To My Family

COMPARISION OF GMM, MAXIMUM LIKELIHOOD AND BAYESIAN ESTIMATIONS IN ESTIMATING STRUCTURAL PARAMETERS OF DSGE MODELS

Graduate School of Social Sciences

TOBB University of Economics and Technology

HASAN HALİT TOPRAK

In Partial Fulfillment of the Requirements for the Degree
of
Master of Science

in

The Deparment of Economics
TOBB University of Economics and Technology

ANKARA

June 2014

I hereby certify that this thesis meets all the requirements of the Graduate School of Social Sciences for a Master's degree.

Prof. Serdar Sayan
Dean, Graduate School of Social Sciences

I have read this thesis and found its content and scope adequate for a Master's thesis in Economics in the Graduate School of Social Sciences.

Assoc. Prof. Bedri Kamil Onur Taş
Supervisor

Assoc. Prof. Hulusi Öğüt
Committee Member

Assist. Prof. Süleyman Cem Karaman
Committee Member

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I have therefore fully cited and referenced all material and results that are not original to this work.

Hasan Halit Toprak

ABSTRACT

COMPARISION OF GMM, MAXIMUM LIKELIHOOD AND BAYESIAN ESTIMATIONS IN ESTIMATING STRUCTURAL PARAMETERS OF DSGE MODELS

Toprak, Hasan Halit
Master of Economics
Supervisor: Assoc. Prof. Bedri Kamil Onur Taş

June 2014

Dynamic stochastic general equilibrium (DSGE) models are macroeconomic models derived from microeconomic principles. These models and estimation methods of their parameters have been very popular among macroeconomists over the past 25 years. Identification of structural parameters of DSGE models is subject of many studies. In this study we compare three estimation methods: Bayesian estimation, maximum likelihood estimation and generalized method of moments estimation, in the cases of (i) if the shocks have an autocorrelated pattern (ii) if data is small.

We generate artificial data at the length of 60 and 180 by using our model with true parameters and obtain the estimated parameters from these estimators. Then, for every estimator, we compare the value of estimated parameters with true ones. The model we use is the three equation New Keynesian model including the Euler condition, Philips curve and monetary policy equations.

As a result of comparison, for all cases, with and without autocorrelation and small and large sample sizes, Bayesian estimation performs best. However, it should be noted that if Dynare allowed us to expand the border of priors for Bayesian estimator, the result might be different.

Keywords: DSGE Models, Bayes. Generalized Method of Moments, Maximum Likelihood, Structural Parameters

ÖZET

RASTSAL GENEL DENGE MODELLERİNİN YAPISAL PARAMETRELERİNİN TAHMİNİNDE GMM, EN ÇOK OLABİLİRLİK VE BAYES TAHMİN METODLARININ KARŞILAŞTIRILMASI

Toprak, Hasan Halit
İktisat Yüksek Lisans
Tez Danışmanı: Doç.Dr. Bedri Kamil Onur Taş

Haziran 2014

Rastsal genel denge modelleri, mikroekonomik prensiplerden elden edilen makroekonomik modellerdir. Bu modeller ve parametre tahmin yöntemleri ekonomistler arasında 25 yıldır popülerliğini koruyor. Bu genel denge modellerinin yapısal parametrelerinin belirlenmesi konusu bir çok akademik çalışmanın ana teması olagelmiştir. Bu çalışmada biz şu 3 tahmin yöntemini (i) şoklar otokorelasyona sahip olursa (ii) ve veri seti küçük ise durumları altında karşılaştırıyoruz: Bayesyan yöntemi, en çok olabilirlik ve GMM.

Gerçek parametreleri kullanarak 60 ve 180 uzunluğunda veri üretip, bu verileri kullanarak bahsi geçen üç tahmin yöntemi ile yapısal parametreleri tahmin ediyoruz. Daha sonra her bir tahmin yöntemi için bu tahmin edilen parametre ile gerçek parametreyi karşılaştırıyoruz. Kullandığımız model, Euler şartı, Philips eğrisi ve para politikası denkleminden oluşan bir Yeni Keynesyen model.

Karşılaştırma sonucunda, bütün durumlarda, otokorelasyon olsun veya olmasın, veri büyüklüğü küçük olsun ya da olmasın, Bayes yöntemi en iyi yöntem olarak görüldü. Fakat şu unutulmamalıdır ki, eğer Dynare Bayes yönteminde kullandığımız öncüllerin sınırlarını genişletmemize izin verseydi, sonuç farklı olabilirdi.

Anahtar Kelimeler: Rastsal Genel Denge Modelleri, Bayes. Genelleştirilmiş Moment Metodu,En Çok Olabilirlik, Yapısal Parametreler

ACKNOWLEDGMENTS

First of all, I would like to send my warmest thanks to my family who were always there for me and who were always generous in their supports all my life and especially to my father whose moral support I deeply felt at all stages of this study.

I would also like to thank Professor Ozan EKŞİ who was helpful in the formation of the main frame of this study and in bringing this thesis to a certain stage, to Professor Bedri Kamil Onur TAŞ who was generous in helping me complete this thesis and in showing me the way in the stressful procedure of PhD applications, to my jury members Professor Hulusi ÖĞÜT and to Professor Süleyman Cem KARAMAN who corrected my mistakes with their valuable opinions and who brought this study to a better point during the process of defending my thesis.

