Identifiability issue in macroeconomic modelling

Ondřej Čížek[1]

Abstract. The effort of central banks to apply DSGE models in practice has increased the complexity of these models. Recently however, it has been pointed out that even medium scale DSGE models chronically suffer from the identification problem, which is a topic that has been traditionally mentioned in macroeconomic modeling only in connection with the simultaneous equations models. The problem of identifiability has, however, far more general nature and apart from other consequences complicates practical econometric estimation of the parameters. The aim of this paper is to investigate the identifiability of the formulated small macroeconomic model by means of which I will show that the problem with identifiability can arise not only as a consequence of the model being too large as is frequently argued nowadays, but as a result of the model being too small. The formulated model falls within the class of linear models with quadratic loss function and its parameters are estimated by the method of maximum likelihood. The approach to analyze identifiability issue is based on the Fisher information matrix.

Keywords: identification, maximum likelihood, Fisher matrix, Kalman filter, DSGE, inflation targeting, macroeconomics.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The goal of this paper is to present key results from the theory of identification and to illustrate how to apply them to the formulated macroeconomic model, which falls within the class of linear models with quadratic loss function. To analyze identifiability issue means to answer the question whether or not different parameter values generate different probabilistic distribution of the observed variables. The chosen approach answers this question without a reference to a particular data set because data generating process (and not a particular data set) is used in the analysis. The analysis of identifiability, however, serves largely for the econometric estimation of the parameters, which is also a topic of this paper. For this reason, data used to estimate parameters by maximum likelihood method are described in chapter 3 immediately after the description of the model equations in chapter 2. Section 4 puts the model in convenient state space form. Chapter 5 then describes the Kalman filter and the next section shows how to use it to compute likelihood function. The outcome from the Kalman filter is also used to compute the Fisher information matrix as the main tool to detect problems with identifiability, which is a topic of the Section 7. Final chapter 8 discusses the results from the analysis of identifiability when applied to the formulated model.

2 Model

The presented model is a modification of the Ball's [1] macroeconomic model very well known in the literature on inflation targeting. Because the model is aimed at analyzing stabilization policy, all the variables represent cyclical component of the original value. There are two main variables in the model representing real economic activity and prices, which are the rate of unemployment (u) and the inflation (π). The first one is modeled by its own lagged value and by the interest rate (r), which is assumed to be under full control of the central bank:

$$u_{t+1} = a \cdot u_t + b \cdot r_{t+1} + \eta_{t+1}, \tag{1}$$

whereas $a \in (0,1)$, $b > 0$ are parameters and η_{t+1} is random shock, which is assumed to be normally distributed with zero mean, constant variance $\operatorname{var}(\eta_t) = \sigma_e^2$ and is also assumed to be independent at time as well as with other random shocks in the model.

The equation (1) is in line with Ball's formulation. I modified, however, the relation between real economic activity and inflation represented by the Phillips curve in order to fit the data for the Czech economy. Ball assumed so called accelerationist Phillips curve in the form that $\pi_{t+1} - \pi_t$ is a function of u_t. The name of this

[1] University of Economics, Department of Econometrics, nám. W. Churchilla 4, Prague 3, cizeko@vse.cz.

type of Phillips curve comes from the fact that inflation raises if unemployment is below its "natural" level. My formulation is based not on this historical experience from the United States, but on the empirical data from the Czech Republic, which suggested negative relation between π_{t+1} and $u_t - u_{t-1}$. This relation can be interpreted in such a way that inflation gets above its "normal" level whenever (monetary) authority is trying to decrease the rate of unemployment. The equation for the inflation is as follows:

$$\pi_{t+1} = -c \cdot (u_t - u_{t-1}) + \varepsilon_{t+1} \tag{2}$$

where $c > 0$ is parameter and ε_{t+1} is random shock fulfilling same assumptions as η_{t+1}.

