Notes on Dynamic Panel Data Models in Terms of Military Spending Determinants Analysis

Jiří Neubauer, Jakub Odehnal, Todor Ivanov

Abstract: The contribution deals with the problem of dynamic panel data modeling. One can find several methods of estimation dynamic data model parameters. We focus on three commonly used approaches to panel data modeling. The first uses a fixed effect model, the second is based on a generalized method of moments, and the third employs a system generalized method of moments. We discuss the properties of the mentioned estimates. The models are applied to time series of military spending in 27 NATO member countries and to other time series describing the economic and security environment in analyzed states in order to identify possible determinants of military spending. The estimated models are compared in terms of a quality of fit and residual analysis.

Keywords: panel data, fixed effect model, generalized method of moments, military spending.

1 Introduction

Panel data analysis is a statistical method, broadly used in social science, econometrics or epidemiology to analyze cross sectional and longitudinal panel data which are two-dimensional. The data are usually collected over time and over the same individuals (countries, persons, factories, ...). This approach can often be found in currently published papers and journals dealing with security issues in terms of modeling of military expenditures (Dunne and Perlo-Freeman, 2003; Dunne et al., 2010; Ali, 2012; Yildirim et al., 2005; Ambler and Neubauer, 2017).

The aim of the modeling of determinants of military spending is to identify factors influencing the amount of money allocated to defense. Empirical studies aimed at identifying military spending determinants classify those determinants into groups of economic factors, security factors, and political factors. The economic environment is usually characterized by the size of the gross domestic product of a country that characterizes its economic level, GDP growth rate or fiscal variables that describe budget surplus (deficit) or state debt. The security environment is often defined by variables that describe the external or internal risks of an armed conflict (measured by dummy variables or possible likelihood of this phenomenon), civil war risks, or military expenditures of potential enemy countries, neighboring countries, or allied countries forming the defense alliance. The political environment is characterized through variables describing the form of government, or, for example, the quality of democracy through quantified values characterizing different counterparts in the form of differences between democracy and autocracy in the countries.

2 Panel models

The *pooling* linear model for panel data is

$$y_{it} = \alpha + \beta' \boldsymbol{X}_{it} + u_{it}, \tag{1}$$

where $i = 1, 2, \dots, n$ is the individual index (group, country, ...), $t = 1, 2, \dots, T$ is the time index and u_{it} is a random zero mean disturbance term, $\boldsymbol{X}_{it}$ is a $k \times 1$ vector of independent variables, β_{it} is a $k \times 1$ vector of parameters (Croissant and Millo, 2008). This model can be estimated by the ordinary least squares method (OLS).

To model individual heterogeneity, let us assume that the error term has two separate components $u_{it} = \mu_i + \epsilon_{it}$, where μ_i is specific to the individual and does not change over time.

$$y_{it} = \alpha + \beta' \boldsymbol{X}_{it} + \mu_i + \epsilon_{it} \tag{2}$$

The error term ϵ_{it} is usually assumed independent of both the regressors $\boldsymbol{X}_{it}$ and the individual component μ_i. If the individual component is correlated with the regressors, it it common to treat the μ_i as next n parameters to be estimated. This is called the *fixed effect* model (Hsiao, 2014; Wooldridge, 2002). If we denote $\alpha_i = \alpha + \mu_i$ we obtain the model

$$y_{it} = \alpha_i + \beta' \boldsymbol{X}_{it} + \epsilon_{it}. \tag{3}$$

This model is sometimes called the *least squares dummy variable* model, it is usually estimated by OLS.

If the individual component μ_i is uncorrelated with the regressors, the model is termed *random effect*, μ_i are not treated as fixed parameters, but as random drawings from a given probability distribution. One of the assumptions related to OLS is that the error term is independently and identically distributed. In the context of panel data it means that $E(u_{it}^2)$ equals a constant σ_u^2 for all i and t, the covariance $E(u_{is}, u_{it})$ is equal to zero for all $s \neq t$ and the covariance $E(u_{jt}, u_{it})$ equals zero for all $j \neq t$. If these assumptions are not met, and they are unlikely to be met in case of panel data, OLS estimator is not the most efficient estimator. To get greater efficiency, generalized least squares (GLS) may be used, taking into account the covariance structure of error term.

