Essays on DSGE Models and Bayesian Estimation

Jae-Yoon Kim

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Economics, Science

Aris Spanos, Chair
Richard A. Ashley
Kwok Ping Tsang
Eric A. Bahel

May 9, 2018
Blacksburg, Virginia

Keywords: Bayesian Inference, Statistical Adequacy, Misspecification Testing

Essays on DSGE Models and Bayesian Estimation

Jae-Yoon Kim

(ABSTRACT)

In the context of empirical modeling learning from data using substantive information, it is always judicious to bring out the statistical model implicitly specified by the substantive (structural, theoretical) model under consideration. The statistical model is simply the set of probabilistic assumptions imposed (often implicitly) on the data. Imposing invalid probabilistic assumptions on a particular data, undermines the reliability of statistical inference, rendering the inference results untrustworthy. Hence, for securing trustworthy evidence one should always validate the implicit statistical model before drawing any inferences from an estimated substantive model. This perspective is used to shed light on a widely used category of macroeconometric models known as Dynamic Stochastic General Equilibrium (DSGE) models. Using U.S. time-series data, the paper demonstrates that a widely used econometric model for the U.S. economy is severely statistically misspecified; almost all of its assumptions are invalid for the data. The paper proceeds to respecify the implicit statistical model behind the structural model with a view to secure its statistical adequacy (validity of its probabilistic assumptions). Using the respecified statistical model, the paper calls into question the literature evaluating the substantive adequacy of current DSGE models, ignoring the fact that such evaluations are untrustworthy because they are based on statistically unreliable procedures.

Essays on DSGE Models and Bayesian Estimation

Jae-Yoon Kim

(GENERAL AUDIENCE ABSTRACT)

For an empirical analysis the statistical model implied in the theoretical model is crucial. The statistical model is simply the set of probabilistic assumptions imposed on the data, and invalid probabilistic assumptions undermines the reliability of statistical inference, rendering the empirical analysis untrustworthy. Hence, for securing trustworthy evidence one should always validate the implicit statistical model before drawing any empirical result from a theoretical model. This perspective is used to shed light on a widely used category of macroeconometric models known as Dynamic Stochastic General Equilibrium (DSGE) Models. Using U.S. time-series data, the paper demonstrates that a widely used econometric model for the U.S. economy is severely statistically misspecified; almost all of its probabilistic assumptions are invalid for the data. The paper proceeds to respecify the implicit statistical model behind the theoretical model with a view to secure its statistical adequacy (validity of its probabilistic assumptions). Using the respecified statistical model, the paper calls into question the literature evaluating the theoretical adequacy of current DSGE models, ignoring the fact that such evaluations are untrustworthy because they are based on statistically unreliable procedures.

Acknowledgements

My deep gratitude goes first to my committee chair Dr. Aris Spanos, who expertly guided me through my graduate education. He continually and convincingly conveyed a spirit of adventure in regard to research and scholarship. Without his guidance and persistent help this dissertation would not have been possible.

I would like to thank my committee members, Dr. Richard Ashley, Dr. Kwok Ping Tsang, and Dr. Eric Bahel. Their valuable and constructive suggestions enabled me to complete my dissertation.

My sincere thanks go to Dr. Jinill Kim at Korea University. His advice and comments helped my studies not to be beside the point. I am also grateful to Dr. Steve Trost, who was very supportive and kindly advised me when I taught economic courses. My appreciation extends to the staff members of our department: Amy standford, Will Bebout, and Lynn Rader for their assistance during my years in the department.

Similar, profound gratitude goes to my Virginia Tech friends: Dongwoo Lee, Bowen Shi, Md. Yasin Arafat, and Karo Solat. The discussions, coffee and tea times, and general help and friendship were all greatly appreciated.

Last but not least, I would like to thank my parents, Sewon Kim and Yanghee Park, who have taught me the important things in life and have supported all my endeavors. My parents and my sister Yuhee Kim have given me great affection and support while patiently enduring my long absence from home.

Contents

List of Figures

List of Tables

Chapter 1

DSGE Modeling

1.1 Introduction

In macroeconomics, Dynamic Stochastic General Equilibrium (DSGE) models are the main tool for both academia and central banking circles for the last three decades because they have been influential in formulating not only microfoundations for macroeconomics from a theoretical perspective, but also offer useful tools for policy analysis from an empirical perspective.

Theoretically, DSGE modeling has been considered as an alternative to the criticism of the "Lucas critique" (see Lucas (1976)), where the economic relations that can be influenced by government policies cannot be structural. To be more concrete, the DSGE relationships consisting of economic equations pertain to more fundamental behavioral relations such as optimal decision rules built on the expectations of economic agents; for example, households and firms. Since any change in government policy will affect the expectations and the optimal decision rules underlying such relationships, will render such equation invariant with respect to any policy changes. In other words, as described by Wickens (2012) and Romer (2012), because the optimal decision rules are contingent on the state of the economy, if the state changes, it alters expectations that leads to the changes of decision rules, and those changes will be accommodated with such economic relations.

In response to the criticism that made a turning point in macroeconomic literature, the macroeconomic models incorporated forward-looking specifications of the structural relations, and assumed rational expectations in analyzing alternative policies (Rotemberg and Woodford, 1997). For instance, Kydland and Prescott (1982) where is generally considered to provide the beginning of DSGE modeling described a real business cycle (RBC) model using microeconomic foundations with exogenous stochastic component such as a technology shock. McCallum (1988, 1999), Leeper and Sims (1994), and Levin (1996) satisfied the Lucas critique with the structural models including monetary policy. According to Herbst and Schorfheide (2015), the term DSGE model covers a broad class of macroeconomic models from the RBC models of Kydland and Prescott (1982) and King et al. (1988) to the New Keynesian models of Rotemberg and Woodford (1997) or Christiano et al. (2005).

From an empirical point of view, DSGE modeling has a distinct advantage over alternative macroeconometric models such as vector autoregression (VAR) model. Since DSGE models are designed from microfoundations they allow to identify structural shocks in a theoretically consistent way, and thus enable them to provide an invariant basis for policy evaluations and forecasting analysis of the key macroeconomic indicators such as output and inflation. Although the early stage of DSGE modeling failed to reproduce stylized facts about macroeconomic data and drew poor forecasting results, Smets and Wouters (2003, 2007); Del Negro et al. (2007); Del Negro and Schorfheide (2004); Christiano et al. (2005) developed the estimation approaches to improve their empirical performances. Due to improved empirical analysis, DSGE modeling has received attention, not only from macroeconomists, but also from policy-making institutions such as central banks.

Although the developed empirical methods ameliorated the fit of DSGE models to the data, there still have been some controversial issues about DSGE modeling such as misspecification. Especially, the recent Great Recession in the United States decoupled an actual phenomenon from the empirical analysis resulted from a DSGE model. Since misspecification in DSGE model pertains to both theoretical and empirical analyses, ensuring the statistical and substantive adequacy becomes an important problem in DSGE modeling. Moreover, we distinguish between statistical misspecification and substantive misspecification that pertains only to the structural model. Before we discuss the various methods used to quantify DSGE models with a view to bring out their vulnerability to departures from the probabilistic assumptions imposed on the data; the implicit statistical model.

1.2 Quantitative Analyses in DSGE Models

1.2.1 Calibration Approach

At the early stage of the quantitative analysis of DSGE models an informal approach, calibration, was implemented without formal statistical methods. The calibration approach restricts parameters in an DSGE model to render the model consistent with long-run growth path and microeconomic observations.

In the followiong paragraphs, we borrow the description of calibration in DeJong and Dave (2011) to elucidate the approach. While the econometric framework provided by Haavelmo (1944) represents theoretical models as complete probability models for statistical analysis and inference such as estimation and testing, calibration uses a parameterized structural model to evaluate a specific quantitative questions such as empirical evaluation of fit of the model equations to economic data and theoretical assessments of economic policy experiments.

Kydland and Prescott (1982) where provide the foundations of the calibration approach conduct a calibration experiment, in place of probability approach concerning estimation and testing, for the quantitative analysis of business cycles. Although there was a practical reason for which formal statistical methods for empirical analysis of DSGE models had yet to be developed, Prescott (1986) describes that in favor of calibration have some connection with a criticism of the probability approach: since the models structured within theoretical framework are incontrovertibly highly abstract, they are necessarily false, and statistical hypothesis testing will reject them.

As mentioned earlier, the calibration approach has developed along the line of Kydland and Prescott (1982), and Kydland and Prescott (1991); Prescott (1986); Cooley and Prescott (1995) involve specific methodology for implementing calibration approaches in application to DSGE models.

1.2.2 Maximum Likelihood Approach

Following the development of calibration approach, econometric framework that formalizes different aspects of the calibration method has also advanced via various studies. In addition, the structural models have been ameliorated, and strong restrictions imposed on the first generation of DSGE models have been relaxed. Consequently, these improvements have enabled empirical studies to utilize established econometric techniques such as the method of Maximum Likelihood and the Bayesian approach to quantify such models. (An and Schorfheide, 2007).

The Maximum Likelihood (ML) approach assumes that the parameters are unknown constants and the data constitutes a truly typical realization of the implicit statistical model. These assumptions define the likelihood function. Given the data, Maximum Likelihood approach estimates the parameters that maximize the likelihood. Based on some regular conditions, the ML-estimator is consistent and asymptotically efficient.

In the process of empirical analysis using ML-estimator in DSGE models, many studies generally linearize the solution of DSGE models due to computational difficulties, and utilize the Kalman filtering algorithm to derive the likelihood function when there are unobservable variables. Ireland (2004) is a representative example of empirical analysis of DSGE models via Maximum Likelihood estimation.

Since the ML estimation utilizes all the statistical information there would be no distortion caused by lack of information, and we are able to identify problems with statistical misspecification by testing the probabilistic assumptions defining the Likelihood function.

However, the ML approach encounters some difficulties when applied to DSGE models. For instance, Fernández-Villaverde (2010) points out that the likelihood of DSGE models is a highly dimensional object, with a dozen or so parameters in the simplest models to almost a hundred in some of complicated model. Since the highly dimensional likelihoods of DSGE models are full of local peaks and cliffs and of nearly flat surfaces, any numerical search algorithms to maximize the likelihood easily get stuck in one of the local maximums and thus mislead the empirical analysis.

1.2.3 Bayesian Approach

Bayesian approach has become the main tool for the empirical analysis of DSGE models ever since Smets and Wouters (2003, 2007) showed that the Bayesian approach was competitive for the quantitative analysis. In addition, a Bayesian framework can address some challenges such as potential model misspecification and identification problems. For instance, using a posterior odds ratio which is Bayesian information for model comparison, this approach enables us to compare the DSGE model with non-nested models such as a VAR model (Watanabe, 2007).

Distinctively, Bayesian approach combines likelihood information with prior information for the empirical analysis, and results in distributional information about the parameters of interest. While the likelihood information involves the one from observations, prior information involves subjective information. Because of the structure of Bayesian approach, the choice of the prior distribution can affect the Bayesian analysis.

As an extension of Bayesian approach, Del Negro and Schorfheide (2004) proposed DSGE-VAR model for evaluating DSGE models. Interestingly, this model utilizes artificial information derived from a DSGE model to form the prior. By changing the number of artificial data generated from the DSGE model and comparing the fit to the actual observations, we can evaluate the DSGE model.

1.3 Bayesian vs. Classical Inference

Bayesian approach is remarkably distinguished from classical (or "frequentist") approach. In this section, we compare and contrast Bayesian inference to the frequentist inference.

In Bayesian analysis, probabilistic statements pertaining to stochastic events are constructed before applying empirical evidence of data, and the ex-ante probabilities are called prior. Moreover, the Bayesian approach to statistical inference accommodates the degree of belief (subjective or rational) interpretation of probability. In contrast, classical approach solely rely on observations.

Both Bayesian and classical framework focus on learning about the parameters of interest. However, unlike classical inference, the mindset of Bayesian inference is different in two crucial aspects:

1) In Bayesian framework, the parameters of interest, say $\boldsymbol{\theta}$, are considered as random variables with their own prior distribution which represents the one's subjective belief how likely the various of values of $\boldsymbol{\theta}$ in the parameter space. On the other hand, classical framework considers $\boldsymbol{\theta}$ as unknown constants without any a priori information.

2) While classical framework views the distribution of the sample are joint with the parameters, Bayesian framework views the distribution of the sample as conditional on $\boldsymbol{\theta}$. In the view of statistical notation, Bayesian framework denote it as $f(\mathbf{x}|\boldsymbol{\theta})$ instead of $f(\mathbf{x};\boldsymbol{\theta})$.

In addition, classical approach collects a sample of observations from the underlying population and estimates the parameters of interest. Then, uncertainty about the accuracy of estimates which is represented by standard error comes from the notion that the analysis is derived from a sample, not for the entire population. Unlike classical notion of sampling uncertainty, however, Bayesian analysis does not have such an uncertainty.

Bayesian approach incorporates prior and likelihood via Bayes' Rule to derive the posterior distribution of parameters of interest. It can be expressed as follows:

$$p(\boldsymbol{\theta}|\mathbf{y},\mathbf{X}) = \frac{p(\boldsymbol{\theta},\mathbf{y}|\mathbf{X})}{p(\mathbf{y}|\mathbf{X})} = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta},\mathbf{X})}{p(\mathbf{y}|\mathbf{X})} \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta},\mathbf{X})$$

Based on the expression above, we can interpret that the posterior distribution is induced by updating prior. Hence, if the data or empirical evidence includes distinctive information from prior which implies there is substantial learning about $\boldsymbol{\theta}$, the posterior will be very different from the prior.

To sum up, classical analysis begins with lack of information about the parameters of interest, and utilize the observations to estimate the parameters. Bayesian analysis begins with prior regarding to the parameters, and utilize the data to update the prior or to derive posterior.

1.4 Statistical Misspecification in Bayesian Approach

Although Bayesian and classical framework have different minds about statistical inference, methodologically, they have a common factor; likelihood function. As mentioned earlier, classical approach solely depends on the likelihood function for its statistical analysis, and Bayesian approach utilize the likelihood function to update the prior. Hence, probabilistic assumptions imposed on the likelihood function will affect to both classical and Bayesian inferences.

Despite the probabilistic assumptions are critical for the statistical inference, its importance and misspecification test have not been emphasized. Moreover, many studies applying Bayesian approach for the empirical analysis focus on finding proper prior, instead of investigating the likelihood function, to improve the empirical performance.

However, those alternatives cannot address the fundamental issues unless the probabilistic assumption are appropriate. For the Bayesian analysis, even if changing the prior gives better empirical performance, justifying the change seems to be dubious because ex post result is not enough to explain the change of ex ante informational content.

Although empirical analyses imply learning from data, it is not difficult to find studies that the implication is ignored. In addition, if there is conflict between the model and empirical result, suspicious questions converge to the theoretical model rather than the specification of econometric model with probabilistic assumptions.

Unlike conventional approaches for misspecification, our studies pay attention to statistical adequacy of econometric models by separating substantive information and statistical information.

1.5 A Brief Overview

In the following chapters, we distinguish between substantive information and statistical information, and then investigate DSGE models in statistical perspective. In chapter 2, the influential work, Smets and Wouters (2007), is describe in the view of substantive information. In chapter 3, we revisit the statistical information of Smets and Wouters (2007) and implement misspecification test to verify its statistical adequacy. In chapter 4, another method to evaluate DSGE models using Bayesian analysis is reviewed in statistical perspective.

Chapter 2

Substantive Information of Smets and Wouters (2007)

2.1 Description of SW (2007) Model

In this chapter, before we begin to explain the SW (2007) model, I divide the procedure deriving the linearized equations in 3 steps as the appendix of SW (2007);

1) Solving the decision problem for each agent,

2) Detrending and transforming the solution condition, and

3) Linearization.

In this model, there are several economic agents such as households, labor union, labor packers, final good producers, intermediate good producers, and government, so we need to see how they are inter-related with one another as solving this model.

For more details on solving the optimization problems see the appendix.

2.1.1 Solving Optimization Problems

Final Goods Producers

The final good Y_t is a composite consisting of a continuum of intermediate good $Y_t(i)$. Hence, final good producers purchase intermediate goods on the market, package Y_t, and then resell the final good to consumers, investors and the government in a perfectly competitive market.

Following is final good producers' profit maximization problem.

$$\max_{Y_t, Y_t(i)} P_t Y_t - \int_0^1 P_t(i) Y_t(i) di \quad \text{s.t} \quad \left[\int_0^1 G\left(\frac{Y_t(i)}{Y_t}; \epsilon_t^p\right) di \right] = 1$$

where P_t and $P_t(i)$ are the price of the final and intermediate goods respectively. We assume that G is a strictly concave and increasing function characterized by $G(1) = 1$. ϵ_t^p is an exogenous process that represents shocks to the aggregator function, and they cause changes in the elasticity of demand and therefore in the mark-up. SW (2007) assume that the exogenous process is constrained by $\epsilon_t^p \in (0, \infty)$.

Using the Lagrange equation for this problem,

$$L = P_t Y_t - \int_0^1 P_t(i) Y_t(i) di + \mu_{f,t} \left(\int_0^1 G\left(\frac{Y_t(i)}{Y_t}\right) di - 1 \right)$$

We have,

$$\frac{Y_t}{\mu_{f,t}} = \frac{\int_0^1 G'\left(\frac{Y_t(i)}{Y_t}\right) \frac{Y_t(i)}{Y_t} di}{P_t}$$

Thus, put this equation to the previous one, we can get

$$Y_t(i) = Y_t G'^{-1} \left[\frac{P_t(i)}{P_t} \int_0^1 G'\left(\frac{Y_t(i)}{Y_t}\right) \frac{Y_t(i)}{Y_t} di \right]$$

Then the assumptions imposed on the function G imply that the demand for input $Y_t(i)$ is decreasing in its relative price, but the elasticity of demand is a positive function of the relative price.

Intermediate Goods Producers

An intermediate good producer i produces his (her) goods using following technology:

$$Y_t(i) = \varepsilon_t^a \overline{K}_t(i)^\alpha [\gamma^t L_t(i)]^{1-\alpha} - \gamma^t \Phi$$

where $\overline{K}_t(i)$ is capital services used in production, $L_t(i)$ is aggregate labor input, Φ is a fixed cost, and γ^t represents the labor-augmenting deterministic growth rate in the economy.

From the firm's profit function:

$$P_t(i)Y_t(i) - W_t L_t(i) - R_t^k \overline{K}_t(i)$$

and it's cost minimization problem will be

$$\min_{L_t(i), \overline{K}_t(i)} W_t L_t(i) + R_t^k \overline{K}_t(i) \quad \text{s.t} \quad Y_t(i) = \varepsilon_t^a \overline{K}_t(i)^\alpha [\gamma^t L_t(i)]^{1-\alpha} - \gamma^t \Phi$$

where W_t is the aggregate nominal wage rate and R_t^k is the rental rate on capital. After deriving the F.O.Cs and combine these two F.O.Cs can be represented that

$$\frac{W_t}{R_t^k} = \frac{\Theta_t(i)\gamma^{(1-\alpha)t}(1-\alpha)\varepsilon_t^a \overline{K}_t(i)^\alpha L_t(i)^{-\alpha}}{\Theta_t(i)\gamma^{(1-\alpha)t}\alpha\varepsilon_t^a \overline{K}_t(i)^{\alpha-1} L_t^{1-\alpha}} = \frac{1-\alpha}{\alpha}\frac{\overline{K}_t(i)}{L_t(i)}$$

Noting that the capital-labor ratio is equal across firms, so we can ignore i, then

$$\frac{W_t}{R_t^k} = \frac{1-\alpha}{\alpha}\frac{\overline{K}_t}{L_t} \quad \Rightarrow \quad \overline{K}_t = \frac{\alpha}{1-\alpha}\frac{W_t}{R_t^k} L_t$$

Since the Lagrangian multiplier $\Theta_t(i)$ is shadow price which means how much the profit or cost will be changed by relaxing the constraint so it has identical implication to marginal cost, $\Theta_t(i)$ equals marginal cost MC_t. Thus, we can derive the MC_t by deriving $\Theta_t(i)$ and the marginal cost is also the same for all firms:

$$\Theta_t = MC_t = \alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)} W_t^{(1-\alpha)} (R_t^k)^\alpha \gamma^{-(1-\alpha)t} (\varepsilon_t^a)^{-1} \tag{2.1}$$

These intermediate goods producers set their prices under the Calvo pricing process with partial indexation. In other words, each firm can set its price with probability $1-\xi_t$ at time t, so if a firm can have chance to set the price, they set the price considering that they would not be able to have chance to optimize the price in the future. Even though some firms cannot set the price at time t with probability ξ_t, they can partially index the price to past inflation. Thus, its optimization problem will be:

$$\max_{\widetilde{P}_t(i)} E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \left[\widetilde{P}_t(i) \left(\Pi_{l=1}^s \pi_{t+l-1}^{l_p} \pi_*^{1-l_p} \right) - MC_{t+s} \right] Y_{t+s}(i)$$

$$\text{s.t} \quad Y_{t+s}(i) = Y_{t+s} G'^{-1} \left(\frac{\widetilde{P}_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s} \right)$$

where

$$\pi_t = \frac{P_t}{P_{t-1}}, \quad \tau_t = \int_0^1 G' \left(\frac{Y_t(i)}{Y_t} \right) \frac{Y_t(i)}{Y_t} di$$

$\widetilde{P}_t(i)$ is the newly set price, ξ_p is the Calvo probability of being allowed to optimize a firm's price, l_p is the degree of indexation to lagged inflation, $\left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \right]$ is the nominal discount factor for firms.

It can be the unconstrained optimization problem:

$$\max_{\widetilde{P}_t(i)} E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \left[\widetilde{P}_t(i) \left(\Pi_{l=1}^s \pi_{t+l-1}^{l_p} \pi_*^{1-l_p} \right) - MC_{t+s} \right] Y_{t+s} G'^{-1} \left(\frac{P_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s} \right)$$

By defining

$$X_{t,s} = \left\{ \begin{array}{l} 1 \quad \text{for} \quad s = 0 \\ (\Pi_{l=1}^s \pi_{t+l-1}^{l_p} \pi_*^{1-l_p}) \quad \text{for} \quad s = 1, ..., \infty \end{array} \right\}$$

we can rewrite the equation as

$$\max_{\widetilde{P}_t(i)} E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \left[\widetilde{P}_t(i) X_{t,s} - MC_{t+s} \right] Y_{t+s} G'^{-1} \left(\frac{P_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s} \right)$$

Differentiating with respect to $\widetilde{P}_t(i)$ give the optimal price for this problem, i.e;

$$E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} Y_{t+s}(i) \left[\widetilde{P}_t(i) X_{t,s} + \left(\widetilde{P}_t(i) X_{t,s} - MC_{t+s} \right) \frac{1}{G'^{-1}(z_{t+s})} \frac{G'(x_{t+s})}{G''(x_{t+s})} \right] = 0 \tag{2.2}$$

Households

Household j controls consumption $C_t(j)$, hours worked $L_t(j)$, bonds $B_t(j)$, investment $I_t(j)$ and capital utilization $Z_t(j)$ to maximize the its objective function:

$$E_t \sum_{s=0}^{\infty} \beta^s \left[\frac{1}{1-\sigma_c} (C_{t+s}(j) - \lambda C_{t+s-1})^{1-\sigma_c} \right] \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_{t+s}(j)^{1+\sigma_l} \right)$$

subject to the following budget constraint:

$$C_{t+s}(j) + I_{t+s}(j) + \frac{B_{t+s}(j)}{\varepsilon_t^b R_{t+s} P_{t+s}} - T_{t+s} \leq$$

$$\frac{B_{t+s-1}(j)}{P_{t+s}} + \frac{W_{t+s}^h(j) L_{t+s}(j)}{P_{t+s}} + \frac{R_{t+s}^k Z_{t+s}(j) K_{t+s-1}(j)}{P_{t+s}} - a(Z_{t+s}(j)) K_{t+s-1}(j) + \frac{Div_{t+s}}{P_{t+s}}$$

and the capital accumulation equation:

$$K_t(j) = (1-\delta) K_{t-1}(j) + \varepsilon_t^i \left[1 - S\left(\frac{I_t(j)}{I_{t-1}(j)} \right) \right] I_t(j)$$

where λ is the parameter capturing external habit formation. ε_t^b can be interpreted in an exogenous premium in the return to bonds reflecting inefficiencies in the financial market sector leading to some premium on deposit rate versus the risk free rate set by the central bank, or a risk premium that households require to hold the one period bond. δ is the depreciation rate, $S(\cdot)$ is the adjustment cost function, with $S(\gamma) = 0,\ S'(\gamma) = 0,\ S''(\cdot) > 0$, and ε_t^q is a stochastic shock to the price of investment relative to consumption goods. T_{t+s} is lump sum taxes or subsidies and Div_{t+s} is the dividends distributed by the intermediate goods producers and the labor unions.

In addition, households can choose the utilization rate of capital. The effective capital amount that households can rent to the firm is:

$$\overline{K}_t(j) = Z_t(j) K_{t-1}(j)$$

Then the income from renting capital services is $R_t^k Z_t(j) K_{t-1}(j)$ and the cost of changing capital utilization is $P_t a(Z_t(j)) K_{t-1}(j)$.

The Lagrange function for the households problem can be expressed by followings:

$$L = E_t \sum_{s=0}^{\infty} \beta^s \left\langle \left[\frac{1}{1-\sigma_c} (C_{t+s}(j) - \lambda C_{t+s-1})^{1-\sigma_c} \right] \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_{t+s}(j)^{1+\sigma_l} \right) \right.$$

$$+\Xi_t \left\{ \frac{B_{t+s-1}(j)}{P_{t+s}} + \frac{W^h_{t+s}(j)L_{t+s}(j)}{P_{t+s}} + \frac{R^k_{t+s}Z_{t+s}(j)K_{t+s-1}(j)}{P_{t+s}} - a(Z_{t+s}(j))K_{t+s-1}(j) + \frac{Div_{t+s}}{P_{t+s}} \right.$$
$$\left. -C_{t+s}(j) - I_{t+s}(j) - \frac{B_{t+s}(j)}{\varepsilon^b_t R_{t+s}P_{t+s}} + T_{t+s} \right\}$$
$$+ \Xi^k_t \left\{ (1-\delta)K_{t-1}(j) + \varepsilon^i_t \left[1 - S\left(\frac{I_t(j)}{I_{t-1}(j)} \right) \right] I_t(j) - K_t(j) \right\} \Bigg\rangle$$

By differentiating with respect to 5 control variables, consumption $C_t(j)$, hours worked $L_t(j)$, bonds $B_t(j)$, investment $I_t(j)$, and capital utilization $Z_t(j)$ respectively and rearranging them, we can have following conditions.

$$(\partial C_t) \quad \Xi_t = \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{1+\sigma_l} \right) (C_t - \lambda C_{t-1})^{-\sigma_c} \tag{2.3}$$

$$(\partial L_t) \quad \left[\frac{1}{1-\sigma_c}(C_t - \lambda C_{t-1})^{1-\sigma_c} \right] \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{1+\sigma_l} \right) (\sigma_c - 1) L_t^{\sigma_l} = -\Xi_t \frac{W^h_t}{P_t} \tag{2.4}$$

$$(\partial B_t) \quad \Xi_t = \beta \varepsilon^b_t R_t E_t \left[\frac{\Xi_{t+1}}{\pi_{t+1}} \right] \tag{2.5}$$

$$\begin{aligned}(\partial I_t) \quad \Xi_t = \Xi^k_t \varepsilon^i_t &\left(1 - S\left(\frac{I_t}{I_{t-1}} \right) - S'\left(\frac{I_t}{I_{t-1}} \right) \frac{I_t}{I_{t-1}} \right) \\ &+ \beta E_t \left[\Xi^k_{t+1} \varepsilon^i_{t+1} S'\left(\frac{I_{t+1}}{I_t} \right) \left(\frac{I_{t+1}}{I_t} \right)^2 \right]\end{aligned} \tag{2.6}$$

$$(\partial \overline{K}_t) \quad \Xi^k_t = \beta E_t \left[\Xi_{t+1} \left(\frac{R^k_{t+1}}{P_{t+1}} Z_{t+1} - a(Z_{t+1}) \right) + \Xi^k_{t+1}(1-\delta) \right] \tag{2.7}$$

$$(\partial Z_t) \quad \frac{R^k_t}{P_t} = a'(Z_t) \tag{2.8}$$

where Ξ_t is the Lagrange multiplier associated with the budget constraint and Ξ^k_t is another Lagrange multiplier associated with capital accumulation constraint. Tobin's Q, $Q_t = \frac{\Xi^k_t}{\Xi_t}$, equals one in the absence of adjustment costs.

Intermediate Labor Union Sector

Similar to the intermediate goods producers' case, in the labor market, intermediate labor union which gathers homogeneous labor services from the households and differentiate the labor services, sets wages subject to a Calvo scheme and offers those labor services to labor packers. Labor used by the intermediate goods producers L_t is a composite made of those differentiated labor services $L_t(l)$. Labor packers buy the differentiated labor services, package L_t, and offer it to the intermediate goods producers.

For the labor packers, its maximization problem is:

$$\max_{L_t, L_t(i)} W_t L_t - \int_0^1 W_t(l) L_t(l) dl$$

$$\text{s.t.} \quad \left[\int_0^1 H\left(\frac{L_t(l)}{L_t}; \epsilon_t^w\right) dl\right] = 1$$

where W_t and $W_t(l)$ are the price of the composite and intermediate labor services respectively. As the assumptions imposed on the function G, SW (2007) assume that H is a strictly concave and increasing function characterized by $H(1) = 1$, and ϵ_t^w is an exogenous process reflecting shocks to the aggregator function that result in changes in the elasticity of demand and therefore in the mark up, and it is constrained by $\epsilon_t^w \in (0, \infty)$.

Using the Lagrange equation for this problem,

$$L = W_t L_t - \int_0^1 W_t(l) L_t(l) dl + \mu_{p,t}\left(\int_0^1 H\left(\frac{L_t(l)}{L_t}\right) dl - 1\right)$$

the F.O.Cs can be rearranged by follows:

$$(\partial L_t(l)) : \quad W_t(l) = \mu_{p,t} H'\left(\frac{L_t(l)}{L_t}\right)\frac{1}{L_t} \Rightarrow H'\left(\frac{L_t(l)}{L_t}\right) = \frac{1}{\mu_{p,t}} W_t(l) L_t$$

Then,

$$\frac{L_t(l)}{L_t} = H'^{-1}\left[\frac{1}{\mu_{p,t}} W_t(l) L_t\right] \Rightarrow L_t(l) = L_t H'^{-1}\left[\frac{1}{\mu_{p,t}} W_t(l) L_t\right]$$

Based on the first F.O.C, we can see that

$$\frac{L_t}{\mu_{p,t}} = \frac{\int_0^1 H'\left(\frac{L_t(l)}{L_t}\right)\frac{L_t(l)}{L_t} dl}{W_t}$$

Thus, put this equation to the previous one, we can get

$$L_t(l) = L_t H'^{-1}\left[\frac{W_t(l)}{W_t}\int_0^1 H'\left(\frac{L_t(l)}{L_t}\right)\frac{L_t(l)}{L_t} dl\right]$$

The labor union can set wages under Calvo pricing scheme with partial indexation as the case of intermediate good producers. In other words, they can optimize their wage with probability $1 - \xi_w$ in every period, if they cannot re-optimize their wage with probability ξ_w, then the wage, $W_t(l)$ will be increased at the deterministic growth rate γ and weighted average of the steady state inflation π_* and of last period's inflation, π_{t-1}. For those can optimize the wage, the problem is to choose a optimal wage $\widetilde{W}_t(l)$ that maximizes the wage income in all states of nature where the labor union is stuck with that wage in the future. Thus their optimization problem would be:

$$\max_{\widetilde{W}_t(l)} E_t \sum_{s=0}^{\infty} \zeta_w^s \left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}}\right] \left(W_{t+s}(l) - W_{t+s}^h\right) L_{t+s}(l)$$

subject to

$$L_{t+s}(l) = \left(\frac{W_{t+s}(l)}{W_{t+s}}\right)^{-\frac{1+\lambda_{w,t+s}}{\lambda_{w,t+s}}} L_{t+s}$$
$$W_{t+s}(l) = \widetilde{W}_t(l)(\Pi_{l=1}^s \gamma \pi_{t+l-1}^{l_w} \pi_*^{1-l_w}) \quad \text{for} \quad s = 1, ..., \infty$$

From the unconstrained problem, the problem will be

$$L = E_t \sum_{s=0}^{\infty} \zeta_w^s \left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}}\right] \left(\widetilde{W}_t(l)(\Pi_{l=1}^s \gamma \pi_{t+l-1}^{l_w} \pi_*^{1-l_w}) - W_{t+s}^h\right) \left(\frac{W_{t+s}(l)}{W_{t+s}}\right)^{-\frac{1+\lambda_{w,t+s}}{\lambda_{w,t+s}}} L_{t+s}$$

and using the definition

$$X_{t,s} = \left\{ \begin{array}{ll} 1 \quad \text{for} \quad s = 0 & \\ (\Pi_{l=1}^s \pi_{t+l-1}^{l_w} \pi_*^{1-l_w}) \quad \text{for} \quad s = 1, ..., \infty & \end{array} \right\}$$

it can be rewrite as follows:

$$L = E_t \sum_{s=0}^{\infty} \zeta_w^s \left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}}\right] \left(\widetilde{W}_t(l) X_{t,s} - W_{t+s}^h\right) \left(\frac{\widetilde{W}_t(l) X_{t,s}}{W_{t+s}}\right)^{-\frac{1+\lambda_{w,t+s}}{\lambda_{w,t+s}}} L_{t+s}$$

Then the F.O.C with respect to $\widetilde{W}_t(l)$ is

$$E_t \sum_{s=0}^{\infty} \zeta_w^s \left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}}\right] L_{t+s}(l) \frac{1}{\lambda_{w,t+s}} \left\{(1 + \lambda_{w,t+s}) W_{t+s}^h - \widetilde{W}_t(l) X_{t,s}\right\} = 0 \qquad (2.9)$$

The aggregate wage expression is

$$W_t = [(1-\zeta_w)\widetilde{W}_t^{\frac{1}{\lambda_{w,t}}} + \zeta_w(\gamma\pi_{t-1}^{l_w}\pi_*^{1-l_w}W_{t-1})^{\frac{1}{\lambda_{w,t}}}]^{\lambda_{w,t}}$$

Intuitively, similar to the intermediate goods producers' case, current wage, W_t, is a weighted average between optimized wage and previous wage. The mark up of the aggregate wage over the wage received by the households is distributed to the households in the form of dividends.

Government Policies

SW (2007) set that the central bank follows a nominal interest rate rule that adjusts its instrument to react the deviations of inflation and output from their respective target levels:

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_R}\left[\left(\frac{\pi_t}{\pi_*}\right)^{\psi_1}\left(\frac{Y_t}{Y_t^*}\right)^{\psi_2}\right]^{1-\rho_R}\left(\frac{Y_t/Y_{t-1}}{Y_t^*/Y_{t-1}^*}\right)^{\psi_3} r_t \tag{2.10}$$

where R^* is the steady state nominal gross rate and Y_t^* is the natural output. ρ_R determines the degree of interest rate smoothing, and r_t is the exogenous monetary policy shock. The government budget constraint is

$$P_tG_t + B_{t-1} = T_t + \frac{B_t}{R_t} \tag{2.11}$$

where T_t is nominal lump-sum taxes (or subsidies), and it is also included in the household's constraint. Government spending is exogenous and it is expressed relative to the steady state output path as $g_t = G_t/(Y\gamma^t)$.