Ball completes the specification of the model by assuming that interest rate is set according to the Taylor rule[2] and then is trying to find such values of the coefficients of this rule that minimize $\text{var}(\pi_t) + \kappa \cdot \text{var}(u_t)$, where $\kappa > 0$ is the weight that is set on stabilization of the unemployment. In this article, I will use slightly more general approach and will assume the monetary authority to minimize quadratic loss function of the form:

$$\underset{0}{E}\left[\sum_{t=0}^{T} \delta^t \cdot \left(\kappa \cdot u_t^2 + \pi_t^2\right)\right] \to MIN \tag{3}$$

which is equivalent to the Ball's approach if the discount factor of the monetary authority $\delta \in (0,1]$ is set to equal one and the horizon for optimization T goes to infinity.

3 Data

The source of data for inflation and rate of unemployment is the Czech Statistical Office. The inflation was calculated from consumer price index (CPI) on a quarterly basis according to $inflat_{t+1} = 100 \cdot (CPI_{t+1} - CPI_t) / CPI_t$. Although CPI is available from 1997 to 2011, I decided to work with data from 1999 because during years 1997 and 1998 the inflation was unstable[3]. There was a significant seasonal pattern in the calculated inflation, which I removed by means of linear regression with dummy variables Q_{it}. To model the increase in DPH at the beginning of the year 1998 from five to nine percentage points and at the beginning of the year 2000 from nine to ten percentage points, I also included another dummy variable DPH_t with value equal to four in the first quarter of the year 1998, one in the first quarter of the year 2000 and zero in all other quarters. All the dummies were statistically significant except the dummy for the third quarter. The cyclical component of the inflation was therefore calculated as residuals from the following regression:

$$inflat_t = \beta_1 \cdot Q_{1t} + \beta_2 \cdot Q_{2t} + \beta_4 \cdot Q_{4t} + \beta_5 \cdot DPH_t + error_t . \tag{4}$$

Data for the rate of unemployment was already seasonally adjusted by the Czech Statistical Office. I used the data from the selective survey of the labor force. The cyclical component of the rate of unemployment was calculated simply as a deviation from the mean value.

4 State space form

Although there are no unobserved variables in the model, it will be useful for the purpose of generalization to write the model in a state space form. If I define the state vector $\mathbf{x_t} = (\pi_t \quad u_t \quad u_{t-1})'$ and the vector of random errors denote as $\mathbf{u_{t+1}} = (\varepsilon_{t+1} \quad \eta_{t+1})'$, then the transition equation can be written in a following manner:

$$\mathbf{x_{t+1}} = \mathbf{A} \cdot \mathbf{x_t} + \mathbf{b} \cdot r_{t+1} + \mathbf{C} \cdot \mathbf{u_{t+1}} , \tag{5}$$

[2] Taylor rule specifies interest rate is a linear function of inflation and unemployment.
[3] The instability was caused by huge disturbances in the foreign exchange market in May 1997, which led to the change in monetary strategy of the central bank to inflation targeting at the beginning of the year 1998.

where $\mathbf{A}=\begin{pmatrix}0 & -c & c\\ 0 & a & 0\\ 0 & 1 & 0\end{pmatrix}$, $\mathbf{b}=\begin{pmatrix}0\\ b\\ 0\end{pmatrix}$, $\mathbf{C}=\begin{pmatrix}1 & 0\\ 0 & 1\\ 0 & 0\end{pmatrix}$.

If I denote the vector of observed variables as $\mathbf{z_t}=\left(\pi_t \quad u_t\right)'$, the measurement equation is as follows:

$$\mathbf{z_t}=\mathbf{D}\cdot\mathbf{x_t}, \tag{6}$$

where $\mathbf{D}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}$.

The quadratic loss function can be written in the following way:

$$E_0\left(\sum_{t=0}^{\infty}\mathbf{x_t}'\cdot\mathbf{K_t}\cdot\mathbf{x_t}\right)\rightarrow MIN, \tag{7}$$

where $\mathbf{K_t}=\delta^t\cdot diag\left(1 \quad \kappa \quad 0\right)$ is the weighting matrix.