I would also like to extend my gratitude to all the faculty members of TOBB UET Economics from whose precious knowledge I made the most of during my graduate and undergraduate programmes.

Many appreciations go to Senem ÜÇBUDAK, the secretary of the institute, who always cleaned up after us during the masters programme.

My heartfelt thanks also go to the head of the Economics Department in Turgut Ozal University, Professor Ramazan TAŞ who made all sorts of conveniences while I was writing my thesis.

And of course my wholehearted thanks and gratitude go to my precious fiancé Vasfiye Betül CAN who, since the day she has become part of my life, has made everything much more easy for me not only in terms of this study but in all aspects of my life.

CONTENTS

Table List

CHAPTER ONE

INTRODUCTION

Dynamic stochastic general equilibrium (DSGE) models are macroeconomic models derived from microeconomic principles. They usually contain optimizing agents, rational expectations, and market clearing processes. The agents interact in an uncertain environment. Because these models are built on microeconomic rules and agents, they are very useful for understanding the economy as a whole and trough which cannels it works.

DSGE models are used in various fields of economics, especially in growth, monetary policy, international trade and finance. They provide great opportunities and useful tools for understanding the source of fluctuations in the economy. They can also help analyzing the shortrun and longrun outcomes of alternative policies. Additionally, these models can be used to evaluate welfare effects of macro policies. For all these advantages, DSGE models have been very popular among macroeconomists over the past 25 years. Because of their usefulness for policy analysis and forecasting, today many central banks and relevant government departments have their own DSGE models.

The studies on econometric analysis of DSDE models have been shaped around evaluating the model and identifying structural parameters of the model. In this study we focus on identification part. Calibration, moment based estimations, impulse response matching are limited information estimators. Likelihood based estimations (e.g. maximum likelihood, Bayesian) are full information estimation methods that are used to estimate structural parameters of DSGE models.

In this study we compare three estimation methods: Bayesian estimation, maximum likelihood estimation and generalized method of moments estimation. Our aim is not to derive conclusions on identification. Our purpose is to summarize three estimators mentioned above briefly for a curios reader who wishes to find basic information about them and to compare the estimation techniques in the following cases: (i) if the shocks have an autocorrelated pattern (ii) if data is small.

In the literature, there is wide variety of study about each estimator, but comparison is done by a few. None of these includes the three estimators that we compare at the same time. In his famous work, Ruge-Murcia (2007) compares Maximum Likelihood, Generalized Method of Moments, Simulated Method of Moments and Indirect Inference. He examines them under the situation of misspecification, small sample case, and compares their computing time. He found that singularity and misspecification affects Maximum Likelihood more severely than others. The result of Monte Carlo analysis shows moment-based techniques more robust to misspecification than likelihood-based techniques. In terms of computing time, GMM is most efficient. Giesen and Scheufele (2013) analyze the small sample properties of full information and limited information estimators in a potentially misspecified DSGE model. They found that if the model is correctly specified, full information estimators performs superior. Canova and Sala (2009) shows that observational equivalence, weak and partial identification make objective functions to have flat surfaces which lead to biases. Identification problems can be detected by using Bayes estimation properly. They also demonstrate that small sample size has negative effect on identification of parameters. Dridi, Guay, and Renault (2007) proposes sequential partial indirect inference (SPII) approach as an alternative to calibration-like estimators in the case of misspecification. Boivin and Giannoni (2006) states that exploiting more information in the model estimation makes more accurate estimation of the model's concept and shocks and also it is important for the conclusions about key structural parameters.

Since we are not interested in evaluating a model or comparing different models of DSGE but to find the most efficient estimator in various cases, we generate artificial data by using our model with true parameters, instead of using real data. By applying the estimation methods to the generated data, we obtain the estimated parameters. Then, for every estimator, we compare the value of estimated parameters with true ones.

The DSGE model we use in the paper is the three equation New Keynesian model. This model includes the Euler condition, Philips curve and monetary policy equations. We use both backward and forward looking terms in the first two equations (following Gali and Gertler 1999), and also autocorrelated error terms.

In order to carry out the GMM estimation we use MATLAB. For Bayesian and Maximum likelihood estimations we use Juillard's (1996) DYNARE package on MATLAB.

The paper is structured as follows. After this introduction, section 2 expresses the model, in section 3 three estimation methods are summarized, section 4 presents the estimation results, and section 5 concludes.

CHAPTER TWO

MODEL

The model is a small scale hybrid New Keynesian model that includes Monetary Policy, Philipps Curve and Euler condition equations:

$$i_t = \lambda_r i_{t-1} + (1-\lambda_r)\left(\lambda_\pi \pi_{t-1} + \lambda_y y_{t-1}\right) + e_{1t}$$

$$\pi_t = \frac{\omega}{1+\omega\beta}\pi_{t-1} + \frac{\beta}{1+\omega\beta}E_t\pi_{t+1} + \frac{(\emptyset+\theta_n)(1-\zeta\beta)(1-\zeta)}{(1+\omega\beta)\zeta}y_t + v_{1t}$$

$$y_t = \frac{h}{1+h}y_{t-1} + \frac{1}{1+h}E_t y_{t+1} - \frac{1}{\emptyset}\left(i_t - E_t\pi_{t+1}\right) + v_{2t}$$

where $v_{1t} = \rho_1 v_{1t-1} + e_{2t}$, $v_{2t} = \rho_2 v_{2t-1} + e_{3t}$ and e_{1t}, e_{2t}, e_{3t} are i.i.d., h is habit persistence parameter in consumption, $\emptyset$ and θ_n are relative risk aversion coefficients for consumption and labor respectively, ω is the degree of price indexation (fraction of backward looking firms), ζ is the degree of Calvo type price stickiness and λ's are policy parameters.

