Chapter 3: Policy Implications at Zero Lower Bound

Having seen how the unobservability generates a greater macroeconomic volatility in the previous section, we discuss the monetary policy implication in this section. We consider the possibility of zero lower bound and assume that the central bank sets an inflation target $\bar{\pi}$, say 0.5% per quarter, so the policy maker sets the policy rate as

$$i_t = i^* + \chi_x \tilde{E}_t x_{t+1} + \chi_\pi (\tilde{E}_t \pi_{t+1} - \bar{\pi}) \tag{7}$$

Where $i^* = \frac{((\alpha_1 + \alpha_2 - 1 - \alpha_3 \phi_x)\frac{1-\lambda_1-\lambda_2}{\lambda_3} + \alpha_3)}{\alpha_3} \bar{\pi}$ which guarantees that one equilibrium inflation is $\bar{\pi}$. Therefore, there are two equilibria: one is a healthy equilibrium $(\bar{x}, \bar{\pi})$, and the other is a deflation equilibrium $(\bar{x}_d, \bar{\pi}_d)$, represented by the graph below.

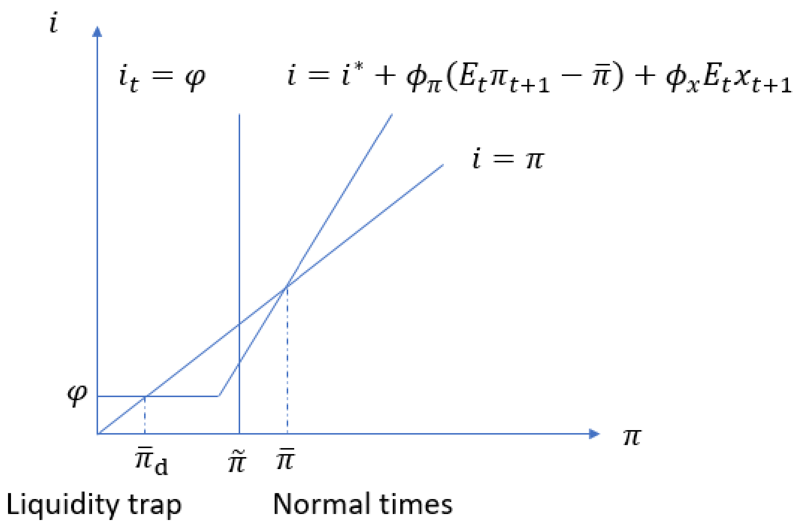

Figure 9. ZLB and two equilibria

In this section, to investigate the role of unobservability on the deflation risk and the policy setting, we only consider the case where agents can observe s_t^0 or s_t^1 but not s_t. It is interesting that greater macroeconomic fluctuations driven by the endogenous expectation switching would push the economy to the deflation equilibrium, in which case a traditional policy making usually fails to avoid deflation risk if policy makers only consider exogenous shocks. We will visualize the effects of two shocks -regime switching shock and endogenous expectation switching shock- in different parameter settings, and then explore how the policy rate threshold is able to avoid as much as possible the economy to fall in a deflationary trap.

3.1. Simulations

Considering zero lower bound, there are two equilibria: the healthy equilibrium $(\bar{x}, \bar{\pi})$ and the deflationary equilibrium $(\bar{x}_d, \bar{\pi}_d)$. When shocks are large enough, the economy will evolve and fall into the deflationary trap $\bar{\pi}_d$, which must be avoided by the policy maker. Policy makers are assumed to use an expected inflation based threshold $\tilde{\pi}$ as an alert: when agents' aggregate inflation expectation is above $\tilde{\pi}$, policy makers use the standard Taylor rule for policy making; when the expected inflation is lower than the threshold $\tilde{\pi}$, policy makers use an aggressive policy rate φ, calibrated as 0.1%, to boost the economy. If the inflation is below $\tilde{\varphi}$ which is calibrated as 0.25%, the economy is counted as be stuck in a deflationary trap.

We do simulations as follows. Imagine that the economy starts from a positive shock, where a half of agents live in each unconnected informational island, meaning that expectation is formed exogenously. After 4000 periods, their information can be exchanged, that is, informational islands are connected, meaning that agents' expectation is formed endogenously. All agents use historical data to update the same forecasts for every step, and continue until 5000 periods. In between, when the expected inflation is less than this threshold, the policy maker would abandon Taylor rule and switch to an aggressive policy rate, say $0.1\% \ll \bar{\pi}$, instead until the expected inflation goes above the threshold. We repeat the economic scenarios 10 times and 50 times. We track the evolution of the economy. We care about several things under different parametrization in threshold value $\tilde{\pi}$ and the connectivity of islands: aggressive policy switching frequency, deflation frequency, time-varying model performance and the convergence time and stability of the convergence path.

