УДК 330.46

AN ECONOMETRIC APPROACH TO ROBUST IDENTIFICATION FOR MODELS OF INVERSE DYNAMIC PROBLEM

D.V. Filchenko
Sumy State University

A new computational approach to identification for models of inverse dynamic problem has been proposed. It is based on robust econometric difference and integral identification algorithms. Their approbation is made on real statistical data for n-industrial open macroeconomic system. All models and sub-models have been tested for adequacy and correspondence with reality.

INTRODUCTION

Dynamic mathematical modeling still remains one of the most urgent and difficult problems in different branches of science. It has a number of complications connected with parameters identification and adequacy of applied methodology. The fundamental achievements in dynamic system analysis were made within the theory of automatic control, the main mathematical instruments of which are the systems of differential and difference equations [1, 2].

There exist two fundamental problems that deal with modeling the evolution of specified dynamic systems: the law of motion identification with the control parameters given (direct dynamic problem, DDP) and control parameters identification with the law of motion given (inverse dynamic problem, IDP) [3]. It is well known that pure mathematical solving of IDP entails the stability problems [4]. Therefore it seems important to work out and approbate robust identification methods for models of IDP.

PROBLEM STATEMENT

Let us suppose that the dynamic element of a system may be described by the system of differential equations

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t), \qquad (1)$$

where $\mathbf{x}(t) = (x_1(t), x_2(t), ..., x_k(t))^/ \in E^k$, $t_0 \leq t \leq T$ is a vector of state variables (continuous functions), $\mathbf{u}(t) = (u_1(t), u_2(t), ..., u_l(t))^/ \in E^l$, $t_0 \leq t \leq T$ is a vector of control variables (piecewise continuous functions), $\mathbf{f}(\cdots) = (f_1(\cdots), f_2(\cdots), ..., f_k(\cdots))^/ \in C^1[t_0, T]$ is s vector of continuously differentiable functions, which specification depends on physical interpretation of the model (1).

We will divide the continuous period $[t_0, T]$ into N discrete moments of time. Suppose that at every point of time $\{t = 0, 1, ..., N\text{-}1\}$ there exists a statistical information $\mathbf{x}_t = (x_0, x_1, ..., x_{N-1})$ about all state variables from row vector $\mathbf{x}(t)$. The task is to find out control functions $\mathbf{u}(t)$, which generate such solution $\{x(t)\}$ of the system (1), that approximately, with a known precision, will satisfy the following congruencies:

$$\mathbf{x}(t) \cong \mathbf{x}_t, \{t = 0, 1, ..., N\text{-}1\}. \qquad (2)$$

Differently stated, we set up a problem of posterior control variables estimation in a dynamic system given the measurement of state variables at defined discrete points of time. This paper is organized as follows. Firstly

we will specify the unknown vector-valued function $\mathbf{f}(\cdots)$ taking macroeconomic system investment development as an example. Secondly, employing econometric tools, we will build up two algorithms of the model (1) identification and examine its simulation and forecast properties. And finally, we will test (1) for adequacy and approbate it using real statistical dataset.

MODEL SPECIFICATION

Consider an open macroeconomic system that consists of n industries. In most applications [5, 6] state variables of such a system are fixed capital stocks $\{x_1, x_2, ..., x_n\}$ per industry and foreign debt x_{n+1}. For the sake of simplicity, fixed capital stocks and foreign debt are treated as accumulative continuous functions. This allows using model (1) henceforth.

The rate of fixed capital change dx_i/dt is the value of net investments in the i-th industry for $\{i = 1, 2, ..., n\}$, while the rate of foreign debt change dx_{n+1}/dt is the value of its gross accumulation. In economic theory [7] these values are used for the analysis of investment activity, but mainly as ratios to gross output (e.g. GDP) rather than in absolute numbers. Therefore, specifying vector-valued function $\mathbf{f}(\cdots)$ as a multiplication of dimensionless values $\mathbf{u}(t)$ and GDP Y, i.e.

$$\mathbf{f}(\cdots) = \mathbf{u}(t)Y(\cdot), \tag{3}$$

we obtain a pure economic interpretation of control variables. Functions $u_1(t), u_2(t), ..., u_n(t)$ are indices of investment activity per industry and $u_{n+1}(t)$ is an index of export-import misbalance.