We simulate the data of output gap, inflation and nominal interest rate from this model at lengths 60 and 180. Consequently, our estimations have small sample implications. In the simulations, we use parameter estimates of Rabanal and Rubio-

Ramirez (2005) from the US data. The priors and their distributions required in Bayesian estimation are also taken from Rabanal and Rubio-Ramirez (2005). The parameters are:

h	$\emptyset$	ω	β	ζ	λ_r	λ_π	λ_y	ρ_1	ρ_2	θ_n
0.85	2.0	0.25	0.985	0.68	0.2	1.55	1.1	0.65	0.65	3.0

In all the parameter estimations written below, we assume true model is known, but not these parameters, except for θ_n. As explained section 3.2, in order to find structural parameters by using GMM, one parameter must be given since we need to import these structural parameters from reduced forms.

CHAPTER THREE

ESTIMATION METHODS

3.1 Bayesian Estimation

First estimation technique that we analyze is Bayesian estimation. Bayesian estimation is relatively new to the others but lately the most preferred one. Among the famous papers using the Bayesian estimation, Schorfheide (2000) proposes a Bayesian econometric procedure for the evaluation and comparison of DSGE models. Rabanal and Rubio-Ramirez (2005) estimates and compares four versions of the New Keynesian monetary model with nominal rigidities by using Bayesian estimation. Adolfson, Laseen, Linde, and Svenson (2011) uses Bayesian estimation for construction optimal policy projections in Ramses, the Sveriges Riksbank's DSGE model.

There are several advantages of using Bayesian estimation of DSGE models. First of all, Bayesian analysis allows one to use prior beliefs that are derived either from macroeconomic or microeconomic studies. For this reason, it can be seen as a version of calibration. This provides great opportunity to take previous literature into consideration. Second, since priors are used for calibration -instead of estimation- one can estimate a wide variety of models under identification problems with Bayesian estimation. Third, using priors helps us avoiding reaching strange points. Likelihoods of DSGE models contain lots of local maxima and minima and they are nearly flat surfaces. Thus, if we put some restrictions to the likelihood functions, as Bayesian approach does by using priors, we may eliminate irrational outcomes. Forth, it also helps eliminating identification problems. If posterior distribution is flat, then usually there is identification problem. In many times, using priority makes enough curvature in the posterior distribution, hence, remove the identification problem.

Basic logic behind the Bayesian estimation is Bayes' theorem: for events A and B, the conditional probability of event A given that B has occurred is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In our context, this equation becomes

$$P(\theta|Y_T) = \frac{P(Y_T|\theta)P(\theta)}{P(Y_T)} \tag{1}$$

where Y_T is the data, θ is parameter set, $P(\theta|Y_T)$ is the posterior distribution that we want to find, $P(\theta)$ is our prior beliefs about parameters, $P(Y_T|\theta)$is our likelihood function that gives us the probability that model assigns to each observation given some parameter values and finally $P(Y_T)$ is the marginal density of the data. Since this marginal density is constant for any parameter set, equation (1) can be written as

$$P(\theta|Y_T) \propto P(Y_T|\theta)P(\theta) \tag{2}$$

The $P(Y_T|\theta)$ term describes the probability of realization of the data given the set of parameters, which can be written in terms of the following likelihood function

$$L(\theta|Y_T) \equiv P(Y_T|\theta),$$

Hence, equation (2) can be written as

$$P(\theta|Y_T) \propto L(\theta|Y_T)P(\theta) \tag{3}$$

In words, the posterior density is proportional to likelihood function that we need to estimate times the prior density that we need to calibrate. The posterior is a mixture of the prior information and the "current information" that is, the data.

The next step is evaluating the likelihood function. When the model is linear in terms of the structural parameters θ, we can use maximum likelihood principles. However, in DSGE models this is not the case most of time. Except in a few cases, there is no analytical or numerical procedure to write down the likelihood function. Moreover, these models frequently involve unobserved or poorly measured state variables. Yet, since DSGE models are linear in terms of the variables, the likelihood may be evaluated with a linear prediction error algorithm like the Kalman filter, which also deals with unobserved state variables. However, in order to use Kalman filter we need to make another assumption: the shocks of the system are normally distributed. Kalman filter is a minimum mean squared error estimator. The prediction errors of this filter are normally distributed. So given this distribution, we can obtain probability distribution (probability of observing the data) given the set of parameters: $P(Y_T|\theta)$. This is also the likelihood of the data given the parameters. Kalman filter is calculated for every θ, each time based on known θ. (Detailed description of the Kalman filter is in the appendix A.) Now that the likelihood function is evaluated by using Kalman filter, we are about to find the distribution of (3). However, the $L(\theta|Y_T)P(\theta)$ term in (3) is not directly a function of θ but the functions of θ in the state equation in the Kalman filter. Hence, instead of maximizing this equation with respect to θ, we use numerical methods such as Metropolis-Hastings algorithm. It works as follows. We draw some parameter set, say θ_A, from the priors and calculate the term in (3). Then assuming that θ_A is the mean of a normal distribution, pick another set of parameters, say θ_B, from this distribution and calculate the value of the (3), which is the posterior density. If θ_B improves the term in (3), reset the new mean of a normal distribution to the θ_B, and pick new draws, and again calculate the density. Otherwise use θ_A as a mean and again continue to pick new draws. When you make enough sampling, build a histogram of the retained values. This "smoothed histogram" will eventually be the posterior distribution. This is the basic mechanism of Bayesian estimation.