- Baseline model

We start with a model with a threshold $\tilde{\pi}$ = 0.003, 4-cycle regime switching and endogenous expectation formation. 4-cycle regime switching denotes that the economy stays in regime s_t^0 for three consecutive periods and then jumps to regime s_t^1 and immediately jumps back to regime s_t^0 after one period. The reason why we set 4-cycle is that we want a relatively frequent regime switching to simulate an economy with frequent structural changes. From graphs of output gap and inflation, we have three interesting findings. First, when islands are connected, the economy converges very fast. The economy converges to the equilibrium very slowly and such equilibrium is a restricted perception equilibrium that is different from the healthy or the deflation equilibrium, and the reason is that the "wrong" expectational feedback matters to the economy and cannot be corrected through shifting to the best-performed model. After endogenous expectation formation is allowed, the economy immediately converges to the area around one equilibrium. Second, large output fluctuations are exactly consistent with policy switching, but inflation does not respond that strongly to policy switching. When the economy is stuck in the low inflation expectation, the policy would be switched to an aggressive policy regime to stay away from the deflationary trap, but the demand would immediately respond to the low interest rate environment leading to boosting economic growth. However, inflation does not have such large reactions because the price level does not directly response to interest rates but the relation between demand and supply. Third, the best-performed model switching is consistent with policy switching. Furthermore, the result shows that agents prefer VAR models during the policy switching. The reason is that when the economy is switched to a different policy regime, the situations would become more complicated for forecasting, and market participants would take some time to figure out what factor would be the good fit to explain what is happening today and use it for forecasting tomorrow. Such complicated scenarios lead agents to behaving in a conservative way, meaning that they give up identifying which is the dominant regime of fundamentals.

The following table shows the aggressive policy switch frequency and deflation frequency in all periods the economy experiences. The number of simulations is how many times the economy goes through 5000 periods and all periods are the number of simulations times 5000 periods. When we simulate the economy 50 times, the economy experiences 54.75% of aggressive policy regime periods and stays in deflationary trap in 27.21% of all periods.

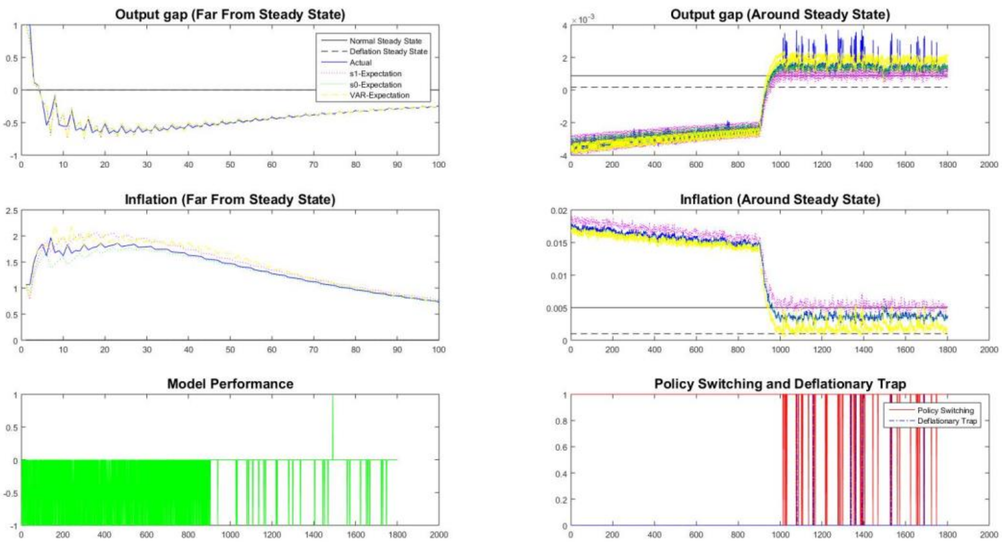

Figure 10. Policy switching threshold $\tilde{\pi}$ = 0.003 and regime switching $z_t = \begin{cases} 0, & t = 4k+1, \dots, 4k+3 \\ 1, & t = 4k \end{cases}$ $(k \in \mathbb{N})$ *and the endogenous expectation formation is allowed after 900 periods. In model performance, 1 denotes that* s_t^1 *model outperforms other models, 0 denotes that* s_t^0 *model outperforms other models and -1 denotes that VAR model outperforms other models. In policy switching and deflationary trap, 1 in policy switching denotes Taylor-rule policy used and 0 in policy switching denotes aggressive policy used; 1 in deflationary trap denotes that the economy is stuck in the deflationary trap and 0 in deflationary trap denotes that the economy is away from deflationary trap.*