5 Optimising control variables and estimating unobserved state

In this chapter I will suppose that the values of the parameters are known and show the results telling us how to solve the problem of minimising the objective function (7) as well as the problem of estimation of the unobserved state vector by the Kalman filter. In the next chapter, it will be shown how to use these result in econometric estimation of the parameters.

The aim of the optimal control problem is to find trajectory of control variables r_t, $t=0,...,T$ that minimize the loss function (7). Derivation of the solution to this problem can be found for example in Chow [2], chapters 8.3 and 8.4. The solution is as follows:

$$r_t=\mathbf{G_t}\cdot\mathbf{x_{t-1|t-1}}, \tag{8}$$

where the matrix $\mathbf{G_t}$ is calculated recursively backwards in time according to the scheme:

$$\mathbf{G_t}=-\left(\mathbf{b}'\mathbf{H_t}\mathbf{b}\right)^{-1}\mathbf{b}'\mathbf{H_t}\mathbf{A}, \tag{9}$$

$$\mathbf{H_{t-1}}=\mathbf{K_{t-1}}+\left(\mathbf{A}+\mathbf{bG_t}\right)'\mathbf{H_t}\left(\mathbf{A}+\mathbf{bG_t}\right), \tag{10}$$

with the initial value $\mathbf{H_T}=\mathbf{K_T}$. The symbol $\mathbf{x_{t-1|t-1}}$ denotes optimal estimation of the unobserved state vector, which is given by the Kalman filter recursions of the form:[4]

$$\mathbf{x_{t|t-1}}=\left(\mathbf{A}+\mathbf{bG_t}\right)\cdot\mathbf{x_{t-1|t-1}}, \tag{11}$$

$$\mathbf{x_{t|t}}=\mathbf{x_{t|t-1}}+\mathbf{P_{t|t-1}}\mathbf{D}'\left(\mathbf{DP_{t|t-1}D}'\right)^{-1}\left(\mathbf{z_t}-\mathbf{Dx_{t|t-1}}\right), \tag{12}$$

where $\mathbf{x_{t|t-1}}\equiv E_{t-1}\left(\mathbf{x_t}\right)$ and matrices $\mathbf{P_{t|t-1}}\equiv E_{t-1}\left(\mathbf{x_t}-\mathbf{x_{t|t-1}}\right)\left(\mathbf{x_t}-\mathbf{x_{t|t-1}}\right)'$ are calculated recursively as follows:

$$\mathbf{P_{t+1|t}}=\mathbf{A}\left[\mathbf{P_{t|t-1}}-\mathbf{P_{t|t-1}}\mathbf{D}'\left(\mathbf{DP_{t|t-1}D}'\right)^{-1}\mathbf{DP_{t|t-1}}\right]\mathbf{A}'+\mathbf{C}\Sigma_{\mathbf{uu}}\mathbf{C}', \tag{13}$$

[4] Because no prior information about initial value of the state vector was available, the Kalman filter recursion was initialized by so called "diffuse prior", that is by setting $\mathbf{x_{0|-1}}=\mathbf{0}$, $\mathbf{P_{0|-1}}=\alpha\cdot\mathbf{I}$, where $\mathbf{I}$ is identity matrix and α is a large number.

where $\Sigma_{\mathbf{uu}} = E(\mathbf{u}\cdot\mathbf{u}')$ is the covariance matrix of the random errors.

6 Econometric estimation

I estimated parameters of the model by the Maximum Likelihood Method (MLE). The probability density function of the observed variables $\mathbf{z}_t$, $t = 1,\dots,n$ can be parameterized by the vector $\boldsymbol{\theta} = (\delta,\ a,\ b,\ c,\ \kappa,\ \sigma_\varepsilon,\ \sigma_\eta)$ and therefore comes from a "family" $\{f(\mathbf{Z}_n \mid \boldsymbol{\theta}), \boldsymbol{\theta}\in\boldsymbol{\Theta}\}$, where $\mathbf{Z}_n = (\mathbf{z}_1^{'},\dots,\mathbf{z}_n^{'})^{'}$ and the set $\boldsymbol{\Theta} = \{x \in R^6 : x_i \in (0,1),\ i = 1,2,\ \textit{and}\ x_j > 0,\ j = 3,\dots,6\}$ is called as a parameter space. The goal is to find $\boldsymbol{\theta}_0$ in the parameter space $\boldsymbol{\Theta}$ that maximizes the likelihood function L defined as $L(\boldsymbol{\theta}) \equiv f(\mathbf{Z}_n \mid \boldsymbol{\theta})$.