3.2 GMM Estimation

Another estimation technique that we employ is Generalized Method of Moments (GMM). GMM is developed by Hansen (1982), and applied to DSGE models by Christiano and Eichenbaum (1992) for the first time. From that time, GMM has been widely used in macroeconomics and finance.

GMM has some attractive properties. Complete knowledge of the distribution of the data is not necessarily required for GMM to be carried out. This is one of the important features that separate GMM from likelihood-based estimations. What one only needs to perform GMM is specified moments which are obtained from an underlying model. In the cases in which there are more moment conditions than model parameters, GMM is very useful to test proposed model specification. Also, in contrast to likelihood-based estimations, normality assumption of the structural shocks is not required in GMM estimation. However, in return, GMM lose efficiency compared to the likelihood-based estimators.

In contrast to ML and Bayesian estimations, GMM is does not use all the system of equations simultaneously. ML and Bayesian estimations directly estimate the structural parameters of the model. However, GMM uses one equation at a time, hence, estimates reduced form coefficients. In order to find structural parameters, GMM requires their mapping from the reduced form estimates.

GMM is based on moment estimation. It uses model equations, separately, to specify some moment conditions. These moment conditions are functions of the model parameters and the data, such that their expectation is zero at the true values of the parameters. As a result, in order to estimate model parameters, GMM minimizes the moment conditions by using the sample data. In the case in which there are more equations (moment conditions) than the unknowns, GMM uses all moment conditions by weighting them.

The equation to be estimated

$$y = X\beta + u,$$

let's define $g_t\,(\beta)$ as

$$g(\beta) = \frac{1}{n} Z'(y - X\beta),$$

where y is a vector of n*1 dependent variable, X a matrix of is n*K independent variable, $(y - X\beta)$ is the moment condition and Z's are instruments that are orthogonal to this moment condition. Z's can be 1 or any variable that is exogenous to $(y - X\beta)$, that is u. The number of Z's, say L, gives us the number of moment conditions used in GMM estimation. If the equation to be estimated is exactly identified, so that L = K, then we have as many equations—the L moment conditions—as we do unknowns—the K coefficients. In this case it is possible to find a $\hat{\beta}$ that solves $g(\beta) = 0$. If the equation is overidentified, however, so that L > K, then we have more equations than we do unknowns, and in general it will not be possible to find a $\hat{\beta}$ that will set all L sample moment conditions to exactly zero. In this case, we take a weighting matrix W and use it to construct a quadratic form in the moment conditions.

This gives us the GMM objective function:

$$J(\beta) = \min\{\mathrm{n}\, g(\beta)' W g(\beta)\},$$

where W is the weighting matrix. A GMM estimator for β is the β^ that minimizes J(β). Deriving and solving the K first order conditions

$$\frac{\partial J(\beta)}{\partial \beta} = 0,$$

$y$ields the GMM estimator

$$\beta\hat{} = (X'ZWZ'X)^{-1}X'ZWZ'y.$$

See Baum et al. (2002) for further details on GMM estimation.

The intuition behind GMM is choosing an estimator for β which sets these L sample moments as close to zero as possible. For weighting matrix, we use Hansen's (1982) optimal weighting matrix which uses inverse of the spectral density of the calculated moments so that more weight is given to moment conditions with less uncertainty. For an initial weighting matrix, we use $(Z'Z)^{-1}$. Since there is no autocorrelation in the error term of this equation, we include White's (1980) kernel based estimator, which accounts for heteroskedasticity in the calculation of a spectral density matrix.

3.3 Maximum Likelihood Estimation

Maximum Likelihood (ML) estimation is another full information estimation method. Among the early studies that use ML estimation are Christiano (1988), Altug (1989), Bencivenga (1992), McGrattan (1994) which are also related to this work. Since then, ML keeps its popularity in estimating structural parameters of DSGE model.

ML has some distinctive features. At first, ML uses all available information. ML requires the construction and evaluation of the likelihood function of the data given the parameters. For this purpose, it uses all the information that is available. Remember that GMM does not require all the information, and we mentioned this as an advantage of GMM. Here we say ML uses all information, this is also an advantage. Is this not a contrast? No, actually not. It depends on the data that you have. If you are confident that your model contains enough information, using ML is an advantage, and it is an efficient method of estimation for you. However, if the model that you use does not have enough information, then this time using GMM is more advantageous for you. Second, in estimating parameters, ML provides a consistent approach. This means that maximum likelihood estimates can be developed for a large variety of estimation problems. Third, ML allows comparing the alternative models easily which cannot be done by using GMM methods.

The basic mechanism of ML is as follows. The joint probability density function of completely specified and independent sample observations is

$$f(y_1, \ldots, y_n|\theta) = \prod_{i=1}^{n} f(y_i|\theta) = L(\theta|y)$$

Intuitively, this is the probability of observing a particular data given some parameter set. Once you calculate this probability for different θ, the one that creates the higher value is more likely the one that generates the data. So given the data, you can create a distribution for θ's (their likelihood) showing their probabilities of generating the data. That is why L(.) uses θ on the left hand side. The main goal of the likelihood function is to find the parameter that makes the observed data most probable. In other words, we are exploring the parameter set that maximizes the likelihood function.