As regards exogenous function $Y(\cdot)$ specification, it merits special attention. Firstly, the model (1), (3) should be closed. Secondly, $Y(\cdot)$ should have a convenient and reasonable functional form. Basically, for the analysis of the first-order effects (production elasticity, marginal productivity, etc.) log-linear form will be sufficient, while the second-order effects (e.g. elasticity of substitution) are usually examined by trans-log forms [8].

In this article we will use two-factor log-linear functional form (Cobb-Douglas class) with the following factors specification:

$$\ln Y = a_0 + a_1 \ln \sum_{i=1}^{n} x_i + a_2 \ln x_{n+1}. \tag{4}$$

As a result the model made up of (1), (3), (4) is closed, and the sub-model (4) allows exploring 'regresand-regressors' system without direct correlation analysis.

It also convenient (for the further research) to use polynomial specification of control vector-valued function $\mathbf{u}(t)$:

$$u_i(t) = b_{i0} + b_{i1} t + b_{i2} t^2 + ... + b_{ik_i} t^{k_i}, \quad \{i = 1, 2, ..., n+1\}. \tag{5}$$

Generally speaking, the order k_i may be chosen with different considerations: to ensure high simulation or forecast properties, to increase the coefficient of determination R^2, etc. Nevertheless the substantial restriction in achieving all of abovementioned goals is that the higher the order k_i in (5), the lower the degree of freedom, and consequently the lower the quality of the model.

MODEL IDENTIFICATION

The question of parameters identification arises naturally (4), (5). Further we will use ordinary least squares (OLS) estimation. In order to apply it, the model made up of (1), (3), (4), and (5) should be transformed into a discrete form (from differential to difference equations). Remember that N is the number of discrete points within $[t_0, T]$. Hence we arrive at the following difference equations for $\{t = 0, 1, ..., N\text{-}2\}$ and $\{I = 1, 2, ..., n+1\}$:

$$x_i(t+1) = x_i(t) + b_{i0}\, Y\big|_t + b_{i1} t\, Y\big|_t + \ldots + b_{ik_i} t^{k_i}\, Y\big|_t + \varepsilon_{k_i}(t)\,, \tag{6}$$

where $\varepsilon_{k_i}(t)$ is a random disturbance, or if presented in a matrix form:

$$
\begin{pmatrix}
Y\big|_{t=0} & 0 & \ldots & 0 \\
Y\big|_{t=1} & Y\big|_{t=0} & \ldots & Y\big|_{t=0} \\
Y\big|_{t=2} & 2Y\big|_{t=2} & \ldots & 2^{k_i} Y\big|_{t=2} \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
Y\big|_{t=N-1} & (N-2)Y\big|_{t=N-1} & \ldots & (N-2)^{k_i} Y\big|_{t=N-1}
\end{pmatrix}
\begin{pmatrix}
b_{i0} \\ b_{i1} \\ b_{i2} \\ \ldots \\ b_{ik_i}
\end{pmatrix}
+
\begin{pmatrix}
\varepsilon_{i0} \\ \varepsilon_{i1} \\ \varepsilon_{i2} \\ \ldots \\ \varepsilon_{ik_i}
\end{pmatrix}
=
\begin{pmatrix}
x_i(1) - x_i(0) \\
x_i(2) - x_i(1) \\
x_i(3) - x_i(2) \\
\ldots\ldots\ldots \\
x_i(N-1) - x_i(N-2)
\end{pmatrix}.
$$

Model (6) is a regression that belongs to the class of regressions through the origin (RTO): its functional form does not contain a constant term that is required for the classical regression models [9].