Resource Constraints

To derive the market clearing condition for the final goods market, we need to integrate households budget constraint across households, and combine it with government budget constraint and the expression for the dividends of intermediate goods for producers and labor unions. Then, finally we can have

$$C_t + I_t + G_t + a(Z_t)K_{t-1} = Y_t \tag{2.12}$$

Exogenous Process

There are seven exogenous processes in this model:

Technology process:

$$lnZ_t = (1 - \rho_z)ln_Z + \rho_z lnZ_{t-1} + \epsilon_{z,t}$$

Investment relative price process:

$$ln\mu_t = (1 - \rho_{\mu_t})ln\mu + \rho_\mu ln\mu_{t-1} + \epsilon_{\mu,t}$$

Intertemporal preference shifter (financial risk premium process):

$$lnb_t = (1 - \rho_b)lnb + \rho_b lnb_{t-1} + \epsilon_{b,t}$$

Government spending process:

$$lng_t = (1 - \rho_g)lng + \rho_g lng_{t-1} + \rho_{ga}lnZ_t - \rho_{ga}lnZ_{t-1} + \epsilon$$

Monetary policy shock:

$$lnr_t = \rho_r lnr_{t-1} + \epsilon_{r,t}$$

Price mark-up shock:

$$ln\lambda_{p,t} = (1 - \rho_p)ln\lambda_p + \rho_p ln\lambda_{p,t-1} - \theta_p\epsilon_{p,t-1} + \epsilon_{p,t}$$

Wage mark-up shock:

$$ln\lambda_{w,t} = (1 - \rho_w)ln\lambda_w + \rho_w ln\lambda_{w,t-1} - \theta_w\epsilon_{w,t-1} + \epsilon_{w,t}$$

the innovations ϵ are distributed as Normal IID:

$$\epsilon_{i,t} \sim N(0, \sigma_i)$$

2.1.2 Detrending and Transforming

Since the model include constant growth rate, γ, we can derive the detrended model by dividing γ. In addition, we can replace the nominal variables in the model with real variables by dividing the price level, P_t. Then, de-trended variables are:

$$y_t = \frac{Y_t}{\gamma^t}, \quad c_t = \frac{C_t}{\gamma^t}, \quad k_t = \frac{K_t}{\gamma^t}, \quad w_t = \frac{W_t}{P_t\gamma^t}, \quad r_t^k = \frac{R_t^k}{P_t}, \quad \xi_t = \Xi_t\gamma^{\sigma_c t}, \quad \widetilde{p}_t = \frac{\widetilde{P}_t}{P_t}$$

$$\overline{\beta}^s = \frac{\beta^s}{\gamma^{\sigma_c s}}, \quad i_t = \frac{I_t}{\gamma^t}, \quad \widetilde{w}_t = \frac{\widetilde{W}_t}{P_t\gamma^t}$$

Intermediate Goods Producers

For intermediate goods producers, the aggregate production function becomes

$$Y_t(i) = \varepsilon_t^a \overline{K}_t(i)^\alpha [\gamma^t L_t(i)]^{1-\alpha} - \gamma^t\Phi \Rightarrow \frac{Y_t(i)}{\gamma^t} = \varepsilon_t^a \left(\frac{\overline{K}_t(i)}{\gamma^t}\right)^\alpha L_t(i)^{1-\alpha} - \Phi$$

$$\therefore y_t(i) = \varepsilon_t^a k_t(i)^\alpha (L_t(i))^{1-\alpha} - \Phi$$

and

$$\overline{K}_t = \frac{\alpha}{1-\alpha}\frac{W_t}{R_t^k}L_t \Rightarrow \frac{\overline{K}_t}{\gamma^t} = \frac{\alpha}{1-\alpha}\frac{\frac{W_t}{\gamma^t P_t}}{\frac{R_t^k}{P_t}}L_t$$

$$\therefore \overline{k}_t = \frac{\alpha}{1-\alpha}\frac{w_t}{r_t^k}L_t$$

The marginal cost becomes:

$$MC_t = \alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)}W_t^{(1-\alpha)}(R_t^k)^\alpha \gamma^{-(1-\alpha)t}(\varepsilon_t^a)^{-1}$$

$$\therefore mc_t = \frac{w_t^{1-\alpha}(r_t^k)^\alpha}{\alpha^\alpha(1-\alpha)^{1-\alpha}\varepsilon_t^a}$$

and the F.O.C,

$$E_t\sum_{s=0}^{\infty}\zeta_p^s\frac{\beta^s\Xi_{t+s}P_t}{\Xi_t P_{t+s}}Y_{t+s}(i)\left[\widetilde{P}_t(i)X_{t,s} + \left(\widetilde{P}_t(i)X_{t,s} - MC_{t+s}\right)\frac{1}{G'^{-1}\left(z_{t+s}\right)}\frac{G'\left(x_{t+s}\right)}{G''\left(x_{t+s}\right)}\right] = 0$$

becomes

$$E_t\sum_{s=0}^{\infty}\zeta_p^s\overline{\beta}^s\frac{\xi_{t+s}}{\xi_t}\gamma^s y_{t+s}(i)$$

$$\times \left[\left(1 + \frac{1}{G'^{-1}\left(z_{t+s}\right)} \frac{G'\left(x_{t+s}\right)}{G''\left(x_{t+s}\right)} \right) \frac{\widetilde{p}_t(i) X_{t,s}}{X_{t+s}^p} - \left(\frac{1}{G'^{-1}\left(z_{t+s}\right)} \frac{G'\left(x_{t+s}\right)}{G''\left(x_{t+s}\right)} \right) mc_{t+s} \right] = 0$$

where

$$X_{t+s}^p = 1 \text{ for } s = 0 \text{ or else } \Pi_{l=1}^s \pi_{t+l}$$

Based on the previous de-trended equations, aggregate profit equation,

$$\Pi_t = P_t Y_t - W_t L_t - R_t^k K_t$$

equals to:

$$\frac{\Pi_t}{P_t \gamma^t} = \frac{P_t Y_t}{P_t \gamma^t} - \frac{W_t L_t}{P_t \gamma^t} - \frac{R_t^k K_t}{P_t \gamma^t} = y_t - w_t L_t - r_t^k k_t$$

$$\Rightarrow \frac{\Pi_t}{P_t \gamma^t} = \left(\frac{1}{mc_t} - 1 \right) \frac{w_t L_t}{1 - \alpha} - \Phi \quad (\because mc_t = \frac{w_t^{1-\alpha} (r_t^k)^\alpha}{\alpha^\alpha (1-\alpha)^{1-\alpha} \varepsilon_t^a})$$

Households

F.O.C with respect to C_t becomes

$$\xi_t = \exp\left(\frac{\sigma_c - 1}{1 + \sigma_l} L_t^{1+\sigma_l} \right) (c_t - \frac{\lambda}{\gamma} c_{t-1})^{-\sigma_c} \tag{2.13}$$

F.O.C with respect to L_t becomes

$$w_t^h = \left(c_t - \left(\frac{\lambda}{\gamma} \right) c_{t-1} \right) L_t^{\sigma_c} \tag{2.14}$$

F.O.C with respect to B_t becomes

$$\xi_t = \overline{\beta} \varepsilon_t^b R_t E_t \left[\frac{\xi_{t+1}}{\pi_{t+1}} \right] \tag{2.15}$$

F.O.C with respect to I_t becomes, $(Q_t = \dfrac{\Xi_t^k}{\Xi_t})$

$$1 = Q_t \varepsilon_t^i \left(1 - S\left(\frac{i_t \gamma}{i_{t-1}} \right) - S'\left(\frac{i_t \gamma}{i_{t-1}} \right) \frac{i_t}{i_{t-1}} \right) + \overline{\beta} E_t \left[\frac{\xi_{t+1}}{\xi_t} \left\{ Q_{t+1} \varepsilon_{t+1}^i S'\left(\frac{i_{t+1}\gamma}{i_t} \right) \left(\frac{i_{t+1}\gamma}{i_t} \right)^2 \right\} \right] \tag{2.16}$$

F.O.C with respect to $\overline{K}_t$ becomes

$$Q_t = \overline{\beta} E_t \left[\frac{\xi_{t+1}}{\xi_t} \left\{ \left(r^k_{t+1} Z_{t+1} - a(Z_{t+1}) \right) + Q_{t+1}(1-\delta) \right\} \right] \tag{2.17}$$

F.O.C with respect to Z_t becomes

$$\frac{R^k_t}{P_t} = a'\left(Z_t\right) \quad \Rightarrow \quad r^k_t = a'(Z_t) \tag{2.18}$$

The F.O.C,

$$E_t \sum_{s=0}^{\infty} \zeta^s_w \left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \right] L_{t+s}(l) \frac{1}{\lambda_{w,t+s}} \left\{ (1+\lambda_{w,t+s}) W^h_{t+s} - \widetilde{W}_t(l) X_{t,s} \right\} = 0$$

becomes

$$E_t \sum_{s=0}^{\infty} \zeta^s_w \overline{\beta}^s \gamma^s \left[\frac{\xi_{t+s}}{\xi_t} \right] L_{t+s}(l) \frac{1}{\lambda_{w,t+s}} \left[(1+\lambda_{w,t+s}) w^h_{t+s} - \frac{(\Pi^s_{l=0} \pi^{l_w}_{t+l-1} \pi^{1-l_w}_*)}{(\Pi^s_{l=1} \pi_{t+l})} \widetilde{w}_t(l) \right] = 0 \tag{2.19}$$

In addition, the equation,

$$\overline{K}_t(j) = Z_t(j) K_{t-1}(j)$$

becomes

$$\overline{k}_t(j) = Z_t(j) \frac{k_{t-1}(j)}{\gamma}$$

Resource Constraint

The resource constraint,

$$C_t + I_t + G_t + a(Z_t)\overline{K}_{t-1} = Y_t$$

becomes

$$\therefore \quad c_t + i_t + Y_* g_t + a(Z_t) \frac{\overline{k}_{t-1}}{\gamma} = y_t \quad \left(\because g_t = \frac{G_t}{Y_* \gamma^t} \right) \tag{2.20}$$

Government Policies

The Taylor rule,

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*} \right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi_*} \right)^{\psi_1} \left(\frac{Y_t}{Y^*_t} \right)^{\psi_2} \right]^{1-\rho_R} \left(\frac{Y_t / Y_{t-1}}{Y^*_t / Y^*_{t-1}} \right)^{\psi_3} r_t$$

becomes

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi_*}\right)^{\psi_1} \left(\frac{y_t}{y_t^*}\right)^{\psi_2}\right]^{1-\rho_R} \left(\frac{y_t/y_{t-1}}{y_t^*/y_{t-1}^*}\right)^{\psi_3} r_t \tag{2.21}$$

2.1.3 Log-linearization

In this section, I just want to introduce to do log-linearization using several methods. For the log-linearization of the equation that is de-trended of F.O.C of household,

$$\xi_t = \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{1+\sigma_l}\right)\left(c_t - \frac{\lambda}{\gamma} c_{t-1}\right)^{-\sigma_c}$$

I take logarithm of the equation first such as

$$ln\xi_t = \left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{1+\sigma_l}\right) - \sigma_c ln\left(c_t - \frac{\lambda}{\gamma} c_{t-1}\right)$$

and then use the Taylor first order approximation to linearize it. For the left hand side,

$$ln\xi_t \approx ln\xi_* + \frac{1}{\xi_*}(\xi_t - \xi_*)$$

For the right hand side,

$$\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{1+\sigma_l}\right) \approx \left\{\frac{\sigma_c - 1}{1+\sigma_l}\left(L_*^{1+\sigma_l} + (1+\sigma_l)L_*^{\sigma_l}(L_t - L_*)\right)\right\}$$

$$-\sigma_c ln\left(c_t - \frac{\lambda}{\gamma} c_{t-1}\right) \approx -\sigma_c\left[ln\left(1 - \frac{\lambda}{\gamma}\right)c_* + \frac{1}{\left(1-\frac{\lambda}{\gamma}\right)c_*}(c_t - c_*) - \frac{\frac{\lambda}{\gamma}}{\left(1-\frac{\lambda}{\gamma}\right)c_*}(c_{t-1} - c_*)\right]$$

Thus,

$$ln\xi_* + \frac{1}{\xi_*}(\xi_t - \xi_*) \approx \left\{\frac{\sigma_c - 1}{1+\sigma_l}\left(L_*^{1+\sigma_l} + (1+\sigma_l)L_*^{\sigma_l}(L_t - L_*)\right)\right\}$$

$$-\sigma_c\left[ln\left(1 - \frac{\lambda}{\gamma}\right)c_* + \frac{1}{\left(1-\frac{\lambda}{\gamma}\right)c_*}(c_t - c_*) - \frac{\frac{\lambda}{\gamma}}{\left(1-\frac{\lambda}{\gamma}\right)c_*}(c_{t-1} - c_*)\right]$$

since

$$ln\xi_* = \left(\frac{\sigma_c - 1}{1+\sigma_l} L_*^{1+\sigma_l}\right) - \sigma_c ln\left(1 - \frac{\lambda}{\gamma}\right)c_*$$

in steady state, so we can get rid of this term in the equation. Therefore,

$$\frac{1}{\xi_*}(\xi_t - \xi_*) \approx (\sigma_c - 1)L_*^{\sigma_l}(L_t - L_*) - \sigma_c \left[\frac{1}{\left(1 - \frac{\lambda}{\gamma}\right) c_*}(c_t - c_*) - \frac{\frac{\lambda}{\gamma}}{\left(1 - \frac{\lambda}{\gamma}\right) c_*}(c_{t-1} - c_*) \right]$$

using the approximation,

$$\widehat{x}_t = log(X_t) - log(X_*) = log(X_t/X_*) \approx \frac{X_t - X_*}{X_*}$$

we can re-write the equation above such as

$$\widehat{\xi}_t \approx (\sigma_c - 1)L_*^{1+\sigma_l}\widehat{L}_t - \sigma_c \left[\left(\frac{\gamma}{\gamma - \lambda}\right)\widehat{c}_t - \left(\frac{\lambda}{\gamma - \lambda}\right)\widehat{c}_{t-1} \right]$$

For another de-trended F.O.C,

$$\xi_t = \overline{\beta}\varepsilon_t^b R_t E_t \left[\frac{\xi_{t+1}}{\pi_{t+1}}\right]$$

I use different way for linear approximation because it includes expectation term. Using the method,

$$lnx_t = lnx_* + \widehat{x}_t \quad \Rightarrow \quad x_t = e^{lnx_* + \widehat{x}_t} = e^{lnx_*} \times e^{\widehat{x}_t} \quad \Rightarrow \quad x_t = x_* e^{\widehat{x}_t}$$

we can re-write the equation such as

$$\xi_* e^{\widehat{\xi}_t} = \overline{\beta}\varepsilon_*^b e^{\widehat{\varepsilon}_t^b} R_* e^{\widehat{R}_t} E_t \left[\frac{\xi_* e^{\widehat{\xi}_{t+1}}}{\pi_* e^{\widehat{\pi}_{t+1}}}\right]$$

and since, in steady state,

$$\xi_* = \overline{\beta}\varepsilon_*^b R_* E_* \left[\frac{\xi_*}{\pi_*}\right]$$

we can remove the term in both side, then

$$e^{\widehat{\xi}_t} = e^{\widehat{\varepsilon}_t^b} e^{\widehat{R}_t} E_t \left[\frac{e^{\widehat{\xi}_{t+1}}}{e^{\widehat{\pi}_{t+1}}}\right]$$

We have not used the approximation yet so far. Approximating the expression $e^{\widetilde{x}_t}$ with a first-order Taylor expansion at the point $\widetilde{x}_t = 0$ yields

$$e^{\widehat{x}_t} \approx 1 + e^0(\widehat{x}_t - 0) = 1 + \widehat{x}_t$$

therefore the equation can be approximated by

$$1 + \widetilde{\xi}_t \approx (1 + \widehat{\varepsilon}_t^b)(1 + \widehat{R}_t)E_t \left\{(1 + \widehat{\xi}_{t+1})(1 - \widehat{\pi}_{t+1})\right\}$$

Since the product of log-deviations from steady states are very small, we can ignore the terms. Then, we would have

$$\widetilde{\xi}_t \approx \widehat{\varepsilon}_t^b + \widehat{R}_t + E_t\widehat{\xi}_{t+1} - E_t\widehat{\pi}_{t+1}$$

Combining with previous approximation,

$$\widehat{\xi}_t \approx (\sigma_c - 1)L_*^{1+\sigma_l}\widehat{L}_t - \sigma_c \left[\left(\frac{\gamma}{\gamma - \lambda}\right)\widehat{c}_t - \left(\frac{\lambda}{\gamma - \lambda}\right)\widehat{c}_{t-1}\right]$$

we can have

$$\begin{aligned} &(\sigma_c - 1)L_*^{1+\sigma_l}\widehat{L}_t - \sigma_c \left[\left(\frac{\gamma}{\gamma - \lambda}\right)\widehat{c}_t - \left(\frac{\lambda}{\gamma - \lambda}\right)\widehat{c}_{t-1}\right] \\ &= (\sigma_c - 1)L_*^{1+\sigma_l}E_t\widehat{L}_{t+1} - \sigma_c \left[\left(\frac{\gamma}{\gamma - \lambda}\right)E_t\widehat{c}_{t+1} - \left(\frac{\lambda}{\gamma - \lambda}\right)\widehat{c}_t\right] + \widehat{\varepsilon}_t^b + \widehat{R}_t - E_t\widehat{\pi}_{t+1} \end{aligned}$$

by re-arranging it,

$$\begin{aligned} \widehat{c}_t &= \left(\frac{\gamma}{\gamma + \lambda}\right)E_t\widehat{c}_{t+1} + \left(\frac{\lambda}{\gamma + \lambda}\right)\widehat{c}_{t-1} - \frac{1}{\sigma_c}\left(\frac{\gamma - \lambda}{\gamma + \lambda}\right)(\widehat{R}_t - E_t\widehat{\pi}_{t+1}) - \frac{1}{\sigma_c}\left(\frac{\gamma - \lambda}{\gamma + \lambda}\right)\widehat{\varepsilon}_t^b \\ &\qquad - \left(\frac{\gamma - \lambda}{\gamma + \lambda}\right)\left(\frac{\sigma_c - 1}{\sigma_c}\right)L_*^{1+\sigma_c}(E_t\widehat{L}_{t+1} - \widehat{L}_t) \\ &= \left(\frac{1}{1 + \frac{\lambda}{\gamma}}\right)E_t\widehat{c}_{t+1} + \left(\frac{\frac{\lambda}{\gamma}}{1 + \frac{\lambda}{\gamma}}\right)\widehat{c}_{t-1} - \frac{1 - \frac{\lambda}{\gamma}}{\sigma_c\left(1 + \frac{\lambda}{\gamma}\right)}(\widehat{R}_t - E_t\widehat{\pi}_{t+1}) - \frac{1 - \frac{\lambda}{\gamma}}{\sigma_c\left(1 + \frac{\lambda}{\gamma}\right)}\widehat{\varepsilon}_t^b \\ &\qquad - \left(\frac{1 - \frac{\lambda}{\gamma}}{1 + \frac{\lambda}{\gamma}}\right)\left(\frac{\sigma_c - 1}{\sigma_c}\right)L_*^{1+\sigma_c}(E_t\widehat{L}_{t+1} - \widehat{L}_t) \end{aligned}$$

From the another F.O.C,

$$w_t^h = \left(c_t - \left(\frac{\lambda}{\gamma}\right)c_{t-1}\right)L_t^{\sigma_c} \quad \Rightarrow \quad w_*^h = \left(1 - \frac{\lambda}{\gamma}\right)c_*L_*^{\sigma_c} \quad \Rightarrow \quad \frac{w_*^h L_*}{c_*} = \left(1 - \frac{\lambda}{\gamma}\right)L_*^{1+\sigma_c}$$

By substituting the equation to the term in the previous one, we can have

$$\widehat{c}_t = \left(\frac{1}{1 + \frac{\lambda}{\gamma}}\right)E_t\widehat{c}_{t+1} + \left(\frac{\frac{\lambda}{\gamma}}{1 + \frac{\lambda}{\gamma}}\right)\widehat{c}_{t-1} - \frac{1 - \frac{\lambda}{\gamma}}{\sigma_c\left(1 + \frac{\lambda}{\gamma}\right)}(\widehat{R}_t - E_t\widehat{\pi}_{t+1}) - \frac{1 - \frac{\lambda}{\gamma}}{\sigma_c\left(1 + \frac{\lambda}{\gamma}\right)}\widehat{\varepsilon}_t^b$$

$$-\left(\frac{(\sigma_c-1)\frac{w_*^h L_*}{c_*}}{\sigma_c\left(1+\frac{\lambda}{\gamma}\right)}\right)(E_t\widehat{L}_{t+1}-\widehat{L}_t)$$

$$\therefore \quad \widehat{c}_t=\left(\frac{1}{1+\frac{\lambda}{\gamma}}\right)E_t\widehat{c}_{t+1}+\left(\frac{\frac{\lambda}{\gamma}}{1+\frac{\lambda}{\gamma}}\right)\widehat{c}_{t-1}-\frac{1-\frac{\lambda}{\gamma}}{\sigma_c\left(1+\frac{\lambda}{\gamma}\right)}(\widehat{R}_t-E_t\widehat{\pi}_{t+1})$$
$$-\frac{1-\frac{\lambda}{\gamma}}{\sigma_c\left(1+\frac{\lambda}{\gamma}\right)}\widehat{\varepsilon}_t^b-\left(\frac{(\sigma_c-1)\frac{w_*^h L_*}{c_*}}{\sigma_c\left(1+\frac{\lambda}{\gamma}\right)}\right)(E_t\widehat{L}_{t+1}-\widehat{L}_t)$$

Using those log-linearization methods, I can have following log-linearized equations.

$$\widehat{y}_t=c_y\widehat{c}_t+i_y\widehat{i}_t+z_y\widehat{z}_t+\varepsilon_t^g \tag{2.22}$$

$$\widehat{c}_t=\frac{h/\gamma}{1+h/\gamma}\widehat{c}_{t-1}+\frac{1}{1+h/\gamma}E_t\widehat{c}_{t+1}+\frac{wl_c(\sigma_c-1)}{\sigma_c(1+h/\gamma}(\widehat{l}_t-E_t\widehat{l}_{t+1})$$
$$-\frac{1-h/\gamma}{(1+h/\gamma)\sigma_c}(\widehat{r}_t-E_t\widehat{\pi}_{t+1})-\frac{1-h/\gamma}{(1+h/\gamma)\sigma_c}\varepsilon_t^b \tag{2.23}$$

$$\widehat{i}_t \;=\; \frac{1}{1+\beta\gamma^(1-\sigma_c)}\widehat{i}_{t-1}\;+\;\frac{\beta\gamma^(1-\sigma_c)}{1+\beta\gamma^(1-\sigma_c)}E_t\widehat{i}_{t+1}\;+\;\frac{1}{\varphi\gamma^2(1+\beta\gamma^(1-\sigma_c))}\widehat{q}_t\;+\;\varepsilon_t^i \tag{2.24}$$

$$\widehat{q}_t=\beta(1-\delta)\gamma^{-\sigma_c)}E_t\widehat{q}_{t+1}-\widehat{r}_t+E_t\widehat{\pi}_{t+1}+(1-\beta(1-\delta)\gamma^{-\sigma_c})E_t\widehat{r}_{t+1}^k-\varepsilon_t^b \tag{2.25}$$

$$\widehat{y}_t=\Phi(\alpha\widehat{k}_t^s+(1-\alpha)\widehat{l}_t+\varepsilon_t^\alpha) \tag{2.26}$$

$$\widehat{k}_t^s=\widehat{k}_{t-1}+\widehat{z}_t \tag{2.27}$$

$$\widehat{z}_t=\frac{1-\psi}{\psi}\widehat{r}_t^k \tag{2.28}$$

$$\widehat{k}_t=\frac{(1-\delta)}{\gamma}\widehat{k}_{t-1}+(1-(1-\delta)/\gamma)\widehat{i}_t+(1-(1-\delta)/\gamma)\varphi\gamma^2(1+\beta\gamma^{(1-\sigma_c)})\varepsilon_t^i \tag{2.29}$$

$$\widehat{\mu}_t^p=\alpha(\widehat{k}_t^s-\widehat{l}_t)-\widehat{w}_t+\varepsilon_t^a \tag{2.30}$$

$$\widehat{\pi}_t=\frac{\beta\gamma^{(1-\sigma_c)}}{1+\iota_p\beta\gamma^(1-\sigma_c)}E_t\widehat{\pi}_{t+1}+\frac{\iota_p}{1+\beta\gamma^{(1-\sigma_c)}}\widehat{\pi}_{t-1}$$
$$-\frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_p)(1-\xi_p)}{(1+\iota_p\beta\gamma^{(1-\sigma_c)})(1+(\Phi-1)\varepsilon_p)\xi_p}\widehat{\mu}_t^p+\varepsilon_t^p \tag{2.31}$$

$$\widehat{r}_t^k = \widehat{l}_t + \widehat{w}_t - \widehat{k}_t^s \tag{2.32}$$

$$\widehat{\mu}_t^w = \widehat{w}_t - \sigma_l \widehat{l}_t - \frac{1}{1-h/\gamma}(\widehat{c}_t - h/\gamma \widehat{c}_{t-1}) \tag{2.33}$$

$$\begin{aligned} &\frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}}(E_t\widehat{w}_{t+1} + E_t\widehat{\pi}_{t+1}) + \frac{1}{1+\beta\gamma(1-\sigma_c)}(\widehat{w}_{t-1} - \iota_w \widehat{\pi}_{t-1}) \\ &\qquad - \frac{1+\beta\gamma^{(1-\sigma_c)}\iota_w}{1+\beta\gamma^{(1-\sigma_c}}\widehat{\pi}_t - \frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_w)(1-\xi_w)}{(1+\beta\gamma^{(1-\sigma_c)})(1+(\lambda_w-1)\epsilon_w)\xi_w}\widehat{\mu}_t^w + \varepsilon_t^w \end{aligned} \tag{2.34}$$

$$\widehat{r}_t = \rho \widehat{r}_{t-1} + (1-\rho)(r_\pi \widehat{\pi}_t + r_y(\widehat{y}_t - \widehat{y}_t^*)) + r\Delta_y((\widehat{y}_t - \widehat{y}_t^*) - (\widehat{y}_{t-1} - \widehat{y}_{t-1})) + \varepsilon_t^r \tag{2.35}$$

In addition, the 7 exogenous shocks evolves as follows.

$$\begin{aligned} \varepsilon_t^a &= \rho_a \varepsilon_{t-1}^a + \eta_t^a \\ \varepsilon_t^b &= \rho_b \varepsilon_{t-1}^b + \eta_t^b \\ \varepsilon_t^g &= \rho_g \varepsilon_{t-1}^a + \rho_{ga}\eta_t^a + \eta_t^g \\ \varepsilon_t^i &= \rho_i \varepsilon_{t-1}^i + \eta_t^i \\ \varepsilon_t^r &= \rho_r \varepsilon_{t-1}^r + \eta_t^r \\ \varepsilon_t^p &= \rho_p \varepsilon_{t-1}^p + \eta_t^p - \mu_p \eta_{t-1}^p \\ \varepsilon_t^w &= \rho_w \varepsilon_{t-1}^w + \eta_t^w - \mu_w \eta_{t-1}^w \end{aligned}$$

Chapter 3

Smets and Wouters (2007) DSGE Model

3.1 Introduction

In empirical macroeconomics, Dynamic Stochastic General Equilibrium (DSGE) model has been widely used to account for and explain the movements of macroeconomic variables. Even though many studies used informal calibration methods for DSGE model with macroeconomic data in the initial stage, during the past two decades, development of the DSGE model and application of formal econometric methods have been made conspicuous progress.

We can distinguish the DSGE model as two groups in the view of econometric techniques such as moment-based estimation methods and likelihood-based inference. For the moment-based estimation methods like generalized methods of moments (GMM), the model is not needed to be fully solved for the estimation. For the likelihood-based estimation, however, it requires a full model solution and latent variables to be integrated out from the joint density of combination of observable and unobservable variables. To solve the DSGE model, many studies used linearization driven by Normal shocks. Even though this linearization approach has some advantage to compute and draw some inferences about the law of motion, this kinds of specifications do not seem to describe some economic activity such as great recession in 2008.

This is the one of the weaknesses of DSGE model. In other words, even though the model can be stylized and explain the implication of theoretical macroeconomic variables, its empirical performances such as forecasting and describing realistic law of motion for all variables are not as good as their theoretical explaining ability.

An alternative way to detour these weaknesses of the DSGE model is unrestricted VAR estimation. This one is more close to statistical model so we would have better results about forecasting or law of motion for the macroeconomic variables. However, since we do not impose any theoretical restrictions on the model, it is difficult to have some economic implication. In addition, the unrestricted VAR model can be seen as reduced form from the structural VAR, there exists some identification problem for the parameters of interest.

In this study, I re-visited one of the influential DSGE model of Smets and Wouters (2007) (hereafter SW (2007)) and examined the model in terms of statistical adequate perspective. One of the reasoning we need to find statistically adequate model is following. In SW (2007), they used Bayesian approach to estimate the parameters of interest and provide some evidence saying that their DSGE model has competitive empirical performance comparing to VAR and Bayesian VAR model. However, if the VAR model is misspecified their comparison is misleading. Also, if the estimation model is statistically inadequate, then by re-specifying the model that is statistically adequate we would have new likelihood function. Using this likelihood function we would have better inference even in Bayesian approach.

This chapter consists of following chapters. In section 2, I would describe the SW (2007) DSGE model and show some derivation of the optimization problem for each agent. Section 3 describe the transition part from macroeconomic theory to econometric matter. For the estimation model I will do the mis-specification test to investigate the estimation driven by theoretical part. In section 4, I will give some evidence that the estimation model is statistically inadequate.

3.2 From Macroeconomics to an Empirical Model

3.2.1 SW (2007) Approach

The DSGE model derived in the previous chapter can be written in the form $E_t\{f(\mathbf{Z}_{t+1}, \mathbf{Z}_t, \mathbf{Z}_{t-1}, \boldsymbol{\varepsilon}_t)\}=0$ taking as solution equations of the type $\mathbf{Z}_t=g(\mathbf{Z}_{t-1}, \boldsymbol{\varepsilon}_t)$ which is called decision rule. More specification about the solution equations can be represented by followings:

$$
\begin{aligned}
\mathbf{Z}_t^* &= M\bar{\mathbf{Z}}(\boldsymbol{\theta}) + M\widehat{\mathbf{Z}}_t + N(\boldsymbol{\theta})\mathbf{X}_t + \boldsymbol{\eta}_t \\
\widehat{\mathbf{Z}}_t &= g_z(\boldsymbol{\theta})\widehat{\mathbf{Z}}_{t-1} + g_\varepsilon(\boldsymbol{\theta}) \\
E(\boldsymbol{\eta}_t\boldsymbol{\eta}^T) &= V(\boldsymbol{\theta}) \\
E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T) &= Q(\boldsymbol{\theta})
\end{aligned}
$$

where $\widehat{\mathbf{Z}}$ are the variables in deviation from steady state, $\bar{\mathbf{Z}}$ is the vector of steady state values and $\boldsymbol{\theta}$ is the vector of structural (or deep) parameters to be estimated.

The second equation is the decision rule which describe a relationship among true endogenous variables, but those variables are not directly observed. The variables we can observe are $\mathbf{Z}_t^*$, and as we can see $\mathbf{Z}_t^*$ is related to $\widehat{\mathbf{Z}}_t$ with errors, $\boldsymbol{\eta}_t$. If the variables have trends the term $N(\boldsymbol{\theta})\mathbf{X}_t$ would capture so that we can deal with the most general case in which the trends depend on the structural parameters.

Using the equations above we need to evaluate likelihood of the DSGE model solution system. A tricky part is that the equations are nonlinear in terms of the structural parameters. However, the equations are linear in terms of the endogenous and exogenous variables so that we can evaluate the likelihood with a linear prediction error algorithm such as Kalman filter. Actually, we can consider the first two equations above as measurement and state (transition) equation, respectively.

For the observations, $t = 1, ..., T$ and with initial values $\mathbf{Z}_1$ and $\mathbf{P}_1$ given, the recursion follows

$$\begin{aligned}
\mathbf{v}_t &= \mathbf{Z}_t^* - \bar{\mathbf{Z}}^* - \mathbf{M}\widehat{\mathbf{Z}}_t - \mathbf{N}\mathbf{X}_t \\
\mathbf{F}_t &= \mathbf{M}\mathbf{P}_t\mathbf{M}^T + \mathbf{V} \\
\mathbf{K}_t &= \mathbf{g}_z\mathbf{P}_t\mathbf{g}_z^T\mathbf{F}_t^{-1} \\
\widehat{\mathbf{Z}}_{t+1} &= \mathbf{g}_z\widehat{\mathbf{Z}}_t + \mathbf{K}_t\mathbf{v}_t \\
\mathbf{P}_{t+1} &= \mathbf{g}_z\mathbf{P}_t(\mathbf{g}_z - \mathbf{K}_t\mathbf{M})^T + \mathbf{g}_\varepsilon\mathbf{Q}\mathbf{g}_\varepsilon^T
\end{aligned}$$

From the Kalman filter recursion, one can derive the log-likelihood given by

$$\ln L(\theta|\mathbf{Z}_T^*) = -\frac{Tk}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}|\mathbf{F}_t| - \frac{1}{2}\mathbf{v}_t^T\mathbf{F}_t^{-1}\mathbf{v}_t$$

where the vector $\boldsymbol{\theta}$ contains the parameter of interest: $\boldsymbol{\theta}, V(\boldsymbol{\theta})$ and $Q(\boldsymbol{\theta})$, and $\mathbf{Z}_t^*$ expresses the set of observable endogenous variables found in the measurement equation. In SW(2007), their measurement equation is that

$$Y_t = \begin{bmatrix} dlGDP_t \\ dlCONS_t \\ dlINV_t \\ dlWAG_t \\ lHOURS_t \\ dlP_t \\ FEDFUNDS_t \end{bmatrix} = \begin{bmatrix} \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{l} \\ \bar{\pi} \\ \bar{r} \end{bmatrix} + \begin{bmatrix} y_t - y_{t-1} \\ c_t - c_{t-1} \\ i_t - i_{t-1} \\ w_t - w_{t-1} \\ l_t \\ \pi_t \\ r_t \end{bmatrix}$$

Since they used Bayesian approach they derive which is called log posterior kernel

$$\ln \kappa(\boldsymbol{\theta}|\mathbf{Z}_t^*) = \ln L(\boldsymbol{\theta}|\mathbf{Z}_t^*) + \ln p(\boldsymbol{\theta})$$

where $p(\boldsymbol{\theta})$ is the priors of parameters to be estimated. Finally, using the numerical methods (they used dynare in Matlab), they estimated the mode of posterior distribution by maximizing the log posterior kernel with respect to $\boldsymbol{\theta}$.

3.2.2 Solution Model to Statistical Model

In the view of statistical perspective, SW(2007) pre-impose some theoretical restrictions such as Gaussian. However, the reliability of any inference based on their approach only be guaranteed after we confirm that the underlying statistical model is adequate based on the data chosen. Thus I need to derive the underlying statistical model first. The model is reduced form of the solution of structural model. An mentioned before, we can express the solution equation as follows.

$$\mathbf{Z}_t = \mathbf{\Pi}\mathbf{Z}_{t-1} + \mathbf{W}\boldsymbol{\varepsilon}_t$$

and this structural solution equations can be decomposed by observable variables and unobservables like followings.

$$\mathbf{d}_t = \mathbf{D}\mathbf{Y}_{t-1} + \mathbf{E}\mathbf{d}_{t-1} + \mathbf{F}\mathbf{v}_t(observable)$$

$$\mathbf{Y}_t = G\mathbf{Y}_{t-1} + \mathbf{H}\mathbf{d}_{t-1} + \mathbf{K}\mathbf{v}_t(unobservable)$$

where $\mathbf{v}_t$ is vector of standard deviation of structural shocks and it has following process, $\mathbf{v}_t = \mathbf{P}\mathbf{v}_{t-1} + \varepsilon_t$, and **D**,**E**,**F**,**G**,**H** and **K** are formed by partitioning of Π.