simulation	Aggressive policy (%)	Deflation (%)
10	59.57%	33.41%
50	54.75%	27.21%

Table 1. $\tilde{\pi}$ = 0.003 and 4-cycle regime switching

- Infrequent Regime Switching Model

Consider the economy experiences the infrequent regime switching with 12-cycle where the economy stays in regime s_t^0 for 11 consecutive periods and in regime s_t^1 for one period. There are two attractive questions for one-life simulation. First, why doesn't the policy switch? Why doesn't the economy fall into the deflationary trap? The straightforward source is due to smaller regime switching shocks resulting in smaller endogenous expectation switching shocks. When macro fluctuations get smaller, the economy would usually go around the healthy equilibrium, and expected inflation would be stable and not smaller than the alert threshold. Therefore, the policy switching does not occur that frequently, making the macroeconomic volatility smaller and more stable, so there are no excessive fluctuations in output and inflation.

Having simulated 10 times and 50 times, from the following table showing the aggressive policy switch frequency and deflation frequency in all periods the economy experiences, we can see that the economy experiences 0.86% of aggressive policy switching and stays in deflationary trap in only 0.42% of all periods when we simulate the economy 10 times. When simulating the economy 50 times, the two numbers are still not changed that much: the probability of the policy maker using an aggressive policy is 1.25% and the economy has the probability of 0.38% in falling into a deflationary trap. Clearly, the infrequent regime switching depresses the endogenous expectation switching shocks and policy switching shocks that generate larger macroeconomic volatility.

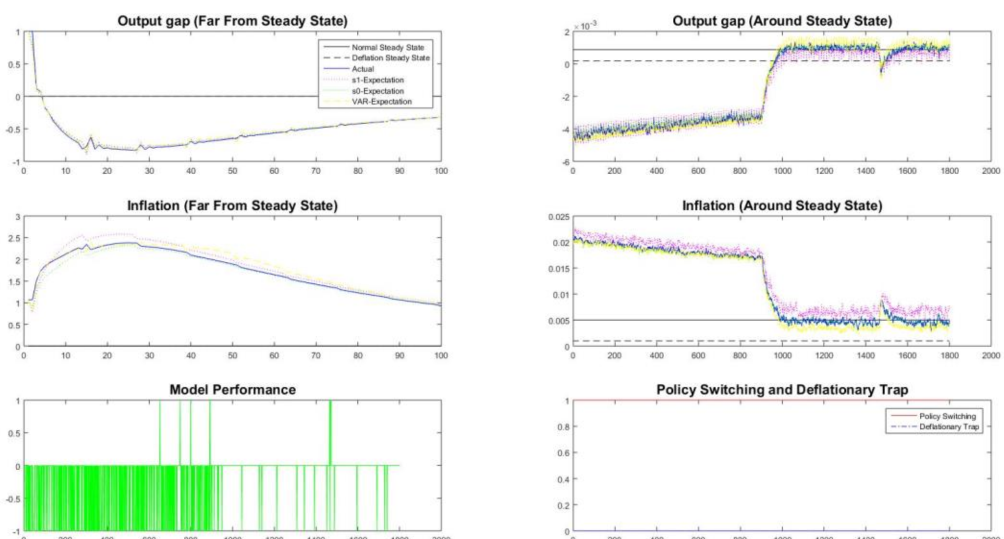

Figure 11. Policy switching threshold $\tilde{\pi}$ = 0.003 and regime switching $z_t = \begin{cases} 0, & t = 12k+1, \ldots, 12k+11 \\ 1, & t = 12k \end{cases}$ $(k \in \mathbb{N})$ and the endogenous expectation formation is allowed after 900 periods. In model performance, 1 denotes that s_t^1 model outperforms other models, 0 denotes that s_t^0 model outperforms other models and -1 denotes that VAR model outperforms other models. In policy switching and deflationary trap, 1 in policy switching denotes Taylor-rule policy used and 0 in policy switching denotes aggressive policy used; 1 in deflationary trap denotes that the economy is stuck in the deflationary trap and 0 in deflationary trap denotes that the economy is away from deflationary trap.

simulation	Aggressive policy (%)	Deflation (%)
10	0.86%	0.42%
50	1.25%	0.38%

Table 2. $\tilde{\pi}$ = 0.003 and 4-cycle regime switching

- A Model with a higher the Expected Inflation Based Threshold

We consider the effect of raising the alert threshold from $\tilde{\pi}$ = 0.003 to $\tilde{\pi}$ = 0.004. When the policy maker realizes that the regime switching can generate more endogenous volatility from connecting informational islands, they must raise an alert threshold $\tilde{\pi}$ to avoid inflation risk. However, there are two side products. First, a higher threshold leads to stronger fluctuations in output and inflation. In higher expected inflation environment, the economy stays in a low policy rate environment, boosting the economic activities and inflation more than in lower expected inflation environment. Second, the unsustainable boom in output lasts longer period. Clearly, the economy is more likely to stay in an aggressive policy environment if the policy maker worries about and want to avoid the deflation spiral.