In ML estimation method, the trick is to derive the likelihood function accurately. It is difficult because of the characteristics of the DSGE models, as explained in section 3.1. In here, we can follow the same pattern as we do in Bayesian likelihood function. We will evaluate the likelihood function again by using Kalman filter. There are some studies that deal with the derivation of likelihood function without using Kalman filter. See Fernández-Villaverde and Rubio-Ramírez (2007) and An and Schorfheide (2007)

After evaluating the likelihood function, the next step is finding the parameter set that maximizes it. Normally, in order to find the value of the parameters that maximize the objective function, we take the first derivative of the likelihood respect to θ, and the value that equates this derivation to zero is called the maximizing value, if the second derivative of the function again respect to θ is negative. Taking the natural logarithm of the objective function makes finding maximizing value easier. However, the likelihood function of the DSGE models is too complicated to perform these simple processes. They usually contain variety of dimensions, therefore taking derivative and then equating it to zero cannot be possible most time. To overcome this problem, in this study, we use Covariance Matrix Adaptation Evolution Strategy (CMA-ES) routine for optimization. Dynare provides us some other options which take less time, however, after some replications, they fail. Therefore, for the model we use, CMA-ES is the best option. The results of CMA-ES routines give us estimated structural parameters of the model.

CHAPTER FOUR

COMPRASION RESULTS

Eight unknown parameters are estimated by using three estimators. We assume that $\boldsymbol{\theta_n}$ is known. For each estimation method, we generate the data of sample size 60 and 180. Each observation corresponds one quarter. Therefore, we create the data for 15 and 45 years which can be considered as small sample size and large sample size, respectively. In generating the data, we use the true values of the parameters which are given in the model section. Every time that we generate the data, we simulate 1000 extra observations and then discard them. We do not generate the data one by one but we did in a continuous manner. This is because in the first case, every data set would be same with the previous one. After simulation, the data sets were checked and it was seen that they were different from each other. All experiments are based on 250 replications. After that, the mean and standard deviation of the replications are saved for comparison.

There are four different cases for each estimator. First we conduct estimations with shocks that are autocorrelated. After that we discard autocorrelation and run the estimations. For autocorrelation we use $\boldsymbol{\rho_1}=\boldsymbol{\rho_2}$=0,65. We examine the cases in which $\boldsymbol{\rho_1}=\boldsymbol{\rho_2}$=0,90 in the Appendix B. Also, for every situation there are two different data sets: sample size of 60 and 180. Therefore we will examine four cases for each estimator.

The results are documented in the tables for all of four cases. In the tables, “mean” is the average value of all 250 replications, “sd” is standard deviation of these replications.

In terms of computing time, in this study, GMM is most efficient as in Ruge-Murcia (2007). ML is the second and Bayes is the worst efficient estimator. This is because ML and Bayes have to solve the whole model. Also, after evaluating the likelihood function, there is no numerical method for finding posterior distribution and the value that maximizes the likelihood. The Dynare uses some algorithms. This also requires longer time.

4.1 Shocks are Autocorrelated

In this case, all the estimation processes (data generation, model declaration and estimation) are done by using the model with autocorrelated shocks.

i. Small Sample

All tree estimators’ results of the model with autocorrelated shocks and the sample size is 60 are below.

Table 1 Shocks Are Aotucorrelated (0.65) And Sample Is Small

		β	$\emptyset$	ζ	λ_r	λ_π	λ_y	Ω	h
	true	0,985	2	0,68	0,2	1,55	1,1	0,25	0,85
	mean	0,9983	1,6169	0,9626	0,7737	3,303	1,4637	0,091	0,7561
GMM	bias	-0,0133	0,3831	-0,2826	-0,5737	-1,753	-0,3637	0,159	0,0939
	sd	0,0058	0,3132	0,047	0,1883	5,1257	7,3481	0,1228	1,1633
	mean	0,6327	2,4736	0,5934	0,214	1,2327	1,1291	0,2497	0,3004
ML	bias	0,3523	-0,4736	0,0866	-0,014	0,3173	-0,0291	0,0003	0,5496
	sd	0,1748	0,1855	0,1518	0,0902	0,1886	0,2284	0,1648	0,2175
	mean	0,9726	1,6147	0,4338	0,2294	1,5678	1,1008	0,2587	0,8128
Bayes	bias	0,0124	0,3853	0,2462	-0,0294	-0,0178	-0,0008	-0,0087	0,0372
	sd	0,0187	0,1251	0,0678	0,0432	0,0335	0,0015	0,0291	0,037

As we can see from the table, the mean of β is nearly equal to its true value in GMM and Bayes, but mean of ML estimates is biased. For $\emptyset$, again the mean of GMM and Bayes are very similar to each other and they estimated 0,4 point below, whereas ML estimated 0,4 point above from the true value. Similarly, none of the mean of estimated ζ is near to its true value, but ML is most close one. For λ_r, λ_π, λ_y and Ω, ML and Bayes are very close to actual value, indeed, the means of Bayes are almost equal to actual values, nevertheless, GMM seems much more biased. For h, again Bayes is the best but this time ML is the worst one. It is important to note that for λ_π and λ_y standard deviation of GMM is very high.

ii. Large Sample

This time the model has again autocorrelated shocks but the data is consisted of 180 observations. The table is below.