How to use results from the Kalman filter in computation of the likelihood function is described for example in Harvey [3], chapter 3.4. The basic idea is that probability density function can be expressed by means of conditional probability densities, which can be easily proved by induction:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{n} f(\mathbf{z}_t \mid \mathbf{Z}_{t-1}, \boldsymbol{\theta}). \tag{14}$$

Because random errors are assumed to be normally distributed, the distribution of $\mathbf{z}_t$ conditioned on $\mathbf{Z}_{t-1}$ is commonly known to be also normal and is therefore fully described by its mean vector and covariance matrix. Because of the relation $\mathbf{z}_t = \mathbf{D}\mathbf{x}_t$, these characteristics depend on the conditional mean $\mathbf{x}_{t|t-1}$ and covariance $\mathbf{P}_{t|t-1}$ of the state vector $\mathbf{x}_t$, which was already computed by the Kalman filter recursion. The characteristics of conditional distribution of $\mathbf{z}_t$ are therefore given by the following equations:

$$\mathbf{z}_{t|t-1} = \mathbf{D}\mathbf{x}_{t|t-1}, \tag{15}$$

$$\mathbf{F}_{t|t-1} \equiv E\left[(\mathbf{z}_t - \mathbf{z}_{t|t-1})(\mathbf{z}_t - \mathbf{z}_{t|t-1})^{'}\right] = \mathbf{D}\mathbf{P}_{t|t-1}\mathbf{D}^{'}. \tag{16}$$

The likelihood function can be therefore calculated as follows:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{n}\left\{\frac{1}{(2\pi)^{\frac{k}{2}}\left|\mathbf{F}_{t|t-1}\right|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\mathbf{z}_t - \mathbf{z}_{t|t-1})^{'}\mathbf{F}_{t|t-1}^{-1}(\mathbf{z}_t - \mathbf{z}_{t|t-1})\right]\right\}, \tag{17}$$

where k is the number of observable variables.

I maximised the logarithmic form of the likelihood function (17) in Matlab with standard numerical algorithms that are included in fminsearch. Maximization of this likelihood function was, however, complicated by the fact that the function is nearly flat in some directions, which brings us to the concept of identifiability of the parameters.

7 Identifiability

The basic prerequisite for making inference about the value of the vector $\boldsymbol{\theta}$ is that different values of $\boldsymbol{\theta}$ implies different processes according to which data are generated, which Rothenberg [5] formalizes as:

$$\left(f(\mathbf{Z}_n \mid \boldsymbol{\theta}) = f(\mathbf{Z}_n \mid \boldsymbol{\theta}_0)\ \ \textit{with probability}\ 1\right) \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0. \tag{18}$$

If (18) holds for every $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, then $\boldsymbol{\theta}_0$ is called to be globally identifiable. If it holds for a nearby neighbourhood of $\boldsymbol{\theta}_0$, then $\boldsymbol{\theta}_0$ is called to be locally identifiable. The model is (locally) identifiable, if every $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$

is (locally) identifiable. If $\boldsymbol{\theta}_0$ is not identifiable, then there is a structure $\boldsymbol{\theta}$ different from $\boldsymbol{\theta}_0$, which cannot be distinguished from $\boldsymbol{\theta}_0$ by observation of the data. Such structures are called observationally equivalent.