In sum, the following table summarizes the three cases above. We can find that infrequent regime switching results in less volatility in the macroeconomy and an aggressive policy is used rarely. From the graph above, we already know that there are no large output fluctuations when an aggressive policy regime is not often reached. In this sense, there are two-level excessive volatility: one is from endogenous expectation switching and the other is from policy regime switching. What is worth emphasizing is that switching to an aggressive policy regime would lead to a considerable but unexpected output volatility. Hence, when the policy maker compromises and raise the alert threshold $\tilde{\pi}$ to maintain inflation expectation to an acceptable level and then low deflation risk which is shown to be reduced roughly from 27% to 16%, the frequency of the policy switching gets higher, that is, from around 55% to 67%, then it is shown in the above graph that macroeconomic fluctuations get larger.

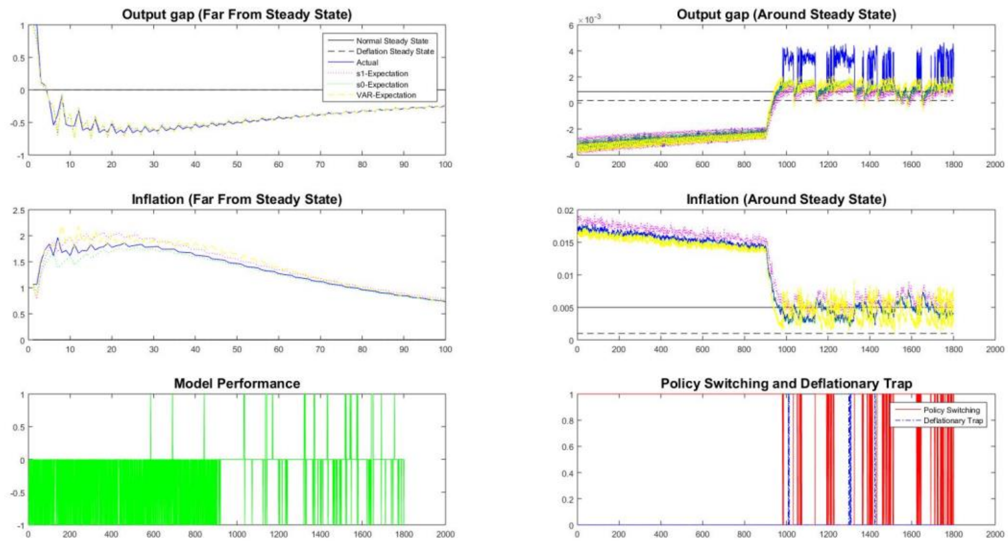

Figure 12. Policy switching threshold $\tilde{\pi}$ = 0.004 and regime switching $z_t = \begin{cases} 0, & t = 4k+1, \dots, 4k+3 \\ 1, & t = 4k \end{cases}$ $(k \in \mathbb{N})$ *and the endogenous expectation formation is allowed after 900 periods. In model performance, 1 denotes that* s_t^1 *model outperforms other models, 0 denotes that* s_t^0 *model outperforms other models and -1 denotes that VAR model outperforms other models. In policy switching and deflationary trap, 1 in policy switching denotes Taylor-rule policy used and 0 in policy switching denotes aggressive policy used; 1 in deflationary trap denotes that the economy is stuck in the deflationary trap and 0 in deflationary trap denotes that the economy is away from deflationary trap.*

	$\tilde{\pi}$ = 0.003 and 4-cycle		$\tilde{\pi}$ = 0.003 and 12-cycle		$\tilde{\pi}$ = 0.004 and 4-cycle	
simulation	Aggressive policy (%)	Deflation (%)	Aggressive policy (%)	Deflation (%)	Aggressive policy (%)	Deflation (%)
10	59.57%	33.41%	0.86%	0.42%	68.67%	11.63%
50	54.75%	27.21%	1.25%	0.38%	66.93%	15.87%

Table 3. Comparison in the frequency of aggressive policy regime switching and the frequency of the deflation between baseline model, the model with infrequent regime switching and the model with a higher alert threshold.

3.2. Discussions

We discuss two topics in this section. First, without unobservability, the economy staying in healthy equilibrium suffers from small shocks, policy makers do not worry that the economy would enter a deflationary spiral. Policy makers' goal is only to maintain the macroeconomic stability. But with unobservability leading larger shocks generated by frequent endogenous expectation switching, policy makers must consider not only to maintain macroeconomic stability, but also to avoid deflation risk even under a small negative shock. Second, the macroeconomics literature has been exploring an interesting but long-lasting question: what is the reason for large macroeconomic fluctuations? Bad policy, bad structural change or bad policy? This questions sometimes is not simple, because often they are endogenous, see Bernanke (2004). We will investigate the effect of unobservability on the two issues.