Therefore it seems naturally to modify model (1) in order to avoid this subtle obstacle. Integrating formula (1) over period of time $[t_0, t]$ will result in

$$\mathbf{x}(t) = \mathbf{x}^*(t_0) + \int_{t_0}^{t} \mathbf{f}(\mathbf{x}(t),\ \mathbf{u}(t),\ t)dt\,.$$

Using (4), (5), we arrive at a discrete counterpart of the model (1):

$$x_i(t+1) = x_i^* + b_{i0}\sum_{j=0}^{t} Y\big|_j + b_{i1}\sum_{j=0}^{t} Y\big|_j + \ldots + b_{ik_i}\sum_{j=0}^{t} j^{k_i}\, Y\big|_j + v_{k_i}(t)\,, \tag{7}$$

where $v_{k_i}(t)$ is a random disturbance, or if presented in a matrix form:

$$
\begin{pmatrix}
1 & 0 & 0 & \ldots & 0 \\
1 & Y\big|_{t=0} & 0 & \ldots & 0 \\
1 & \sum\limits_{t=0}^{1} Y\big|_t & \sum\limits_{t=0}^{1} tY\big|_t & \ldots & \sum\limits_{t=0}^{1} t^{k_i} Y\big|_t \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
1 & \sum\limits_{t=0}^{N\text{-}2} Y\big|_t & \sum\limits_{t=0}^{N\text{-}2} tY\big|_t & \ldots & \sum\limits_{t=0}^{N\text{-}2} t^{k_i} Y\big|_t
\end{pmatrix}
\begin{pmatrix}
x_i^* \\ b_{i0} \\ b_{i1} \\ \ldots \\ b_{ik_i}
\end{pmatrix}
+
\begin{pmatrix}
v_i^* \\ v_{i0} \\ v_{i1} \\ \ldots \\ v_{ik_i}
\end{pmatrix}
=
\begin{pmatrix}
x_i(0) \\ x_i(1) \\ x_i(2) \\ \ldots\ldots \\ x_i(N-1)
\end{pmatrix}.
$$

In contrast to difference identification scheme (6), integral identification scheme (7) is a regression with an intercept. Despite that in econometric literature one can encounter some ambiguities concerning RTO, we will use both scheme (6) and (7) in order to compare them from the view-point of simulation and forecast properties.

We still have to identify the order k_i of (5). In this article they will be obtained under conditions of forecast confidence intervals minimization. For the forecast value of state variable $x_i(N)$ the latter is defined as follows:

$$x_i(N) - \delta \cdot t_\alpha < x_i(N) < x_i(N) + \delta \cdot t_\alpha , \qquad (8)$$

where $x_i(N)$ is a point forecast; t_α is a percentile of Student's distribution with α, level of significance; δ is a standard error of forecasting [10].

ESTIMATED RESULTS

Approbation of the models (6) and (7) is based on two-industrial Danish economy in **1966-1997**. All statistical information is available (see **[11, 12]**). Let the first industry consist of manufacturing and agricultural branches and the second one of services branches of economy.

According to (4) Denmark's GDP is approximated with two regressors: fixed capital stock $\sum_{i=1}^{n} x_i$ and foreign debt x_{n+1}. The obtained OLS-estimations are:

$$\ln Y = \underset{\text{(s.e.)}}{0.3859} + \underset{(0.0428)}{0.8145} \ln(x_1 + x_2) + \underset{(0.0252)}{0.1120} \ln x_3, \quad R^2 = 0.9968, \qquad (9)$$

where numbers in brackets are standard errors of regression coefficients. All coefficients, except the first one, appear to be statistically significant (we use Student's test for the number of freedom $l = 29$ and level of significance $\alpha = 5\%$). Such results are quite natural in econometric literature [8] when using log-linear functional forms for the purpose of regression analysis and together with high coefficient of determination R^2 imply that the model is quite fulfilled with specified factors.

Since elasticity of production for factor x is defined by the formula

$$E_x = \frac{\partial \ln Y}{\partial \ln x},$$

it follows from (9) that the elasticity of production for summarized fixed capital stock $x_1 + x_2$ is 0.8145, while the elasticity of production for foreign debt x_3 is 0.1120. Consequently, in the analyzed period of time the impact of foreign debt on Denmark's GDP was not essential (although it appears to be statistically significant). Another conclusion is that the diminishing return to scale was present ($a_1 + a_2 = 0.9265 < 1$) at that time.