From these definitions, one can see that the problem of identifiability arises because of similarity of the probability density functions. It is, however, impossible to deal with the question of identifiability right from these definitions. The userful tool for analysing local identifiability is the following theorem, the proof of which can be found in Rothenberg [5]:

Theorem 1. *Let $\boldsymbol{\theta}_0$ be a regular point of the information matrix $\mathbf{R}_n(\boldsymbol{\theta})$.Then $\boldsymbol{\theta}_0$ is locally identifiable if and only if $\mathbf{R}_n(\boldsymbol{\theta}_0)$ is non-singular.*

A point is called regular if it belongs to an open neighborhood where the rank of the matrix does not change. Without this assumption the condition is only sufficient for local identification. The singularity of the information matrix means that the likelihood function is flat at $\boldsymbol{\theta}_0$, which can happen either because some parameters do not affect likelihood at all or different parameters have the same effect on the likelihood function. Iskrev [4] formalized both these situations as follows:

1. changing θ_i does not change the likelihood, so the i-th diagonal element of the information matrix
   $\Delta_i \equiv E\left(\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i}\right)^2$ is equal to zero,
2. effect of changing θ_i is offset by changing other parameters at $\boldsymbol{\theta}$, so the multiple correlation coefficient
   $\rho_i \equiv corr\left(\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i};\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{-i}}\right)$ is equal to one, where $\boldsymbol{\theta}_{-i} \equiv (\theta_1,...,\theta_{i-1},\theta_{i+1},...,\theta_p)$, p being the number of parameters.

Iskrev argues that parameters are unidentified if economic features they represent are either unimportant on their own ($\Delta_i = 0$) or they are redundant given the other features in the model ($\rho_i = 1$). The case of $\rho_i = 1$ then consider as a sign of "overparameterization" of the model, which is particularly relevant for large DSGE models that are often criticized for being too rich in features. From this point of view, theory of identification has important economic modelling aspect and most common suggestion for reducing identifiability problem is to reduce the model. Identification problem can also arise because of data deficiency (not enough or inadequate observable variables, not enough observations), or because of statistical methodology, which is, however, not our case as the method of maximum likelihood uses all the available information about the data generating process, which is described by the probability density function.

Iskrev points out that in most macroeconomic models the information matrix is albeit regular however nearly singular, which is a case that he called weak local identifiability and the reasons for this to happen can be formally written as $\Delta_i \approx 0$ or $\rho_i \approx 1$. By this reasoning he adopted a particular measure of identification

$$s_i^2 \equiv \Delta_i \cdot \left(1-\rho_i^2\right) \tag{19}$$

and showed that $\left(1/s_i^2\right)$ is equal to the i-th diagonal element of $\left(\mathbf{R}_n(\boldsymbol{\theta})\right)^{-1}$ provided that $\mathbf{R}_n(\boldsymbol{\theta})$ is regular.

How to compute information matrix derives for example Harvey [3] in chapter 3.4. The element in the i-th row and j-th column is computed according to:

$$\left[\mathbf{R}_n(\boldsymbol{\theta})\right]_{ij} = \frac{1}{2}\sum_{t=1}^{T} tr\left(\mathbf{F}_{t|t-1}^{-1}\frac{\partial \mathbf{F}_{t|t-1}}{\partial \theta_i}\mathbf{F}_{t|t-1}^{-1}\frac{\partial \mathbf{F}_{t|t-1}}{\partial \theta_j}\right) + E\left(\sum_{t=1}^{T}\frac{\partial\left(\mathbf{z}_t - \mathbf{z}_{t|t-1}\right)'}{\partial \theta_i}\mathbf{F}_{t|t-1}^{-1}\frac{\partial\left(\mathbf{z}_t - \mathbf{z}_{t|t-1}\right)}{\partial \theta_j}\right). \tag{20}$$

where $\partial \mathbf{F}_{t|t-1}/\partial\theta_i$ and $\partial\left(\mathbf{z}_t - \mathbf{z}_{t|t-1}\right)/\partial\theta_j$ were computed numerically.

The mean value at the right side of this equation is unfortunately impossible to compute except very special cases and so the usual approach is to drop the expectation operator in the second term.