3.2.1. Policy Makers' Dilemma with Unobservability

When fundamental shocks are small enough and there is the unobservability problem, the policy maker does not need to worry about the deflation risk. The only thing that policy makers care about is how to maintain the macroeconomic stability. The solution to this problem is suggested by Bernanke (2004), who argues that to maintain macroeconomic stability policy makers must avoid a bad cycle: attempts to achieve higher output lower interest rates, and then it leads to an increase in inflation, and policy makers tighten the policy for rising inflation and then make a sharp contraction in output. However, once we consider that the economy has the unobservability problem, the unobservability in the dominant regimes would generate a substantial endogenous volatility, which results in that a small negative fundamental shock can be amplified to be a very large one, probably making the economy fall into a deflationary spiral. Therefore, policy makers must use an aggressive policy to protect the economy from deflation risk in advance. But the problem if policy makers take this step is that when the inflation expectation sometimes is a little low but the current inflation is still good, the low policy rate would boost the economic activities and inflation very quickly and strongly. Since the threshold is higher, the economy is more likely to be in the aggressive policy regime so that the unstable and large volatility in output would last longer. In this case, policy makers face a dilemma: to avoid deflation risk, policy makers must raise the alert threshold which leading to macroeconomic instability. It is worth pointing out that such dilemma does not exist if there is no unobservability problem.

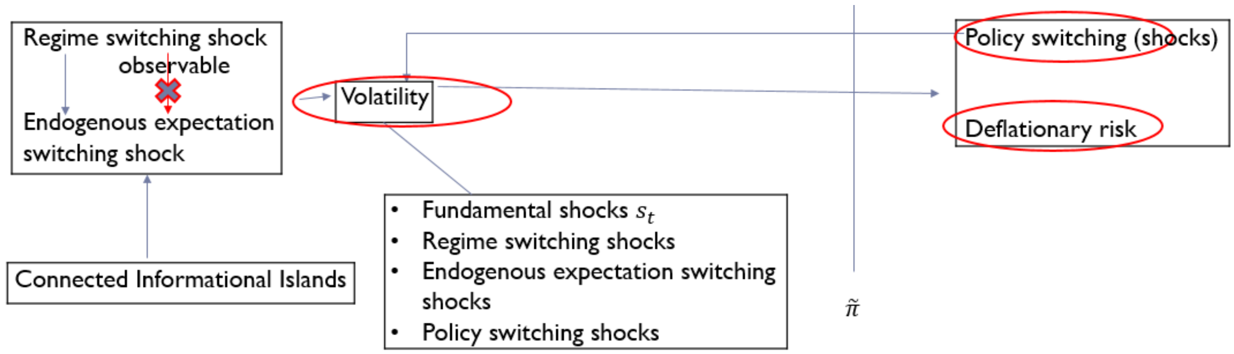

Figure 13. The relationship among macroeconomic volatility, policy switching and deflation risk

3.2.2. Volatility: Bad Luck or Bad Policy?

Much traditional literature widely discusses where the substantial macroeconomic volatility comes from. There are three typical sources: bad luck, structural change and bad policy. Until now, looking for a starting point of macroeconomic volatility is far from settled. For example, researchers have been exploring why the volatility in the macroeconomy gets smaller after 1980, which, a reduced volatility, is opposite from what we care about under zero lower bound. Some economists agree with good luck. Ahmed, Levin and Wilson (2004) shows that 50%~75% of the reduction of output volatility since 1980s is from good luck. However, they also argue that almost all the reduction of inflation volatility since 1980s is from good policy. Stock and Watson (2003) draw a similar conclusion with Ahmed, Levin and Wilson (2004), but they also investigate the relation between US and G-7 countries that part of the reduction in volatility is from the reduction in common international shocks between countries. Some other researchers emphasize the importance of structural changes. Kahn, McConnell and Perez-Quiros (2002) find that the improvement in inventory management techniques is the most important reason that output becomes less volatile because a decline in the inventory volatility smooths the output volatility. Dynan, Elmendorf and Sichel (2005) explain that the output volatility is reduced through efficiency-improved financial market. They found that the financial services are more and more dependent on online transaction platforms which has been improved aggressively so that people are easier to have access to financial markets and easier to get loans or other financial services, which smooths consumption and then output. Some economists think that good luck or good structural changes result from good policies. Clarida, Galí and Gertler (2000) argue that before 1979, Fed did raise the nominal interest rate but the real interest rate still declined because the expected inflation goes up more quickly than nominal interest rates do. The policy maker raised the nominal interest rate and real interest rate aggressively after 1980s so that the high-inflation economy then cooled down. Bernanke (2004) argues that there is a bad cycle before 1979: initially, policy makers lower interest rates to attempt to achieve higher output, and then inflation would rise very quickly which is unexpected, so the policy would be tightened, leading to a sharp contraction in output, and policy makers must lower interest rates again. This cycle is going on again and again, making output more volatile. He argues that the reduced impact of oil and commodity shocks after 1980s is due to the stable and low inflation expectation resulted from a successful policy after Volker took office as the Chairman of the Federal Reserve Board.