Assuming that **D** has inverse matrix, Poudyal(2012) showed that we can have following equations.

$$\mathbf{d}_t = [\mathbf{D}\mathbf{G}\mathbf{D}^{-1} + \mathbf{E}]\mathbf{d}_{t-1} + \mathbf{D}(\mathbf{H} - \mathbf{G}\mathbf{D}^{-1}E)\mathbf{d}_{t-2} + \mathbf{e}_t$$

$$\mathbf{e}_t = \mathbf{F}\mathbf{v}_t + \mathbf{D}(\mathbf{K} - \mathbf{G}\mathbf{D}^{-1}\mathbf{F})\mathbf{v}_{t-1}$$

$$\mathbf{v}_t = \mathbf{P}\mathbf{v}_{t-1} + \boldsymbol{\varepsilon}_t$$

and from the equations above we can derive

$$\mathbf{d}_t = \Psi_1\mathbf{d}_{t-1} + \Psi_2\mathbf{d}_{t-2} + \mathbf{u}_t$$

This is the statistical model we can derive from the DSGE model. Since each terms in the form of vector, it is VAR(2) model. In general, the VAR(2) model is assumed to have Normal distribution for the error probability. In other words, the variables have Normal and

Markov(2) process in probability perspective. This VAR is called unconstrained VAR model implying that they do not impose any theoretical restriction.

However there exist some critical points we need to consider. Specifically, we cannot guarantee that the VAR(2) model is validate based on the data in our hands and if the VAR(2) model is invalidate based on the data we the inferences using this VAR(2) could be misleading. Indeed, SW(2007) compared their Bayesian estimates to the several VAR models to persuade how their DSGE model has competitive performance. Even though their results showed that their model is good and better than VAR and Bayesian VAR models in several aspects such as marginal likelihood and out-of-sample prediction performance, I would say that their inference could be suspicious if their compared VAR model is statistically inadequate. In following chapter, I explore the data and analyze their statistical adequacy.

3.3 Mis-specification Test for VAR(2) Model

3.3.1 Exploratory Data Analysis(EDA)

The Probabilistic Reduction (PR) approach suggests the use of graphical techniques and the Exploratory Data Analysis(EDA) is an integral part of statistical modeling (Spanos, 2006). The EDA provide useful information regarding to statistical information in the data. SW(2007) uses quarterly data 1947 Q1 through 2004 Q4. I attached description and sources of the data in appendix. When I plot the data I checked both level and grow rate although SW(2007) used growth rate data only. The reason why I plot the level data as well as grow rate is that if a variable has trend then log-difference would also have trend as Niraj(2006) showed. For example, if $X_t = \delta_0 + \delta_1 t$, log-difference $x_t = (X_t/X_{t-1})$ is still a function of time variable t. Followings are the plots of observed variables for level and growth rate.

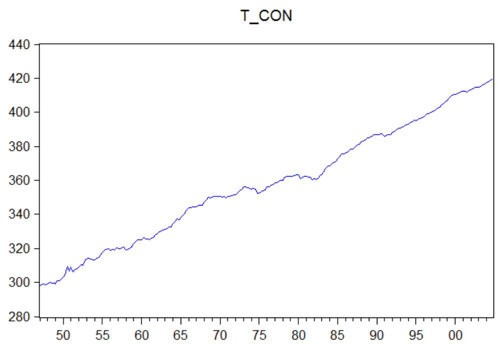

Figure 3.1: Consumption (Level)

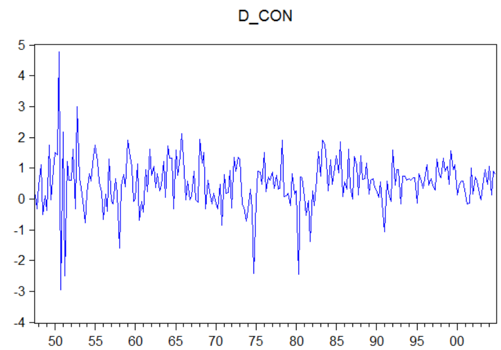

Figure 3.2: Consumption (Growth Rate)

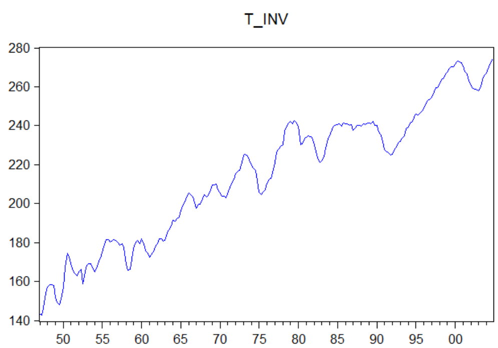

Figure 3.3: Investment (Level)

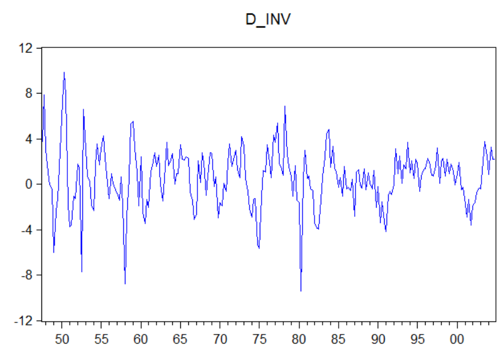

Figure 3.4: Investment (Growth Rate)

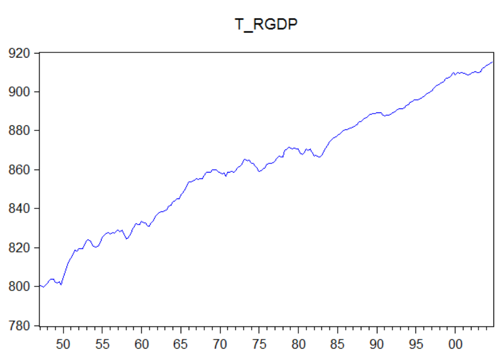

Figure 3.5: Output (Level)

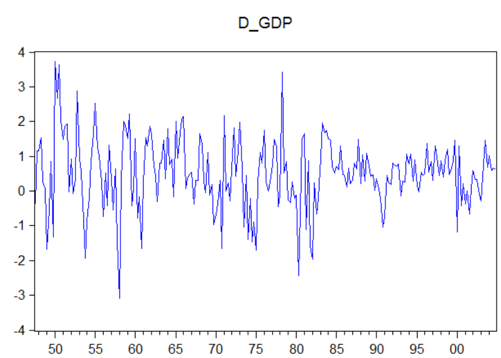

Figure 3.6: Output (Growth Rate)

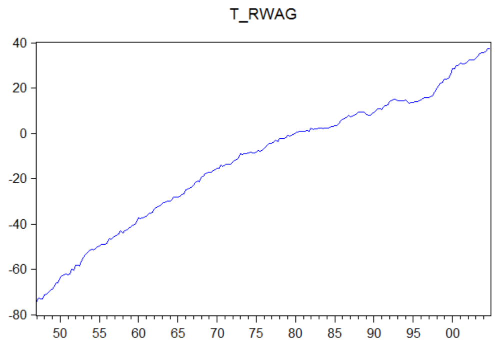

Figure 3.7: Real Wage (Level)

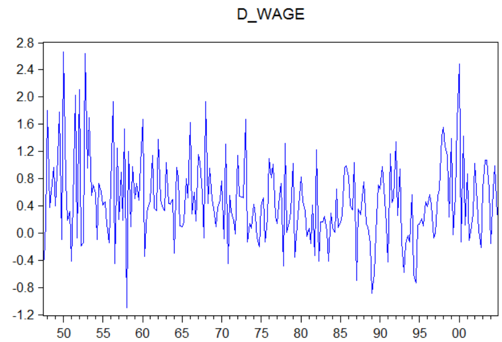

Figure 3.8: Real Wage (Growth Rate)

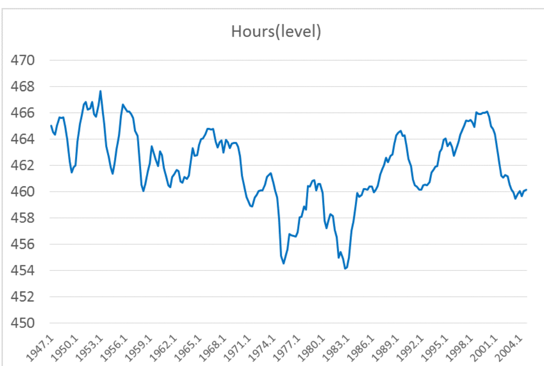

Figure 3.9: Hours (Level)

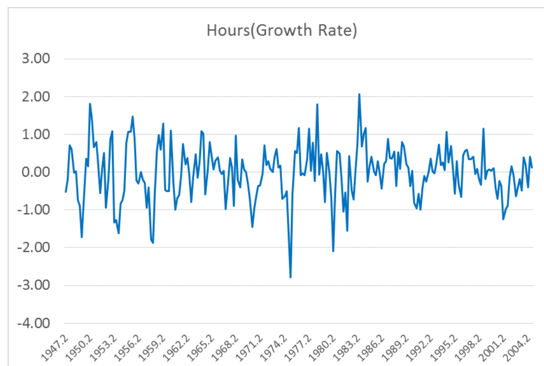

Figure 3.10: Hours (Growth Rate)

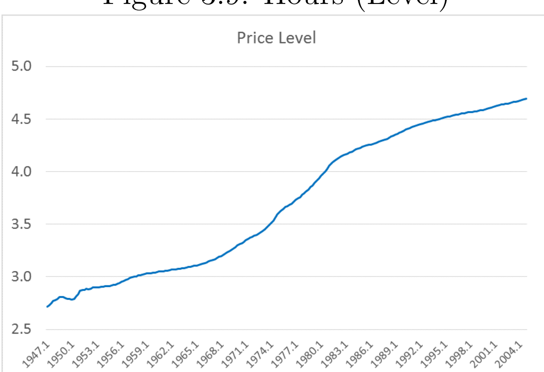

Figure 3.11: Price(level)

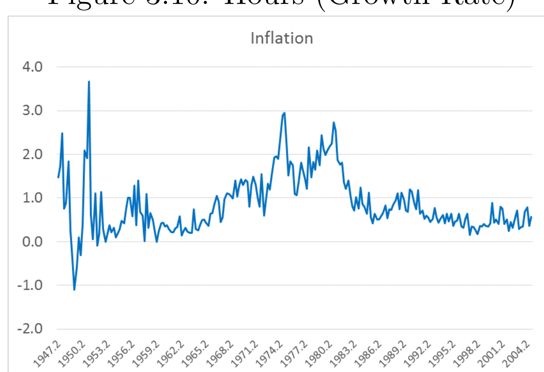

Figure 3.12: Inflation (Growth Rate)

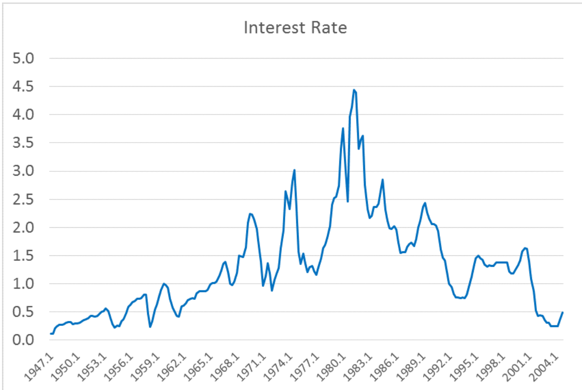

Figure 3.13: Interest Rate(level)

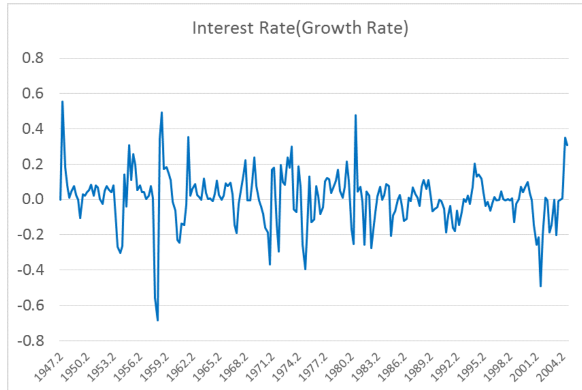

Figure 3.14: Interest Rate (Growth Rate)

As we can see, t-plots in level shows trends that some of them are linear, other are non-linear. These mean that a model without a trend in mean, variance or both is possibly mis-specified. Furthermore, we can verify that there exist positive dependence in the variables of investment, hours, and interest rate. Also, interest rate shows strong volatility in level. For the log-difference variables, it is obvious that the variance is changing over time for most of variables. For the inflation, we can suspect that it has non-linear trend.

3.3.2 Probabilistic Reduction and VAR Model

The solution system of SW(2007) can be transformed to the reduced model which would be estimable. However, the parameter space of the reduced form is determined by the structural model, and this restriction can be very restrictive in the view of statistical data generating mechanism. Moreover, the parameter space could be impossible, then we cannot learn anything from the data rather the result would be misleading. To find out statistically adequate model so that we can learn from data, we need to set up statistical generating mechanism that is free from any structural restriction. I view the SW(2007)'s model as structural VAR with two lags and compare it with statistical model without structural restriction as follows;

$$\mathbf{Z_t} = \mathbf{a_0} + \mathbf{A_1^T Z_{t-1}} + \mathbf{A_2^T Z_{t-2}} + \mathbf{u_t}, \quad \mathbf{u_t} \sim N(\mathbf{0}, \Omega), \quad t \in \mathbb{N}$$

After statistical adequacy of the model is guaranteed, we can ensure that our inference based on the model is reliable. In addition, we can test that the coefficients estimated from theoretically restricted are the same the estimates from statistically adequate one.

In the view of parameterization of the stochastic process $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ whose probabilistic structure is given by the joint distribution $D(\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_T; \Theta)$ where T is sample size, we can turn this model following three sets of assumptions using Spanos(1990)'s Probabilistic Reduction (PR) approach:

Specification of VAR(2)

Distribution(D)	Normal
Dependence(I)	Markov(2)
Heterogeneity(H)	Homogeneous

Based on the Markov(2) and the homogeneity assumption, we can reduce the dimension of the joint distribution $D(\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_T; \Theta)$ as $D(\mathbf{Z}_t, \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}; \Theta), t \in \mathbb{N}$, and this can be decomposed by product of the conditional distribution and the marginal distribution such as

$$D(\mathbf{Z}_t, \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}; \Theta) = D(\mathbf{Z}_t | \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}; \Theta_1) \times D(\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}; \Theta_2), \quad t \in \mathbb{N}$$

where $\Theta, \Theta_1, \Theta_2$ are the parameters of the joint distribution, the conditional distribution, and the marginal distribution, respectively. From the normality assumption, we can see that both conditional distribution and marginal distribution follows normal distribution. Using these, probabilistic assumption of VAR(2) model can be represented by

Statistical Generating Mechanism: $\mathbf{Z_t} = \mathbf{a_0} + \mathbf{A_1'} \mathbf{Z_{t-1}} + \mathbf{A_2'} \mathbf{Z_{t-2}} + \mathbf{u_t}, \quad t \in \mathbb{N}$

(1)	Normality:	$\mathrm{D}(\mathbf{Z_t}, \mathbf{Z_{t-1}}, ..., \mathbf{Z_1}; \theta)$ is Normal
(2)	Linearity:	$\mathrm{E}(\mathbf{Z}_t \vert \sigma(\mathbf{Z}^0_{t-1})) = \mathbf{a_0} + \mathbf{A_1'} \mathbf{Z_{t-1}} + \mathbf{A_2'} \mathbf{Z_{t-2}}$
(3)	Homosked.:	$\mathrm{Var}(\mathbf{Z_t} \vert \sigma(\mathbf{Z^0_{t-1}})) = \Omega$ is free of $\mathbf{Z^0_{t-1}} := (\mathbf{Z_{t-1}}, ..., \mathbf{Z_1})$
(4)	Markov:	$\{\mathbf{Z}_t, t \in \mathbb{N}\}$ is a Markov(2) process
(5)	t-invariance:	$\Theta_1 := (\mathbf{a_0}, \mathbf{A_1}, \mathbf{A_2}, \Omega)$ are t-invariant for all $t \in \mathbb{N}$

From the joint distribution we can directly derive the parameter of the VAR(2) model (Anderson, 2003) such as

$$\mathbf{A_1} = \mathbf{\Sigma_{12}}(\mathbf{\Sigma_{11}} - \mathbf{\Sigma_{12}}\mathbf{\Sigma_{11}^{-1}}\mathbf{\Sigma_{12}'})^{-1} - \mathbf{\Sigma_{13}}\mathbf{\Sigma_{11}}\mathbf{\Sigma_{12}'}(\mathbf{\Sigma_{11}} - \mathbf{\Sigma_{12}}\mathbf{\Sigma_{11}^{-1}}\mathbf{\Sigma_{12}'})^{-1}$$

$$\mathbf{A_2} = -\mathbf{\Sigma_{12}}\mathbf{\Sigma_{11}}\mathbf{\Sigma_{12}}(\mathbf{\Sigma_{11}} - \mathbf{\Sigma_{12}'}\mathbf{\Sigma_{11}^{-1}}\mathbf{\Sigma_{12}})^{-1} + \mathbf{\Sigma_{13}}(\mathbf{\Sigma_{11}}) - \mathbf{\Sigma_{12}'}\mathbf{\Sigma_{11}^{-1}}\mathbf{\Sigma_{12}})^{-1}$$

$$\mathbf{\Omega} = \mathbf{\Sigma_{11}} - \begin{bmatrix} \mathbf{\Sigma_{12}} & \mathbf{\Sigma_{13}} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma_{11}} & \mathbf{\Sigma_{12}} \\ \mathbf{\Sigma_{12}}' & \mathbf{\Sigma_{11}} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma_{12}} \\ \mathbf{\Sigma_{13}} \end{bmatrix}, \quad \mathbf{a_0} = [\boldsymbol{\mu} - \mathbf{A_1}\boldsymbol{\mu} - \mathbf{A_2}\boldsymbol{\mu}]$$

3.3.3 Mis-specification Tests(M-S)

Based on the VAR(2) model, we need to examine if the VAR(2) model has statistical adequacy. In other words, our assumption for the VAR(2) should be valid for the data if the model is statistically adequate. We can investigate it through mis-specification(M-S) test. Following Spanos and McGuirk(2001), I would M-S tests for the Normal linear regression model through F-tests on the significance of auxiliary regressions' coefficients. They also suggests to use the fitted values of the null model as the regressors of the auxiliary regression instead of using all the regressors because it can reduce the number of regressors so that we can have more degree of freedom. In this case because there are 7 variables with 2 lags, I would lose more than 14 degree of freedom unless I don't use fitted values of the null model in M-S testing.

Here is the procedure what I did for M-S test. I estimated VAR(2) model equation by equation using OLS and derive the residuals. I regressed the residuals on past dependent variables, time variables, squared dependent variables, and squared past dependent variables to check whether the independent variables shows significance in this regression. If it some variables show significant result we can suspect that the original estimation overlooked the feature of variables.

I implement M-S test jointly which means testing several assumptions simultaneously to avoid erroneous diagnoses for both regression and skedastic function as well as separate M-S test. Then for the consumption equation, the auxiliary regression function and skedastic function that I used are

$$\widehat{u}_{xt} = \gamma_0 + \gamma_1 \widehat{u}_{xt-1} + \gamma_2 \widehat{u}_{xt-2} + \gamma_3 \widehat{x}_t + \gamma_4 \widehat{x}_t^2 + \gamma_5 t_1 + \gamma_6 t_2 + \gamma_7 t_3 + v_t$$

$$\widehat{u}_{xt}^2 = \delta_0 + \delta_1 \widehat{x}_t + \delta_2 \widehat{x}_t^2 + \delta_3 \widehat{x}_{t-1}^2 + \delta_4 \widehat{x}_{t-2}^2 + \delta_5 t_1 + \delta_6 t_2 + \delta_7 t_3 + v_t$$

where the t_1, t_2 and t_3 are the Chebyshev orthogonal polynomial value such as $t_1 = x$, $t_2 = 2x^2 - 1$, $t_3 = 4x^3 - 3x$, and x implies dependent variables of each regression such as consumption, investment, output and etc. Followings are the tables that summarize the hypothesis tests for the regression and skedastic function.

Regression Function For M-S Tests

Assumption	Null Hypothesis	D.o.F
(1) Linearity	$H_0 : \gamma_4 = 0$	F(1,220)
(2) t-invarince	$H_0 : \gamma_5 = \gamma_6 = \gamma_7 = 0$	F(3,220)
(3) Independence	$H_0 : \gamma_1 = \gamma_2 = 0$	F(2,220)

Skedastic Function For M-S Tests

Assumption	Null Hypothesis	D.o.F
(1) Heteroskedasticity	$H_0 : \delta_1 = \delta_2 = 0$	F(2,220)
(2) 2nd order Independence	$H_0 : \delta_3 = \delta_4 = 0$	F(2,220)
(3) 2nd order t-invariance	$H_0 : \delta_5 = \delta_6 = \delta_7 = 0$	F(3,220)
(4) Hetero. & 2nd Indep.	$H_0 : \delta_1 = \delta_2 = \delta_3 = \delta_4 = 0$	F(4,220)
(5) Hetero. & 2nd t-invar.	$H_0 : \delta_1 = \delta_2 = \delta_5 = \delta_6 = \delta_7 = 0$	F(5,220)

Therefore if the M-S test shows some rejection for any assumption of it, I would be able to say the model is statistically mis-specified. Followings are the results of M-S test using the SW(2007)'s data.

M-S Test Results for VAR(2)

Assumption(s)	$\ln y_t$	$\ln c_t$	$\ln i_t$	$\ln w_t$	$\ln l_t$	$\ln \pi_t$	$\ln r_t$
Reg. Function							
(1) Linearity	0.0068	0.9223	**9.9244**	**14.0023**	0.2838	0.0088	0.1330
	(0.934)	(0.338)	**(0.002)**	**(0.000)**	(0.595)	(0.925)	(0.716)
(2) t-invariance	0.5280	0.9936	**3.4579**	**5.4352**	2.0164	0.0550	1.5375
	(0.663)	(0.397)	**(0.017)**	**(0.001)**	(0.113)	(0.983)	(0.206)
(3) Independence	1.0618	0.1830	0.0450	2.0358	0.6277	0.8071	**6.5547**
	(0.348)	(0.833)	(0.956)	(0.133)	(0.535)	(0.447)	**(0.002)**
Sked. Function							
(1) Heteroskedasticity	**5.5386**	0.1306	2.4171	2.3071	2.3927	**16.9237**	0.1312
	(0.005)	(0.878)	(0.092)	(0.102)	(0.094)	**(0.000)**	(0.877)
(2) 2nd order Indep.	2.5734	0.4181	**3.9813**	2.8667	1.2303	0.6686	**13.8826**
	(0.079)	(0.659)	**(0.020)**	(0.059)	(0.294)	(0.513)	**(0.000)**
(3) 2nd order t-invar.	1.3349	0.8913	1.9244	1.2440	1.1791	**33.3970**	0.6713
	(0.264)	(0.446)	(0.127)	(0.295)	(0.319)	**(0.000)**	(0.570)
(4) Hetero. & 2nd Indep.	**3.5452**	0.3367	2.2098	1.5259	1.6316	**12.1615**	**10.5921**
	(0.008)	(0.853)	(0.069)	(0.196)	(0.167)	**(0.000)**	**(0.000)**
(5) Hetero. & 2nd t-invar.	**2.3280**	0.6971	**2.9847**	2.0088	1.8143	**20.7729**	0.7208
	(0.044)	(0.626)	**(0.013)**	(0.078)	(0.111)	**(0.000)**	(0.608)

Based on the result I have from the M-S test, I would like to say that there are several assumptions that the tests reject. In other words, the estimated model, VAR(2), cannot explain several features that the data has. Consistent with our graphical analysis, I could find several problems in skedastic function such as Heteroskedastic and Heterogeneity in output, investment, inflation, and interest rate variables. I have tested Normality assumption using the methods of Anderson-Darling, Cramer-von Mises. The test results conflicts each other but I would suspect that the normality assumption is not appropriate because the results show some departures from the dependence and identically distributed assumption that cannot compatible with normality distribution assumption.

3.4 Re-specification

Re-specification is done by taking into account the statistical information that is captured by M-S test. For the re-specification, one needs to start from the beginning by postulating a new model that is possibly statistically adequate based on the results from M-S test. Thus I suspect that departure from linearity and t-invariant assumption and second order dependence and heteroskedasticity.

To model heteroskedasticity, I consider heterogeneous Studnet's t distributed VAR model. This model would capture the heteroskedastic and non-normality feature of the data. To model the t-variance feature, I would add a polynomial of t in the estimation equation.

Based on this approach I modeled hetero St-VAR model with different order of t polynomials, different lags and different degree of freedoms. From the postulated models, I regressed the model and did M-S test again using the residuals to check that the model is statistically adequate. After trying several models, I have found that St-VAR(2) with 3rd order trend polynomial is statistically adequate. Followings are plots of residuals of St-VAR(2) model with 3rd order t polynomial.

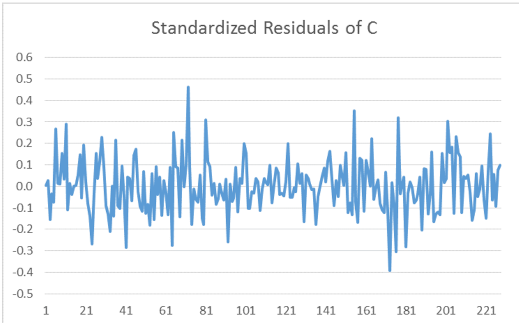

Figure 4.15: Consumption

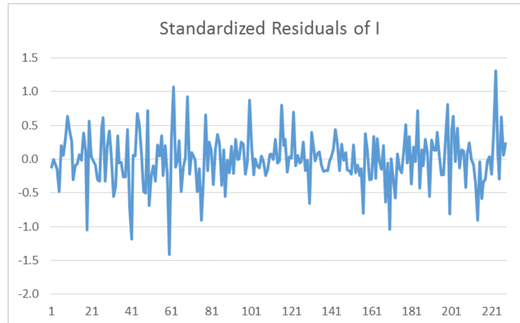

Figure 4.16: Investment

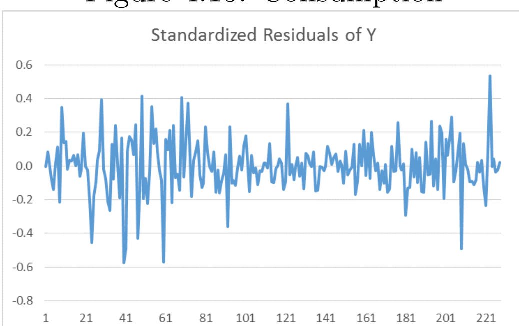

Figure 4.17: Real GDP

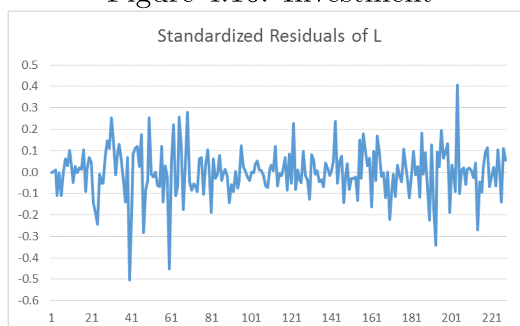

Figure 4.18: Hours

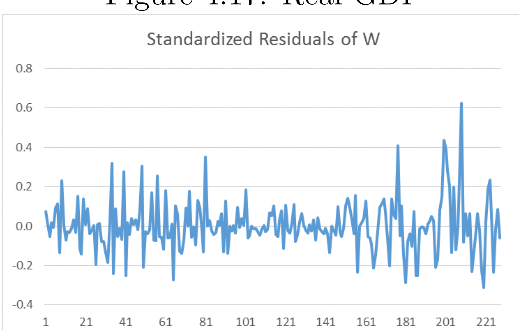

Figure 4.19: Real Wage

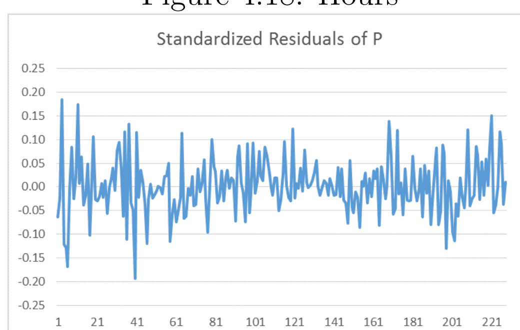

Figure 4.20: Inflation

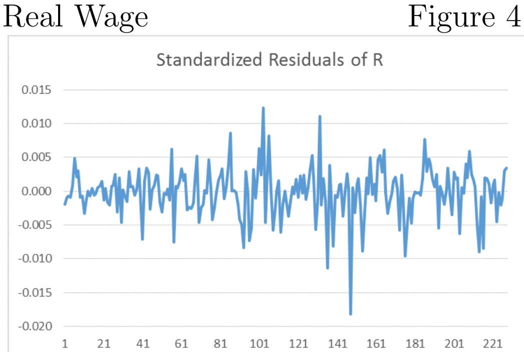

Figure 4.21: Interest Rate

and I have got following results from the M-S test of the model.

Table 4.1: M-S Test Results for Heterogeneous St-VAR(2)

Assumption(s)	$\ln y_t$	$\ln c_t$	$\ln i_t$	$\ln w_t$	$\ln l_t$	$\ln \pi_t$	$\ln r_t$
(1) Linearity	0.394	0.756	0.011	0.277	0.104	0.514	1.077
	(0.531)	(0.386)	(0.916)	(0.599)	(0.748)	(0.474)	(0.301)
(2) t-invariance	0.101	0.205	0.158	0.178	0.009	0.508	0.130
	(0.904)	(0.815)	(0.854)	(0.837)	(0.991)	(0.602)	(0.878)
(3) Independence	0.689	0.304	1.186	0.760	0.297	1.769	1.623
	(0.503)	(0.738)	(0.307)	(0.469)	(0.744)	(0.173)	(0.200)
(4) Heteroskedasticity	1.548	1.934	0.713	1.730	1.252	2.744	0.390
	(0.215)	(0.147)	(0.492)	(0.180)	(0.288)	(0.067)	(0.677)
(5) 2nd order Indep.	0.457	0.384	2.474	0.718	2.443	1.924	0.671
	(0.634)	(0.681)	(0.087)	(0.489)	(0.089)	(0.148)	(0.512)
(6) 2nd order t-invar.	4.773	0.272	2.604	8.469	1.800	12.547	3.541
	(0.009)	(0.762)	(0.076)	(0.000)	(0.168)	(0.000)	(0.031)

Even though I still found out couple of departures from the 2nd order t-invariant assumption with GDP and inflation, it seems that these departures caused by outlier observations. I propose that the heterogenous St-VAR(2) model is our estimation model which is statistically adequate and following is the estimation results.

Table 4.2: St-VAR(2; $\nu = 5$) with 3rd Order Trend Polynomial

$$\Delta \mathbf{Z}_t = \text{Const.} + \boldsymbol{\beta}_1 \text{trends} + \boldsymbol{\beta}_2 \Delta \mathbf{Z}_{t-1} + \boldsymbol{\beta}_3 \Delta \mathbf{Z}_{t-2} + \boldsymbol{\varepsilon}_t$$

	$\ln y_t$	$\ln c_t$	$\ln i_t$	$\ln w_t$	$\ln l_t$	$\ln \pi_t$	$\ln r_t$
Const.	0.566	0.773	0.709	0.376	-0.104	-0.021	-0.0042
	(0.122)	(0.099)	(0.291)	(0.088)	(0.084)	(0.046)	(0.003)
t	-0.099	0.0001	-0.080	-0.117	-0.037	0.0390	-0.0036
	(0.070)	(0.056)	(0.141)	(0.048)	(0.050)	(0.020)	(0.002)

	$\ln y_t$	$\ln c_t$	$\ln i_t$	$\ln w_t$	$\ln l_t$	$\ln \pi_t$	$\ln r_t$
t_2	-0.080	-0.191	-0.102	0.169	-0.144	-0.103	0.0009
	(0.075)	(0.062)	(0.164)	(0.053)	(0.054)	(0.024)	(0.002)
t_3	0.053	-0.022	0.155	0.083	-0.098	-0.011	0.0018
	(0.051)	(0.042)	(0.106)	(0.036)	(0.036)	(0.015)	(0.001)
$\ln y_{t-1}$	0.015	0.204	0.319	0.030	0.152	-0.028	0.0012
	(0.073)	(0.059)	(0.180)	(0.055)	(0.050)	(0.030)	(0.002)
$\ln c_{t-1}$	0.145	-0.250	0.105	0.045	0.133	0.102	0.0019
	(0.081)	(0.064)	(0.198)	(0.060)	(0.055)	(0.033)	(0.002)
$\ln i_{t-1}$	0.095	0.030	0.379	-0.004	0.046	-0.005	0.0023
	(0.025)	(0.020)	(0.055)	(0.019)	(0.017)	(0.010)	(0.001)
$\ln w_{t-1}$	-0.089	-0.055	0.119	0.021	-0.046	0.060	-0.0002
	(0.077)	(0.062)	(0.190)	(0.056)	(0.053)	(0.032)	(0.002)
$\ln l_{t-1}$	0.129	0.092	0.211	-0.018	0.066	-0.027	0.0034
	(0.099)	(0.078)	(0.240)	(0.072)	(0.070)	(0.040)	(0.003)
$\ln \pi_{t-1}$	0.185	-0.317	0.751	0.173	0.214	0.534	0.0075
	(0.143)	(0.119)	(0.356)	(0.107)	(0.100)	(0.054)	(0.004)
$\ln r_{t-1}$	-2.399	-4.808	-7.347	-3.367	0.856	1.466	0.3575
	(1.949)	(1.562)	(4.765)	(1.421)	(1.339)	(0.806)	(0.006)
$\ln y_{t-2}$	0.105	0.132	-0.240	-0.024	0.098	-0.075	0.0037
	(0.101)	(0.081)	(0.243)	(0.073)	(0.068)	(0.040)	(0.003)
$\ln c_{t-2}$	0.055	0.013	-0.016	0.072	-0.012	0.036	-0.0038
	(0.109)	(0.088)	(0.266)	(0.081)	(0.072)	(0.044)	(0.003)
$\ln i_{t-2}$	0.035	0.032	0.183	-0.024	0.054	0.024	0.0003
	(0.033)	(0.027)	(0.078)	(0.025)	(0.022)	(0.014)	(0.001)
$\ln w_{t-2}$	-0.281	-0.096	-0.637	0.127	-0.236	0.077	-0.0034
	(0.102)	(0.083)	(0.237)	(0.074)	(0.067)	(0.040)	(0.003)
$\ln l_{t-2}$	-0.139	-0.129	-0.141	0.029	-0.183	0.037	-0.0003
	(0.124)	(0.101)	(0.295)	(0.093)	(0.084)	(0.050)	(0.003)
$\ln \pi_{t-2}$	-0.372	-0.060	-1.110	-0.121	-0.309	0.302	-0.0026
	(0.168)	(0.140)	(0.415)	(0.129)	(0.116)	(0.067)	(0.005)
$\ln r_{t-2}$	-7.600	-3.342	-9.708	1.705	-3.656	0.696	-0.2243
	(2.356)	(1.959)	(5.938)	(1.787)	(1.647)	(0.971)	(0.065)

Chapter 4

DSGE-VAR(λ) for a Model Evaluation Revisited

4.1 Introduction

Among the various criteria to evaluate the Dynamic Stochastic General Equilibrium (DSGE) models, the DSGE-VAR(λ) framework proposed by Del Negro and Schorfheide (2004) (hereafter DNS (2004)) has a competitive advantage over the others, because we can acquire knowledge about the practical validity of the DSGE parameters from this framework. The validity of the estimated model lies in the reliability of the framework, and our questions begin from this standpoint.

In this study, we inquire whether the estimated model provided by DNS (2004) is statistically adequate and about the way in which we can address the problem unless its statistical adequacy is validated. Furthermore, we deal with the question pertaining to the way in which the model evaluation changes in the framework along with the use of a statistically adequate one.