Table 2 Shocks Are Aotucorrelated (0.65) And Sample Is Large

		β	$\emptyset$	ζ	λ_r	λ_π	λ_y	Ω	h
	true	**0,985**	**2**	**0,68**	**0,2**	**1,55**	**1,1**	**0,25**	**0,85**
	mean	0,9993	1,2106	0,9726	0,8777	1,8389	1,76	0,1021	0,4794
GMM	bias	-0,0143	0,7894	-0,2926	-0,6777	-0,2889	-0,66	0,1479	0,3706
	sd	0,0026	0,1371	0,0267	0,1	5,4723	6,4263	0,1242	0,2682
	mean	0,6608	2,482	0,5992	0,0968	1,2767	1,299	0,2828	0,3673
ML	bias	0,3242	-0,482	0,0808	0,1032	0,2733	-0,199	-0,0328	0,4827
	sd	0,1878	0,1791	0,1175	0,0816	0,1984	0,2162	0,1609	0,2322
	mean	0,9949	1,3276	0,5356	0,142	1,7244	1,1001	0,2515	0,8103
Bayes	bias	-0,0099	0,6724	0,1444	0,058	-0,1744	-1E-04	-0,0015	0,0397
	sd	0,0449	0,3905	0,0767	0,0272	0,16	0,003	0,0231	0,0593

The results of GMM, Bayes and ML for β, ζ, λ_y *and* Ω are similar to those are found the sample size of 60. Estimated $\emptyset$ is biased in all three of them, but the least in ML. For λ_r, λ_π and h Bayes is the best but this time GMM is the second good and ML is the worst. Again, GMM lines are more flat. It can be noticed that compared to 60 sample size, results of 180 sample size have lower standard deviation.

4.2 Shocks are not Autocorrelated

In this case, all the estimation processes (data generation, model declaration and estimation) are done by using the model having shocks that are not autocorrelated.

i. Small Sample

The results are documented the table below.

Table 3 Shocks Are Not Aotucorrelated And Sample Is Small

		B	$\emptyset$	ζ	λ_r	λ_π	λ_y	Ω	h
	true	**0,985**	**2**	**0,68**	**0,2**	**1,55**	**1,1**	**0,25**	**0,85**
GMM	mean	0,9982	1,7323	0,9487	0,5252	1,5933	1,3536	0,0796	0,4998
	bias	-0,0132	0,2677	-0,2687	-0,3252	-0,0433	-0,2536	0,1704	0,3502
	sd	0,0072	0,232	0,0583	0,2397	2,16	3,4119	0,129	0,5406
ML	mean	0,8244	2,4618	0,6369	0,2096	1,6069	1,1556	0,1154	0,4323
	bias	0,1606	-0,4618	0,0431	-0,0096	-0,0569	-0,0556	0,1346	0,4177
	sd	0,0692	0,095	0,0285	0,0225	0,0792	0,0801	0,0465	0,1266
Bayes	mean	1,003	1,8982	0,5182	0,4119	1,5547	1,1156	-0,0678	0,8491
	bias	-0,018	0,1018	0,1618	-0,2119	-0,0047	-0,0156	0,3178	0,0009
	sd	0,0463	0,1532	0,0912	0,171	0,0106	0,0552	0,3153	0,0016

β is estimated nearly equal to its true value by GMM and Bayes, the result of ML for β is biased nearly 0,16. For $\emptyset$, the most biased one is ML and the least is Bayes. The closest mean to true value of ζ belongs to ML, while GMM and Bayes are more biased. ML also estimates λ_r with very little bias and GMM is the worst. For λ_π, λ_y and h, Bayes estimates almost equal to actual values, nevertheless, GMM seems much more biased, and the means of ML estimations are less biased, compared to GMM. For Ω, this time ML is the best but Bayes is the worst one.

ii. Large Sample

The results are documented the table below.

Table 4 Shocks Are Not Aotucorrelated And Sample Is Large

		β	$\emptyset$	ζ	λ_r	λ_π	λ_y	Ω	h
	true	**0,985**	**2**	**0,68**	**0,2**	**1,55**	**1,1**	**0,25**	**0,85**
	mean	0,9993	1,4214	0,9695	0,6818	1,6222	1,353	0,1119	0,2228
GMM	bias	-0,0143	0,5786	-0,2895	-0,4818	-0,0722	-0,253	0,1381	0,6272
	sd	0,0024	0,1672	0,033	0,2027	2,6398	3,1718	0,1417	0,1961
	mean	0,7561	2,3632	0,4858	0,0997	1,2009	1,3332	0,2016	0,268
ML	bias	0,2289	-0,3632	0,1942	0,1003	0,3491	-0,2332	0,0484	0,582
	sd	0,1161	0,2206	0,0657	0,0735	0,3153	0,2176	0,1103	0,1866
	mean	0,9721	1,9353	0,437	0,1525	1,6249	1,1189	0,1026	0,8265
Bayes	bias	0,0129	0,0647	0,243	0,0475	-0,0749	-0,0189	0,1474	0,0235
	sd	0,0259	0,0302	0,0659	0,0168	0,0255	0,0143	0,029	0,0178

Increasing sample size in non-autocorrelated model doesn't change relative comparison among the three estimators for parameters β, $\emptyset$, ζ, λ_y , Ω, h. However, this time λ_r is estimated best by Bayes and λ_π is estimated best by GMM

CHAPTER FIVE

CONCLUSION

This study compares Bayesian, General Methods of Moments and Maximum Likelihood techniques that are used to estimate structural parameters of Dynamic Stochastic General Equilibrium, in different cases. The model we use includes the Euler condition, Philips curve and monetary policy equations. The data is generated artificially at the length of 60 and 180.