However, the question itself is not easy to be completely settled. The fundamental problem is the endogeneity problem. The fundamental shocks would be reduced along with a good policy implemented by central banks, and good structural changes may happen as well. Conversely, if the economy has a good luck or good structural change, the central banks will have more potential policy tools to achieve the dual mandate. The worse the economic situations (bad luck or bad structural changes), in a higher probability the policy would be a bad policy due to a more limited number of feasible policy tools. So, the fundamental question is: where is the "starting point"?

We put forth a potential mechanism that can exogenously generate excessive macroeconomic volatility under unobservability. There are two layers of "amplifiers" for macroeconomic fluctuations: endogenous expectation switching and policy strategy switching. The mechanism is as follows. Unobservability makes agents endogenously choose best-performed models to form expectations leading to the first level of excessive volatility amplified by endogenous expectation formation. Upon seeing such large volatility, the policy maker must set a higher expected inflation threshold to avoid the deflation risk, and such a higher threshold makes the policy regime switching more frequently, resulting in an even larger, named as the second level of, macroeconomic volatility.

Based on our analysis above, a large macroeconomic fluctuation does not necessarily result from a bad luck. The endogenous expectation formation gives a first level of amplified volatility. The dilemma faced by policy makers indicates that sometimes an excessive macroeconomic volatility does not imply that a bad policy is made because monetary policy also takes a responsibility of avoiding deflation risk.

Chapter 4: Conclusion

Our paper provides a theoretical explanation for excessive macroeconomic fluctuations with unobservability in regime switching where agents consider to endogenously choose a best-performed model for expectation formation. We discuss three cases: no regime switching, regime switching with agents' ability to observe the dominant regime and regime switching with agents' inability to observe the dominant regime. We find that without the regime switching the agents in most cases endogenously use the "correct" forecasting model to form expectation so that the macroeconomic volatility cannot be amplified. However, when the regime switching without unobservability is introduced, there is a larger economic fluctuation resulting from the regime switching. However, agents that can observe the two regimes and the dominant regime will use the "right" model to form expectation in most cases, so there is no frequent endogenous expectation switching for more fluctuations. When the regime switching with unobservability is considered, since agents cannot observe which regime switching is dominant one, agents will shift their expectation back and forth based on the two candidate models' performance, giving rise to a large increase in the macroeconomic volatility.

Furthermore, we explore the monetary policy implications under unobservability. We consider the zero lower bound constraint and then the problem of deflation risk. Since it is indicated in the third chapter that the unobservability of regime switching can generate the larger endogenous volatility in the macroeconomy, the policy maker must consider a policy-related solution for those fluctuations. Our recommended policy solution is setting an aggregate expected inflation based alert threshold, and the standard Taylor-rule based monetary policy is adopted if the aggregate expected inflation is above the threshold; otherwise, an aggressive low policy rate is used. We have four findings. First, the output fluctuation pattern is consistent with policy switching, but inflation is not. Second, the switching of agents' expectation always goes along with the policy regime. Third, infrequent regime switching does lead to less policy regime switching and then lower volatility of output and inflation. fourth, raising the threshold generates stronger fluctuations in both output and inflation, and output peaks last longer than the case where the threshold is lower.

Moreover, we discuss two interesting issues. First, with unobservability, policy makers face a dilemma: maintain macroeconomic stability and avoid inflation risk. It is worth pointing out that such dilemma does not necessarily exist without unobservability. Regime switching will lead agents to change their expectation frequently and generate more volatility and policy makers must raise the threshold to avoid the deflationary spiral and the economy is more frequently switched between the normal policy regime and the aggressive policy regime, giving rise to a strong macroeconomic instability. So, if policy makers do not raise the threshold, the economy would be exposed under a higher deflation risk. Second, large macroeconomic fluctuations are not necessarily from bad luck or bad policy. We argue that unobservability can generate endogenous volatility which is not from bad luck, and policy makers raising the threshold to avoid the deflationary spiral but make the economy more frequently switched between the normal policy regime and the aggressive policy regime, giving rise to a strong macroeconomic instability.