Through the application of misspecification tests developed by Spanos (2017), this paper shows that the statistical assumptions imposed on the Normal VAR such as homoskedasticity and normality are invalidated with the macroeconomic time series data. To address the departures that were identified by the misspecification tests, we propose the Student's t VAR (StVAR) model as a statistically adequate one. Moreover, we suggest DSGE-StVAR(λ) for the DSGE model evaluation.

In this framework, $\widehat{\lambda}$ indicates accordance of the DSGE model moments to the likelihood function. Our analysis reveals that DSGE-StVAR(λ) gives rise to a higher $\widehat{\lambda}$ than DSGE-VAR(λ) in the case of the simple New Keynesian model with the same sample periods. Interpreting the result, we conclude that an inadequate estimated model underestimates the suitability of the theoretical model to the empirical evidence. In other perspective, an increased coincidence between theoretical moments and the likelihood established by correcting the estimated model can be considered as a theoretically parsimonious DSGE model, which would be adequate to follow the feature of empirical evidence if we set the estimated model in a statistically adequate manner.

The contributions of our study can be distinguished in two aspects. First, we pay attention to the importance of the likelihood function in Bayesian analysis. Most empirical studies dealing with the DSGE model rely on Bayesian estimation approach, especially considering the notions forwarded by Smets and Wouters (2003), and the way to apply the Bayesian approach in macroeconomics has been developed in the direction of finding proper priors. Litterman et al. (1979) and Litterman (1980) introduce the so-called Minnesota prior which assumes that each variable follows a random walk process, and the assumption is based on the statistical reason rather than the theoretical one. Ingram and Whiteman (1994) employ priors from a DSGE model for VAR; thus, it helps to analyze the relationship between the observed time series and the macroeconomic variables. However, in this case, we come back to square one with respect to the empirical work. We investigate the likelihood function derived by the statistical assumptions and their validity with the data. It is obvious that this perspective can be extended to the other Bayesian approach in macroeconomic studies.

Subsequently, we reconsider the implication of $\widehat{\lambda}$ in the DSGE-VAR(λ) framework. As mentioned earlier, $\widehat{\lambda}$ represents how well the moments derived by the DSGE model correspond to the likelihood function. However, the equation $\widehat{\lambda} = \infty$ does not necessarily imply that the theoretical model captures the feature of real world precisely if the likelihood function is induced by a statistically inadequate model. It reiterates the importance of statistical adequacy in this framework. Therefore, statistical adequacy should precede model evaluation.

The importance of the statistical model in the DSGE-VAR(λ) framework has been pointed out by other literature. Consolo et al. (2009) also test the statistical adequacy of the VAR model in DNS (2004). However, our approach to address it is quite different from their one. They add more macroeconomic data into the statistical model by employing Factor Augmented VAR (FAVAR). From the perspective of re-specification, it is difficult to state that FAVAR can be a way to address the misspecified statistical model, because we need to address the departures from the given data without adding the new data. This is the reason for which we suggest the StVAR as the statistically adequate model for this study.

Certain commentators of the study by Del Negro et al. (2007), while focusing on the role of model evaluation of the DSGE-VAR(λ) framework, pointed out the adequacy of its statistical model as well. Although the authors stated that the statistical assumptions imposed on their estimated model were set for computational convenience, we cannot be certain about how large the cost of convenience would be in this regard.

This chapter is organized as follows. Section 2 contains a description of the DSGE-VAR(λ) framework. Section 3 contains the misspecification tests of the DNS(2004) model and re-specification. In Section 4, we re-evaluate the simple New Keynesian model with the statistically adequate model. Section 5 presents the conclusion of this chapter.

4.2 DSGE-VAR(λ) for Evaluating DSGE Models

In this section, we introduce the notion of DSGE-VAR(λ) and describe the way in which this framework can be utilized for evaluating DSGE models.

4.2.1 DSGE-VAR(λ) Framework

The DSGE-VAR(λ) framework is introduced in DNS (2004), and its application of model evaluation has been paid attention to since the publication of the works published by Del Negro and Schorfheide (2006) and Del Negro et al. (2007). The concise outline of the framework as a process of model evaluation can be intuitively explained as follows.

Based on the case where the DSGE model parameters are time invariant, we generated a sequence of structural shocks such as technology, monetary policy, and interest rate. By embedding them in the DSGE model, we can acquire a number of artificial observations. Since DSGE models are closely related to VAR representation (Sims, 1980; Fernandez-Villaverde et al., 2007; Ravenna, 2007; Giacomini, 2013), autocovariances estimated by a VAR with the dummy observations approximate the autocovariances of the DSGE model. As long as the DSGE parameter space is nested in the VAR parameter space, the estimation of the DSGE model can be regarded as an estimation of a VAR with cross-equation restrictions. But, by including a hyperparameter λ, the DSGE-VAR(λ) framework makes provisions for deviations from the restrictions imposed on the VAR. Two extreme cases, where $\lambda = \infty$ and $\lambda = 0$, can be interpreted as the restrictions that are strictly imposed and are completely ignored, respectively, in the estimation of the VAR parameters. Then, repeating this procedure by changing the parameter values for the DSGE model and finding a value of λ that highly correlates to the data is the way in which the DSGE-VAR(λ) framework evaluates the DSGE model.

From the econometric perspective, the DSGE-VAR(λ) framework is based on the Bayesian approach. In this approach, the prior represents information about the VAR parameters that are restricted by the information obtained from the DSGE model. In other words, the prior mean is derived by the artificial observations generated by the theoretical model. For the prior of the VAR coefficients, the hyperparameter λ appears at its variance in inverse terms. Following this, the variance decreases as the λ increases. Hence, it implies that the prior places all of its mass on the theoretical mean if the $\lambda = \infty$ and vice versa. Therefore, if the fit improves as the λ decreases, it implies the theoretical cross-equation restrictions are relaxed; this can be considered to be a supportive evidence that establishes that the restrictions induced by the theoretical model are incompatible with the data, i.e. the model specification is invalid.

In a nutshell, we can regard the λ of the DSGE-VAR(λ) framework as a measurement that reveals how effectively the DSGE model specification can be supported by the data. Namely, the DSGE part of DSGE-VAR(λ) represents a theoretical model setup with restrictions, while the VAR part represents its unrestricted statistical counterpart or empirical evidence. Moreover, λ is determined by the degree of fit between them. Finally, if we find a model that provides a higher value of λ, it indicates that the model is capable of explaining or describing the real world. Thus, we can use this model for policy analysis and forecasting.

In DNS (2004) and Del Negro et al. (2007), they presume that their VAR and error correction models in the DSGE-VAR(λ) framework are statistically adequate. However, we cannot be certain that the estimated models are statistically adequate without any misspecification test. If the VAR representation do not validate its statistical adequacy, any inference from the DSGE-VAR(λ) framework cannot be guaranteed. Furthermore, the output, such as policy analysis and forecasting from the results, will be misleading in such a case. Indeed, Consolo et al. (2009), Poudyal and Spanos (2013), and Spanos (2015) have pointed out the statistical inadequacy of the VAR models with macroeconomic time series data. Thus, we are also highly suspicious about its statistical adequacy pertaining to the Normal VAR in DNS (2004). We will illustrate the details of its statistical adequacy based on the notion of statistical identification developed by Spanos (1994) in section 3. Prior to that, we need to provide a detailed exposition of the DSGE-VAR(λ) framework in terms of the econometric standpoint.

4.2.2 Bayesian Analysis of DSGE-VAR(λ)

As mentioned earlier, the implementation of the DSGE-VAR(λ) framework is based on the Bayesian estimation. In the description of the estimation procedure, we borrow terms and notations from DNS (2004).

VAR Representation

A covariance stationary process has an infinite order moving average (MA) (Wold, 1938), and we can consider the DSGE model as the MA processes for output growth, inflation, and interest rate with restrictions. Then, we can consider a VAR model as a statistical counterpart since MA processes can be represented by employing autoregressions under suitable conditions.[1]

If we consider a p-th order VAR model:

$$\boldsymbol{y_t} = \boldsymbol{\Phi_0} + \boldsymbol{\Phi_1 y_{t-1}} + \boldsymbol{\Phi_2 y_{t-2}} + \cdots + \boldsymbol{\Phi_p y_{t-p}} + \boldsymbol{u_t}, \quad \boldsymbol{u_t} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma_u}) \tag{4.1}$$

where $\boldsymbol{y_t}$ denotes a $n \times 1$ vector of observable variables, $\boldsymbol{u_t}$ represents a vector of one-step-ahead forecast errors which is assumed to follow independent, identically normal distributed variables. With T number of samples ($t = 1, 2, \cdots, T$), the VAR representation above can be rewritten as follows:

$$\boldsymbol{Y} = \boldsymbol{X\Phi} + \boldsymbol{U} \tag{4.2}$$

where $\mathbf{Y}$ is the $T \times n$ matrix with rows $\boldsymbol{y_t}'$, $\mathbf{X}$ represents the $T \times k$, $(k = 1 + np)$ matrix with rows $\boldsymbol{x_t}' = [\boldsymbol{1}, \boldsymbol{y'_{t-1}}, \cdots, \boldsymbol{y'_{t-p}}]$, $\mathbf{U}$ is the $T \times n$ matrix with rows $\boldsymbol{u_t}'$, and $\boldsymbol{\Phi} = [\boldsymbol{\Phi_0}, \boldsymbol{\Phi_1}, \cdots, \boldsymbol{\Phi_p}]$.

Likelihood Function

From the expression of the Equation (4.2) and a multivariate normal distribution assumption of $\boldsymbol{u_t}$, its likelihood function can be derived as follows:

$$p(\boldsymbol{Y}|\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}) \propto |\boldsymbol{\Sigma_u}|^{-T/2} exp\left\{-\frac{1}{2}\text{tr}\left[\boldsymbol{\Sigma_u}^{-1}(\boldsymbol{Y} - \boldsymbol{X\Phi})'(\boldsymbol{Y} - \boldsymbol{X\Phi})\right]\right\} \tag{4.3}$$

[1] A VAR representation of the DSGE model can be derived in a different manner. For instance, Fernandez-Villaverde et al. (2007) show that state-space form of the DSGE model can be transformed to VAR(∞) under the condition called "poor man's invertibility."

where tr[·] denotes the trace of a matrix. Let's say $\boldsymbol{\theta}$ is a vector of the DSGE model parameters. Since $\boldsymbol{\theta}$ has a much lower dimension than the VAR parameter vector, it can be regarded that the DSGE model imposes restrictions on the VAR($\boldsymbol{p}$) model.

Prior Distribution

This part draws a distinction between the DSGE-VAR(λ) framework and the other Bayesian VAR. As we mentioned before, the artificial observations are generated in this stage. Specifically, suppose that in the equation $T^* = \lambda T$, the number of artificial observations ($\boldsymbol{Y}^*, \boldsymbol{X}^*$) are generated through the DSGE model with the structural parameters $\boldsymbol{\theta}$. With the structural shocks assuming normal i.i.d. errors, its likelihood function will be the same as Equation (4.3), except the artificial sample, i.e. as follows:

$$p(\boldsymbol{Y^*(\theta)|\Phi, \Sigma_u}) \propto |\boldsymbol{\Sigma_u}|^{-\lambda T/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma_u}^{-1}\left(\boldsymbol{Y^* - X^*\Phi}\right)'\left(\boldsymbol{Y^* - X^*\Phi}\right)\right]\right\} \qquad (4.4)$$

As described by DNS (2004), if we consider that the actual observations are augmented with the artificial observations, the likelihood function for the combined sample will be the product of Equation (4.3) and (4.4) that can be shown as follows:

$$p(\boldsymbol{Y^*(\theta), Y|\Phi, \Sigma_u}) = p(\boldsymbol{Y^*(\theta)|\Phi, \Sigma_u}) \times p(\boldsymbol{Y|\Phi, \Sigma_u}) \qquad (4.5)$$

From the Bayesian perspective, the first term on the right-hand side of Equation (4.5) can be regarded as a prior density induced from the DSGE model for the statistical parameter Φ and Σ_u. In other words, it contains the information about the VAR parameters relating to the theoretical restrictions that are derived from the sample of artificial observations.

For the artificial observations, $p(\boldsymbol{Y^*(\theta)|\Phi, \Sigma_u})$, DNS (2004) use the sample moments, $\boldsymbol{\Gamma^*_{yy}(\theta)} = E_\theta[\boldsymbol{y_t y_t}']$, $\boldsymbol{\Gamma^*_{yx}(\theta)} = E_\theta[\boldsymbol{y_t x_t}']$, $\boldsymbol{\Gamma^*_{xy}(\theta)} = E_\theta[\boldsymbol{x_t y_t}']$ and $\boldsymbol{\Gamma^*_{xx}(\theta)} = E_\theta[\boldsymbol{x_t x_t}']$ instead of $\boldsymbol{Y^{*\prime}Y^*}$, $\boldsymbol{Y^{*\prime}X^*}$, $\boldsymbol{X^{*\prime}Y^*}$ and $\boldsymbol{X^{*\prime}X^*}$ to avoid stochastic variation from repeated sampling. Furthermore, they add the improper prior $p(\boldsymbol{\Phi, \Sigma_u}) \propto |\boldsymbol{\Sigma_u}|^{-(n+1)/2}$. Then, the prior density of Equation (4.4) can be converted to the following equation.

$$p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{\theta}, \lambda) = c^{-1}(\boldsymbol{\theta})|\boldsymbol{\Sigma_u}|^{-\frac{\lambda T+n+1}{2}} \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\lambda T \boldsymbol{\Sigma_u}^{-1}(\boldsymbol{\Gamma}^*_{yy}(\boldsymbol{\theta}) - \boldsymbol{\Phi}'\boldsymbol{\Gamma}^*_{xy}(\boldsymbol{\theta}) - \boldsymbol{\Gamma}^*_{yx}(\boldsymbol{\theta})\boldsymbol{\Phi} + \boldsymbol{\Phi}'\boldsymbol{\Gamma}^*_{xx}(\boldsymbol{\theta})\boldsymbol{\Phi})\right]\right\} \quad (4.6)$$

where $c^{-1}(\theta)$ represents a normalizing constant to ensure that the aforementioned prior density integrates to form one. Then, this prior density conditional on θ has the well-known Inverted Wishart(IW)-Normal(N) form such as the following.

$$(\boldsymbol{\Sigma_u}|\boldsymbol{\theta}, \lambda) \sim IW(\lambda T \boldsymbol{\Sigma}^*_{\boldsymbol{u}}(\boldsymbol{\theta}), \lambda T - k - n) \quad (4.7)$$

$$(\mathrm{vec}(\boldsymbol{\Phi})|\boldsymbol{\Sigma_u}, \boldsymbol{\theta}, \lambda) \sim N(\mathrm{vec}(\boldsymbol{\Phi}^*(\boldsymbol{\theta})), \boldsymbol{\Sigma_u} \otimes (\lambda T \boldsymbol{\Gamma}^*_{xx}(\boldsymbol{\theta}))^{-1}) \quad (4.8)$$

where

$$\boldsymbol{\Phi}^*(\boldsymbol{\theta}) = \boldsymbol{\Gamma}^*_{xx}(\boldsymbol{\theta})^{-1}\boldsymbol{\Gamma}^*_{xy}(\boldsymbol{\theta})$$

and

$$\boldsymbol{\Sigma}^*_{\boldsymbol{u}}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}^*_{yy}(\boldsymbol{\theta}) - \boldsymbol{\Gamma}^*_{yx}(\boldsymbol{\theta})\boldsymbol{\Gamma}^*_{xx}(\boldsymbol{\theta})^{-1}\boldsymbol{\Gamma}^*_{xy}(\boldsymbol{\theta})$$

As mentioned above, if $\lambda = \infty$, then $(\lambda T \boldsymbol{\Gamma}^*_{xx}(\boldsymbol{\theta}))^{-1}$ in Equation (4.8) becomes zero. Hence, the prior for the coefficients spots on the theoretical values are observed. These are the description of the prior for the statistical parameters. The full prior is $p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}, \boldsymbol{\theta}|\lambda)$, and its specification is set as the hierarchical structure mentioned below.

$$p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}, \boldsymbol{\theta}|\lambda) = p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{\theta}, \boldsymbol{\lambda}) \times p(\boldsymbol{\theta})$$

It consists of a prior for the VAR parameters conditional on the DSGE parameters and a prior for the DSGE parameters. We will describe the prior for the DSGE parameters in the following section.

Posterior Distribution

The full posterior distribution in the DSGE-VAR(λ) framework is $p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}, \boldsymbol{\theta}|\boldsymbol{Y})$ and we can factorize the posterior into two parts as follows.

$$p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}, \boldsymbol{\theta}|\boldsymbol{Y}, \lambda) = p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{Y}, \boldsymbol{\theta}, \lambda) \times p(\boldsymbol{\theta}|\boldsymbol{Y}, \lambda) \quad (4.9)$$

The first term on the right-hand side in the aformentioned equation represents the posterior density of the VAR parameters considering that the DSGE structural parameters and the second term represent the posterior density of the DSGE model parameters.

Applying the likelihood function and the prior that we have ascertained from Equation (4.3), (4.7) and (4.8), we can obtain the posterior density of the VAR parameters, $p(\Phi, \Sigma_u|Y, \theta, \lambda)$. Since the prior distribution has the Inverted Wishart-Normal form, which is conjugate to the likelihood function, the posterior distribution also has the Inverted Wishart-Normal form such as the following.

$$(\boldsymbol{\Sigma_u}|\boldsymbol{Y}, \boldsymbol{\theta}, \lambda) \sim IW((\lambda+1)T\widetilde{\boldsymbol{\Sigma_u}}(\boldsymbol{\theta}), (\lambda+1)T-k, n) \tag{4.10}$$

$$(\boldsymbol{\Phi}|\boldsymbol{Y}, \boldsymbol{\Sigma_u}, \boldsymbol{\theta}, \lambda) \sim N(\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}), \boldsymbol{\Sigma_u} \otimes (\lambda T\boldsymbol{\Gamma}^*_{\boldsymbol{xx}}(\boldsymbol{\theta}) + \boldsymbol{X'X})^{-1}) \tag{4.11}$$

where

$$\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}) = (\lambda T\boldsymbol{\Gamma}^*_{\boldsymbol{xx}}(\boldsymbol{\theta}) + \boldsymbol{X'X})^{-1}(\lambda T\boldsymbol{\Gamma}^*_{\boldsymbol{xy}}(\boldsymbol{\theta}) + \boldsymbol{X'Y})$$

and

$$\begin{aligned}\widetilde{\boldsymbol{\Sigma}}_u(\boldsymbol{\theta}) = \frac{1}{(\lambda+1)T}\Big[&(\lambda T\boldsymbol{\Gamma}^*_{\boldsymbol{yy}}(\boldsymbol{\theta}) + \boldsymbol{Y'Y}) \\ &-(\lambda T\boldsymbol{\Gamma}^*_{\boldsymbol{yx}}(\boldsymbol{\theta}) + \boldsymbol{Y'X})(\lambda T\boldsymbol{\Gamma}^*_{\boldsymbol{xx}}(\boldsymbol{\theta}) + \boldsymbol{X'X})^{-1}(\lambda T\boldsymbol{\Gamma}^*_{\boldsymbol{xy}}(\boldsymbol{\theta}) + \boldsymbol{X'Y})\Big]\end{aligned}$$

The other term on the right-hand side in Equation (4.9), which is the posterior density of the DSGE model parameters, $p(\boldsymbol{\theta}|\boldsymbol{Y})$, can be derived by the following equation.

$$p(\boldsymbol{\theta}|\boldsymbol{Y}, \lambda) \propto p(\boldsymbol{Y}|\boldsymbol{\theta}, \lambda) \times p(\boldsymbol{\theta}) \tag{4.12}$$

where the likelihood function for the structural parameters, $p(\boldsymbol{Y}|\boldsymbol{\theta}, \lambda)$, is provided by the following equation.

$$p(\boldsymbol{Y}|\boldsymbol{\theta}, \lambda) = \frac{p(\boldsymbol{Y}|\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}, \boldsymbol{\theta}, \lambda) \times p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{\theta}, \lambda)}{p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{Y}, \boldsymbol{\theta}, \lambda)} \tag{4.13}$$

Finding $\widehat{\lambda}$ and Model Evaluation

As we know from the fact that the equation $T^* = \lambda T$, the effective sample size of the artificial observations, is determined by the hyperparameter λ, subsequently the empirical performance of the DSGE-VAR(λ) framework is affected by the value of λ. In the opposite manner, if we can determine the value of $\widehat{\lambda}$ that provides the highest marginal density using data-driven procedure, we can interpret that it indicates the empirical evidence about the DSGE model misspecification and the degree of relaxation pertaining to the restrictions. Then, a large value of $\widehat{\lambda}$ can be interpreted, as the moments derived from the theoretical model fit well to the time series data that we have procured. Hypothetically, if $\widehat{\lambda} = \infty$, we would interpret that the theoretical model captures the empirical features precisely, while if $\widehat{\lambda} = 0$, it would imply that the theoretical model cannot be supported by the data at all.

For the formal notation, the optimal values of λ is as follows:

$$\widehat{\lambda} = arg \max_{\lambda \in \Lambda} p_\lambda(\boldsymbol{Y}) \tag{4.14}$$

And, $p_\lambda(Y)$ can be derived by the following equation:

$$p_\lambda(Y) = \int p_\lambda(\boldsymbol{Y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{4.15}$$

4.2.3 Revisiting the Likelihood Function

This section presents the description of the DSGE-VAR(λ) framework. By applying the Bayesian approach, this framework facilitates the process of learning the structural parameters and evaluating the theoretical model. However, we need to be cautious that those inferences will be reliable only if the statistical adequacy of the likelihood function that is applied in the framework is secured. If we employ an estimated model that is statistically inadequate, the model evaluation in the framework would be misleading. In spite of its significance in the empirical work, verification of the estimated model's statistical adequacy is an aspect that tends to be ignored. In DNS (2004), they do not mention about its statistical adequacy of their VAR(4) model.

In fact, Consolo et al. (2009) and Poudyal and Spanos (2013) have pointed out the statistical inadequacy of VAR models with macroeconomic time series data in their studies, and our questions also begin from this viewpoint. Therefore, the model evaluation with the statistically unidentified model could be misleading. We need to investigate whether the estimated model is statistically adequate with the data that we have obtained and re-specify the model if we verify that the model is not statistically adequate. The following section explains the procedure investigating its statistical adequacy.

4.3 Statistical Identification of the DSGE Model

As mentioned in the previous section, the DSGE-VAR(λ) framework as a method of evaluating the DSGE model has been applied on the premise that the VAR model is statistically adequate. In other words, any inference from this framework will be misleading unless it has a statistically identified model. To emphasize and prove the importance of a statistically adequate model, Spanos (1990) introduced the notion of statistical identification that is distinguished from structural identification. From a statistical perspective, a structural model can be regarded as a reduced form model with restrictions, and the issue pertaining to structural identification relates to determining the unique value of the structural parameters that are defined by the restrictions. In contrast, statistical identification relates to process of selecting a statistically well-defined model as a reduced form among the set of permissible models. Thus, a statistically identified model has data-oriented specification with the aim of allowing the feature of the observed data to fulfill a central role in the empirical analysis.

In Consolo et al. (2009) and Poudyal and Spanos (2013), although their perception of statistical identification and the ways to address the issues are a bit different, they insist in general that a Normal VAR model specification fails to correspond with the macroeconomic time series data for their empirical analysis.

In this section, we will examine the statistical adequacy of the Normal VAR(4) applied in DNS (2004). To investigate its statistical adequacy, we will implement the misspecification tests suggested by Spanos (2017).

4.3.1 Exploratory Data Analysis (EDA)

Exploratory data analysis with graphical techniques can provide us with beneficial hints relating to the statistical model specification. Thus, it is not only fundamental but also important as an integral part of statistical modeling. Del Negro and Schorfheide (2004) used the data ranging from the quarter 1955:Q3 to 2001:Q3, but we extend the data range to the period of 2016:Q3 in this study to improve the power of the misspecification tests and the standard error of the estimates.[2]

As shown in Figure 3.5, the interest rate has been very close to the zero (zero lower bound, ZLB) since the great recession periods in the latter half of 2008. During those periods, we can suspect that the economic bahavior could be different from the normal situation, but Givens (2017) showed that the estimates of the DSGE model parameters in Christiano et al. (2005) are generally robust based on the U.S. macroeconomic data from 1997 to 2015 that includes the ZLB periods. Thus, we expect that the extension of the sample range would not affect our result seriously.

To attain a better idea about the probabilistic structures of these macroeconomic time series data, we have also plotted the level data for each variable as well as the ones used in Del Negro and Schorfheide (2004). Thus, Figure 3.1, Figure 3.3, and Figure 3.5 present the t-plots of the level data while Figure 3.2, Figure 3.4, and Figure 3.6 depict the t-plots of

[2]The data for real output is obtained from the Bureau of Economic Analysis (BEA, Real Gross Domestic Product: Seasonally Adjusted at Annual Rate (SAAR) and Billions of Chained 2009 dollars. For inflation, we have used the Consumer Price Index-All Urban Consumers (1982-1984=100, seasonally adjusted, average of three months) data from the Bureau of Labor Statistics (BLS). The original data for the interest rate that is the Federal Funds Effective Rate (FFER, annual rate) is obtained from the Federal Reserve Banks (FRB), and we have computed the average of the federal funds rate (business days only) during the first month of the quarter, as constructed in Clarida et al. (2000) and Del Negro and Schorfheide (2004).

the log-differenced and logged data for the real Gross Domestic Product (GDP), Consumer Price Index (CPI), and interest rate, respectively.

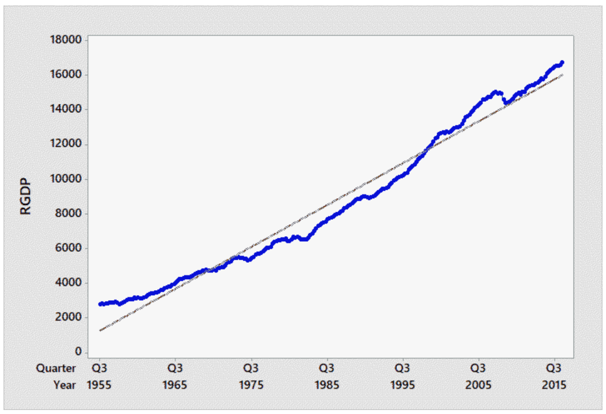

Figure 3.1: Level of Real GDP

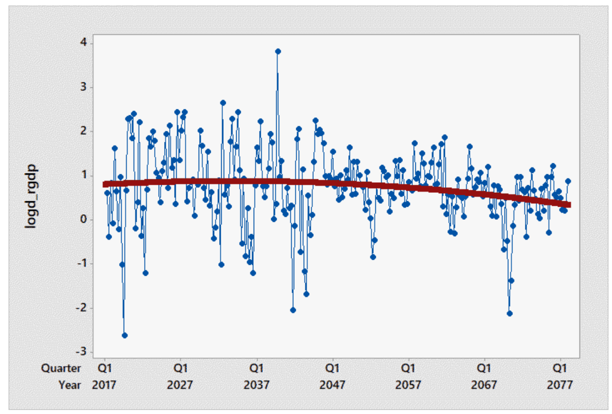

Figure 3.2: Log-differenced Real GDP

First of all, as we can see from the figures, all of the series have non-linear trends in level and this fact leads us a critical fact that we cannot get detrended data by differencing the logged variables. Poudyal (2012) explained it by giving the following simple example: If a random variable has a linear trend, t, such that $Y_t = \alpha_0 + \alpha_1 t$, its log-differenced one, $y_t = ln(Y_t/Y_{t-1})$, is still a function of the trend, t. What is worse that it appears non-linearly in its log-differenced one. This could be the reason for which we can observe the (non-linear) heterogeneous mean in the right column of the figures above. Moreover, Geweke (1993) and Andreou and Spanos (2003) concluded that the macroeconomic series were not difference stationary but trend stationary as opposed to the notions forwarded by Nelson and Plosser (1982).

Subsequently, we can find positive dependence lasting for a certain period of time in the log-differenced CPI and logged interest rate (Figure 3.4 and 3.6). In addition, for all series in the logged and log-differenced variables, we can find changing variance across the time. For instance, the log-differenced real GDP has been less volatile since the 1990s.

The model specification for estimation with the observable variables that are log-differenced real GDP (Δlnx_t), log-differenced CPI (ΔlnP_t), and logged interest rate (lnR_t^a) implemented

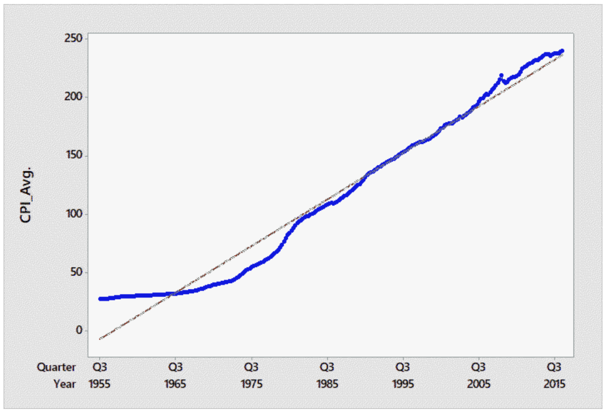

Figure 3.3: Level of CPI

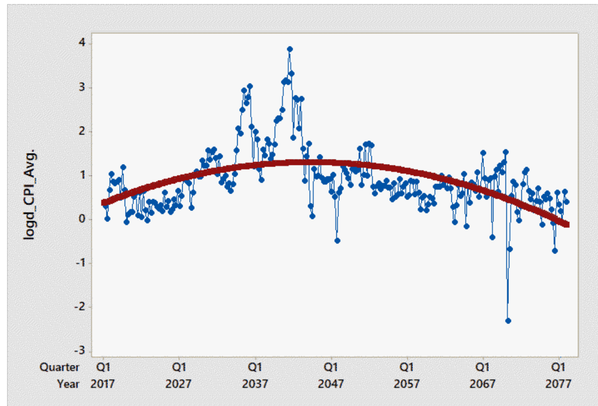

Figure 3.4: Log-differenced CPI

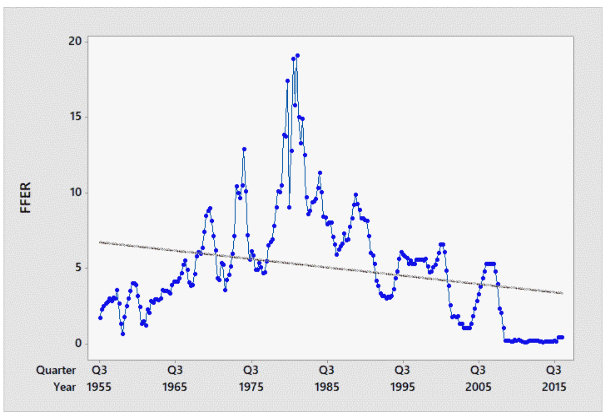

Figure 3.5: Level of Interest Rate

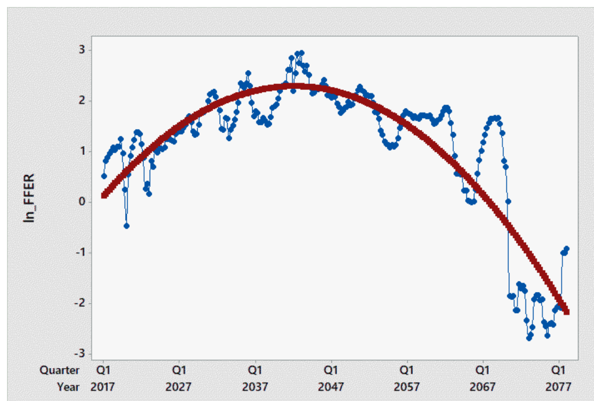

Figure 3.6: Logged Interest Rate

in Del Negro and Schorfheide (2004) is a Normal VAR model with 4 lags, which is as follows.

$$\boldsymbol{y_t} = \boldsymbol{\Phi_0} + \boldsymbol{\Phi_1 y_{t-1}} + \boldsymbol{\Phi_2 y_{t-2}} + \boldsymbol{\Phi_3 y_{t-3}} + \boldsymbol{\Phi_4 y_{t-4}} + \boldsymbol{u_t}, \quad \boldsymbol{u_t} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma_u}) \tag{4.16}$$

where $y_t = [\Delta\ln\boldsymbol{x_t}, \Delta\ln\boldsymbol{P_t}, \ln\boldsymbol{R_t}^a]^T$. In fact, this model specification is the output from an amalgam of structural restrictions and statistical assumptions. The parameter spaces are restricted by the structural information and the homogeneous mean as well as the homogeneous or homoskedastic variance assumptions are imposed. However, based on the implications deduced from the EDA, we can suspect that departures from the model assumptions potentially exist in this regard. To probe these potential departures in a systematic manner, we have employed misspecification tests in the probabilistic reduction approach developed by Spanos

(2017).

4.3.2 Probabilistic Reduction (PR) Approach

Spanos (2017) proposed the PR approach that provides a statistical model based on a purely probabilistic structure. Its underlying perspective lies within London School of Economics (LSE) approach that suggests letting 'the data tell itself' as stated by Denis Sargan and David Hendry.[3] In this approach, we distinguish the statistical (chance regularities) information and the substantive ones determined from a structural model by postulating a statistical model that is free from structural restrictions. Following this, we categorize the kinds of probabilistic assumptions that are imposed on the statistical model.

Suppose a purely statistical model corresponding to the Equation (4.16) such as the following equation:

$$\boldsymbol{y_t} = \boldsymbol{\Psi_0} + \boldsymbol{\Psi_1 y_{t-1}} + \boldsymbol{\Psi_2 y_{t-2}} + \boldsymbol{\Psi_3 y_{t-3}} + \boldsymbol{\Psi_4 y_{t-4}} + \boldsymbol{u_t}, \quad \boldsymbol{u_t} \sim N(\boldsymbol{0}, \boldsymbol{\Omega_u}) \tag{4.17}$$

As Spanos (1990) suggested, this equation can be regarded as a reduction from the Haavelmo or the joint distribution $D(\boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_T}; \boldsymbol{\phi})$ consisting of the 3-dimensional multivariate stochastic process $\{\boldsymbol{y_t}, t \in \mathbb{N}\}$ with the sample size T. He divided the probabilistic assumptions into three categories such as dependence, heterogeneity, and distribution to transform the stochastic process into a statistical model. In our case, the parameterization for the VAR(4) model is drawn from a reduction of the joint distribution $D(\boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_T}; \boldsymbol{\phi})$ with the Markov (M), Stationary (S), and Normal (N) distribution assumptions that are imposed on the stochastic process $\{y_t, t \in \mathbb{N}\}$, i.e.

$$\begin{aligned}
D(\boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_T}; \boldsymbol{\phi}) \quad &= \quad D(\boldsymbol{y_1}, \boldsymbol{y_2}, \boldsymbol{y_3}, \boldsymbol{y_4}; \boldsymbol{\varphi_1}) \textstyle\prod_{t=5}^{T} D_t(\boldsymbol{y_t} | \boldsymbol{y_{t-1}}^0; \boldsymbol{\varphi_t}) \\
&\overset{M}{=} \quad D(\boldsymbol{y_1}, \boldsymbol{y_2}, \boldsymbol{y_3}, \boldsymbol{y_4}; \boldsymbol{\varphi_1}) \textstyle\prod_{t=5}^{T} D(\boldsymbol{y_t} | \boldsymbol{y_{t-1}}, \boldsymbol{y_{t-2}}, \boldsymbol{y_{t-3}}, \boldsymbol{y_{t-4}}; \boldsymbol{\varphi_t}) \\
&\overset{M\&S}{=} \quad D(\boldsymbol{y_1}, \boldsymbol{y_2}, \boldsymbol{y_3}, \boldsymbol{y_4}; \boldsymbol{\varphi_1}) \textstyle\prod_{t=5}^{T} D(\boldsymbol{y_t} | \boldsymbol{y_{t-1}}, \boldsymbol{y_{t-2}}, \boldsymbol{y_{t-3}}, \boldsymbol{y_{t-4}}; \boldsymbol{\varphi})
\end{aligned}$$

[3]Indeed, the key contributions of the LSE tradition revolve around Haavelmo's statement "to build models that explain what has been observed" (Spanos, 2014)

for $(\boldsymbol{y_1}, \cdots, \boldsymbol{y_T}) \in \mathbb{R}^{3T}$. While the first equality does not entail any assumptions, the second equality follows from the Markov process assumption, whereas the third follows from the Stationary assumption. This reduction with the Normality assumption that is imposed on $D(\boldsymbol{y_t}|\boldsymbol{y_{t-1}}, \boldsymbol{y_{t-2}}, \boldsymbol{y_{t-3}}, \boldsymbol{y_{t-4}}; \boldsymbol{\varphi})$ gives rise to the VAR(4) model which is stated below.