In our model and data, GMM estimates more accurately in small samples rather than large samples. The existence of autocorrelation does not change this. It is also obvious that GMM performs better if autocorrelation does not exist, both in small sample and large sample. The length of data does not significantly affect Bayesian estimation, regardless of existence of autocorrelation in error terms. Autocorrelation has negative effects on Bayes in small sample; however, in large sample four parameters are estimated better and four are estimated worse. ML performs better in small samples both with autocorrelation and without autocorrelation. Autocorrelation affects negatively ML in small sample rather than large sample size case.

In conclusion, when we include autocorrelated error terms to the model, the best estimator is Bayesian. Similarly, when we conduct the estimations with short data, again Bayesian is the least biased. However, it should be noted, if Dynare had allowed us to expand the border of priors for Bayesian estimator, the result might be different.

References

Adolfson, M., Laséen, S., Lindé, J., and Svensson, L. E. 2011. "Optimal Monetary Policy in an Operational Medium-Sized DSGE Model," *Journal of Money, Credit and Banking 43(7),* 1287-1331.

Altug, S. 1989. "Time-to-Build and Aggregate Fluctuations: Some New Evidence." *International Economic Review 30*, 889–920

An, S., and Schorfheide, F. 2007. "Bayesian Analysis of DSGE Models,*"Econometric reviews" 26(2-4)*, 113-172.

Bencivenga, V.R., 1992. "An Econometric Study of Hours and Output Variation with Preference Shocks," *International Economic Review 33*, 449–471.

Boivin, J., and Giannoni, M. 2006. "DSGE Models in a Data-Rich Environment," *National Bureau of Economic Research,* (No. w12772)

Canova, F., and Sala, L. 2009. "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics, 56(4)*, 431-449.

Christiano, L.J. 1988. "Why Does Inventory Investment Fluctuate So Much?" *Journal of Monetary Economics*, *21*, 247–280

Christiano, L. J., and Eichenbaum, M. 1992. "Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations," *The American Economic Review*, 430-450.

Christopher F Baum, Mark E. Schaffer and Steven Stillman. 2002. "Instrumental Variables and GMM: Estimation and Testing," *Boston College Working Papers in Economics 545*, revised 14 Feb 2003.

Dridi, R., Guay, A., Renault, E. 2007 "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models," *Journal of Econometrics,136(2)*, 397-430.

Fernández-Villaverde, J., and Rubio-Ramirez, J. 2007. "How Structural are Structural parameters?".

Galí, Jordi and Mark Gertler. 1999. "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics, 44*, 195-222

Giesen, S., and Scheufele, R. 2013. "Effects of Incorrect Specification on the Finite Sample Properties of Full and Limited Information Estimators in DSGE Models," *Halle Institute for Economic Research*, No. 8.

Hansen, L. P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators" *Econometrica: Journal of the Econometric Society*, 1029-1054.

Juillard, Michel, 1996. "Dynare : A Program for the Resolution and Simulation of Dynamic Models with Forward Variables Through the Use of a Relaxation Algorithm," *CEPREMAP Working Papers (Couverture Orange) 9602*, CEPREMAP.

Lubik, Thomas A., and Frank Schorfheide. 2007. "Do Central Banks Respond to Exchange Rate Movements? A Structural Investigation," *Journal of Monetary Economics 54.4*, 1069-1087.

McGrattan, Ellen R. 1994. "The Macroeconomic Effects of Distortionary Taxation," *Journal of Monetary Economics 33.3*, 573-601.

Rabanal, P. and Rubio-Ramírez, J. F. 2005, "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach", *Journal of Monetary Economics" 52(6),* 1151-1166.

Ruge-Murcia, F. J. 2007. "Methods to Estimate Dynamic Stochastic General Equilibrium Models," *Journal of Economic Dynamics and Control, 31(8)*, 2599-2636.
Schorfheide, F. 2000. "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics, 15(6)*, 645-670.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticit" *Econometrica 48*, 817-838

APPENDIX

A-Kalman Filter

Kalman filter is an algorithm that uses a series of measurements observed over time, containing noise (random variations) and other inaccuracies, and produces estimates of unknown variables that tend to be more precise than those based on a single measurement alone. So it is a statistically optimal estimate of the underlying system state. The Kalman filter model assumes that:

The true state at time k is evolved from the state at (k−1) according to

$$x_k = F_k x_{k-1} + B_k u_{k-1} + w_k$$

where $w_k \sim N(0, Q_k)$

At time k an observation (or measurement) zk of the true state xk is made according to $z_k = H_k x_k + v_k$ and v_k is the observation noise which is assumed to be $v_k \sim N(0, R_k)$.