Appendix

A. Benchmark Parameter Calibration: Galí (2015)

Parameters	Description	Calibrated Value
$\boldsymbol{\alpha_1}$	Backward-looking coefficient in IS curve	0
$\boldsymbol{\alpha_2}$	Forward-looking coefficient in IS curve	1
$\boldsymbol{\alpha_3}$	Inverse elasticity of intertemporal substitution	1
$\boldsymbol{\lambda_1}$	Backward-looking coefficient in Phillips curve	0
$\boldsymbol{\lambda_2}$	Discount factor	0.99
$\boldsymbol{\lambda_3}$	Slope of the NKPC (composite parameter for price rigidity)	0.0572
$\boldsymbol{\rho_e}$	Persistence parameter in demand shock	0.8
$\boldsymbol{\rho_u}$	Persistence parameter in supply shock	0.8
$\boldsymbol{\sigma_e}$	Standard deviation in demand shock	0.0001
$\boldsymbol{\sigma_u}$	Standard deviation in demand shock	0.0001
$\boldsymbol{\chi_x}$	Central bank's output gap response	1.5
$\boldsymbol{\chi_\pi}$	Central bank's output gap response	1.5

B. Derivation for Coefficients' Update

According to PLM 1 $y_t = B_{t-1}s_t + v_t$, where v_t is iid with mean 0. When new time-t data comes into agents' information set, B_t will be updated. Using the moment method, we have

$$Ey_t s_t{}' = B_t E s_t s_t{}'$$

Where $Ey_t s_t' = E(\xi_{1t}y_{t-1} + \xi_{2t}s_t + \xi_{3t}s_t^0)s_t' = \xi_{1t}Ey_{t-1}s_t' + \xi_{2t}Es_t s_t' + \xi_{3t}Es_t^0 s_t' = B_t E s_t s_t{}'$, then we have

$$B_t = \xi_{1t}Ey_{t-1}s_t'(Es_t s_t{}')^{-1} + \xi_{2t} + \xi_{3t}Es_t^0 s_t'(Es_t s_t{}')^{-1}$$

Using the sample data available until time t, we obtain $Es_t s_t' = \frac{1}{t}\sum_{i=1}^{t} s_i\, s_i{}'$, $Es_t^0 s_t' = \frac{1}{t}\sum_{i=1}^{t} s_i^0\, s_i{}'$ and $Ey_{t-1}s_t' = \frac{1}{t}\sum_{i=1}^{t} y_{i-1}\, s_i{}'$. Finally, we have $B_t = \xi_{1t}r_{yst} + \xi_{2t} + \xi_{3t}r_{sst}$, where $r_{yst} = (\sum_{i=1}^{t} y_{i-1}\, s_i{}')(\sum_{i=1}^{t} s_i\, s_i{}')^{-1}$ and $r_{ss0t} = (\sum_{i=1}^{t} s_i^0\, s_i{}')(\sum_{i=1}^{t} s_i\, s_i{}')^{-1}$.

Applying the similar method, we can obtain $B_t^0 = \xi_{1t} r_{ys0t} + \xi_{2t} r_{ss0t} + \xi_{3t}$ where $r_{ys0t} = (\sum_{i=1}^{t} y_{i-1}\, s_i^{0\prime})(\sum_{i=1}^{t} s_i^0\, s_i^{0\prime})^{-1}$ and $r_{ss0t} = (\sum_{i=1}^{t} s_i\, s_i^{0\prime})(\sum_{i=1}^{t} s_i^0\, s_i^{0\prime})^{-1}$ and $C_t = \xi_{1t} + \xi_{2t} r_{syt} + \xi_{3t} r_{sy0t}$, where $r_{sy0t} = (\sum_{i=1}^{t} s_i y_{i-1}')(\sum_{i=1}^{t} y_{i-1} y_{i-1}')^{-1}$ and $r_{sy0t} = (\sum_{i=1}^{t} s_i^0 y_{i-1}')(\sum_{i=1}^{t} y_{i-1} y_{i-1}')^{-1}$

Reference

Ahmed, S., Levin, A., & Wilson, B. A. (2004). Recent US macroeconomic stability: good policies, good practices, or good luck?. *Review of economics and statistics*, *86*(3), 824-832.

An, S., & Schorfheide, F. (2007). Bayesian analysis of DSGE models. *Econometric reviews*, *26*(2-4), 113-172.

Basu, S., & Bundick, B. (2015). *Endogenous volatility at the zero lower bound: implications for stabilization policy* (No. w21838). National Bureau of Economic Research.

Bernanke, B. S. (2004). The Great Moderation. Remarks at the Meetings of the Eastern Economic Association. Washington DC February 20.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307-327.

Bollerslev, T. (2010). Glossary to ARCH (GARCH) in Volatility and Time Series Econometrics Essays in Honor of Robert Engle. Mark Watson, Tim Bollerslev and Jerrey Russell (eds.), Oxford University Press.

Branch, W. A., & Evans, G. W. (2007). Model uncertainty and endogenous volatility. *Review of Economic Dynamics*, *10*(2), 207-237.