Table 3.1: Normal Vector Autoregressive (VAR(4)) Model

Statistical GM: $\boldsymbol{y_t} = \boldsymbol{\Psi_0} + \boldsymbol{\Psi_1 y_{t-1}} + \boldsymbol{\Psi_2 y_{t-2}} + \boldsymbol{\Psi_3 y_{t-3}} + \boldsymbol{\Psi_4 y_{t-4}} + \boldsymbol{u_t}$		
(1) Normality	:	$D(\boldsymbol{y_t}|\boldsymbol{y^0_{t-1}}; \boldsymbol{\phi})$ for $\boldsymbol{y_{t-1}}^0 \equiv (\boldsymbol{y_{t-1}}, \cdots \boldsymbol{y_1})$
(2) Linearity	:	$E(\boldsymbol{y_t}|\boldsymbol{\sigma(y_{t-1}}^0)) = \boldsymbol{\Psi_0} + \boldsymbol{\Psi_1 y_{t-1}} + \boldsymbol{\Psi_2 y_{t-2}} + \boldsymbol{\Psi_3 y_{t-3}} + \boldsymbol{\Psi_4 y_{t-4}} + \boldsymbol{u_t}$
(3) Homoskedasticity	:	$Var(\boldsymbol{y_t}|\boldsymbol{\sigma(y_{t-1}}^0)) = \boldsymbol{\Omega_u}$ is free of $\boldsymbol{y_{t-1}}^0$
(4) Markov	:	$\{\boldsymbol{y_t}, \in \mathbb{N}\}$ is a Markov process
(5) t-invariance	:	$\boldsymbol{\Phi} \equiv (\boldsymbol{\Psi_0}, \boldsymbol{\Psi_1}, \boldsymbol{\Psi_2}, \boldsymbol{\Psi_3}, \boldsymbol{\Psi_4}, \boldsymbol{\Omega_u})$ are t-invariianct $\forall t \in \mathbb{N}$

4.3.3 Misspecification (M-S) Tests for VAR Models

From the PR approach, we ascertain the statistical and the substantive information, then establish the statistical model without any structural restrictions by imposing only probabilistic assumptions on the observable variables or stochastic process. Following this, using the statistical model VAR(4), we should check that the probabilistic assumptions imposed on the statistical model are valid for the data to ensure its statistical adequacy. If the statistical adequacy is not validated, we cannot guarantee the procurement of any information, both statistical and substantive, from the model.

As a firm method to ensure the statistical adequacy of the model, we conducted out misspecification (M-S) tests developed and utilized by Spanos (2017) and Poudyal and Spanos (2013). In the M-S tests, we examined the residuals from the VAR(4) model through the auxiliary regressions based on the principle that the residuals from a statistically adequate model are characterized by truly random one. Therefore, we can interpret that any departure of the residuals from a random normal distribution is an indication of misspecification of the statistical model or the model is not statistically identified.

The auxiliary regressions are designed to capture any significant departures from the null hypothesis stating that the residuals are random. Since the model assumptions are generally interrelated, testing the assumptions separately can result in misdiagnosis. Hence, we conducted the joint M-S tests, as suggested by Spanos (2017), by setting up the following auxiliary regression equations.

$$\widehat{u}_{i,t} = \gamma_0 + \gamma_1\widehat{u}_{i,t-1} + \gamma_2\widehat{u}_{i,t-2} + \gamma_3\widehat{i}_t + \gamma_4\widehat{i}_t^2 + \gamma_5 t_1 + \gamma_6 t_2 + \gamma_7 t_3 + v_t \tag{4.18}$$

$$\widehat{u}_{i,t}^2 = \delta_0 + \delta_1\widehat{i}_t + \delta_2\widehat{i}_t^2 + \delta_3\widehat{i}_{t-1}^2 + \delta_4\widehat{i}_{t-2}^2 + \delta_5 t_1 + \delta_6 t_2 + \delta_7 t_3 + v_t \tag{4.19}$$

where $i \in \{\Delta\ln\boldsymbol{x_t}, \Delta\ln\boldsymbol{P_t}, \ln\boldsymbol{R_t}^a\}$, $\widehat{u}_{i,t}$ is the residual, $\widehat{i}_i$ is the fitted value of the equation i_t, and the t_1, t_2 and t_3 are the Chebyshev orthogonal polynomials such as $t_1 = x$, $t_2 = 2x^2 - 1$, $t_3 = 4x^3 - 3x$.[4] Equation (4.18), the regression function, includes lags of the residuals, squared fitted value, and polynomial terms as regressors to test the dependence, linearity, and t-invariance of the mean. For instance, if the test rejects the null hypothesis ,'$\gamma_4 = 0$', it not only means that $\widehat{u}_{i,t}$ is not random but also that the linearity assumption is invalid. Similarly, Equation (4.19), the skedastic function, includes the fitted (squared) values with the lags and orthogonal polynomials to investigate the homoskedasticity, dependence, and t-invariance of the variance. The following tables summarize the hypothesis tests that we implemented for the M-S tests and present their results.

Table 3.3 shows the F-test statistics and their p-values in the parentheses. Prior to checking the results in the tables, note that the p-value is affected by the number of observations, as it decreases when the sample size increases. Therefore, we need to choose the rigorous rejection criterion that corresponds to the sample size. In our study, since we have more than 200 observations, 1% significant level is suitable for the rejection criterion rather than 5% or 10% significant level.

From the results in Table 3.3, we can verify that some of the assumptions imposed on the Normal VAR(4) model are invalid based on the data. Especially, the homoskedastic as-

[4] The domain of x is $x \in [-1, 1]$.

Table 3.2: M-S Tests for VAR(4) Model

Assumption	Null Hypothesis
Joint (Reg. Function)	$H_0 : \gamma_1 = \gamma_2 = \gamma_4 = \gamma_5 = \gamma_6 = \gamma_7 = 0$
Non-linearity	$H_0 : \gamma_4 = 0$
Mean Heterogeneity	$H_0 : \gamma_5 = \gamma_6 = \gamma_7 = 0$
Temporal Depdendence	$H_0 : \gamma_1 = \gamma_2 = 0$
Joint (Sked. Function)	$H_0 : \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = \delta_7 = 0$
Heteroskedasticity	$H_0 : \delta_1 = \delta_2 = 0$
Variance Dependence	$H_0 : \delta_3 = \delta_4 = 0$
Variance Heterogeneity	$H_0 : \delta_5 = \delta_6 = \delta_7 = 0$

Table 3.3: M-S Test Results for Normal VAR(4) Model

Assumptions	$\Delta\ln\boldsymbol{x_t}$	$\Delta\ln\boldsymbol{P_t}$	$\ln\boldsymbol{R_t}^a$
Joint (Reg. Function)	1.331 [0.244]	**2.999 [0.008]***	0.861 [0.525]
Linearity	1.228 [0.221]	**3.420 [0.001]***	0.351 [0.554]
t-invarince	1.671 [0.174]	1.648 [0.179]	1.281 [0.282]
Independence	0.397 [0.673]	0.502 [0.606]	0.579 [0.561]
Joint (Sked. Function)	**7.506 [0.000]***	**5.819 [0.000]***	**3.514 [0.001]***
Heteroskedasticity	**4.695 [0.010]***	**5.595 [0.004]***	**4.469 [0.013]***
2nd order Independence	1.368 [0.257]	1.975 [0.141]	2.386 [0.094]
2nd order t-invariance	**12.820 [0.000]***	1.027 [0.381]	2.342 [0.074]

* represents significance at 1%

sumption is strictly rejected for all the variables of the model. Table 3.4 shows the results of the normal distribution test. We implemented three different tests for the distribution of the residuals; $\Delta\ln\boldsymbol{x_t}$ and $\ln\boldsymbol{R_t}^a$ show severe departure from the normal distribution assumption in all three distribution tests.

Table 3.4: Normality Test Results for Normal VAR(4) Model

Tests	$\Delta\ln \boldsymbol{x_t}$	$\Delta\ln \boldsymbol{P_t}$	$\ln \boldsymbol{R_t}^a$
Anderson-Darling (AD)	**1.214 [0.004]***	0.519 [0.185]	**1.505 [0.001]***
Shapiro-Wilk (SW)	**0.983 [0.008]***	0.991 [0.178]	**0.976 [0.001]***
Cramer-von Mises (W2)	**0.207 [0.004]***	0.087 [0.166]	**0.236 [0.002]***

* represents significance at 1%

Consequently, it is our valid conclusion that the Normal VAR(4) model is not statistically adequate, and this outcome is consistent with the earlier findings by Consolo et al. (2009) and Poudyal and Spanos (2013) with respect to the fact that the VAR models in their studies are also statistically invalid. Hence, any inference drawn from the regression analysis of the VAR(4) model can be misleading. Moreover, the assessment of the DSGE model via the DSGE-VAR(λ) framework on the premise that the normal VAR(4) model is statistically adequate cannot assure its evaluation. Therefore, we need to reconsider the probabilistic assumptions that are imposed on the stochastic process $\{y_t, t \in \mathbb{N}\}$ and re-specify the statistical model that captures the departures detected in the M-S tests.

4.3.4 Re-specification for Statistical Identification

As a strategy for statistical identification, Consolo et al. (2009) adopt a Factor Augmented VAR (FAVAR) model including additional economic factors that the vector of observable variables, y_t, cannot entirely capture. Thus, they apply the DSGE-FAVAR(λ) model for model evaluation. However, the strategy in question and the FAVAR as a statistically identified model are dubious for certain reasons. First, the statistical identification process should be operated within the domain of the given statistical information instead of adding variables from outside, because including new variables leads to the different vector of the stochastic process. Second, adding new variables cannot be the way to rectify the departures. Intuitively, provided that both of them are statistically adequate, it is obvious that a FAVAR model would have certain advantages over a VAR model for inferences, because it has more statistical information. However, more statistical information does not relate to the statistical adequacy. Lastly, the tests conducted by Consolo et al. (2009) for statistical identification of the FAVAR model occur separately. As we mentioned previously, since the statistical assumptions are interrelated, testing them individually can give rise to misleading diagnoses.

With regard to the appropriate method for statistical identification, we re-specify the statistical model by taking into account the features that were not identified by the Normal VAR(4) model. Although the M-S tests give us certain directions about where we can go to attain the statistically adequate model, it does not tell us the process by which we can get there. In other words, the rejection of homoskedasticity in Table 3.3 indicates that a potentially statistically adequate model makes provisions for heteroskedasticity, but it does not show us what model would allow one to do so. Thus, we need to repeat the M-S tests after establishing a candidate statistical model that is presumed to be statistically adequate.

To postulate a statistically adequate model, we summarize all the relevant information that we procured from the EDA and M-S tests as follows:

- The vector of the stochastic process is mean-heterogeneous in levels
- Homogeneous and no temporal dependence of the mean are valid
- Homoskedasticity is strictly rejected in all three equations
- Non-linearity and the 2nd order t-invariance are indicated in $\Delta\ln\boldsymbol{P_t}$ and $\Delta\ln\boldsymbol{x_t}$ respectively
- Normality is strongly rejected in $\Delta\ln\boldsymbol{x_t}$ and $\ln\boldsymbol{R_t}^a$

Based on this information, we adopted the following strategies to rectify the departures.

- Owing to the mean-heterogeneous of the stochastic process in level, log-differenced and logged variables also seem to have trends in their process. Hence, a statistically adequate VAR model might have trends.

- To model heteroskedasticity, we take the Student's t distribution assumption, as suggested by Spanos (1994) and Poudyal and Spanos (2013), and the choice of the distribution assumption is also affected by the non-normality indicated by the M-S tests and following analysis.

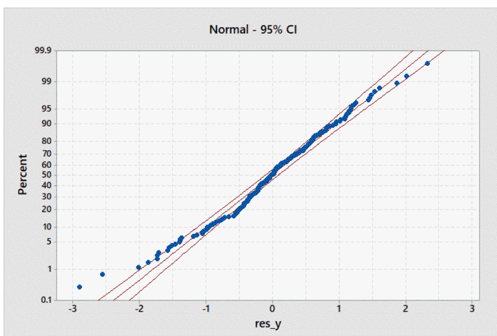

Figure 3.7: Probability Plot (Δlnx_t)

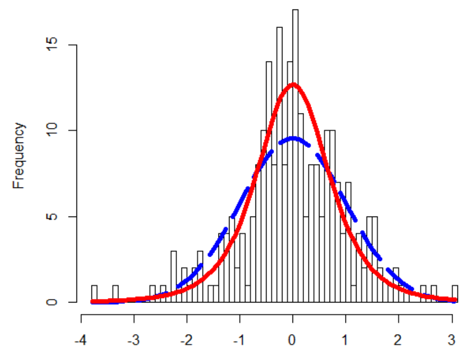

Figure 3.8: Histogram (Δlnx_t)

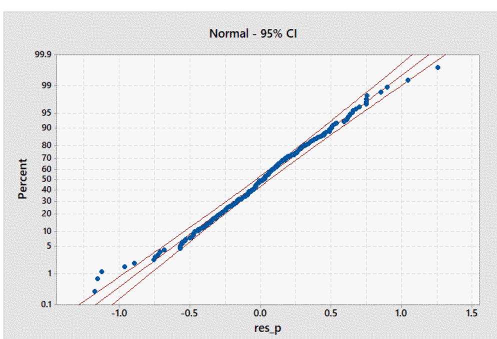

Figure 3.9: Probability Plot (ΔlnP_t)

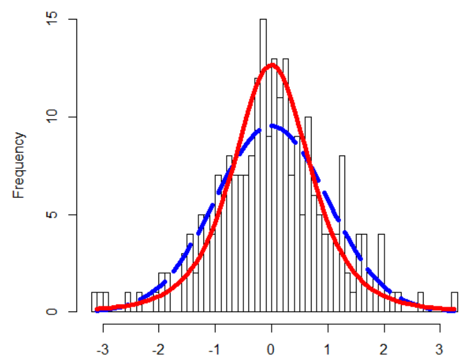

Figure 3.10: Histogram (ΔlnP_t)

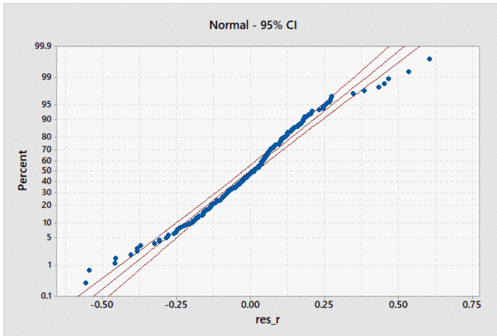

Figure 3.11: Probability Plot (lnR_t^a)

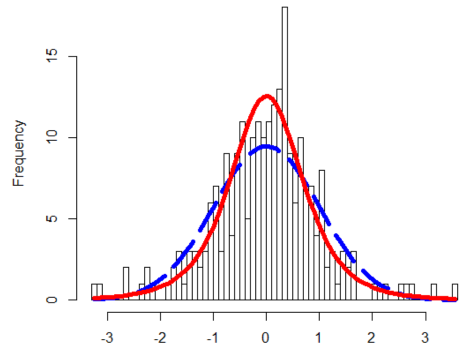

Figure 3.12: Histogram (lnR_t^a)

The figures 3.7, 3.9, and 3.11 present the probability plots of the normal distribution, and the figures 3.8, 3.10, and 3.12 present the histograms of the residuals. In the histograms, the long-dashed (blue) lines represent the normal distribution corresponding to the residuals and the solid (red) lines represent the student's t distribution of them.[5] As observed from the histograms, although they are bell-shaped, their peaks are higher, and tails are fatter than the normal distribution. These facts make the Student's t distribution assumption more reasonable.

- Although the M-S tests also indicate other departures such as the 2nd order t-invariance of $\Delta\ln\boldsymbol{x_t}$ and linearity of $\Delta\ln\boldsymbol{P_t}$, we consider a candidate statistical model capturing heteroskedasticity as the first action and then observe the ways in which the test results change.

Our candidate statistical models are as follows:

$$\text{St-VAR}(p;\nu) : y_t = \boldsymbol{\Psi_0} + \boldsymbol{\Psi_1 y_{t-1}} + \cdots + \boldsymbol{\Psi_{t-p+1} y_{t-p+1}} + \boldsymbol{\Psi_{t-p} y_{t-p}} + \boldsymbol{u_t} \tag{4.20}$$

[5]The variance of the student's t with zero mean and unit standard deviation is $\frac{v}{v-2}$ and its distribution shows fatter tail and lower peak compared to the standard normal distribution. However, we need to use the identical mean and variance for appropriate comparison of the distributional fit. As we can see in the figures 3.8, 3.10, and 3.12, identical variance shows that the student's t distribution is leptokurtic.(Spanos, 1999).

with different lagged variables. With the models, we repeat the M-S tests and fine-tune them to identify the one that passes all the M-S tests. Our finding is the Student's t VAR(2; $\nu = 3$) with linear trend, which as follows:

$$\boldsymbol{y_t} = \boldsymbol{\delta_0} + \boldsymbol{\delta_1 t} + \boldsymbol{\Psi_1 y_{t-1}} + \boldsymbol{\Psi_2 y_{t-2}} + \boldsymbol{u_t} \tag{4.21}$$

Table 3.5: M-S Test Results for Heterogeneous St-VAR(2)

Assumptions	$\Delta\ln\boldsymbol{x_t}$	$\Delta\ln\boldsymbol{P_t}$	$\ln\boldsymbol{R_t}^a$
Joint (Reg. Function)			
Linearity	0.152 [0.697]	0.872 [0.351]	0.147 [0.702]
t-invarince	3.716 [0.012]	1.648 [0.719]	1.761 [0.155]
Independence	2.820 [0.062]	4.419 [0.013]	3.428 [0.034]
Joint (Sked. Function)			
Heteroskedasticity	0.656 [0.520]	1.740 [0.178]	0.895 [0.410]
2nd order Independence	2.617 [0.075]	0.894 [0.410]	1.006 [0.367]
2nd order t-invariance	2.014 [0.113]	1.125 [0.340]	1.171 [0.322]
Student's Distribution	1.883 [0.107]	2.619 [0.043]	2.509 [0.049]

As mentioned earlier, since we have more than 200 observations, we take the 1 % significant level as the rejection criterion, and in this way, no departure is found from the M-S tests. Hence, we can interpret from the results that the heterogeneous StVAR(2) model is statistically adequate based on the data that we have procured. In other words, the statistical model is finally identified.

4.4 Analysis of the DSGE-VAR with the Student's t

In this section, we will re-investigate the DSGE-VAR(λ) framework with the re-specified statistical model, i.e. we apply the DSGE-StVAR(λ) model for the DSGE model evaluation.

Since we verified that the heterogeneous StVAR(2) model is statistically adequate, the model evaluation based on the DSGE-StVAR(λ) will be statistically reliable.

4.4.1 Modifying the Likelihood Function for the Structural Parameters

The critical difference from the previous statistical model is its distributional assumption. The change from the Normal to Student's t leads to the different functional form of the likelihood function which is the Student's t probability density function (pdf). To derive it, at first, we consider the joint distribution of the Student's t from our statistical model such as the following.

$$\boldsymbol{Z_t} = \begin{bmatrix} \boldsymbol{y_t} & (d_1 \times 1) \\ \boldsymbol{y}_{t-1}^{t-2} & (d_2 \times 1) \end{bmatrix}, \quad \boldsymbol{\mu_z} = \begin{bmatrix} \boldsymbol{\mu}_{z,t} & (d_1 \times 1) \\ \boldsymbol{\mu}_{z,t-1}^{z,t-2} & (d_2 \times 1) \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11}(d_1 \times d_1) & \boldsymbol{\Sigma_{12}}(d_1 \times d_2) \\ \boldsymbol{\Sigma_{12}}^T(d_2 \times d_1) & \boldsymbol{\Sigma_{22}}(d_2 \times d_2) \end{bmatrix}$$

where d_1 and d_2 represent the dimensions of the vector $\boldsymbol{y_t}$ and $\boldsymbol{y}_{t-1}^{t-2}$ respectively. From the joint distribution, we can derive the conditional student's t pdf as follows.

$$D(\boldsymbol{y_t}|\boldsymbol{\sigma}(\boldsymbol{y_{t-1}^{t-p}});\boldsymbol{\psi_1}) = \left[\frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu+d_2}{2}\right)}\right] \frac{\sqrt{det(\Lambda_{11})}}{(\pi\nu c_t)^{\frac{d_1}{2}}} \left(1 + \frac{(\boldsymbol{y_t} - \boldsymbol{\Phi^T}\boldsymbol{y_{t-1}^{t-p}})^T \boldsymbol{\Lambda_{11}}(\boldsymbol{y_t} - \boldsymbol{\Phi^T}\boldsymbol{y_{t-1}^{t-p}})}{\nu c_t}\right)^{-\frac{d+\nu}{2}} \tag{4.22}$$

where $\boldsymbol{\Lambda_{11} = \Sigma_u^{-1} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)}^{-1}$, $c_t = (1+(y_{t-1}^{t-2}-\mu_{z,t-1}^{z,t-2})^T Q(y_{t-1}^{t-2}-\mu_{z,t-1}^{z,t-2}))$, $\boldsymbol{Q} = \dfrac{1}{\nu}\boldsymbol{\Sigma_{22}}^{-1}$ and ν is the degree of freedom.

Following this, the previous likelihood function described in Equation (4.3) is replaced by Equation (4.22) with T number of observations. This change also affects the second term in Equation (4.5) for Bayesian analysis. In DNS (2004), since the priors are conjugate for the likelihood function which is normal pdf, they can provide advantages such that its posterior distribution can be derived analytically and the computational time is reduced in their Bayesian estimation process.

In spite of the same priors as Equation (4.7) and (4.8), however, since the likelihood function has been changed to the Student's t pdf, the priors are not conjugate any longer. Hence, we do not have an analytic form of the posterior (conditional) distributions. This change prevents us from applying Gibbs sampler that requires a known form of conditional posterior kernel for the set of parameters. But, we can apply the Metropolis-Hastings (MH) algorithm, which is a more general sampling method of Markov Chain Monte Carlo (MCMC) for Bayesian analysis. Although previous studies in the literature also apply the MH algorithm, the difference is that we need to use the algorithm for estimating both $p(\boldsymbol{\theta}|\boldsymbol{Y}, \lambda)$ and $p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{Y}, \boldsymbol{\theta}, \lambda)$ in Equation (4.9), while they use the algorithm for estimating $p(\boldsymbol{\theta}|\boldsymbol{Y}, \lambda)$ only.

From the model evaluation perspective, we place more weight on determining the optimal value of λ ($\widehat{\lambda}$) based on the DSGE-StVAR(λ) model. To find the value of $\widehat{\lambda}$ that gives the highest marginal density of $p_\lambda(Y)$, we need to ascertain the marginal density of Equation (4.15) for each λ and to find in what value of λ the $p_\lambda(Y)$ gives the highest density.

To implement this process, we need to know both $p(\boldsymbol{Y}|\boldsymbol{\theta}, \lambda)$ and $p(\boldsymbol{\theta})$. While $p(\boldsymbol{\theta})$ is given as shown in Table 4.6, we need to derive the $p(\boldsymbol{Y}|\boldsymbol{\theta}, \lambda)$.

Derivation of the likelihood function for the structural parameters is induced by Equation (4.13). However, as we mentioned before, we do not have an analytic form of the posterior density which is the denominator of the equation. To avoid the concerned difficulty, we apply the posterior kernel instead of the full density of the posterior distribution as follows.

$$p(\boldsymbol{Y}|\boldsymbol{\theta}, \lambda) \propto \frac{p(\boldsymbol{Y}|\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}, \boldsymbol{\theta}, \lambda) \times p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{\theta}, \lambda)}{p^*(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{Y}, \boldsymbol{\theta}, \lambda)} \tag{4.23}$$

where $p^*(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{Y}, \boldsymbol{\theta}, \lambda)$ represents the posterior kernel. By using the posterior kernel, we attain a proportional relationship to $p(\boldsymbol{Y}|\boldsymbol{\theta}, \lambda)$ instead of equality as in the case of Equation (4.13). However, as we can observe in Equation (4.12), its posterior is estimated in a proportional relationship. Therefore, we can expect a proportional relationship in the likelihood function will not affect the result.

Table 4.6: Prior Distributions for DSGE Model Parameters

Parameters	Range	Density	Mean	SD
$ln\gamma$	$\mathbb{R}$	Normal	0.500	0.250
$ln\pi^*$	$\mathbb{R}$	Normal	1.000	0.500
lnr^*	$\mathbb{R}_+$	Gamma	0.500	0.250
κ	$\mathbb{R}_+$	Gamma	0.300	0.150
τ	$\mathbb{R}_+$	Gamma	2.000	0.500
ψ_1	$\mathbb{R}_+$	Gamma	1.500	0.250
ψ_2	$\mathbb{R}_+$	Gamma	0.125	0.100
ρ_R	$[0,1)$	Beta	0.500	0.200
ρ_g	$[0,1)$	Beta	0.800	0.100
ρ_z	$[0,1)$	Beta	0.300	0.100
σ_R	$\mathbb{R}_+$	Inv.Gamma	0.251	0.139
σ_g	$\mathbb{R}_+$	Inv.Gamma	0.630	0.323
σ_z	$\mathbb{R}_+$	Inv.Gamma	0.875	0.430

Based on this reasoning, the specification of the likelihood function for the structural parameters is derived as follows.

$$p(\boldsymbol{Y}|\boldsymbol{\theta},\lambda) \propto \prod_{t=1}^{T}\left[\left(\frac{\Gamma\left(\frac{v+d}{2}\right)}{\Gamma\left(\frac{v+d_2}{2}\right)}\right)\right](\pi\nu c_t)^{-d_1/2}(2\pi)^{-k/2}|(\lambda T\boldsymbol{\Gamma}^*_{\boldsymbol{xx}}(\boldsymbol{\theta}))^{-1}|^{-1/2}\times$$
$$\left[2^{(\lambda T-k)}M/2\pi^{M(M-1)/4}\prod_{i=1}^{M}\Gamma\left(\frac{(\lambda T-k)+1-i}{2}\right)\right]^{-1}|\lambda T\boldsymbol{\Sigma}^*_{\boldsymbol{u}}(\boldsymbol{\theta})|^{(\lambda T-k)/2} \quad (4.24)$$

As we can see, the derived likelihood function for the structural parameters is free from the statistical parameters. In other words, the likelihood function only includes the structural parameters $\boldsymbol{\theta}$ and the hyperparameter λ.[6]

[6] In the likelihood function, there still exists certain statistical parameters in the term c_t that represents the statistical parameters of the joint distribution. Since our statistical parameters of interest are the one in the conditional distribution, we treated the statistical parameters in c_t as given.

4.4.2 Finding $\widehat{\lambda}$ Using the DSGE-StVAR(λ) Model

For a given λ, the product of the likelihood function (4.24) and the priors of the DSGE model parameters, as shown in Table 4.6, will get $p_\lambda(Y|\theta) \cdot p(\theta)$, and we applied the harmonic mean estimator (Herbst and Schorfheide, 2015; Geweke, 1999) to calculate the marginal density as done in previous literature.

For practical implementation, we assumed a finite set of the hyperparameter λ, $\Lambda = \{0.11, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2, 5, \infty\}$. To get the proper prior, the minimum value of the lambda should satisfy the condition that $\lambda T \geq k + n$ or $\lambda \geq \frac{k+n}{T}$, and it is 0.11 in our examination. Then, for each λ, we attained the $p_\lambda(Y|\theta) \cdot p(\theta)$ via the Random Walk MH algorithm with 300,000 iterations. The first 270,000 draws are burn-ins, and we only kept the remaining 30,000 draws for the estimation. After that, we calculated the marginal density $p_\lambda(Y)$ using the modified harmonic mean estimator. Details of this procedure is described in the appendices.[7]

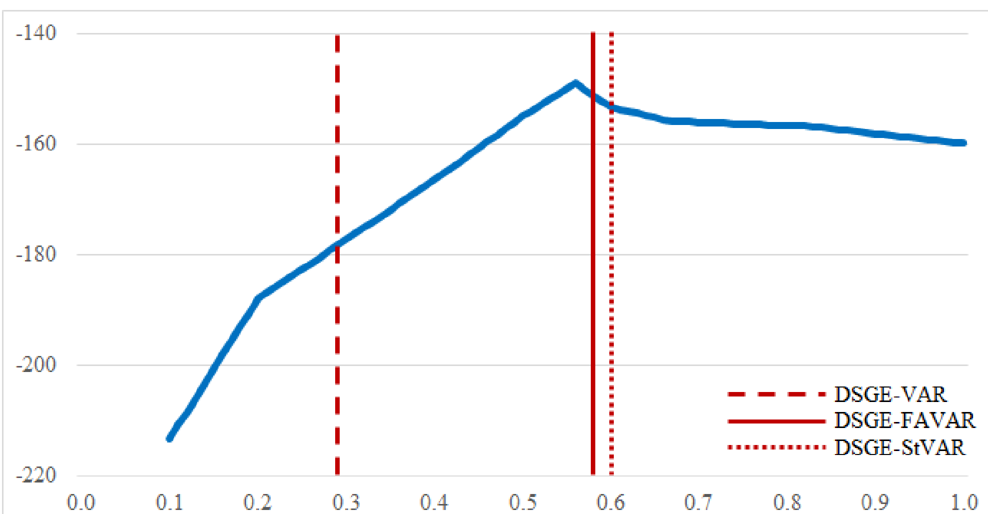

Figure 4.13: Marginal Densities of DSGE-StVAR(2)

[7]For the Random Walk MH algorithm, we set the acceptance rate to be between 20% and 30%. This acceptance rate is regarded as the proper rate for the sampling. In addition, we were required to transform λ to $\lambda^* = \frac{1}{\lambda+1}$ for practical implementation. Following this, we used $\lambda^* = 1$ when we computed the case where $\lambda = \infty$.

The above figure summarizes our analysis. The (red) vertical lines represent the $\widehat{\lambda}$ of different models for the same sample periods as DNS (2004) that is 1955:Q3 to 2001:Q3. Concretely, the dashed line represents that the $\widehat{\lambda} = 0.6$ with the DSGE-VAR(4) model in DNS (2004), the solid line represents that the $\widehat{\lambda} = 1.4$ with the DSGE-FAVAR(4,4) model in Consolo et al. (2009), and the dotted line represents that the $\widehat{\lambda} = 1.5$ with the DSGE-StVAR(2) model. Based on these results, our inferences are as follows.

We can observe that the statistically adequate model provides the significantly higher optimal value of λ. Since the $\widehat{\lambda}$ measures the degree of accordance between the moments of the DSGE model that are used to form the prior and the likelihood function, we interpret that a statistically inadequate model underestimates the suitability of the theoretical model. Furthermore, this inference can be extended to two different directions.

First, we can employ this framework based on the simple New Keynesian model, but by correcting the statistical model, the fitness of the theoretical model to the empirical evidence increases significantly. It shows that statistical adequacy is as important as improving a theoretical model for empirical studies. Going a step further, we suggest that, instead of adding structural frictions into a DSGE model, employing a statistically adequate model would give rise to better inferences for empirical studies.

Second, we suggest that, for a meaningful evaluation of the DSGE model, a statistically adequate estimated model should precede this framework. Higher value of $\widehat{\lambda}$ in the DSGE-StVAR can be regarded, as the StVAR provides a better approximation of the DSGE model than the VAR. Therefore, it would be worthy to evaluate the models within the StVAR boundaries. After that, we can evaluate the theoretical part by comparing the heights (margianl density) with different parameter values. Adolfson et al. (2008) compare the VAR and Vector Error Correction model (VECM) in the DSGE-VAR framework and show that the VECM is a better approximation of their model by investigating the marginal likelihoods. We observe that it relates to a higher value of $\widehat{\lambda}$ of the VECM than the VAR. Thus, the result has some coherence with regard to our inference.

However, we are not insisting that only a statistically adequate model is critical for macroeconomic empirical work. As we can see in Figure 4.13 above, the kinked (blue) line represents the marginal densities of each λ for the entire sample periods: 1955:Q3 to 2016:Q3. Based on the result, we can infer the followings: $\widehat{\lambda} = 1.25$ is still far from $\widehat{\lambda} = \infty$. It can be interpreted that our basic New Keynesian model fails to include numerous empirical evidences. In addition, the optimal value of λ slightly decreases with the full sample, and this can be regarded as additional empirical evidence support the lack of theoretical portion of the DSGE model explaining the real world.

4.5 Conclusion

For empirical studies of the DSGE models, model evaluation is inevitable. Thus, finding the appropriate macroeconomic model for each economy has been given due attention for a long time. Accordingly, DNS (2004) and Del Negro et al. (2007) propose the DSGE-VAR framework as a tool for evaluating the DSGE models. From the econometric perspective, it is a Bayesian VAR model with priors from the DSGE models, and the hyperparameter λ measures the fit between the theoretical information and the statistical information. However, this measurement would be reliable when the statistical model is identified as pointed out by Spanos (1990).

Although the VAR model in DNS (2004) is assumed to be statistically adequate, as Consolo et al. (2009) and Poudyal and Spanos (2013) pointed out, we verified that the VAR model is invalid with macroeconomic time series data that we obtained. The M-S tests indicate that the statistical assumptions imposed on the VAR model such as homoskedasticity and Normality are strictly rejected. Thus, we re-specified the model to remedy the statistical adequacy of the estimated model. The heterogeneous StVAR(2) model is our suggested statistically adequate model.

With the new estimated model, we evaluated the fundamental New Keynesian DSGE model and determined the optimal value of λ. The results suggest the following. A statistically unidentified model underestimates the suitability of the theoretical model to the empirical evidence. It could be an evidence of the fact that the theoretically parsimonious model is effective enough to provide the empirical evidence if the estimated model is statistically adequate. Therefore, acquiring a statistically adequate model is as important as developing a theoretical model.

However, we do not undervalue the importance of developing the theoretical model. The optimal value of λ is still small, and it implies that the fundamental New Keynesian model failed to consider several empirical information. Thus, a room for theoretical development still exist.

References

Adolfson, Malin, Stefan Laséen, Jesper Lindé, and Mattias Villani (2008) "Evaluating an Estimated New Keynesian Small Open Economy Model," *Journal of Economic Dynamics and Control*, Vol. 32, pp. 2690–2721.

An, Sungbae and Frank Schorfheide (2007) "Bayesian analysis of DSGE models," *Econometric reviews*, Vol. 26, pp. 113–172.

Andreou, Elena and Aris Spanos (2003) "Statistical Adequacy and the Testing of Trend Versus Difference Stationarity," *Econometric Reviews*, Vol. 22, pp. 217–237.

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

Clarida, Richard, Jordi Gali, and Mark Gertler (2000) "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *The Quarterly Journal of Economics*, Vol. 115, pp. 147–180.

Consolo, Agostino, Carlo A Favero, and Alessia Paccagnini (2009) "On the Statistical Identification of DSGE Models," *Journal of Econometrics*, Vol. 150, pp. 99–115.