Notice that the filter assumes that parameters such as F, B and H, i.e. the model, are known. Practical implementation of the Kalman Filter is often difficult due to the inability in getting a good estimate of the noise covariance matrices Qk and Rk. There are some methods to do that. It depends on the code you use. The Kalman filter can be written as a single equation; however it is most often conceptualized as two distinct phases: "Predict" and "Update". The predict phase uses the state estimate from the previous time step to produce an estimate of the state at the current time step. This predicted state estimate is also known as the a priori state estimate. Then use this state estimate to predict current observation, that is z_k. If current observation is different than your estimate, use this difference (known as Kalman Gain) to refine your current state estimate. As a result, current state estimate is a combination of your prior estimate and the gain you obtain from the current observation. So it is a kind of error correction algorithm.

Predict

Predicted (a priori) state estimate $\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + B_k u_{k-1}$

Predicted (a priori) estimate covariance (of X) $P_{k|k-1} = F_k P_{k-1|k-1} + F_k^T + Q_k$

Update

Innovation or measurement residual $\tilde{y}_k = z_k - H_k \hat{x}_{k|k-1}$

Innovation (or residual) covariance $S_k = H_k P_{k|k-1} H_k^T + R_k$

Optimal Kalman gain $K_k = P_{k|k-1} H_k^T + S_k^{-1}$

Updated (a posteriori) state estimate $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k$

Updated (a posteriori) estimate covariance $P_{k|k} = (I - K_k H_k) P_{k|k-1}$

For instance if R is big (the variance of the observation data), Kalman Gain is small. That is your update of priori estimate of the state variable by using the observation is small. The Kalman filter is a recursive estimator. This means that only the estimated state from the previous time step and the current measurement are needed to compute the estimate for the current state.

B- Autocorrelation In Error Terms Is 0.9

In this case, we examine the situation in which error terms have 0.9 autocorrelations. The results are represented below in table 1 and table 2. Comparing to 0.65 autocorrelation, some parameters are estimated better and some are worse. However, it is very obvious that for almost every parameter, now the result is worse than the case in which there is no autocorrelation. Almost all of the estimated parameters are more biased. This is true for all estimators and for both small and large samples. In small samples, estimated parameters are much more biased. Like 0.65 autocorrelation case, in here GMM is affected more severely by 0.9 autocorrelation than likelihood based approaches.

For sample size of 60, β is estimated nearly equal to its true value by GMM and Bayes, the mean of β in ML is biased nearly 0.36 which is very high. For $\emptyset$, GMM is the most biased one, and Bayes is the least. ML has the closest mean to true value of ζ, while GMM and Bayes are more biased. Bayes estimates. For $\boldsymbol{\lambda_r}$, $\boldsymbol{\lambda_\pi}$, $\boldsymbol{\lambda_y}$, Ω and h, with the least bias and GMM is the worst.

In the case of large sample, relative comparison of every parameter is the same with those of sample size of 60. However, it is obvious that, almost all parameters are less biased.

Table 5 Shocks Are Aotucorrelated (0.95) And Sample Is Small

		β	$\emptyset$	ζ	λ_r	λ_π	λ_y	Ω	h
	true	**0,985**	**2**	**0,68**	**0,2**	**1,55**	**1,1**	**0,25**	**0,85**
	mean	1,0143	1,2588	1,0013	0,6801	3,4752	2,3567	-0,8534	0,5477
GMM	bias	-0,0293	0,7412	-0,3213	-0,4801	-1,9252	-1,2567	1,1034	0,3023
	sd	0,0035	0,256	0,0472	0,1724	2,1234	1,524	0,423	0,235
	mean	0,6209	2,5952	0,5788	0,4144	1,3129	1,2149	0,1246	0,5786
ML	bias	0,3641	-0,5952	0,1012	-0,2144	0,2371	-0,1149	0,1254	0,2714
	sd	0,1369	0,1705	0,0908	0,06	0,1953	0,1285	0,1206	0,1907
	mean	0,9616	1,7277	0,3893	0,3156	1,6492	1,1014	0,2694	0,7419
Bayes	bias	0,0234	0,2723	0,2907	-0,1156	-0,0992	-0,0014	-0,0194	0,1081
	sd	0,2715	0,3125	0,2318	0,0154	0,5016	0,284	0,1424	0,2225

Table 6 Shocks Are Aotucorrelated (0.95) And Sample Is Large

		β	$\emptyset$	ζ	λ_r	λ_π	λ_y	Ω	H
	true	**0,985**	**2**	**0,68**	**0,2**	**1,55**	**1,1**	**0,25**	**0,85**
	mean	1,0051	1,5468	1,0258	0,4013	2,9046	1,3974	-0,1734	0,5785
GMM	bias	-0,0201	0,4532	-0,3458	-0,2013	-1,3546	-0,2974	0,4234	0,2715
	sd	0,0021	0,513	0,144	0,0981	1,981	2,1343	0,2564	0,1823
	mean	0,6043	2,5316	0,5888	0,4012	1,4487	1,2365	0,1655	0,6455
ML	bias	0,3807	-0,5316	0,0912	-0,2012	0,1013	-0,1365	0,0845	0,2045
	sd	0,2251	0,0934	0,1209	0,0725	0,2315	0,1824	0,1296	0,2353
	mean	0,9661	1,8177	0,3413	0,0971	1,6364	1,1145	0,2539	0,6624
Bayes	bias	0,0189	0,1823	0,3387	0,1029	-0,0864	-0,0145	-0,0039	0,1876
	sd	0,092	0,1425	0,0124	0,0214	0,2385	0,0291	0,034	0,4125