Branch, W. A., & Evans, G. W. (2011). Monetary policy and heterogeneous expectations. *Economic Theory*, *47*(2-3), 365-393.

Clarida, R., Galí, J., & Gertler, M. (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *The Quarterly journal of economics*, *115*(1), 147-180.

Cochrane, J. H. (2009). Can learnability save new-Keynesian models? *Journal of Monetary Economics*, *56*(8), 1109-1113.

Dieobold, F. X. (1986). Modeling the persistence of conditional variances: A comment. *Econometric Reviews*, *5*(1), 51-56.

Dynan, K. E., Elmendorf, D. W., & Sichel, D. E. (2006). Can financial innovation help to explain the reduced volatility of economic activity?. *Journal of monetary Economics*, *53*(1), 123-150.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The journal of finance*, *48*(5), 1749-1778.

Fernández-Villaverde, J., & Rubio-Ramírez, J. F. (2007). Estimating macroeconomic models: A likelihood approach. *The Review of Economic Studies*, *74*(4), 1059-1087.

Fernández-Villaverde, J., & Rubio-Ramírez, J. (2010). *Macroeconomics and volatility: Data, models, and estimation* (No. w16618). National Bureau of Economic Research.

Fuhrer, J. C. (2000). Habit formation in consumption and its implications for monetary-policy models. *American Economic Review*, *90*(3), 367-390.

Galí, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press.

Gomes, O. (2017). Heterogeneous wage setting and endogenous macro volatility. *Journal of Economic Interaction and Coordination*, *12*(1), 27-57.

Aït-Sahalia, Y., Hansen, L. P., & Scheinkman, J. A. (2010). Operator methods for continuous-time Markov processes. *Handbook of financial econometrics*, *1*, 1-66.

Justiniano, A., & Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations. *American Economic Review*, *98*(3), 604-41.

Khan, M. S. (1977). The variability of expectations in hyperinflations. *Journal of Political Economy*, *85*(4), 817-827.

Kahn, J. A., McConnell, M. M., & Perez-Quiros, G. (2002). On the Causes of the Increased Stability of the US Economy.*Federal Reserve Bank of New York Economic Policy Review*,*8*(1), 183-202.

Kim, C. J., & Nelson, C. R. (1999). Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics and Statistics*, *81*(4), 608-616.

King, R. G., & Rebelo, S. T. (1999). Resuscitating real business cycles. *Handbook of macroeconomics*, *1*, 927-1007.

Klein, B. (1977). The demand for quality-adjusted cash balances: Price uncertainty in the US demand for money function. *Journal of Political Economy*, *85*(4), 691-715.

Lansing, K. J. (2009). Time-varying US inflation dynamics and the New Keynesian Phillips curve. *Review of Economic Dynamics*, *12*(2), 304-326.

Levine, P., Pearlman, J., Perendia, G., & Yang, B. (2012). Endogenous persistence in an estimated DSGE model under imperfect information. *The Economic Journal*, *122*(565), 1287-1312.

Lubik, T. A., & Schorfheide, F. (2004). Testing for indeterminacy: an application to US monetary policy. *American Economic Review*, *94*(1), 190-217.

Lucas, R. E. (1973). Some international evidence on output-inflation tradeoffs. *The American Economic Review*, *63*(3), 326-334.

Mankiw, N. G., & Reis, R. (2002). Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve. *The Quarterly Journal of Economics*, *117*(4), 1295-1328.

Marcet, A., & Nicolini, J. P. (2003). Recurrent hyperinflations and learning. *American Economic Review*, *93*(5), 1476-1498.

McConnell, M. M., & Perez-Quiros, G. (2000). Output fluctuations in the United States: What has changed since the early 1980's?. *American Economic Review*, *90*(5), 1464-1476.

Milani, F. (2014). Learning and time-varying macroeconomic volatility. *Journal of Economic Dynamics and Control*, *47*, 94-114.

Navarro, G. (2014). Financial crises and endogenous volatility. *manuscript, New York University*.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.

Sensier, M., & Dijk, D. V. (2004). Testing for volatility changes in US macroeconomic time series. *Review of Economics and Statistics, 86*(3), 833-839.

Sentana, E. (1995). Quadratic ARCH models. *The Review of Economic Studies*, *62*(4), 639-661.

Shephard, N. (2008). Stochastic volatility models. The New Palgrave Dictionary of Economics. Eds. Steven N. Durlauf and Lawrence E. Blume.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, *50*(3), 665-690.

Smets, F., & Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European economic association*, *1*(5), 1123-1175.

Smets, F., & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American economic review*, *97*(3), 586-606.

Stock, J. H., & Watson, M. W. (2003). Has the business cycle changed? Evidence and explanations. *Monetary policy and uncertainty: adapting to a changing economy*, 9-56.

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

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, *18*(5), 931-955.