Cooley, Thomas F and Edward C Prescott (1995) "Economic growth and business cycles," *Frontiers of business cycle research*, Vol. 1.

DeJong, David N and Chetan Dave (2011) *Structural macroeconometrics*: Princeton University Press.

Del Negro, Marco and Frank Schorfheide (2004) "Priors from General Equilibrium Models for VARS," *International Economic Review*, Vol. 45, pp. 643–673.

——— (2006) "How Good Is What You've Got? DGSE-VAR as a Toolkit for Evaluating DSGE Models," *Economic Review-Federal Reserve Bank of Atlanta*, Vol. 91, pp. 21–37.

Del Negro, Marco, Frank Schorfheide, Frank Smets, and Rafael Wouters (2007) "On the Fit of New Keynesian Models," *Journal of Business & Economic Statistics*, Vol. 25, pp. 123–143.

Fernández-Villaverde, Jesús (2010) "The econometrics of DSGE models," *SERIEs*, Vol. 1, pp. 3–49.

Fernandez-Villaverde, Jesus, Juan Rubio-Ramirez, and Thomas J Sargent (2007) "ABCs (and Ds) of Understanding VARs," *American Economic Review*, Vol. 97, pp. 1021–1026.

Geweke, J. (1993) "Bayesian Treatment of the Independent Student-t Linear Model," *Journal of Applied Econometrics*, Vol. 8, pp. S19–S40.

Geweke, John (1999) "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, Vol. 18, pp. 1–73.

Giacomini, Raffaella (2013) "The Relationship between DSGE and VAR Models," in *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims*: Emerald Group Publishing Limited, pp. 1–25.

Givens, Gregory E (2017) "Do Data Revisions Matter for DSGE Estimation?" *Journal of Money, Credit and Banking*, Vol. 49, pp. 1385–1407.

Haavelmo, Trygve (1944) "The probability approach in econometrics," *Econometrica: Journal of the Econometric Society*, pp. iii–115.

Herbst, Edward P and Frank Schorfheide (2015) *Bayesian Estimation of DSGE Models*: Princeton University Press.

Ingram, Beth F and Charles H Whiteman (1994) "Supplanting the 'Minnesota' Prior: Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors," *Journal of Monetary Economics*, Vol. 34, pp. 497–510.

Ireland, Peter N (2004) "Technology shocks in the new Keynesian model," *Review of Economics and Statistics*, Vol. 86, pp. 923–936.

King, Robert G, Charles I Plosser, and Sergio T Rebelo (1988) "Production, growth and business cycles: I. The basic neoclassical model," *Journal of monetary Economics*, Vol. 21, pp. 195–232.

Kydland, Finn E and Edward C Prescott (1982) "Time to build and aggregate fluctuations," *Econometrica: Journal of the Econometric Society*, pp. 1345–1370.

——— (1991) "The econometrics of the general equilibrium approach to business cycles," *The Scandinavian Journal of Economics*, pp. 161–178.

Leeper, Eric M and Christopher A Sims (1994) "Toward a modern macroeconomic model usable for policy analysis," *NBER macroeconomics annual*, Vol. 9, pp. 81–118.

Levin, Andrew (1996) "A comparison of alternative monetary policy rules in the Federal Reserve Board's Multi-Country Model," in *BIS Conference papers*, Vol. 2, pp. 340–366.

Litterman, Robert B et al. (1979) "Techniques of Forecasting Using Vector Autoregressions,"Technical report.

Litterman, Robert B (1980) "A Bayesian Procedure for Forecasting with Vector Autoregressions," *Working paper.*

Lucas, Robert E (1976) "Econometric policy evaluation: A critique," in *Carnegie-Rochester conference series on public policy*, Vol. 1, pp. 19–46, Elsevier.

McCallum, Bennet T (1988) "Robustness properties of a rule for monetary policy," in *Carnegie-Rochester conference series on public policy*, Vol. 29, pp. 173–203, Elsevier.

McCallum, Bennett T (1999) "Issues in the design of monetary policy rules," *Handbook of macroeconomics*, Vol. 1, pp. 1483–1530.

Nelson, Charles R and Charles R Plosser (1982) "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, Vol. 10, pp. 139–162.

Newton, Michael A and Adrian E Raftery (1994) "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 3–48.

Poudyal, Niraj (2012) "Confronting Theory with Data: the Case of DSGE Modeling," Ph.D. dissertation, Virginia Polytechnic Institute and State University.

Poudyal, Niraj and Aris Spanos (2013) "Confronting Theory with Data: Model Validation and DSGE Modeling,"Technical report, Working paper, Department of Economics, Virginia Tech, USA.

Prescott, Edward C (1986) "Theory ahead of business-cycle measurement," in *Carnegie-Rochester conference series on public policy*, Vol. 25, pp. 11–44, Elsevier.

Ravenna, Federico (2007) "Vector Autoregressions and Reduced Form Representations of DSGE Models," *Journal of Monetary Economics*, Vol. 54, pp. 2048–2064.

Romer, David (2012) *Advanced Macroeconomics*: McGraw-Hill.

Rotemberg, Julio J and Michael Woodford (1997) "An optimization-based econometric framework for the evaluation of monetary policy," *NBER macroeconomics annual*, Vol. 12, pp. 297–346.

Schorfheide, Frank (2000) "Loss Function-based Evaluation of DSGE models," *Journal of Applied Econometrics*, Vol. 15, pp. 645–670.

Sims, Christopher A (1980) "Macroeconomics and Reality," *Econometrica: Journal of the Econometric Society*, pp. 1–48.

——— (2002) "Solving Linear Rational Expectations Models," *Computational Economics*, Vol. 20, pp. 1–20.

Smets, Frank and Raf Wouters (2003) "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, Vol. 1, pp. 1123–1175.

Smets, Frank and Rafael Wouters (2007) "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *American economic review*, Vol. 97, pp. 586–606.

Spanos, Aris (1990) "The Simultaneous-equations Model Revisited: Statistical Adequacy and Identification," *Journal of Econometrics*, Vol. 44, pp. 87 – 105.

——— (1994) "On Modeling Heteroskedasticity: the Student's t and Elliptical Linear Regression Models," *Econometric Theory*, Vol. 10, pp. 286–315.

——— (1999) *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*: Cambridge University Press.

——— (2014) "Reflections on the LSE Tradition in Econometrics: a Student's Perspective," *Œconomia. History, Methodology, Philosophy*, pp. 343–380.

——— (2015) "Revisiting Haavelmo's Structural Econometrics: Bridging the Gap between Theory and Data," *Journal of Economic Methodology*, Vol. 22, pp. 171–196.

——— (2017) "Mis-specification Testing in Retrospect," *Journal of Economic Surveys*.

Watanabe, Toshiaki (2007) "The application of DSGE-VAR model to macroeconomic data in Japan,"Technical report, ESRI Discussion Paper Series 225-E.

Wickens, Michael (2012) *Macroeconomic theory: a dynamic general equilibrium approach*: Princeton University Press.

Wold, Herman (1938) "A Study in the Analysis of Stationary Time Series," Ph.D. dissertation, Almqvist & Wiksell.

Appendices

Appendix A: The Likelihood Function for the Structural Parameters

In our analysis, since our likelihood function is student's t pdf, we do not have a known form of the posterior distribution, $P(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{Y}, \boldsymbol{\theta}, \lambda)$, and it is challenging to get the likelihood function for the structural parameters induced by

$$p(\boldsymbol{Y}|\boldsymbol{\theta}, \lambda) = \frac{p(\boldsymbol{Y}|\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}, \boldsymbol{\theta}, \lambda) \times p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{\theta}, \lambda)}{p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{Y}, \boldsymbol{\theta}, \lambda)} \tag{A.1}$$

Instead of the full posterior distribution, we apply the posterior kernel of it, $P^*(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{Y}, \boldsymbol{\theta}, \lambda)$ that are proportional to the full posterior distribution. Then,

$$p(\boldsymbol{Y}|\boldsymbol{\theta}, \lambda) \propto \frac{p(\boldsymbol{Y}|\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}, \boldsymbol{\theta}, \lambda) \times p(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{\theta}, \lambda)}{p^*(\boldsymbol{\Phi}, \boldsymbol{\Sigma_u}|\boldsymbol{Y}, \boldsymbol{\theta}, \lambda)} \tag{A.2}$$

The numerator consists of the multiplication of the equation (4.22) and the prior distributions such as Normal distribution for Ψ and Inverted Wishart distribution for Σ_u, i.e.

$$(2\pi)^{-k/2}|\boldsymbol{\Sigma_u}(\lambda T\boldsymbol{\Gamma^*_{xx}}(\boldsymbol{\theta}))^{-1}|^{-1/2} exp\left(-\frac{1}{2}(\boldsymbol{\Psi}^T - \boldsymbol{\Phi^*}(\boldsymbol{\theta}))^T(\boldsymbol{\Sigma_u}(\lambda T\boldsymbol{\Gamma^*_{xx}}(\boldsymbol{\theta}))^{-1})^{-1}(\boldsymbol{\Psi}^T - \boldsymbol{\Phi^*}(\boldsymbol{\theta}))\right)$$

for $\boldsymbol{\Psi}$, and

$$\left(2^{(\lambda T-k)M/2}\pi^{M(M-1)/4}\prod_{i=1}^{M}\Gamma\left(\frac{(\lambda T-k)+1-i}{2}\right)\right)^{-1}$$
$$\times |\lambda T\boldsymbol{\Sigma^*_u}(\boldsymbol{\theta})|^{(\lambda T-k)/2}|\boldsymbol{\Sigma_u}|^{-((\lambda T-k)+M+1)/2} exp\left(-\frac{1}{2}tr(\lambda T\boldsymbol{\Sigma^*_u}(\boldsymbol{\theta})\boldsymbol{\Sigma_u}^{-1})\right)$$

for $\boldsymbol{\Sigma_u}$. The posterior kernel is

$$|\boldsymbol{\Sigma_u}|^{-(1+\frac{\lambda T-k)+M+1}{2}}\left(1+\frac{(\boldsymbol{y_t}-\boldsymbol{\Psi^T y^{t-p}_{t-1}})^{\boldsymbol{T}}\boldsymbol{\Sigma_u^{-1}}(\boldsymbol{y_t}-\boldsymbol{\Psi^T y^{t-p}_{t-1}})}{\nu c_t}\right)^{-\left(\frac{d+\nu}{2}\right)}\times$$
$$exp\left(-\frac{1}{2}(\boldsymbol{\Psi^T}-\boldsymbol{\Phi^*}(\boldsymbol{\theta}))^T(\Sigma_u(\lambda T\boldsymbol{\Gamma^*_{xx}}(\boldsymbol{\theta}))^{-1})^{-1}(\boldsymbol{\Psi^T}-\boldsymbol{\Phi^*}(\boldsymbol{\theta}))\right)\times exp\left(-\frac{1}{2}tr(\lambda T\boldsymbol{\Sigma^*_u}(\boldsymbol{\theta})\boldsymbol{\Sigma_u}^{-1})\right)$$

Finally, the likelihood function for the structural parameters is

$$p(\boldsymbol{Y}|\boldsymbol{\theta},\lambda) \propto \prod_{t=1}^{T}\left[\left(\frac{\Gamma\left(\frac{v+d}{2}\right)}{\Gamma\left(\frac{v+d_2}{2}\right)}\right)\right](\pi\nu c_t)^{-d_1/2}(2\pi)^{-k/2}|(\lambda T\boldsymbol{\Gamma}^*_{\boldsymbol{xx}}(\boldsymbol{\theta}))^{-1}|^{-1/2}\times$$
$$\left[2^{(\lambda T-k)}M/2\pi^{M(M-1)/4}\prod_{i=1}^{M}\Gamma\left(\frac{(\lambda T-k)+1-i}{2}\right)\right]^{-1}|\lambda T\boldsymbol{\Sigma}^*_{\boldsymbol{u}}(\boldsymbol{\theta})|^{(\lambda T-k)/2} \quad \text{(A.3)}$$

where only includes the structural parameters.

Appendix B: Procedures to Find the $\widehat{\lambda}$

We follow the procedures that described in DNS (2004). Firstly, we assume a finite set of the hyperparameter λ such as $\Lambda = \{\lambda_{min}, \cdots, \lambda_{max}\}$. For each $\lambda \in \Lambda$, we use the Metropolis Hastings algorithm as Schorfheide (2000) to generate draws from $p_\lambda(\boldsymbol{Y}|\boldsymbol{\theta}) \times p(\boldsymbol{\theta})$. The likelihood function for the structural parameters $p_\lambda(\boldsymbol{Y}|\boldsymbol{\theta})$ is derived in the equation (A.3) and its factors such as $\boldsymbol{\Gamma}^*_{\boldsymbol{xx}}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}^*_{\boldsymbol{u}}(\boldsymbol{\theta})$ are induced from the following steps.

A solved DSGE model with the algorithm described in Sims (2002) can be represented by a transition equation form such as the following equation.

$$\boldsymbol{s_t} = \boldsymbol{T}(\boldsymbol{\theta})\boldsymbol{s_{t-1}} + \boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{\epsilon_t} \quad \text{(B.1)}$$

Then, we can relate the transition equation to the measurement equation in DNS (2004) as follows.

$$\boldsymbol{y_t} = \boldsymbol{Z}(\boldsymbol{\theta})\boldsymbol{s_t} + \boldsymbol{D}(\boldsymbol{\theta}) + \boldsymbol{v_t} \quad \text{(B.2)}$$

We choose the $\boldsymbol{s_t}$ that gives $\boldsymbol{v_t} = \boldsymbol{0}$ as DNS (2004). If we define the variance-covariance matrices of the shocks as

$$E[\boldsymbol{v_t}\boldsymbol{v_t}'] = \boldsymbol{\Sigma_{vv}}(\boldsymbol{\theta}), \quad E[\boldsymbol{\epsilon_t}\boldsymbol{\epsilon_t}'] = \boldsymbol{\Sigma_{\epsilon\epsilon}}(\boldsymbol{\theta}), \quad E[\boldsymbol{\epsilon_t}\boldsymbol{v_t}'] = \boldsymbol{\Sigma_{\epsilon v}}(\boldsymbol{\theta})$$

then the population moments, $\boldsymbol{\Gamma}^*_{\boldsymbol{yy}}(\boldsymbol{\theta})$, $\boldsymbol{\Gamma}^*_{\boldsymbol{yx}}(\boldsymbol{\theta})$, $\boldsymbol{\Gamma}^*_{\boldsymbol{xx}}(\boldsymbol{\theta})$, can be derived as follows.

$$E[\boldsymbol{y_t}\boldsymbol{y_t}'] = \boldsymbol{Z}(\boldsymbol{\theta})\boldsymbol{\Omega_{ss}}\boldsymbol{Z}(\boldsymbol{\theta})' + \boldsymbol{Z}(\boldsymbol{\theta})\boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{\Sigma_{\epsilon v}}(\boldsymbol{\theta}) + (\boldsymbol{Z}(\boldsymbol{\theta})\boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{\Sigma_{\epsilon v}}(\boldsymbol{\theta}))' + \boldsymbol{\Sigma_{vv}}(\boldsymbol{\theta}) + \boldsymbol{D}(\boldsymbol{\theta})\boldsymbol{D}(\boldsymbol{\theta})' \quad \text{(B.3)}$$

$$E[\boldsymbol{y_t y_{t-p}}'] = \boldsymbol{Z(\theta)T(\theta)^p(\Omega_{ss}Z(\theta)' + R(\theta)\Sigma_{ev}(\theta)) + D(\theta)D(\theta)'} \tag{B.4}$$

where $\boldsymbol{\Omega_{ss}} = E[\boldsymbol{s_t s_t}']$ and it can be obtained by solving the Lyapunov equation: $\boldsymbol{\Omega_{ss}} = \boldsymbol{T(\theta)\Omega_{ss}T(\theta)' + R(\theta)\Sigma_{\epsilon\epsilon}R(\theta)'}$ or $vec(\boldsymbol{\Omega_{ss}}) = [\boldsymbol{I - T(\theta) \otimes T(\theta)}]^{-1}\boldsymbol{vec(R(\theta)R(\theta)')}$.

From the draws, we implement harmonic mean estimator to obtain numerical approximations of the marginal density $p_\lambda(\boldsymbol{Y})$. Among the marginal densities, find the $\widehat{\lambda}$ that gives the highest data density.

Appendix C: Harmonic Mean Estimator of Marginal Likelihood

Herbst and Schorfheide (2015) introduce various methods for computing the marginal likelihood. Among the methods, the modified harmonic mean estimator proposed by Geweke (1999) is generally used in the DSGE literature.

Suppose that $g(\boldsymbol{\theta})$ is a probability density function. Then we can see that the following equality is satisfied.

$$E\left(\frac{g(\boldsymbol{\theta})}{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}\right) = \int \frac{g(\boldsymbol{\theta})}{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})} p(\boldsymbol{\theta}|\boldsymbol{Y})d\boldsymbol{\theta} = \int \frac{g(\boldsymbol{\theta})}{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}\frac{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{Y})}d\boldsymbol{\theta}$$

$$= \int \frac{g(\boldsymbol{\theta})}{p(\boldsymbol{Y})}d\boldsymbol{\theta} = \frac{1}{p(\boldsymbol{Y})}\int g(\boldsymbol{\theta})d\boldsymbol{\theta} = \frac{1}{p(\boldsymbol{Y})}$$

Thus, we can estimate the marginal likelihood as follows.

$$p(\boldsymbol{Y}) = E\left(\frac{g(\boldsymbol{\theta})}{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}\right)^{-1} \approx \left(\frac{1}{N}\sum_{i=1}^{N}\frac{g(\theta_i)}{p(\boldsymbol{Y}|\theta_i)p(\theta_i)}\right)^{-1}$$

If we set $g(\theta) = p(\theta)$ (Newton and Raftery, 1994), then the equation above is simplified as

$$p(\boldsymbol{Y}) \approx \left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{p(\boldsymbol{Y}|\theta_i)}\right)^{-1}$$

where is called the harmonic mean estimator. However, if $p(\boldsymbol{Y}|\theta_i)$ is close to 0 then its reciprocal goes explosive which makes the estimator unstable. To avoid this, Geweke (1999) proposed to set $g(\theta)$ as truncate normal such as

$$g(\boldsymbol{\theta}) = \tau^{-1}(2\pi)^{-k/2}|\boldsymbol{\Sigma}|^{-1/2}exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right)I[(\boldsymbol{\theta}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\leq F^{-1}_{\chi^2_k}(\tau)]$$

where k represent the dimension of θ and I is the indicator function that returns 1 if the inequality is satisfied and zero otherwise.

Appendix D: The Sims representation of the Simple New Keynesian Model

In DNS (2004), a system of equations derived from the simple New Keynesian model is log-linearized as follows.

$$\widetilde{x}_t = E_t\widetilde{x}_t - \frac{1}{\tau}(R_t - E_t\widetilde{\pi}_{t+1}) + (1-\rho_g)\widetilde{g}_t + \rho_z\frac{1}{\tau}\widetilde{z}_t \tag{D.1}$$

$$\widetilde{\pi}_t = \frac{\gamma}{r^*}E_t\widetilde{\pi}_{t+1} + \kappa(\widetilde{x}_t - \widetilde{g}_t) \tag{D.2}$$

$$\widetilde{R}_t = \rho_R\widetilde{R}_{t-1} + (1-\rho_R)(\psi_1\widetilde{\pi}_t + \psi_2\widetilde{x}_t) + \epsilon_{R,t} \tag{D.3}$$

$$\widetilde{g}_t = \rho_g\widetilde{g}_{t-1} + \epsilon_{g,t} \tag{D.4}$$

$$\widetilde{z}_t = \rho_z\widetilde{z}_{t-1} + \epsilon_{z,t} \tag{D.5}$$

As Sims (2002), the linearized system of equations can be represented in the following form

$$\boldsymbol{\Gamma_0\widetilde{Z}_t = \Gamma_1\widetilde{Z}_{t-1} + C + \Psi\epsilon_t + \Pi\eta_t}$$

$t = 1, ..., T$, where C is a vector of constants, ϵ_t represents an exogenously evolving random disturbance, and η_t represents an expectational error which is treated as determined by the model solution instead of exogenously given, with the condition $E_t(\boldsymbol{\eta_{t+1}}) = 0$ for all t. Then the matrices are

$$\widetilde{\boldsymbol{Z}}_t = \begin{bmatrix} \widetilde{x}_t \\ \widetilde{\pi}_t \\ \widetilde{R}_t \\ \widetilde{R}^*_t \\ \widetilde{g}_t \\ \widetilde{z}_t \\ E_t\widetilde{x}_{t+1} \\ E_t\widetilde{\pi}_{t+1} \end{bmatrix}, \quad \boldsymbol{\epsilon}_t = \begin{bmatrix} \epsilon_{R,t} \\ \epsilon_{g,t} \\ \epsilon_{z,t} \end{bmatrix}, \quad \boldsymbol{\eta}_t = \begin{bmatrix} \widetilde{x}_t - E_{t-1}\widetilde{x}_t \\ \widetilde{\pi}_t - E_{t-1}\widetilde{\pi}_t \end{bmatrix}$$

$$\boldsymbol{\Gamma}_0 = \begin{bmatrix} 1 & 0 & \tau^{-1} & 0 & -(1-\rho_g) & -\rho_z\tau^{-1} & -1 & -\tau^{-1} \\ -\kappa & 1 & 0 & 0 & \kappa & 0 & 0 & -\frac{\gamma}{r^*} \\ 0 & 0 & 1 & -(1-\rho_R) & 0 & 0 & 0 & 0 \\ -\Psi_2 & -\Psi_1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{\Gamma}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_R & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_g & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Psi} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{\Pi} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Appendix E: Derivation of Smets and Wouters (2007) Model

Smets and Wouters(2007) (hereafter SW(2007)), set up a DSGE model which is based on Christiano, Eichenbaum, and Evan's paper in 2005 (CEE(2005)). In this paper, they present and estimate DSGE model with some frictions of wages and prices for U.S., and these frictions follows the way Calvo proposed in 1983. They estimate the model parameters using Bayesian approach.

Model

In this model there are several agents, households, labor union, labor packers, final good producers, intermediate good producers, government, and I will explain how they are interrelated with each other as solving the model.

Final Goods Producers

The final good Y_t consists of a continuum of intermediate good $Y_t(i)$. Thus, final good producers buy intermediate goods on the market, package Y_t, and then they resell the final good to consumers, investors and the government in a perfectly competitive market.

Based on the setup, final good producers face following profit maximization problem.

$$\max_{Y_t, Y_t(i)} P_t Y_t - \int_0^1 P_t(i) Y_t(i) di \quad \text{s.t} \quad \left[\int_0^1 G\left(\frac{Y_t(i)}{Y_t}; \epsilon_t^p \right) di \right] = 1$$

where P_t and $P_t(i)$ are the price of the final and intermediate goods respectively. G is assumed that it is a strictly concave and increasing function characterized by $G(1) = 1$. ϵ_t^p is an exogenous process reflecting shocks to the aggregator function. Then they change the elasticity of demand and therefore in the mark-up. It is constrained by $\epsilon_t^p \in (0, \infty)$.

The Lagrange equation for this problem,

$$L = P_t Y_t - \int_0^1 P_t(i) Y_t(i) di + \mu_{f,t} \left(\int_0^1 G\left(\frac{Y_t(i)}{Y_t} \right) di - 1 \right)$$

we can have First Order Necessary Condition(F.O.N.C or F.O.C)s as follows:

$$(\partial Y_t): \quad P_t = \frac{\mu_{f,t}}{Y_t}\int_0^1 G'\left(\frac{Y_t(i)}{Y_t}\right)\frac{Y_t(i)}{Y_t}di$$

$$(\partial Y_t(i)): \quad P_t(i) = \mu_{f,t} G'\left(\frac{Y_t(i)}{Y_t}\right)\frac{1}{Y_t}$$

These F.O.Cs can be rearranged by follows:

$$(\partial Y_t(i)): \quad P_t(i) = \mu_{f,t} G'\left(\frac{Y_t(i)}{Y_t}\right)\frac{1}{Y_t} \Rightarrow G'\left(\frac{Y_t(i)}{Y_t}\right) = \frac{1}{\mu_{f,t}}P_t(i)Y_t$$

Then,

$$\frac{Y_t(i)}{Y_t} = G'^{-1}\left[\frac{1}{\mu_{f,t}}P_t(i)Y_t\right] \Rightarrow Y_t(i) = Y_t G'^{-1}\left[\frac{1}{\mu_{f,t}}P_t(i)Y_t\right]$$

From the first F.O.C, we can see that

$$\frac{Y_t}{\mu_{f,t}} = \frac{\int_0^1 G'\left(\frac{Y_t(i)}{Y_t}\right)\frac{Y_t(i)}{Y_t}di}{P_t}$$

Thus, put this equation to the previous one, we can get

$$Y_t(i) = Y_t G'^{-1}\left[\frac{P_t(i)}{P_t}\int_0^1 G'\left(\frac{Y_t(i)}{Y_t}\right)\frac{Y_t(i)}{Y_t}di\right]$$

The assumptions imposed on the function G imply that the demand for input $Y_t(i)$ is decreasing in its relative price, while the elasticity of demand is a positive function of the relative price.

Intermediate Goods Producers

An intermediate goods producer i produces his (her) goods using following technology:

$$Y_t(i) = \varepsilon_t^a \overline{K}_t(i)^\alpha[\gamma^t L_t(i)]^{1-\alpha} - \gamma^t\Phi$$

where $\overline{K}_t(i)$ is capital services used in production, $L_t(i)$ is aggregate labor input, Φ is a fixed cost, and γ^t represents the labor-augmenting deterministic growth rate in the economy. ε_t^a is total factor productivity which follows the process:

$$ln\varepsilon_t^a = (1-\rho_z)ln\varepsilon^a + \rho_z ln\varepsilon_{t-1}^a + \eta_t^a, \quad \eta_t^a \sim N(0, \sigma_a)$$

Since the firm's profit is given by:

$$P_t(i)Y_t(i) - W_t L_t(i) - R_t^k \overline{K}_t(i)$$

it's cost minimization problem will be

$$\min_{L_t(i), \overline{K}_t(i)} W_t L_t(i) + R_t^k \overline{K}_t(i) \quad \text{s.t} \quad Y_t(i) = \varepsilon_t^a \overline{K}_t(i)^\alpha [\gamma^t L_t(i)]^{1-\alpha} - \gamma^t \Phi$$

where W_t is the aggregate nominal wage rate and R_t^k is the rental rate on capital. For the Lagrangian function,

$$L = W_t L_t(i) + R_t^k \overline{K}_t(i) + \Theta_t(i) \left\{ Y_t(i) - \varepsilon_t^a \overline{K}_t(i)^\alpha [\gamma^t L_t(i)]^{1-\alpha} + \gamma^t \Phi \right\}$$

it yields following F.O.Cs:

$$(\partial L_t(i)): \quad \Theta_t(i)\gamma^{(1-\alpha)t}(1-\alpha)\varepsilon_t^a \overline{K}_t(i)^\alpha L_t(i)^{-\alpha} = W_t$$

$$(\partial \overline{K}_t(i)): \quad \Theta_t(i)\gamma^{(1-\alpha)t}\alpha\varepsilon_t^a \overline{K}_t(i)^{\alpha-1} L_t(i)^{1-\alpha} = R_t^k$$

Combining these two F.O.Cs can be represented that

$$\frac{W_t}{R_t^k} = \frac{\Theta_t(i)\gamma^{(1-\alpha)t}(1-\alpha)\varepsilon_t^a \overline{K}_t(i)^\alpha L_t(i)^{-\alpha}}{\Theta_t(i)\gamma^{(1-\alpha)t}\alpha\varepsilon_t^a \overline{K}_t(i)^{\alpha-1} L_t^{1-\alpha}} = \frac{1-\alpha}{\alpha}\frac{\overline{K}_t(i)}{L_t(i)}$$

Noting that the capital-labor ratio is equal across firms, so we can ignore i, then

$$\frac{W_t}{R_t^k} = \frac{1-\alpha}{\alpha}\frac{\overline{K}_t}{L_t} \quad \Rightarrow \quad \overline{K}_t = \frac{\alpha}{1-\alpha}\frac{W_t}{R_t^k} L_t$$

Since the Lagrangian multiplier $\Theta_t(i)$ is shadow price which means how much the profit or cost will be changed by relaxing the constraint so it has identical implication to marginal cost, $\Theta_t(i)$ equals marginal cost MC_t. Thus, we can derive the MC_t by deriving $\Theta_t(i)$ and the marginal cost is also the same for all firms:

$$\Theta_t = MC_t = \alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)}W_t^{(1-\alpha)}(R_t^k)^\alpha \gamma^{-(1-\alpha)t}(\varepsilon_t^a)^{-1} \tag{4.25}$$

These intermediate goods producers set their prices under the Calvo pricing process with partial indexation. In other words, each firm can set its price with probability $1-\xi_t$ at time t, so if a firm can have chance to set the price, they set the price considering that they would not be able to have chance to optimize the price in the future. Even though some firms cannot set the price at time t with probability ξ_t, they can partially index the price to past inflation. Thus, its optimization problem will be:

$$\max_{\widetilde{P}_t(i)} E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \left[\widetilde{P}_t(i)\left(\Pi_{l=1}^s \pi_{t+l-1}^{l_p} \pi_*^{1-l_p}\right) - MC_{t+s}\right] Y_{t+s}(i)$$

$$\text{s.t} \quad Y_{t+s}(i) = Y_{t+s} G'^{-1}\left(\frac{\widetilde{P}_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s}\right)$$

where

$$\pi_t = \frac{P_t}{P_{t-1}}, \quad \tau_t = \int_0^1 G'\left(\frac{Y_t(i)}{Y_t}\right)\frac{Y_t(i)}{Y_t} di$$

$\widetilde{P}_t(i)$ is the newly set price, ξ_p is the Calvo probability of being allowed to optimize a firm's price, l_p is the degree of indexation to lagged inflation, $\left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}}\right]$ is the nominal discount factor for firms.

It can be the unconstrained optimization problem:

$$\max_{\widetilde{P}_t(i)} E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \left[\widetilde{P}_t(i)\left(\Pi_{l=1}^s \pi_{t+l-1}^{l_p} \pi_*^{1-l_p}\right) - MC_{t+s}\right] Y_{t+s} G'^{-1}\left(\frac{P_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s}\right)$$

By defining

$$X_{t,s} = \left\{ \begin{array}{ll} 1 \quad \text{for} \quad s=0 & \\ (\Pi_{l=1}^s \pi_{t+l-1}^{l_p} \pi_*^{1-l_p}) \quad \text{for} \quad s = 1, ..., \infty & \end{array} \right\}$$

we can rewrite the equation as

$$\max_{\widetilde{P}_t(i)} E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \left[\widetilde{P}_t(i) X_{t,s} - MC_{t+s}\right] Y_{t+s} G'^{-1}\left(\frac{P_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s}\right)$$

Differentiating with respect to $\widetilde{P}_t(i)$ give the optimal price for this problem, i.e;

$$E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \left[X_{t,s} Y_{t+s}(i) + \left(\widetilde{P}_t(i) X_{t,s} - MC_{t+s} \right) Y_{t+s} \frac{\frac{X_{t,s}}{P_{t+s}} \tau_{t+s}}{G'' \left(G'^{-1} \left(\frac{\widetilde{P}_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s} \right) \right)} \right] = 0$$

$$E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} Y_{t+s}(i) \left[X_{t,s} + \left(\widetilde{P}_t(i) X_{t,s} - MC_{t+s} \right) \frac{1}{G'^{-1}(z_{t+s})} \frac{\frac{X_{t,s}}{P_{t+s}} \tau_{t+s}}{G''\left(G'^{-1}(z_{t+s})\right)} \right] = 0$$

$$\text{because } Y_{t+s}(i) = Y_{t+s} G'^{-1} \left(\frac{P_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s} \right) \Rightarrow Y_{t+s} = \frac{Y_{t+s}(i)}{G'^{-1} \left(\frac{P_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s} \right)}.$$

Then,

$$E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} Y_{t+s}(i) \left[\widetilde{P}_t(i) X_{t,s} + \left(\widetilde{P}_t(i) X_{t,s} - MC_{t+s} \right) \frac{1}{G'^{-1}(z_{t+s})} \frac{\frac{\widetilde{P}_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s}}{G''\left(G'^{-1}(z_{t+s})\right)} \right] = 0$$

$$E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} Y_{t+s}(i) \left[\widetilde{P}_t(i) X_{t,s} + \left(\widetilde{P}_t(i) X_{t,s} - MC_{t+s} \right) \frac{1}{G'^{-1}(z_{t+s})} \frac{z_{t+s}}{G''\left(G'^{-1}(z_{t+s})\right)} \right] = 0$$

$$E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} Y_{t+s}(i) \left[\widetilde{P}_t(i) X_{t,s} + \left(\widetilde{P}_t(i) X_{t,s} - MC_{t+s} \right) \frac{1}{G'^{-1}(z_{t+s})} \frac{G'(x_{t+s})}{G''(x_{t+s})} \right] = 0 \tag{4.26}$$

Digression

For differentiation of inverse function w.r.t $\widetilde{P}_t(i)$ part, we can think of followings:

$$\frac{\partial}{\partial \widetilde{P}_t(i)} K = \frac{\partial K}{\partial \widetilde{P}_t(i)} \quad \text{where} \quad K = \frac{\widetilde{P}_t(i) X_{t,s}}{P_{t+s}} \tau_{t+s}$$

$$\Rightarrow \frac{\partial}{\partial \widetilde{P}_t(i)} G'\left(G'^{-1}(K)\right) = \frac{\partial K}{\partial \widetilde{P}_t(i)}$$

$$\Rightarrow G''\left(G'^{-1}(K)\right)\left(G''^{-1}(K)\right) \frac{\partial K}{\partial \widetilde{P}_t(i)} = \frac{\partial K}{\partial \widetilde{P}_t(i)} \Rightarrow G''\left(G'^{-1}(K)\right)\left(G''^{-1}(K)\right) = 1$$

$$\Rightarrow G''^{-1}(K) = \frac{1}{G''(G'^{-1}(K))}$$

$$\Rightarrow \frac{\partial G'^{-1}(K)}{\partial \widetilde{P}_t(i)} \frac{\widetilde{P}_t(i)}{\partial K} = \frac{1}{G''(G'^{-1}(K))}$$

Therefore,

$$\frac{\partial G'^{-1}(K)}{\partial \widetilde{P}_t(i)} = \frac{1}{G''(G'^{-1}(K))} \times \frac{\partial K}{\partial \widetilde{P}_t(i)} = \frac{\frac{\partial K}{\partial \widetilde{P}_t(i)}}{G''(G'^{-1}(K))} = \frac{\frac{X_{t,s}}{P_{t+s}}\tau_{t+s}}{G''\left(G'^{-1}\left(\frac{\widetilde{P}_t(i)X_{t,s}}{P_{t+s}}\tau_{t+s}\right)\right)} \blacksquare$$

The aggregate price index is in this case would be:

$$P_t = (1-\xi_p)\widetilde{P}_t(i)G'^{-1}\left(\frac{P_t(i)\tau_t}{P_t}\right) + \xi_p \pi_{t-1}^{l_p}\pi_*^{1-l_p}P_{t-1}G'^{-1}\left(\frac{\pi_{t-1}^{l_p}\pi_*^{1-l_p}P_{t-1}\tau_t}{P_t}\right)$$

Intuitively, we can think that the price level P_t is the weighted average between optimized price and previous price with partial indexation.