The exponential form of equation (9) is

$$Y = 1.4710 \cdot (x_1 + x_2)^{0.8145} x_3^{0.1120} .$$

Let us now turn to the polynomial (5) identification. Below there are results of OLS estimation for difference identification scheme (6)

$$u_1(t) = 0.1469 - 0.0037\,t\,, \quad u_2(t) = 0.221964 - 0.006613\,t + 0.000042\,t^2 \quad (10)$$
$$\text{(s.e.)} \quad (0.0118) \quad (0.0005) \qquad \text{(s.e.)} \quad (0.0448) \quad (0.0043) \quad (0.0001)$$

and integral identification scheme **(7)**

$$u_1(t) = 0.1543 - 0.0041\,t\,, \quad u_2(t) = 0.2278 - 0.0062\,t + 0.000012\,t^2\,. \quad (11)$$
$$\text{(s.e.)} \quad (0.0033) \quad (0.0002) \qquad \text{(s.e.)} \quad (0.0167) \quad (0.0019) \quad (0.000051)$$

All coefficients, except the last two for $u_2(t)$ in (10) and the last for $u_2(t)$ in (11), are statistically significant, providing the high level of approximation for state variables x_1 and x_2. Particularly, coefficients of determination R^2 are 0.9986 and 0.9982 for scheme (6) and 0.9979 and 0.9966 for scheme (7).

As the experiment reveals, optimal order k_i for each of polynomials (5) appears to be between 0 and 3. Interval forecasts (8) for state variables x_1 and x_2 are 1197674.539 ± 20073.928 and 1800344.834 ± 36369.454 for scheme (6) versus 11883527.018 ± 34599.568 and 1764626.422 ± 56578.455 for scheme (7). Obviously, confidence intervals for the forecasts based on difference identification algorithm are narrower: 1.68 and 2.02 versus 2.92 and 3.21 (as percentages of point forecasts). Further approbation, as well as abovementioned one, confirms that difference identification scheme (6) is more preferable for forecast purposes, while integral identification scheme (7) ensures better simulation properties.

TESTING FOR ADEQUACY

The procedure of model verification comprises two stages: statistical analysis of precision and testing for adequacy. The first was demonstrated in the previous section and revealed excellent simulation and forecast precisions. The latter is based on examining model's residuals, specifically, on testing their correspondence with normal distribution and all assumptions of the classical regression [8].

Firstly, let us consider sub-model (9). Usually it is worth to start with testing for multicollinearity. It may happen that correlation between fixed capital stock and foreign debt is rather strong, especially when fixed capital accumulation stems from foreign investments, while the part of export in balance of payment deficit is negligible. Using correlation analysis of 'refressors' system for sub-model (9), one can conclude that according to Fisher-Yates criterion [10] with level of significance $\alpha = 5\%$ the problem of multicollinearity does not occur.

The hypothesis about normal distributed residuals ε_i ($i = 0, 1, \ldots, N\text{-}1$) in (9) is tested according to Fisher criterion. It answers the following question: whether estimation $\hat{A}$ of skewness and estimation $\hat{E}$ of kurtosis differ significantly from their mathematical expectations, which in case of normal distribution equal to zero. The rule-of-thumb described in [10] for $\hat{A} = 0.54$ ($\sigma_{\hat{A}}^2 = 0.41$) and $\hat{E} = 0.03$ ($\sigma_{\hat{E}}^2 = 0.81$) implies that values $\hat{A}$ and $\hat{E}$ can be considered statistically insignificant.

The following four conditions to validate are known as Gauss-Markov assumptions. The first one claims that mathematical expectation of residuals ε_i ($i = 0, 1, \ldots, N\text{-}1$) in (9) is equal to zero. Direct computation gives $\sum \hat{\varepsilon}_i = 0$.

The second assumption (homoscedasticity) we will prove using Breush-Pagan test [9], which is grounded on significance verification for squared residuals model

$$\hat{\varepsilon}^2 = \delta_0 + \delta_1(x_1 + x_2) + \delta_2 x_3 + \varepsilon^* .$$

When finding OLS-estimators $\hat{\delta}_0$, $\hat{\delta}_1$, $\hat{\delta}_2$ of unknown coefficients δ_0, δ_1, δ_2, one can compare Fisher statistics $F = 2.85$-statistics and its critical value $F_{cr} = 3.33$. As a result ($F < F_{cr}$) the auxiliary squared residuals model is statistically insignificant.