8 Results

The information matrix was nearly singular at all randomly chosen points in the parameter space and the third diagonal element of $\left(\mathbf{R}_n(\boldsymbol{\theta})\right)^{-1}$ corresponding to the parameter b in the model was incomparably higher than other diagonal elements, which indicates very weak identification of this parameter. At first glance, this result seems rather strange as it says that the sensitivity of unemployment to the interest rate is not important in the process of generating data. This can be explained by the fact that the interest rate is not a part of the data set because in the presented model it is seen as a control variable of the central bank and not as an observable indicator of a state variable. Lower sensitivity of unemployment to interest rate is compensated by a more aggressive strategy for setting interest rate by the central bank, which produces practically the same data for unemployment and inflation because of which it is nearly impossible to identify the parameter b in the model. So in this case the problem with identification of parameter b is caused by insufficient number of observable variables.

When treating parameter b as a known constant there were still two relatively large diagonal elements in $\left(\mathbf{R}_n(\boldsymbol{\theta})\right)^{-1}$ at all randomly chosen points in the parameter space and these elements corresponded to the parameters δ and κ in the loss function of the central bank. The problem with weak identification was solved only after treating any of these parameters as a known constant, which suggests that these two parameters play very similar role in the model. This result is rather surprising as economic interpretation of these parameters is quite different. This can be intuitively explained by the fact that in the presented model the inflation is controlled only through unemployment and therefore there is a very close relationship between these two variables, which causes parameters δ and κ to play similar role. Modeling inflation in a more complicated way would thus solve the problem with weak identification of the parameters from another part of the model (from the los function in our case), which can be regarded as analogy to so called identification paradox known from simultaneous equations models. My contribution of this analysis is that weak identification of the parameters (δ and κ in our case) is not necessarily a consequence of "overparameterization" of the model (i.e. that the model is too large) as is frequently argued nowadays in the context of large DSGE models, but a result of omitting important relationships in the model (i.e. that the model is too small). So, in my opinion, the frequently cited suggestion to reduce the model not only needn't reduce the problem with identifiability, but can make things even worse.

When fixing $b = 0.2$ and $\delta = 1$, the other five parameters were estimated by MLE as follows: $a = 0.73$, $c = 1.33$, $\kappa = 0.007$, $\sigma_\varepsilon = 0.57$, $\sigma_\eta = 0.34$. I also performed simulations with these parameter values and generated 1000 data sets each having 100 observations. Then I estimated all seven parameters by MLE 1000-times using standard numerical algorithms based on derivatives when maximizing likelihood function with initial values of the parameters equal to the true ones. The most surprising was that the standard error of the most weakly identified parameter b computed from these simulations was only 0.17 while for example computed standard error for a much better identified parameter a was 0.6. This result does not mean that the true maximum was ordinarily located near the true value of this parameter. In fact, the true maximum was not found at all. The explanation for the obtained result is that the parameter b is very weakly identified and so the likelihood is nearly flat in that direction, which is the reason why numerical algorithm practically did not proceed in this direction. Not to mention the fact that there had to be problems with numerical precision as numerical algorithms based on derivatives use ordinarily inverse of the hessian of the likelihood (or some approximation of it), which corresponds to the (nearly singular) information matrix. These results show that it is not only time-consuming, but also practically impossible to assess the reliability of the estimated parameters using simulation techniques.

Acknowledgements

Financial support of IGA F4/1/2012 is gratefully acknowledged by the author.

References

[1] Ball, L.: Efficient rules for monetary policy. *International Finance*, Wiley Blackwell, vol. 2(1), April, 1999, 63-83.

[2] Chow, G: *Analysis and control of dynamic economic systems.* John Wiley & Sons, New York, 1975.

[3] Harvey, A.C.: *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1989.

[4] Iskrev, N.: Evaluating the strength of identification in DSGE models. An a priori approach. W*orking paper w201032*, Banco de Portugal, Economics and Research Department, 2010.

[5] Rothenberg, T.: Identification in parametric models. *Econometrica*, 39, 1971, 577-591.