Households

Household j chooses consumption $C_t(j)$, hours worked $L_t(j)$, bonds $B_t(j)$, investment $I_t(j)$ and capital utilization $Z_t(j)$ to maximize the following objective function:

$$E_t \sum_{s=0}^{\infty} \beta^s \left[\frac{1}{1-\sigma_c}(C_{t+s}(j) - \lambda C_{t+s-1})^{1-\sigma_c}\right] \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_{t+s}(j)^{1+\sigma_l}\right)$$

subject to the budget constraint:

$$C_{t+s}(j) + I_{t+s}(j) + \frac{B_{t+s}(j)}{\varepsilon_t^b R_{t+s} P_{t+s}} - T_{t+s} \leq$$

$$\frac{B_{t+s-1}(j)}{P_{t+s}} + \frac{W_{t+s}^h(j)L_{t+s}(j)}{P_{t+s}} + \frac{R_{t+s}^k Z_{t+s}(j)K_{t+s-1}(j)}{P_{t+s}} - a(Z_{t+s}(j))K_{t+s-1}(j) + \frac{Div_{t+s}}{P_{t+s}}$$

and the capital accumulation equation:

$$K_t(j) = (1-\delta)K_{t-1}(j) + \varepsilon_t^i \left[1 - S\left(\frac{I_t(j)}{I_{t-1}(j)}\right)\right] I_t(j)$$

where λ is the parameter capturing external habit formation. We can consider ε_t^b as an exogenous premium in the return to bonds reflecting inefficiencies in the financial market sector leading to some premium on deposit rate versus the risk free rate set by the central bank, or a risk premium that households require to hold the one period bond. δ is the depreciation rate, $S(\cdot)$ is the adjustment cost function, with $S(\gamma) = 0,\ S'(\gamma) = 0,\ S''(\cdot) > 0$, and ε_t^q is a stochastic shock to the price of investment relative to consumption goods. T_{t+s} is lump sum taxes or subsidies and Div_{t+s} is the dividends distributed by the intermediate goods producers and the labor unions.

In addition, households choose the utilization rate of capital. The amount of effective capital that households can rent to the firm is:

$$\overline{K}_t(j) = Z_t(j)K_{t-1}(j)$$

Then the income from renting capital services is $R_t^k Z_t(j)K_{t-1}(j)$ and the cost of changing capital utilization is $P_t a(Z_t(j))K_{t-1}(j)$.

The Lagrange function for the households problem can be expressed by followings:

$$L = E_t \sum_{s=0}^{\infty} \beta^s \left\langle \left[\frac{1}{1-\sigma_c}(C_{t+s}(j) - \lambda C_{t+s-1})^{1-\sigma_c}\right] \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_{t+s}(j)^{1+\sigma_l}\right)\right.$$

$$+\Xi_t \left\{ \frac{B_{t+s-1}(j)}{P_{t+s}} + \frac{W_{t+s}^h(j)L_{t+s}(j)}{P_{t+s}} + \frac{R_{t+s}^k Z_{t+s}(j)K_{t+s-1}(j)}{P_{t+s}} - a(Z_{t+s}(j))K_{t+s-1}(j) + \frac{Div_{t+s}}{P_{t+s}}\right.$$

$$\left. -C_{t+s}(j) - I_{t+s}(j) - \frac{B_{t+s}(j)}{\varepsilon_t^b R_{t+s} P_{t+s}} + T_{t+s} \right\}$$

$$\left. + \Xi_t^k \left\{ (1-\delta)K_{t-1}(j) + \varepsilon_t^i \left[1 - S\left(\frac{I_t(j)}{I_{t-1}(j)}\right)\right] I_t(j) - K_t(j) \right\} \right\rangle$$

By differentiating with respect to 5 control variables, consumption $C_t(j)$, hours worked $L_t(j)$, bonds $B_t(j)$, investment $I_t(j)$, and capital utilization $Z_t(j)$ respectively, we can get F.O.Cs as follows:

F.O.C with respect to $C_t(j)$ is

$$(C_t(j) - \lambda C_{t-1})^{-\sigma_c} \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t(j)^{1+\sigma_l}\right) - \Xi_t = 0$$

By rearranging it, we can see that

$$\Xi_t = \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t(j)^{1+\sigma_l}\right)(C_t(j) - \lambda C_{t-1})^{-\sigma_c}$$

F.O.C with respect to $L_t(j)$ is

$$\left[\frac{1}{1-\sigma_c}(C_t(j) - \lambda C_{t+s-1})^{1-\sigma_c}\right]\exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t(j)^{1+\sigma_l}\right)(\sigma_c - 1)L_t(i)^{\sigma_l} + \Xi_t \frac{W_t^h(j)}{P_t} = 0$$

Then,

$$\left[\frac{1}{1-\sigma_c}(C_t(j) - \lambda C_{t+s-1})^{1-\sigma_c}\right]\exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t(j)^{1+\sigma_l}\right)(\sigma_c - 1)L_t(i)^{\sigma_l} = -\Xi_t \frac{W_t^h(j)}{P_t}$$

F.O.C with respect to $B_t(j)$ is

$$-\Xi_t \frac{1}{\varepsilon_t^b R_t P_t} + \beta E_t\left[\Xi_{t+1}\frac{1}{P_{t+1}}\right] = 0$$

By rearranging it,

$$\Xi_t = \beta \varepsilon_t^b R_t E_t\left[\Xi_{t+1}\frac{P_t}{P_{t+1}}\right] = \beta \varepsilon_t^b R_t E_t\left[\Xi_{t+1}\frac{1}{\pi_{t+1}}\right]$$

F.O.C with respect to $I_t(j)$ is

$$\begin{aligned} -\Xi_t + \Xi_t^k \varepsilon_t^i &\left\{-S'\left(\frac{I_t(j)}{I_{t-1}(j)}\right)\frac{I_t(j)}{I_{t-1}(j)} + \left[1 - S\left(\frac{I_t(j)}{I_{t-1}(j)}\right)\right]\right\} \\ &+\beta E_t\left\{\Xi_{t+1}^k \varepsilon_{t+1}^i \left[S'\left(\frac{I_{t+1}(j)}{I_t(j)}\right)\frac{I_{t+1}(j)}{I_t(j)^2}\right] I_{t+1}(j)\right\} = 0 \end{aligned}$$

By rearranging it,

$$\begin{aligned} \Xi_t = \Xi_t^k \varepsilon_t^i &\left\{1 - S\left(\frac{I_t(j)}{I_{t-1}(j)}\right) - S'\left(\frac{I_t(j)}{I_{t-1}(j)}\right)\frac{I_t(j)}{I_{t-1}(j)}\right\} \\ &+\beta E_t\left\{\Xi_{t+1}^k \varepsilon_{t+1}^i S'\left(\frac{I_{t+1}(j)}{I_t(j)}\right)\left(\frac{I_{t+1}(j)}{I_t(j)}\right)^2\right\} \end{aligned}$$

F.O.C with respect to $\overline{K}_t(j)$ is

$$-\Xi_t^k + \beta E_t\left[\Xi_{t+1}\left(\frac{R_{t+1}^k Z_{t+1}(j)}{P_{t+1}} - a(Z_{t+1}(j))\right) + \Xi_{t+1}^k(1-\delta)\right] = 0$$

By rearranging it,

$$\Xi_t^k = \beta E_t \left[\Xi_{t+1} \left(\frac{R_{t+1}^k Z_{t+1}(j)}{P_{t+1}} - a(Z_{t+1}(j)) \right) + \Xi_{t+1}^k (1-\delta) \right]$$

F.O.C with respect to $Z_t(j)$ is

$$\Xi_t \left(\frac{R_t^k K_{t-1}(j)}{P_t} - a'(Z_t(j)) K_{t-1}(j) \right) = 0$$

By rearranging it,

$$\frac{R_t^k}{P_t} = a'(Z_t)$$

In equilibrium households will make identical choices for consumption, hours worked, bonds, investment and capital utilization, so we can ignore the j index. Then,

$$(\partial C_t) \quad \Xi_t = \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{1+\sigma_l} \right) (C_t - \lambda C_{t-1})^{-\sigma_c} \tag{4.27}$$

$$(\partial L_t) \quad \left[\frac{1}{1-\sigma_c} (C_t - \lambda C_{t-1})^{1-\sigma_c} \right] \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{1+\sigma_l} \right) (\sigma_c - 1) L_t^{\sigma_l} = -\Xi_t \frac{W_t^h}{P_t} \tag{4.28}$$

$$(\partial B_t) \quad \Xi_t = \beta \varepsilon_t^b R_t E_t \left[\frac{\Xi_{t+1}}{\pi_{t+1}} \right] \tag{4.29}$$

$$\begin{aligned}(\partial I_t) \quad \Xi_t = \Xi_t^k \varepsilon_t^i & \left(1 - S\left(\frac{I_t}{I_{t-1}} \right) - S'\left(\frac{I_t}{I_{t-1}} \right) \frac{I_t}{I_{t-1}} \right) \\ & + \beta E_t \left[\Xi_{t+1}^k \varepsilon_{t+1}^i S'\left(\frac{I_{t+1}}{I_t} \right) \left(\frac{I_{t+1}}{I_t} \right)^2 \right]\end{aligned} \tag{4.30}$$

$$(\partial \overline{K}_t) \quad \Xi_t^k = \beta E_t \left[\Xi_{t+1} \left(\frac{R_{t+1}^k}{P_{t+1}} Z_{t+1} - a(Z_{t+1}) \right) + \Xi_{t+1}^k (1-\delta) \right] \tag{4.31}$$

$$(\partial Z_t) \quad \frac{R_t^k}{P_t} = a'(Z_t) \tag{4.32}$$

where Ξ_t and Ξ_t^k are the Lagrange multiplier associated with the budget and capital accumulation constraint respectively. Tobin's Q is $Q_t = \frac{\Xi_t^k}{\Xi_t}$ and equals one in the absence of adjustment costs.

Intermediate Labor Union Sector

Similar to the case of intermediate good producers, in the labor market, intermediate labor union which gathers homogeneous labor services from the households and differentiate the labor services, sets wages subject to a Calvo scheme and offers those labor services to labor packers. Labor used by the intermediate goods producers L_t is a composite made of those differentiated labor services $L_t(l)$. Labor packers buy the differentiated labor services, package L_t, and offer it to the intermediate goods producers.

For the labor packers, its maximization problem is:

$$\max_{L_t, L_t(i)} W_t L_t - \int_0^1 W_t(l) L_t(l) dl$$

$$\text{s.t.} \quad \left[\int_0^1 H\left(\frac{L_t(l)}{L_t}; \epsilon_t^w\right) dl \right] = 1$$

where W_t and $W_t(l)$ are the price of the composite and intermediate labor services respectively. As the assumptions imposed on the function G, H is assumed that it is a strictly concave and increasing function characterized by $H(1) = 1$, and ϵ_t^w is an exogenous process reflecting shocks to the aggregator function then they result in changes in the elasticity of demand and therefore in the mark up. In addition, it is constrained by $\epsilon_t^w \in (0, \infty)$.

Using the Lagrange equation for this problem,

$$L = W_t L_t - \int_0^1 W_t(l) L_t(l) dl + \mu_{p,t} \left(\int_0^1 H\left(\frac{L_t(l)}{L_t}\right) dl - 1 \right)$$

Then F.O.Cs are:

$$(\partial L_t): \quad W_t = \frac{\mu_{p,t}}{L_t} \int_0^1 H'\left(\frac{L_t(l)}{L_t}\right) \frac{L_t(l)}{L_t} dl$$

$$(\partial L_t(l)): \quad W_t(l) = \mu_{p,t} H'\left(\frac{L_t(l)}{L_t}\right) \frac{1}{L_t}$$

These F.O.Cs can be rearranged by follows:

$$(\partial L_t(l)): \quad W_t(l) = \mu_{p,t} H'\left(\frac{L_t(l)}{L_t}\right) \frac{1}{L_t} \Rightarrow H'\left(\frac{L_t(l)}{L_t}\right) = \frac{1}{\mu_{p,t}} W_t(l) L_t$$

Then,

$$\frac{L_t(l)}{L_t} = H'^{-1}\left[\frac{1}{\mu_{p,t}}W_t(l)L_t\right] \Rightarrow L_t(l) = L_t H'^{-1}\left[\frac{1}{\mu_{p,t}}W_t(l)L_t\right]$$

Based on the first F.O.C, we can see that

$$\frac{L_t}{\mu_{p,t}} = \frac{\int_0^1 H'\left(\frac{L_t(l)}{L_t}\right)\frac{L_t(l)}{L_t}dl}{W_t}$$

Thus, put this equation to the previous one, we can get

$$L_t(l) = L_t H'^{-1}\left[\frac{W_t(l)}{W_t}\int_0^1 H'\left(\frac{L_t(l)}{L_t}\right)\frac{L_t(l)}{L_t}dl\right]$$

The labor union can set wages under Calvo pricing scheme with partial indexation as intermediate goods producers case. In other words, they can optimize their wage with probability $1-\xi_w$ in every period, if they cannot re-optimize their wage with probability ξ_w, then the wage, $W_t(l)$ will be increased at the deterministic growth rate γ and weighted average of the steady state inflation π_* and of last period's inflation, π_{t-1}. For those can optimize the wage, the problem is to choose a optimal wage $\widetilde{W}_t(l)$ that maximizes the wage income in all states of nature where the labor union is stuck with that wage in the future. Thus their optimization problem would be:

$$\max_{\widetilde{W}_t(l)} E_t\sum_{s=0}^{\infty}\zeta_w^s\left[\frac{\beta^s\Xi_{t+s}P_t}{\Xi_t P_{t+s}}\right]\left(W_{t+s}(l)-W_{t+s}^h\right)L_{t+s}(l)$$

subject to

$$L_{t+s}(l) = \left(\frac{W_{t+s}(l)}{W_{t+s}}\right)^{-\frac{1+\lambda_{w,t+s}}{\lambda_{w,t+s}}}L_{t+s}$$

$$W_{t+s}(l) = \widetilde{W}_t(l)(\Pi_{l=1}^s\gamma\pi_{t+l-1}^{l_w}\pi_*^{1-l_w}) \quad \text{for} \quad s=1,...,\infty$$

From the unconstrained problem, the problem will be

$$L = E_t\sum_{s=0}^{\infty}\zeta_w^s\left[\frac{\beta^s\Xi_{t+s}P_t}{\Xi_t P_{t+s}}\right]\left(\widetilde{W}_t(l)(\Pi_{l=1}^s\gamma\pi_{t+l-1}^{l_w}\pi_*^{1-l_w})-W_{t+s}^h\right)\left(\frac{W_{t+s}(l)}{W_{t+s}}\right)^{-\frac{1+\lambda_{w,t+s}}{\lambda_{w,t+s}}}L_{t+s}$$

and using the definition

$$X_{t,s} = \left\{ \begin{array}{ll} 1 \quad \text{for} \quad s=0 & \\ (\Pi_{l=1}^{s} \pi_{t+l-1}^{l_w} \pi_*^{1-l_w}) \quad \text{for} \quad s = 1, ..., \infty & \end{array} \right\}$$

it can be rewrite as follows:

$$L = E_t \sum_{s=0}^{\infty} \zeta_w^s \left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \right] \left(\widetilde{W}_t(l) X_{t,s} - W_{t+s}^h \right) \left(\frac{\widetilde{W}_t(l) X_{t,s}}{W_{t+s}} \right)^{-\frac{1+\lambda_{w,t+s}}{\lambda_{w,t+s}}} L_{t+s}$$

Then the F.O.C with respect to $\widetilde{W}_t(l)$ is

$$E_t \sum_{s=0}^{\infty} \zeta_w^s \left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \right] \left\{ X_{t,s} L_{t+s}(l) + (\widetilde{W}_t(l) X_{t,s} - W_{t+s}^h) \left(-\frac{1+\lambda_{w,t+s}}{\lambda_{t+s}} \right) \right.$$
$$\left. \times \left(\frac{\widetilde{W}_t(l) X_{t,s}}{W_{t+s}} \right)^{\frac{1+\lambda_{w,t+s}}{\lambda_{w,t+s}} - 1} \frac{X_{t,s}}{W_{t+s}} L_{t+s} \right\} = 0$$

and it can be rearranged by follows:

$$E_t \sum_{s=0}^{\infty} \zeta_w^s \left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \right] \left\{ X_{t,s} L_{t+s}(l) + (\widetilde{W}_t(l) X_{t,s} - W_{t+s}^h) \left(-\frac{1+\lambda_{w,t+s}}{\lambda_{t+s}} \right) L_{t+s}(l) \frac{1}{\widetilde{W}_t(l)} \right\} = 0$$

$$E_t \sum_{s=0}^{\infty} \zeta_w^s \left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \right] \left\{ \widetilde{W}_t(l) X_{t,s} L_{t+s}(l) + (\widetilde{W}_t(l) X_{t,s} - W_{t+s}^h) \left(-\frac{1+\lambda_{w,t+s}}{\lambda_{t+s}} \right) L_{t+s}(l) \right\} = 0$$

$$E_t \sum_{s=0}^{\infty} \zeta_w^s \left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} \right] L_{t+s}(l) \frac{1}{\lambda_{w,t+s}} \left\{ (1+\lambda_{w,t+s}) W_{t+s}^h - \widetilde{W}_t(l) X_{t,s} \right\} = 0 \tag{4.33}$$

The aggregate wage expression is

$$W_t = [(1-\zeta_w) \widetilde{W}_t^{\frac{1}{\lambda_{w,t}}} + \zeta_w (\gamma \pi_{t-1}^{l_w} \pi_*^{1-l_w} W_{t-1})^{\frac{1}{\lambda_{w,t}}}]^{\lambda_{w,t}}$$

Intuitively, similar to the intermediate goods producers' case, current wage, W_t, is a weighted average between optimized wage and previous wage. The mark up of the aggregate wage over the wage received by the households is distributed to the households in the form of dividends.

Government Policies

The central bank follows a nominal interest rate rule that adjusts its instrument to respond to the deviations of inflation and output from their respective target levels:

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi_*}\right)^{\psi_1} \left(\frac{Y_t}{Y_t^*}\right)^{\psi_2}\right]^{1-\rho_R} \left(\frac{Y_t/Y_{t-1}}{Y_t^*/Y_{t-1}^*}\right)^{\psi_3} r_t \tag{4.34}$$

where R^* is the steady state nominal gross rate and Y_t^* is the natural output. The parameter ρ_R determines the degree of interest rate smoothing, and r_t is the exogenous monetary policy shock. The government budget constraint is

$$P_t G_t + B_{t-1} = T_t + \frac{B_t}{R_t} \tag{4.35}$$

where T_t is nominal lump-sum taxes (or subsidies), and it is also included in household's constraint. Government spending is exogenous and it is expressed relative to the steady state output path as $g_t = G_t/(Y\gamma^t)$.

Resource Constraints

To derive the market clearing condition for the final goods market, we need to integrate households budget constraint across households, and combine it with government budget constraint and the expression for the dividends of intermediate goods for producers and labor unions. Then, finally we can have

$$C_t + I_t + G_t + a(Z_t)K_{t-1} = Y_t \tag{4.36}$$

Exogenous Process

There are seven exogenous processes in this model:

- Technology process:

$$lnZ_t = (1-\rho_z)ln_Z + \rho_z lnZ_{t-1} + \epsilon_{z,t}$$

- Investment relative price process:

$$ln\mu_t = (1 - \rho_{\mu_t})ln\mu + \rho_\mu ln\mu_{t-1} + \epsilon_{\mu,t}$$

- Intertemporal preference shifter (financial risk premium process):

$$lnb_t = (1 - \rho_b)lnb + \rho_b lnb_{t-1} + \epsilon_{b,t}$$

- Government spending process:

$$lng_t = (1 - \rho_g)lng + \rho_g lng_{t-1} + \rho_{ga} lnZ_t - \rho_{ga} lnZ_{t-1} + \epsilon$$

- Monetary policy shock:

$$lnr_t = \rho_r lnr_{t-1} + \epsilon_{r,t}$$

- Price mark-up shock:

$$ln\lambda_{p,t} = (1 - \rho_p)ln\lambda_p + \rho_p ln\lambda_{p,t-1} - \theta_p\epsilon_{p,t-1} + \epsilon_{p,t}$$

- Wage mark-up shock:

$$ln\lambda_{w,t} = (1 - \rho_w)ln\lambda_w + \rho_w ln\lambda_{w,t-1} - \theta_w\epsilon_{w,t-1} + \epsilon_{w,t}$$

the innovations ϵ are distributed as Normal IID:

$$\epsilon_{i,t} \sim N(0, \sigma_i)$$

Detrending

Since the model include constant growth rate, γ, we can derive the detrended model by dividing γ. In addition, we can replace the nominal variables in the model with real variables by dividing the price level, P_t. Then, de-trended variables are:

$$y_t = \frac{Y_t}{\gamma^t}, \quad c_t = \frac{C_t}{\gamma^t}, \quad k_t = \frac{K_t}{\gamma^t}, \quad w_t = \frac{W_t}{P_t\gamma^t}, \quad r_t^k = \frac{R_t^k}{P_t}, \quad \xi_t = \Xi_t\gamma^{\sigma_c t}, \quad \widetilde{p}_t = \frac{\widetilde{P}_t}{P_t}$$

$$\overline{\beta}^s = \frac{\beta^s}{\gamma^{\sigma_c s}}, \quad i_t = \frac{I_t}{\gamma^t}, \quad \widetilde{w}_t = \frac{\widetilde{W}_t}{P_t\gamma^t}$$

Intermediate Goods Producers

For intermediate goods producers, the aggregate production function becomes

$$Y_t(i) = \varepsilon_t^a \overline{K}_t(i)^\alpha [\gamma^t L_t(i)]^{1-\alpha} - \gamma^t \Phi \Rightarrow \frac{Y_t(i)}{\gamma^t} = \varepsilon_t^a \left(\frac{\overline{K}_t(i)}{\gamma^t} \right)^\alpha L_t(i)^{1-\alpha} - \Phi$$

$$\therefore y_t(i) = \varepsilon_t^a k_t(i)^\alpha (L_t(i))^{1-\alpha} - \Phi$$

and

$$\overline{K}_t = \frac{\alpha}{1-\alpha} \frac{W_t}{R_t^k} L_t \Rightarrow \frac{\overline{K}_t}{\gamma^t} = \frac{\alpha}{1-\alpha} \frac{\frac{W_t}{\gamma^t P_t}}{\frac{R_t^k}{P_t}} L_t$$

$$\therefore \overline{k}_t = \frac{\alpha}{1-\alpha} \frac{w_t}{r_t^k} L_t$$

The marginal cost becomes:

$$MC_t = \alpha^{-\alpha} (1-\alpha)^{-(1-\alpha)} W_t^{(1-\alpha)} (R_t^k)^\alpha \gamma^{-(1-\alpha)t} (\varepsilon_t^a)^{-1}$$

$$\Rightarrow \frac{MC_t}{P_t} = mc_t = \frac{1}{\alpha^\alpha (1-\alpha)^{1-\alpha} \varepsilon_t^a} \left(\frac{W_t}{P_t \gamma^t} \right)^{1-\alpha} \left(\frac{R_t^k}{P_t} \right)^\alpha$$

$$\therefore mc_t = \frac{w_t^{1-\alpha} (r_t^k)^\alpha}{\alpha^\alpha (1-\alpha)^{1-\alpha} \varepsilon_t^a}$$

and the F.O.C,

$$E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} Y_{t+s}(i) \left[\widetilde{P}_t(i) X_{t,s} + \left(\widetilde{P}_t(i) X_{t,s} - MC_{t+s} \right) \frac{1}{G'^{-1}(z_{t+s})} \frac{G'(x_{t+s})}{G''(x_{t+s})} \right] = 0$$

$$\Rightarrow E_t \sum_{s=0}^{\infty} \zeta_p^s \frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}} Y_{t+s}(i)$$
$$\times \left[\left(1 + \frac{1}{G'^{-1}(z_{t+s})} \frac{G'(x_{t+s})}{G''(x_{t+s})} \right) \widetilde{P}_t(i) X_{t,s} - \left(\frac{1}{G'^{-1}(z_{t+s})} \frac{G'(x_{t+s})}{G''(x_{t+s})} \right) MC_{t+s} \right] = 0$$

becomes

$$E_t \sum_{s=0}^{\infty} \zeta_p^s \overline{\beta}^s \frac{\xi_{t+s}}{\xi_t} \gamma^s y_{t+s}(i) \frac{P_t}{P_{t+s}}$$
$$\times \left[\left(1 + \frac{1}{G'^{-1}(z_{t+s})} \frac{G'(x_{t+s})}{G''(x_{t+s})} \right) \widetilde{P}_t(i) X_{t,s} - \left(\frac{1}{G'^{-1}(z_{t+s})} \frac{G'(x_{t+s})}{G''(x_{t+s})} \right) MC_{t+s} \right] = 0$$

$$E_t \sum_{s=0}^{\infty} \zeta_p^s \overline{\beta}^s \frac{\xi_{t+s}}{\xi_t} \gamma^s y_{t+s}(i)$$
$$\times \left[\left(1 + \frac{1}{G'^{-1}(z_{t+s})} \frac{G'(x_{t+s})}{G''(x_{t+s})} \right) \frac{\widetilde{p}_t(i) X_{t,s}}{X_{t+s}^p} - \left(\frac{1}{G'^{-1}(z_{t+s})} \frac{G'(x_{t+s})}{G''(x_{t+s})} \right) mc_{t+s} \right] = 0$$

where

$$X_{t+s}^p = 1 \text{ for } s = 0 \text{ or else } \Pi_{l=1}^s \pi_{t+l}$$

Based on the previous de-trended equations, aggregate profit equation,

$$\Pi_t = P_t Y_t - W_t L_t - R_t^k K_t$$

equals to:

$$\frac{\Pi_t}{P_t \gamma^t} = \frac{P_t Y_t}{P_t \gamma^t} - \frac{W_t L_t}{P_t \gamma^t} - \frac{R_t^k K_t}{P_t \gamma^t} = y_t - w_t L_t - r_t^k k_t$$
$$\Rightarrow \varepsilon_t^a k_t^\alpha L_t^{1-\alpha} - \Phi - w_t L_t - r_t^k k_t \quad (\because y_t = \varepsilon_t^a k_t^\alpha L_t^{1-\alpha} - \Phi)$$
$$\Rightarrow \varepsilon_t^a k_t^\alpha L_t^{1-\alpha} - \Phi - w_t L_t - r_t^k \frac{\alpha}{1-\alpha} \frac{w_t}{r_t^k} L_t \quad (\because k_t = \frac{\alpha}{1-\alpha} \frac{w_t}{r_t^k} L_t)$$
$$\Rightarrow \left[\varepsilon_t^a k_t^\alpha L_t^{-\alpha} - w_t - r_t^k \frac{\alpha}{1-\alpha} \frac{w_t}{r_t^k} \right] L_t - \Phi = \left[\varepsilon_t^a k_t^\alpha L_t^{-\alpha} - \frac{1}{1-\alpha} w_t \right] L_t - \Phi$$
$$\Rightarrow \left[\varepsilon_t^a \left(\frac{\alpha}{1-\alpha} \frac{w_t}{r_t^k} L_t \right)^\alpha L_t^{-\alpha} - \frac{1}{1-\alpha} w_t \right] L_t - \Phi = \left[\frac{\varepsilon_t^a \alpha^\alpha (1-\alpha)^{1-\alpha}}{w_t^{1-\alpha} (r_t^k)^\alpha} - 1 \right] \frac{w_t L_t}{1-\alpha} - \Phi$$
$$\Rightarrow \frac{\Pi_t}{P_t \gamma^t} = \left(\frac{1}{mc_t} - 1 \right) \frac{w_t L_t}{1-\alpha} - \Phi \quad (\because mc_t = \frac{w_t^{1-\alpha} (r_t^k)^\alpha}{\alpha^\alpha (1-\alpha)^{1-\alpha} \varepsilon_t^a})$$

Households

F.O.C with respect to C_t becomes

$$\Xi_t = \exp\left(\frac{\sigma_c - 1}{1 + \sigma_l} L_t^{1+\sigma_l} \right) (C_t - \lambda C_{t-1})^{-\sigma_c} \quad \Rightarrow \quad \xi_t = \exp\left(\frac{\sigma_c - 1}{1 + \sigma_l} L_t^{1+\sigma_l} \right) \left(\frac{C_t}{\gamma^t} - \frac{\lambda}{\gamma} \frac{C_{t-1}}{\gamma^{t-1}} \right)^{-\sigma_c}$$

$$\therefore \quad \xi_t = \exp\left(\frac{\sigma_c - 1}{1 + \sigma_l} L_t^{1+\sigma_l} \right) (c_t - \frac{\lambda}{\gamma} c_{t-1})^{-\sigma_c} \tag{4.37}$$

F.O.C with respect to L_t becomes

$$\left[\frac{1}{1 - \sigma_c} (C_t - \lambda C_{t-1})^{1-\sigma_c} \right] \exp\left(\frac{\sigma_c - 1}{1 + \sigma_l} L_t^{1+\sigma_l} \right) (\sigma_c - 1) L_t^{\sigma_l} = -\Xi_t \frac{W_t^h}{P_t}$$

$$\Rightarrow \frac{1}{1-\sigma_c}(C_t-\lambda C_{t-1})(C_t-\lambda C_{t-1})^{-\sigma}\exp\left(\frac{\sigma_c-1}{1+\sigma_l}L_t^{1+\sigma_l}\right)(\sigma_c-1)L_t^{\sigma_l} = -\Xi_t\frac{W_t^h}{P_t}$$
$$\Rightarrow \frac{1}{1-\sigma_c}(C_t-\lambda C_{t-1})(\sigma_c-1)L_t^{\sigma_l}\Xi_t = -\Xi_t\frac{W_t^h}{P_t}$$
$$\Rightarrow \frac{1}{1-\sigma_c}\frac{1}{\gamma^t}(C_t-\lambda C_{t-1})(\sigma_c-1)L_t^{\sigma_l} = -\frac{W_t^h}{P_t\gamma^t}$$
$$\therefore \quad w_t^h = \left(c_t-\left(\frac{\lambda}{\gamma}\right)c_{t-1}\right)L_t^{\sigma_c} \tag{4.38}$$

F.O.C with respect to B_t becomes

$$\Xi_t = \beta\varepsilon_t^b R_t E_t\left[\frac{\Xi_{t+1}}{\pi_{t+1}}\right] \;\Rightarrow\; \xi_t\gamma^{-\sigma_c t} = \overline{\beta}\gamma^{\sigma_c}\varepsilon_t^b R_t E_t\left[\frac{\xi_{t+1}\gamma^{-\sigma_c(t+1)}}{\pi_{t+1}}\right]$$
$$\therefore \quad \xi_t = \overline{\beta}\varepsilon_t^b R_t E_t\left[\frac{\xi_{t+1}}{\pi_{t+1}}\right] \tag{4.39}$$

F.O.C with respect to I_t becomes, ($Q_t = \dfrac{\Xi_t^k}{\Xi_t}$)

$$\Xi_t = \Xi_t^k\varepsilon_t^i\left(1-S\left(\frac{I_t}{I_{t-1}}\right)-S'\left(\frac{I_t}{I_{t-1}}\right)\frac{I_t}{I_{t-1}}\right)+\beta E_t\left[\Xi_{t+1}^k\varepsilon_{t+1}^i S'\left(\frac{I_{t+1}}{I_t}\right)\left(\frac{I_{t+1}}{I_t}\right)^2\right]$$
$$\Rightarrow 1 = Q_t\varepsilon_t^i\left(1-S\left(\frac{i_t\gamma}{i_{t-1}}\right)-S'\left(\frac{i_t\gamma}{i_{t-1}}\right)\frac{i_t}{i_{t-1}}\right)+\beta E_t\left[Q_{t+1}\frac{\Xi_{t+1}}{\Xi_t}\varepsilon_{t+1}^i S'\left(\frac{i_{t+1}\gamma}{i_t}\right)\left(\frac{i_{t+1}\gamma}{i_t}\right)^2\right]$$
$$\therefore \quad 1 = Q_t\varepsilon_t^i\left(1-S\left(\frac{i_t\gamma}{i_{t-1}}\right)-S'\left(\frac{i_t\gamma}{i_{t-1}}\right)\frac{i_t}{i_{t-1}}\right)+\overline{\beta}E_t\left[\frac{\xi_{t+1}}{\xi_t}\left\{Q_{t+1}\varepsilon_{t+1}^i S'\left(\frac{i_{t+1}\gamma}{i_t}\right)\left(\frac{i_{t+1}\gamma}{i_t}\right)^2\right\}\right] \tag{4.40}$$

F.O.C with respect to $\overline{K}_t$ becomes

$$\Xi_t^k = \beta E_t\left[\Xi_{t+1}\left(\frac{R_{t+1}^k}{P_{t+1}}Z_{t+1}-a(Z_{t+1})\right)+\Xi_{t+1}^k(1-\delta)\right]$$
$$\Rightarrow Q_t = \overline{\beta}\gamma^{-\sigma_c}E_t\left[\frac{\Xi_{t+1}}{\Xi_t}\left(\frac{R_{t+1}^k}{P_{t+1}}Z_{t+1}-a(Z_{t+1})\right)+Q_{t+1}\frac{\Xi_{t+1}}{\Xi_t}(1-\delta)\right]$$
$$\Rightarrow Q_t = \overline{\beta}\gamma^{-\sigma_c}E_t\left[\frac{\Xi_{t+1}}{\Xi_t}\left\{\left(\frac{R_{t+1}^k}{P_{t+1}}Z_{t+1}-a(Z_{t+1})\right)+Q_{t+1}(1-\delta)\right\}\right]$$
$$\therefore \quad Q_t = \overline{\beta}E_t\left[\frac{\xi_{t+1}}{\xi_t}\left\{\left(r_{t+1}^k Z_{t+1}-a(Z_{t+1})\right)+Q_{t+1}(1-\delta)\right\}\right] \tag{4.41}$$

F.O.C with respect to Z_t becomes

$$\frac{R_t^k}{P_t} = a'(Z_t) \quad \Rightarrow \quad r_t^k = a'(Z_t) \tag{4.42}$$

The F.O.C,

$$E_t \sum_{s=0}^{\infty} \zeta_w^s \left[\frac{\beta^s \Xi_{t+s} P_t}{\Xi_t P_{t+s}}\right] L_{t+s}(l) \frac{1}{\lambda_{w,t+s}} \left\{(1+\lambda_{w,t+s}) W_{t+s}^h - \widetilde{W}_t(l) X_{t,s}\right\} = 0$$

becomes

$$E_t \sum_{s=0}^{\infty} \zeta_w^s \overline{\beta}^s \left[\frac{\xi_{t+s}}{\xi_t}\right] L_{t+s}(l) \frac{1}{\lambda_{w,t+s}} \left\{(1+\lambda_{w,t+s}) \frac{W_{t+s}^h}{P_{t+s}} - \frac{\widetilde{W}_t(l)}{P_{t+s}} X_{t,s}\right\} = 0$$

$$\Rightarrow E_t \sum_{s=0}^{\infty} \zeta_w^s \overline{\beta}^s \left[\frac{\xi_{t+s}}{\xi_t}\right] L_{t+s}(l) \frac{1}{\lambda_{w,t+s}} \left\{(1+\lambda_{w,t+s}) \frac{W_{t+s}^h}{P_{t+s}\gamma^t} - \frac{\widetilde{W}_t(l)}{P_{t+s}\gamma^t} X_{t,s}\right\} = 0$$

$$\therefore \quad E_t \sum_{s=0}^{\infty} \zeta_w^s \overline{\beta}^s \gamma^s \left[\frac{\xi_{t+s}}{\xi_t}\right] L_{t+s}(l) \frac{1}{\lambda_{w,t+s}} \left[(1+\lambda_{w,t+s}) w_{t+s}^h - \frac{(\Pi_{l=0}^s \pi_{t+l-1}^{l_w} \pi_*^{1-l_w})}{(\Pi_{l=1}^s \pi_{t+l})} \widetilde{w}_t(l)\right] = 0 \tag{4.43}$$

In addition, the equation,

$$\overline{K}_t(j) = Z_t(j) K_{t-1}(j)$$

becomes

$$\overline{k}_t(j) = Z_t(j) \frac{k_{t-1}(j)}{\gamma}$$

Resource Constraint

The resource constraint,

$$C_t + I_t + G_t + a(Z_t)\overline{K}_{t-1} = Y_t$$

becomes

$$\frac{1}{\gamma^t}(C_t + I_t + G_t + a(Z_t)\overline{K}_{t-1}) = \frac{Y_t}{\gamma^t}$$

$$\therefore \quad c_t + i_t + Y_* g_t + a(Z_t)\frac{\overline{k}_{t-1}}{\gamma} = y_t \quad \left(\because g_t = \frac{G_t}{Y_*\gamma^t}\right) \tag{4.44}$$