The next assumption is about non-correlative values of ε: $cov(\varepsilon_i, \varepsilon_j) = 0$ for all $i \neq j$. Using Durbin-Watson test [8], we can find out whether the first-order autocorrelation is present in (9): d-statistics is computed as follows:

$$d = \frac{\sum\limits_{t=1}^{N-1} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum\limits_{t=0}^{N-1} \hat{\varepsilon}_t^2} .$$

It equals **0.35**, that is lower than critical value $d_l = 1.31$. This confirms presence of the first-order positive autocorrelation. The latter frequently occurs in time-series analysis and may detect wrong factors specification. Nevertheless, presence of autocorrelation in many dynamic series is considered to be possible.

The fourth Gauss-Markov assumption (non-stochastic regressors) means that $cov(x_{ik}, \varepsilon_i) = 0$ for all values of i and k. Correlation analysis in 'factors-disturbances' system gives strong evidence that there is no correlation between regressors and random disturbances (all partial coefficients of correlation appears to be lower than Fisher-Yates critical value $r_{cr} = 0.349$).

Finally, let us turn to models (6)-(7). Testing for adequacy is made within the framework outlined earlier. In difference identification scheme (6), as well as in integral one (7), the first-order positive autocorrelation is revealed. Other assumptions of classical regression are fulfilled except the first Gauss-Markov condition for scheme (6). RTO is a special class of regressions which require special attention, especially when computing coefficient of determination R^2 and standard errors for estimated coefficients. Thus an additional regression analysis with some robust correction should be made. On the whole, its results correspond to earlier analysis.

CONCLUSIONS

Robust identification for models of IDP is a problem that occurs in different branches connected with dynamic systems and optimal control. In this article two econometric algorithms were elaborated and tested. Difference identification scheme appeared to be more useful for forecast analysis, while integral one demonstrated better simulation properties. The approbation of all models and sub-models was made using an open macroeconomic system investment development as an example. All models and sub-models corresponded to reality and demonstrated the possibility of their practical application. One of the essential parts of econometric analysis, procedure of verification, also confirmed adequacy of proposed models.

SUMMARY

**ЕКОНОМІЧНИЙ ПІДХІД ДО ІДЕНТИФІКАЦІЇ МОДЕЛЕЙ
ОБЕРНЕНИХ ЗАДАЧ ДИНАМІКИ**

Д.В. Фільченко

Стаття присвячена проблемі ідентифікації моделей оберненої задачі динаміки. Запропонований новий підхід, що базується на економетричному аналізі і є робастним методом для практичного застосування. Особлива увага зосереджена на порівняльному аналізі двох алгоритмів: за різницевою та інтегральною схемами. Також описано приклад їх можливого застосування в макроекономічному моделюванні. Всі моделі апробовані на реальних статистичних даних та перевірені на адекватність.

LITERATURE

1. Болтянский В.Г. Математические методы оптимального управления. – М.: Наука, 1969. – 408 с.
2. Feldbaum A.A. Optimal control systems. – New York: Academic Press Inc., 1965. – 296 p.
3. Красовский Н.Н. Управление динамической системой. – М.: Наука, 1985. – 518 с.
4. Osipov Yu. S., Kryazhimskii A.V. Inverse problem of ordinary differential equations: dynamical solutions. – London: Gordon and Beach, 1995. – 220 p.
5. Назаренко О.М. Динамічне моделювання інвестиційного розвитку та оптимальної макроекономічної інвестиційної політики // Механизм регулирования экономикой. – 2006. – №4. – С. 51-60.
6. Obstfeld M., Rogoff K. Foundations of international macroeconomics. – The MIT press, 1996. – 785 p.
7. Шевчук В.О. Міка: теорія і практика. – Львів: Каменяр, 2003. – 560 с.
8. Greene W.H. Econometric analysis. Fifth Edition. — New Jersey: Prentice Hall Upper Saddle River, 2003. — 802 p.
9. Wooldridge J. Introductory Econometrics: A Modern Approach. — New York: Southwestern Publishers, 2000. — 805 p.
10. Назаренко О. М. Основи економетрики: Підручник. – Київ: Центр навчальної літератури, 2004. – 392 с.
11. Denmark Statistics: http://www.dst.dk.
12. OECD, Statistic database: http://www.oecd.org.

Фільченко Д.В., аспірант, СумДУ, м. Суми