Government Policies

The Taylor rule,

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi_*}\right)^{\psi_1} \left(\frac{Y_t}{Y_t^*}\right)^{\psi_2}\right]^{1-\rho_R} \left(\frac{Y_t/Y_{t-1}}{Y_t^*/Y_{t-1}^*}\right)^{\psi_3} r_t$$

becomes

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi_*}\right)^{\psi_1} \left(\frac{y_t}{y_t^*}\right)^{\psi_2}\right]^{1-\rho_R} \left(\frac{y_t/y_{t-1}}{y_t^*/y_{t-1}^*}\right)^{\psi_3} r_t \tag{4.45}$$

Log-linearized Model

For the log-linearization of the equation that is de-trended of F.O.C of household,

$$\xi_t = \exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{1+\sigma_l}\right)\left(c_t - \frac{\lambda}{\gamma} c_{t-1}\right)^{-\sigma_c}$$

I take logarithm of the equation first such as

$$ln\xi_t = \left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{1+\sigma_l}\right) - \sigma_c ln\left(c_t - \frac{\lambda}{\gamma} c_{t-1}\right)$$

and then use the Taylor first order approximation to linearize it. For the left hand side,

$$ln\xi_t \approx ln\xi_* + \frac{1}{\xi_*}(\xi_t - \xi_*)$$

For the right hand side,

$$\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{1+\sigma_l}\right) \approx \left\{\frac{\sigma_c - 1}{1+\sigma_l}\left(L_*^{1+\sigma_l} + (1+\sigma_l)L_*^{\sigma_l}(L_t - L_*)\right)\right\}$$

$$-\sigma_c ln\left(c_t - \frac{\lambda}{\gamma} c_{t-1}\right) \approx -\sigma_c\left[ln\left(1-\frac{\lambda}{\gamma}\right)c_* + \frac{1}{\left(1-\frac{\lambda}{\gamma}\right)c_*}(c_t - c_*) - \frac{\frac{\lambda}{\gamma}}{\left(1-\frac{\lambda}{\gamma}\right)c_*}(c_{t-1} - c_*)\right]$$

Thus,

$$ln\xi_* + \frac{1}{\xi_*}(\xi_t - \xi_*) \approx \left\{\frac{\sigma_c - 1}{1+\sigma_l}\left(L_*^{1+\sigma_l} + (1+\sigma_l)L_*^{\sigma_l}(L_t - L_*)\right)\right\}$$

$$-\sigma_c\left[ln\left(1-\frac{\lambda}{\gamma}\right)c_* + \frac{1}{\left(1-\frac{\lambda}{\gamma}\right)c_*}(c_t - c_*) - \frac{\frac{\lambda}{\gamma}}{\left(1-\frac{\lambda}{\gamma}\right)c_*}(c_{t-1} - c_*)\right]$$

since

$$ln\xi_* = \left(\frac{\sigma_c - 1}{1+\sigma_l} L_*^{1+\sigma_l}\right) - \sigma_c ln\left(1 - \frac{\lambda}{\gamma}\right) c_*$$

in steady state, so we can get rid of this term in the equation. Therefore,

$$\frac{1}{\xi_*}(\xi_t - \xi_*) \approx (\sigma_c - 1)L_*^{\sigma_l}(L_t - L_*) - \sigma_c \left[\frac{1}{\left(1-\frac{\lambda}{\gamma}\right)c_*}(c_t - c_*) - \frac{\frac{\lambda}{\gamma}}{\left(1-\frac{\lambda}{\gamma}\right)c_*}(c_{t-1} - c_*)\right]$$

using the approximation,

$$\widehat{x}_t = log(X_t) - log(X_*) = log(X_t/X_*) \approx \frac{X_t - X_*}{X_*}$$

we can re-write the equation above such as

$$\widehat{\xi}_t \approx (\sigma_c - 1)L_*^{1+\sigma_l}\widehat{L}_t - \sigma_c\left[\left(\frac{\gamma}{\gamma - \lambda}\right)\widehat{c}_t - \left(\frac{\lambda}{\gamma-\lambda}\right)\widehat{c}_{t-1}\right]$$

For another de-trended F.O.C,

$$\xi_t = \overline{\beta}\varepsilon_t^b R_t E_t\left[\frac{\xi_{t+1}}{\pi_{t+1}}\right]$$

I use different way for linear approximation because it includes expectation term. Using the method,

$$lnx_t = lnx_* + \widehat{x}_t \quad \Rightarrow \quad x_t = e^{lnx_* + \widehat{x}_t} = e^{lnx_*} \times e^{\widehat{x}_t} \quad \Rightarrow \quad x_t = x_* e^{\widehat{x}_t}$$

we can re-write the equation such as

$$\xi_* e^{\widehat{\xi}_t} = \overline{\beta}\varepsilon_*^b e^{\widehat{\varepsilon}_t^b} R_* e^{\widehat{R}_t} E_t\left[\frac{\xi_* e^{\widehat{\xi}_{t+1}}}{\pi_* e^{\widehat{\pi}_{t+1}}}\right]$$

and since, in steady state,

$$\xi_* = \overline{\beta}\varepsilon_*^b R_* E_*\left[\frac{\xi_*}{\pi_*}\right]$$

we can remove the term in both side, then

$$e^{\widehat{\xi}_t} = e^{\widehat{\varepsilon}_t^b} e^{\widehat{R}_t} E_t\left[\frac{e^{\widehat{\xi}_{t+1}}}{e^{\widehat{\pi}_{t+1}}}\right]$$

We have not used the approximation yet so far. Approximating the expression $e^{\widetilde{x}_t}$ with a first-order Taylor expansion at the point $\widetilde{x}_t = 0$ yields

$$e^{\widehat{x}_t} \approx 1 + e^0(\widehat{x}_t - 0) = 1 + \widehat{x}_t$$

therefore the equation can be approximated by

$$1 + \widetilde{\xi}_t \approx (1 + \widehat{\varepsilon}_t^b)(1 + \widehat{R}_t)E_t\left\{(1 + \widehat{\xi}_{t+1})(1 - \widehat{\pi}_{t+1})\right\}$$

Since the product of log-deviations from steady states are very small, we can ignore the terms. Then, we would have

$$\widetilde{\xi}_t \approx \widehat{\varepsilon}_t^b + \widehat{R}_t + E_t\widehat{\xi}_{t+1} - E_t\widehat{\pi}_{t+1}$$

Combining with previous approximation,

$$\widehat{\xi}_t \approx (\sigma_c - 1)L_*^{1+\sigma_l}\widehat{L}_t - \sigma_c\left[\left(\frac{\gamma}{\gamma - \lambda}\right)\widehat{c}_t - \left(\frac{\lambda}{\gamma - \lambda}\right)\widehat{c}_{t-1}\right]$$

we can have

$$\begin{aligned}(\sigma_c - 1)L_*^{1+\sigma_l}\widehat{L}_t - \sigma_c\left[\left(\frac{\gamma}{\gamma - \lambda}\right)\widehat{c}_t - \left(\frac{\lambda}{\gamma - \lambda}\right)\widehat{c}_{t-1}\right]\\ = (\sigma_c - 1)L_*^{1+\sigma_l}E_t\widehat{L}_{t+1} - \sigma_c\left[\left(\frac{\gamma}{\gamma - \lambda}\right)E_t\widehat{c}_{t+1} - \left(\frac{\lambda}{\gamma - \lambda}\right)\widehat{c}_t\right] + \widehat{\varepsilon}_t^b + \widehat{R}_t - E_t\widehat{\pi}_{t+1}\end{aligned}$$

by re-arranging it,

$$-\sigma_c\left(\frac{\gamma + \lambda}{\gamma - \lambda}\right)\widehat{c}_t = (\sigma_c - 1)L_*^{1+\sigma_l}(E_t\widehat{L}_{t+1} - \widehat{L}_t) - \sigma_c\left(\frac{\gamma}{\gamma - \lambda}\right)E_t\widehat{c}_{t+1} - \sigma_c\left(\frac{\lambda}{\gamma - \lambda}\right)\widehat{c}_{t-1} + \widehat{R}_t - E_t\widehat{\pi}_{t+1} + \widehat{\varepsilon}_t^b$$

Then,

$$\begin{aligned}\widehat{c}_t &= \left(\frac{\gamma}{\gamma + \lambda}\right)E_t\widehat{c}_{t+1} + \left(\frac{\lambda}{\gamma + \lambda}\right)\widehat{c}_{t-1} - \frac{1}{\sigma_c}\left(\frac{\gamma - \lambda}{\gamma + \lambda}\right)(\widehat{R}_t - E_t\widehat{\pi}_{t+1}) - \frac{1}{\sigma_c}\left(\frac{\gamma - \lambda}{\gamma + \lambda}\right)\widehat{\varepsilon}_t^b\\ &\quad - \left(\frac{\gamma - \lambda}{\gamma + \lambda}\right)\left(\frac{\sigma_c - 1}{\sigma_c}\right)L_*^{1+\sigma_c}(E_t\widehat{L}_{t+1} - \widehat{L}_t)\\ &= \left(\frac{1}{1 + \frac{\lambda}{\gamma}}\right)E_t\widehat{c}_{t+1} + \left(\frac{\frac{\lambda}{\gamma}}{1 + \frac{\lambda}{\gamma}}\right)\widehat{c}_{t-1} - \frac{1 - \frac{\lambda}{\gamma}}{\sigma_c\left(1 + \frac{\lambda}{\gamma}\right)}(\widehat{R}_t - E_t\widehat{\pi}_{t+1}) - \frac{1 - \frac{\lambda}{\gamma}}{\sigma_c\left(1 + \frac{\lambda}{\gamma}\right)}\widehat{\varepsilon}_t^b\end{aligned}$$

$$-\left(\frac{1-\frac{\lambda}{\gamma}}{1+\frac{\lambda}{\gamma}}\right)\left(\frac{\sigma_c-1}{\sigma_c}\right)L_*^{1+\sigma_c}(E_t\widehat{L}_{t+1}-\widehat{L}_t)$$

From the another F.O.C,

$$w_t^h=\left(c_t-\left(\frac{\lambda}{\gamma}\right)c_{t-1}\right)L_t^{\sigma_c}\quad\Rightarrow\quad w_*^h=\left(1-\frac{\lambda}{\gamma}\right)c_*L_*^{\sigma_c}\quad\Rightarrow\quad\frac{w_*^hL_*}{c_*}=\left(1-\frac{\lambda}{\gamma}\right)L_*^{1+\sigma_c}$$

By substituting the equation to the term in the previous one, we can have

$$\widehat{c}_t=\left(\frac{1}{1+\frac{\lambda}{\gamma}}\right)E_t\widehat{c}_{t+1}+\left(\frac{\frac{\lambda}{\gamma}}{1+\frac{\lambda}{\gamma}}\right)\widehat{c}_{t-1}-\frac{1-\frac{\lambda}{\gamma}}{\sigma_c\left(1+\frac{\lambda}{\gamma}\right)}(\widehat{R}_t-E_t\widehat{\pi}_{t+1})-\frac{1-\frac{\lambda}{\gamma}}{\sigma_c\left(1+\frac{\lambda}{\gamma}\right)}\widehat{\varepsilon}_t^b$$

$$-\left(\frac{(\sigma_c-1)\frac{w_*^hL_*}{c_*}}{\sigma_c\left(1+\frac{\lambda}{\gamma}\right)}\right)(E_t\widehat{L}_{t+1}-\widehat{L}_t)$$

$$\therefore\quad\widehat{c}_t=\left(\frac{1}{1+\frac{\lambda}{\gamma}}\right)E_t\widehat{c}_{t+1}+\left(\frac{\frac{\lambda}{\gamma}}{1+\frac{\lambda}{\gamma}}\right)\widehat{c}_{t-1}-\frac{1-\frac{\lambda}{\gamma}}{\sigma_c\left(1+\frac{\lambda}{\gamma}\right)}(\widehat{R}_t-E_t\widehat{\pi}_{t+1})$$

$$-\frac{1-\frac{\lambda}{\gamma}}{\sigma_c\left(1+\frac{\lambda}{\gamma}\right)}\widehat{\varepsilon}_t^b-\left(\frac{(\sigma_c-1)\frac{w_*^hL_*}{c_*}}{\sigma_c\left(1+\frac{\lambda}{\gamma}\right)}\right)(E_t\widehat{L}_{t+1}-\widehat{L}_t)$$

For the following equation,

$$1=Q_t\varepsilon_t^i\left(1-S\left(\frac{i_t\gamma}{i_{t-1}}\right)-S'\left(\frac{i_t\gamma}{i_{t-1}}\right)\frac{i_t\gamma}{i_{t-1}}\right)+\overline{\beta}E_t\left[\frac{\xi_{t+1}}{\xi_t}\left\{Q_{t+1}\varepsilon_{t+1}^iS'\left(\frac{i_{t+1}\gamma}{i_t}\right)\left(\frac{i_{t+1}\gamma}{i_t}\right)^2\right\}\right]$$

it can be

$$1=Q_*e^{\widehat{Q}_t}\varepsilon_*^ie^{\widehat{\varepsilon}_t^i}\left(1-S\left(\frac{i_*e^{\widehat{i}_t}\gamma}{i_*e^{\widehat{i}_{t-1}}}\right)-S'\left(\frac{i_*e^{\widehat{i}_t}\gamma}{i_*e^{\widehat{i}_{t-1}}}\right)\frac{i_*e^{\widehat{i}_t}\gamma}{i_*e^{\widehat{i}_{t-1}}}\right)$$

$$+\overline{\beta}E_t\left[\frac{\xi_*e^{\widehat{\xi}_{t+1}}}{\xi_*e^{\widehat{\xi}_t}}\left\{Q_*e^{\widehat{Q}_{t+1}}\varepsilon_*^ie^{\widehat{\varepsilon}_{t+1}^i}S'\left(\frac{i_*e^{\widehat{i}_{t+1}}\gamma}{i_*e^{\widehat{i}_t}}\right)\left(\frac{i_*e^{\widehat{i}_{t+1}}\gamma}{i_*e^{\widehat{i}_t}}\right)^2\right\}\right]$$

$$=Q_*\varepsilon_*^ie^{\widehat{Q}_t+\widehat{\varepsilon}_t^i}\left(1-S\left(\frac{\gamma e^{\widehat{i}_t}}{e^{\widehat{i}_{t-1}}}\right)-S'\left(\frac{\gamma e^{\widehat{i}_t}}{e^{\widehat{i}_{t-1}}}\right)\frac{e^{\widehat{i}_t}\gamma}{e^{\widehat{i}_{t-1}}}\right)$$

$$+\overline{\beta}E_t\left[\frac{e^{\widehat{\xi}_{t+1}}}{e^{\widehat{\xi}_t}}\left\{Q_*\varepsilon_*^ie^{\widehat{Q}_{t+1}+\widehat{\varepsilon}_{t+1}}S'\left(\frac{\gamma e^{\widehat{i}_{t+1}}}{e^{\widehat{i}_t}}\right)\left(\frac{e^{\widehat{i}_{t+1}}\gamma}{e^{\widehat{i}_t}}\right)^2\right\}\right]$$

For $S\left(\frac{\gamma e^{\widehat{i}_t}}{e^{\widehat{i}_{t-1}}}\right)$ and $S'\left(\frac{\gamma e^{\widehat{i}_t}}{e^{\widehat{i}_{t-1}}}\right)$,

$$S\left(\frac{\gamma e^{\widehat{i}_t}}{e^{\widehat{i}_{t-1}}}\right) \approx S\left(\frac{\gamma(1+\widehat{i}_t)}{1+\widehat{i}_{t-1}}\right) \approx S(\gamma) + \gamma S'(\gamma)\widehat{i}_t - \gamma S'(\gamma)\widehat{i}_{t-1} \approx 0$$

and

$$S'\left(\frac{\gamma e^{\widehat{i}_t}}{e^{\widehat{i}_{t-1}}}\right) \approx S'\left(\frac{\gamma(1+\widehat{i}_t)}{1+\widehat{i}_{t-1}}\right) \approx S'(\gamma) + \gamma S''(\gamma)\widehat{i}_t - \gamma S''(\gamma)\widehat{i}_{t-1} \approx \gamma S''(\gamma)(\widehat{i}_t - \widehat{i}_{t-1})$$

Then, the approximated equation would be,

$$\begin{aligned} 1 \approx & \left(1+\widehat{Q}_t+\widehat{\varepsilon}^i_t\right)\left[1-\gamma S''(\gamma)\left(\widehat{i}_t-\widehat{i}_{t-1}\right)\left\{\gamma(1-\widehat{i}_t-\widehat{i}_{t-1})\right\}\right] \\ & +\overline{\beta}E_t\left[\left(1+\widehat{\xi}_{t+1}-\widehat{\xi}_t\right)\left\{\left(1+\widehat{Q}_{t+1}+\widehat{\varepsilon}_{t+1}\right)\left(\gamma S''(\gamma)(\widehat{i}_{t+1}-\widehat{i}_t)\right)\left(\gamma^2(1+2\widehat{i}_{t+1}-2\widehat{i}_t)\right)\right\}\right] \end{aligned}$$

then can re-arrange the equation as follows;

$$S''(\gamma)\gamma^2(1+\overline{\beta})\widehat{i}_t \approx S''(\gamma)\gamma^2\widehat{i}_{t-1} + \overline{\beta}\gamma^3 E_t\widehat{i}_{t+1} + \widehat{Q}_t + \widehat{\varepsilon}^i_t$$

$$\therefore \quad \widehat{i}_t = \frac{1}{(1+\overline{\beta})}\left(\widehat{i}_{t-1} + \overline{\beta}\gamma E_t\widehat{i}_{t+1} + \frac{1}{S''(\gamma)\gamma^2}\widehat{Q}_t + \frac{1}{S''(\gamma)\gamma^2}\widehat{\varepsilon}^i_t\right)$$

For the intermediate goods producers' production function,

$$y_t = \varepsilon^a_t k^\alpha_t (L_t)^{1-\alpha} - \Phi$$

to do linear approximation

$$y_t + \Phi = \varepsilon^a_t k^\alpha_t (L_t)^{1-\alpha}$$

Then taking logarithm and subtract from the log of steady state:

$$ln(y_t+\Phi) - ln(y_*+\Phi) = \widehat{y_t+\Phi} = \widehat{\varepsilon}^a_t + \alpha\widehat{k}_t + (1-\alpha)\widehat{L}_t$$

We need to transform $\widehat{y_t+\Phi}$ to $\widehat{y}_t$,

$$\widehat{y_t+\Phi} \approx \frac{(y_t+\Phi)-(y_*+\Phi)}{y_*+\Phi} = \frac{y_t-y_*}{y_*}\frac{y_*}{y_*+\Phi} = \widehat{y}_t\frac{y_*}{y_*+\Phi}$$

then

$$\widehat{y}_t \frac{y_*}{y_* + \Phi} = \widehat{\varepsilon}_t^a + \alpha \widehat{k}_t + (1-\alpha)\widehat{L}_t$$

$$\widehat{y}_t = \frac{y_* + \Phi}{y_*}(\widehat{\varepsilon}_t^a + \alpha \widehat{k}_t + (1-\alpha)\widehat{L}_t) = 1 + \frac{\Phi}{y_*}(\widehat{\varepsilon}_t^a + \alpha \widehat{k}_t + (1-\alpha)\widehat{L}_t)$$

$$\therefore \quad \widehat{y}_t = \Phi^*(\widehat{\varepsilon}_t^a + \alpha \widehat{k}_t + (1-\alpha)\widehat{L}_t)$$

where Φ^* is one plus the share of fixed costs in production reflecting the presence of fixed costs in production.

Log-linearization of effective capital amount equation would be

$$\overline{k}_t = Z_t \frac{k_{t-1}}{\gamma} \quad \Rightarrow \quad \widehat{\overline{k}}_t = \widehat{Z}_t + \widehat{k}_{t-1}$$

and this approximation,

$$\overline{k}_t = \frac{\alpha}{1-\alpha}\frac{w_t}{r_t^k}L_t \quad \Rightarrow \quad \widehat{\overline{k}}_t = \widehat{w}_t - \widehat{r}_t^k + \widehat{L}_t \quad \Rightarrow \quad \widehat{r}_t^k = \widehat{L}_t + \widehat{w}_t - \widehat{\overline{k}}_t$$

implies that the rental rate of capital is negatively related to the capital-labor ratio and positively to the real wage.

Marginal cost equation,

$$mc_t = \frac{w_t^{1-\alpha}(r_t^k)^\alpha}{\alpha^\alpha (1-\alpha)^{1-\alpha}\varepsilon_t^a}$$

becomes

$$mc_* e^{\widehat{mc}_t} = \frac{w_*^{1-\alpha}(r_*^k)^\alpha e^{\widehat{w}_t(1-\alpha)} e^{\widehat{r}_t^k \alpha}}{\alpha^\alpha (1-\alpha)^(1-\alpha)\varepsilon_t^a e^{\widehat{\varepsilon}_t^a}} \quad \Rightarrow \quad e^{\widehat{mc}_t} = \frac{e^{\widehat{w}_t(1-\alpha)} e^{\widehat{r}_t^k \alpha}}{e^{\widehat{\varepsilon}_t^a}} \approx 1 + \widehat{mc}_t = 1 + (1-\alpha)\widehat{w}_t + \alpha \widehat{r}_t^k - \widehat{\varepsilon}_t^a$$

Therefore,

$$\widehat{mc}_t = (1-\alpha)\widehat{w}_t + \alpha \widehat{r}_t^k - \widehat{\varepsilon}_t^a$$

For the Taylor rule,

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi_*}\right)^{\psi_1} \left(\frac{y_t}{y_t^{flex}}\right)^{\psi_2}\right]^{1-\rho_R} \left(\frac{y_t / y_{t-1}}{y_t^{flex}/y_{t-1}^{flex}}\right)^{\psi_3} r_t$$

it can be expressed by

$$\frac{R^* e^{\widehat{R}_t}}{R^*} = \left(\frac{R^* e^{\widehat{R}_{t-1}}}{R^*}\right)^{\rho_R} \left[\left(\frac{\pi_* e^{\widehat{\pi}_t}}{\pi_*}\right)^{\psi_1} \left(\frac{y_* e^{\widehat{y}_t}}{y_*^{flex} e^{\widehat{y}_t^{flex}}}\right)^{\psi_2}\right]^{1-\rho_R} \left(\frac{y_* e^{\widehat{y}_t}/y_* e^{\widehat{y}_{t-1}}}{y_*^{flex} e^{\widehat{y}_t^{flex}}/y_*^{flex} e^{\widehat{y}_{t-1}^{flex}}}\right)^{\psi_3} r_* e^{\widehat{r}_t}$$

and it equals

$$e^{\widehat{R}_t} = e^{\rho_R \widehat{R}_{t-1}} e^{\psi_1 \widehat{\pi}_t (1-\rho_R)} \left(\frac{y_*}{y_*^{flex}}\right)^{\psi_2(1-\rho_R)} \left(\frac{e^{\widehat{y}_t}}{e^{\widehat{y}_t^{flex}}}\right)^{\psi_2(1-\rho_R)} \left(\frac{e^{\psi_3(\widehat{y}_t - \widehat{y}_{t-1})}}{e^{\psi_3(\widehat{y}_t^{flex} - \widehat{y}_{t-1}^{flex})}}\right) r_* e^{\widehat{r}_t}$$

it becomes

$$e^{\widehat{R}_t} = e^{\rho_R \widehat{R}_{t-1}} e^{\psi_1 \widehat{\pi}_t (1-\rho_R)} \left(\frac{e^{\widehat{y}_t}}{e^{\widehat{y}_t^{flex}}}\right)^{\psi_2(1-\rho_R)} \left(\frac{e^{\psi_3(\widehat{y}_t - \widehat{y}_{t-1})}}{e^{\psi_3(\widehat{y}_t^{flex} - \widehat{y}_{t-1}^{flex})}}\right) e^{\widehat{r}_t} \quad \left(\because \left(\frac{y_*}{y_*^{flex}}\right)^{\psi_2(1-\rho_R)} r_* = 1\right)$$

then

$$e^{\widehat{R}_t} = e^{\rho_R \widehat{R}_{t-1} + \psi_1 \widehat{\pi}_t (1-\rho_R) + \psi_2(1-\rho_R)(\widehat{y}_t - \widehat{y}_t^{flex}) + \psi_3((\widehat{y}_t - \widehat{y}_{t-1}) - (\widehat{y}_t^{flex} - \widehat{y}_{t-1}^{flex})) + \widehat{r}_t}$$

Applying the Taylor approximation,

$$1 + \widehat{R}_t \approx 1 + \rho_R \widehat{R}_{t-1} + \psi_1 \widehat{\pi}_t (1-\rho_R) + \psi_2(1-\rho_R)(\widehat{y}_t - \widehat{y}_t^{flex}) + \psi_3((\widehat{y}_t - \widehat{y}_{t-1}) - (\widehat{y}_t^{flex} - \widehat{y}_{t-1}^{flex})) + \widehat{r}_t$$

$$\therefore \widehat{R}_t \approx \rho_R \widehat{R}_{t-1} + (1-\rho_R)(\psi_1 \widehat{\pi}_t + \psi_2(\widehat{y}_t - \widehat{y}_t^{flex})) + \psi_3((\widehat{y}_t - \widehat{y}_{t-1}) - (\widehat{y}_t^{flex} - \widehat{y}_{t-1}^{flex})) + \widehat{r}_t$$

For the resource constraint,

$$c_t + i_t + y_* g_t + a(Z_t)\frac{\overline{k}_{t-1}}{\gamma} = y_t$$

it can be re-write as

$$c_* e^{\widehat{c}_t} + i_* e^{\widehat{i}_t} + y_* g_* e^{\widehat{g}_t} + a(Z_* e^{\widehat{Z}_t})\frac{k_* e^{\widehat{k}_{t-1}}}{\gamma} = y_* e^{\widehat{y}_t}$$

the approximation becomes

$$c_*(1+\widehat{c}_t) + i_*(1+\widehat{i}_t) + y_* g_*(1+\widehat{g}_t) + a(Z_*(1+\widehat{Z}_t))\frac{\overline{k}_*(1+\widehat{\overline{k}}_{t-1})}{\gamma} \approx y_*(1+\widehat{y}_t)$$

$$c_*\widehat{c}_t + i_*\widehat{i}_t + y_* g_* \widehat{g}_t + \frac{a'(1)\widehat{Z}_t \overline{k}_*}{\gamma} \approx y_* \widehat{y}_t \quad \Rightarrow \quad c_*\widehat{c}_t + i_*\widehat{i}_t + y_* g_* \widehat{g}_t + r_*^k k_* \widehat{Z}_t \approx y_* \widehat{y}_t$$

$$\therefore \quad \widehat{y}_t = \frac{c_*}{y_*}\widehat{c}_t + \frac{i_*}{y_*}\widehat{i}_t + g_* \widehat{g}_t + \frac{r_*^k k_*}{y_*}\widehat{Z}_t$$

From the equation,

$$\frac{1}{\lambda_{p,t}}\left(\widetilde{p}_t-(1+\lambda_{p,t})mc_t\right)y_t+E_t\sum_{s=1}^{\infty}\zeta_p^s\bar{\beta}^s\gamma^s\frac{\xi_{t+s}}{\xi_t}\frac{1}{\lambda_{p,t+s}}\left(\widetilde{p}_t\frac{\Pi_{l=1}^s\pi_{t+l-1}^{\iota_p}\pi_*^{1-\iota_p}}{\Pi_{l=1}^s\pi_{t+l}}-(1+\lambda_{p,t+s})mc_{t+s}\right)y_{t+s}=0$$

multiply by $\lambda_{p,t}$ and $\dfrac{1}{y_t}$, then

$$\widetilde{p}_t-(1+\lambda_{p,t})mc_t+E_t\sum_{s=1}^{\infty}\zeta_p^s\bar{\beta}^s\gamma^s\frac{\xi_{t+s}}{\xi_t}\frac{\lambda_{p,t}}{\lambda_{p,t+s}}\left(\widetilde{p}_t\frac{\Pi_{l=1}^s\pi_{t+l-1}^{\iota_p}\pi_*^{1-\iota_p}}{\Pi_{l=1}^s\pi_{t+l}}-(1+\lambda_{p,t+s})mc_{t+s}\right)\frac{y_{t+s}}{y_t}=0$$

taking total differentiation at steady state, then

$$\begin{aligned}d(\widetilde{p}_t)-mc_*d(\lambda_{p,t})-(1+\lambda_{p,*})d(mc_t)+E_t\sum_{s=1}^{\infty}\zeta_p^s\bar{\beta}^s\gamma^s\Bigg[\frac{1}{\xi_*}(1-(1+\lambda_{p,*})mc_*)d(\xi_{t+s})\\
-\frac{1}{\xi_*}(1-(1+\lambda_{p,*})mc_*)d(\xi_t)+\frac{1}{\lambda_{p,*}}(1-(1+\lambda_{p,*})mc_*)d(\lambda_{p,t})-\frac{1}{\lambda_{p,*}}(1-mc_*)d(\lambda_{p,t+s})+d(\widetilde{p}_t)\\
+d\left(\frac{\Pi_{l=1}^s\pi_{t+l-1}^{\iota_p}\pi_*^{1-\iota_p}}{\Pi_{l=1}^s\pi_{t+l}}\right)+\frac{1}{y_*}(1-(1+\lambda_{p,*})mc_*)d(y_{t+s})-\frac{1}{y_*}(1-(1+\lambda_{p,*})mc_*)d(y_t)\\
-(1+\lambda_{p,*})d(mc_{t+s})\Bigg]=0\end{aligned}$$

Since $1-(1+\lambda_{p,*})mc_*=0 \quad\Leftrightarrow\quad (1+\lambda_{p,*})mc_*=1$ then we have

$$\begin{aligned}d(\widetilde{p}_t)-mc_*d(\lambda_{p,t})-(1+\lambda_{p,*})d(mc_t)+E_t\sum_{s=1}^{\infty}\zeta_p^s\bar{\beta}^s\gamma^s\Bigg[-mc_*d(\lambda_{p,t+s})+d(\widetilde{p}_t)+\\
+d\left(\frac{\Pi_{l=1}^s\pi_{t+l-1}^{\iota_p}\pi_*^{1-\iota_p}}{\Pi_{l=1}^s\pi_{t+l}}\right)-(1+\lambda_{p,*})d(mc_{t+s})\Bigg]=0\end{aligned}$$

for the equation of $t+1$ at time t,

$$\begin{aligned}\zeta_p\overline{\beta}\gamma\Bigg\{d(\widetilde{p}_{t+1})-mc_*d(\lambda_{p,t+1})-(1+\lambda_{p,*})d(mc_{t+1})+E_t\sum_{s=1}^{\infty}\zeta_p^s\bar{\beta}^s\gamma^s\Bigg[-mc_*d(\lambda_{p,t+1+s})+d(\widetilde{p}_{t+1})+\\
+d\left(\frac{\Pi_{l=1}^s\pi_{t+l}^{\iota_p}\pi_*^{1-\iota_p}}{\Pi_{l=1}^s\pi_{t+1+l}}\right)-(1+\lambda_{p,*})d(mc_{t+1+s})\Bigg]\Bigg\}=0\end{aligned}$$

By taking difference, we can have

$$\frac{1}{1-\zeta_p\overline{\beta}\gamma}d(\widetilde{p}_t)-(1+\lambda_{p,*})d(mc_t)-mc_*d(\lambda_{p,t})-\frac{\zeta_p\overline{\beta}\gamma}{1-\zeta\overline{\beta}\gamma}d(\widetilde{p}_{t+1})+\frac{\zeta_p\overline{\beta}\gamma}{1-\zeta\overline{\beta}\gamma}\left(\frac{\iota_p}{\pi_*}d(\pi_t)-\frac{1}{\pi_*}d(\pi_{t+1})\right)=0$$

then using the approximation method,

$$\frac{dx_t}{x} \approx \frac{x_t - x}{x} \approx log\left(\frac{x_t}{x}\right) \approx \widehat{x}_t \quad \therefore \quad dx_t \approx x\widehat{x}_t$$

it becomes

$$\frac{1}{1-\zeta_p\overline{\beta}\gamma}\widehat{\widehat{p}}_t - (1+\lambda_{p,*})mc_*\widehat{mc}_t - mc_*\lambda_{p,*}\widehat{\lambda}_{p,t} - \frac{\zeta_p\overline{\beta}\gamma}{1-\zeta\overline{\beta}\gamma}\widehat{\widehat{p}}_{t+1} + \frac{\zeta_p\overline{\beta}\gamma}{1-\zeta\overline{\beta}\gamma}\left(\iota_p\widehat{\pi}_t - \widehat{\pi}_{t+1}\right) = 0$$

using the equation,

$$\widehat{\widehat{p}}_t = \frac{\zeta_p}{1-\zeta_p}\left(\widehat{\pi}_t - \iota_p\widehat{\pi}_{t-1}\right)$$

we can re-write previous equation as

$$\frac{1}{1-\zeta_p\overline{\beta}\gamma}\frac{\zeta_p}{1-\zeta_p}\left(\widehat{\pi}_t - \iota_p\widehat{\pi}_{t-1}\right) - \widehat{mc}_t - mc_*\lambda_{p,*}\widehat{\lambda}_{p,t} - \frac{\zeta_p\overline{\beta}\gamma}{1-\zeta\overline{\beta}\gamma}\frac{\zeta_p}{1-\zeta_p}\left(\widehat{\pi}_{t+1} - \iota_p\widehat{\pi}_t\right) + \frac{\zeta_p\overline{\beta}\gamma}{1-\zeta\overline{\beta}\gamma}\left(\iota_p\widehat{\pi}_t - \widehat{\pi}_{t+1}\right) = 0$$

and it can be rearranged by

$$\begin{aligned}\frac{(1-\zeta_p\overline{\beta}\gamma)(1-\zeta_p)}{\zeta_p}\left(\widehat{mc}_t + mc_*\lambda_{p,*}\widehat{\lambda}_{p,t}\right) &= \left(\widehat{\pi}_t - \iota_p\widehat{\pi}_{t-1}\right) - \zeta_p\overline{\beta}\gamma\left(\widehat{\pi}_{t+1} - \iota_p\widehat{\pi}_t\right) + (\overline{\beta}\gamma - \zeta_p\overline{\beta}\gamma)\left(\iota_p\widehat{\pi}_t - \widehat{\pi}_{t+1}\right) \\ &= (1+\overline{\beta}\gamma\iota_p)\widehat{\pi}_t - \iota_p\widehat{\pi}_{t-1} - \overline{\beta}\gamma\widehat{\pi}_{t+1}\end{aligned}$$

$$\therefore \quad \widehat{\pi}_t = \frac{\iota_p}{1+\overline{\beta}\gamma\iota_p}\widehat{\pi}_{t-1} + \frac{\overline{\beta}\gamma}{1+\overline{\beta}\gamma\iota_p}\widehat{\pi}_{t+1} + \frac{(1-\zeta_p\overline{\beta}\gamma)(1-\zeta_p)}{\zeta_p(1+\overline{\beta}\gamma\iota_p)}\left(\widehat{mc}_t + \frac{\lambda_{p,*}}{1+\lambda_{p,*}}\widehat{\lambda}_{p,t}\right)$$

Note. One of the different feature in SW 2007 from the SW 2003 is that they use Kimball(1995)'s more general aggregator instead of the Dixit-Stiglitz aggregator in the intermediate goods and labor market. However, if the Kimball's aggregator is the case of $G(x) = x^{1/(1+\lambda)}$ and $\frac{(1+\frac{G''}{G'})}{(2+\frac{G'''}{G''})} = 1$ then we have the identical New Keynesian Phillips Curve (NKPC).