A *dynamic linear panel* data model can be written in the form

$$y_{it} = \rho y_{i,t-1} + \beta' \boldsymbol{X}_{it} + \mu_i + \epsilon_{it}. \tag{4}$$

Using fixed model estimator for the dynamic panel model (4) we obtain estimates which are biased Hsiao (2014). In this case, the generalized method of moments is recommended. To eliminate the individual effect, the first difference of the model (4) is computed

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \beta' \Delta \boldsymbol{X}_{it} + \Delta \epsilon_{it}. \tag{5}$$

The error term $\Delta\epsilon_{it}$ is autocorrelated and also correlated with lagged dependent variable $\Delta y_{i,t-1}$. Generalized method of moments (GMM) approach is used to get estimates of equation (5), see Arellano and Bond (1991). Least squares are inconsistent because $\Delta\epsilon_{it}$ is correlated with Δy_{it-1}. It can be shown that $y_{i,t-2}$ is an instrument for $\Delta y_{i,t-1}$ (Anderson and Hsiao, 1981). The GMM estimator uses the fact that the number of valid instruments is growing with t

- $t = 3$: y_{i1},
- $t = 4$: y_{i1}, y_{i2},
- $t = 5$: y_{i1}, y_{i2}, y_{i3}.

The matrix of instruments is

$$\boldsymbol{Z}_i = \begin{pmatrix} y_{i1} & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \boldsymbol{X}'_{i3} \\ 0 & y_{i1} & y_{i2} & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \boldsymbol{X}'_{i4} \\ 0 & 0 & 0 & y_{i1} & y_{i2} & y_{i3} & \dots & 0 & 0 & 0 & 0 & \boldsymbol{X}'_{i5} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & y_{i1} & y_{i2} & \dots & y_{i,T-2} & \boldsymbol{X}'_{i,T} \end{pmatrix}. \tag{6}$$

The moment conditions are: $\sum_{i=1}^n \boldsymbol{Z}'_i \boldsymbol{e}_i(\beta)$ where $\boldsymbol{e}_i(\beta)$ is the vector of residuals for individual i. The GMM estimator minimize

$$\left(\sum_{i=1}^n \boldsymbol{e}_i(\beta)' \boldsymbol{Z}_i \right) \boldsymbol{A} \left(\sum_{i=1}^n \boldsymbol{Z}'_i \boldsymbol{e}_i(\beta) \right), \tag{7}$$

where $\boldsymbol{A}$ is the weighting matrix of the moments.

One-step estimators are computed using a known weighting matrix

$$\boldsymbol{A}^{(1)} = \left(\sum_{i=1}^{n} \boldsymbol{Z}_i' \boldsymbol{H}^{(1)} \boldsymbol{Z}_i\right)^{-1}, \quad \boldsymbol{H}^{(1)} = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -1 & 2 \end{pmatrix}. \tag{8}$$

Two-steps estimators are obtained using $\boldsymbol{H}_i^{(2)} = \sum_{i=1}^{n} \boldsymbol{e}_i^{(1)} \boldsymbol{e}_i^{(1)'}$ where $\boldsymbol{e}_i^{(1)}$ are the residuals of the one step estimate.

Blundell and Bond (1998) showed that the lagged levels are valid but weak instruments for first differenced variables, especially if the variables are close to a random walk. Their modification of the estimator includes lagged levels as well as lagged differences. More precisely, they proved that $\Delta y_{it-2} = y_{it-2} - y_{it-3}$ is a valid instrument. The estimator is obtained using the residual vector in difference and in level

$$\boldsymbol{e}_i^{+} = (\Delta \boldsymbol{e}_i, \boldsymbol{e}_i)$$

and the matrix of instruments

$$\boldsymbol{Z}_i^{+} = \begin{pmatrix} y_{i1} & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \Delta \boldsymbol{X}_{i3}' \\ 0 & y_{i2} & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \Delta \boldsymbol{X}_{i4}' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & y_{i,T-2} & 0 & 0 & \dots & 0 & \Delta \boldsymbol{X}_{iT}' \\ 0 & 0 & \dots & \dots & 0 & \Delta y_{i2} & 0 & \dots & 0 & \boldsymbol{X}_{i3}' \\ 0 & 0 & \dots & \dots & 0 & 0 & \Delta y_{i3} & \dots & 0 & \boldsymbol{X}_{i4}' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta y_{i,T-1} & \boldsymbol{X}_{iT}' \end{pmatrix}. \tag{9}$$

This estimator is usually denoted as the System GMM estimator (GMM-SYS).

3 Data description

To quantify the determinants of military expenditures, the authors selected data defining economic, security and political risks of the respective countries. In order to analyze economic environment as a determinant of military expenditures, the following variables were monitored: budget balance as a percentage of GDP, foreign debt as a percentage of GDP, economic conditions measured by the GDP and GDP growth rate and risk of inflation. For security risk analysis, the following variables were used: the risk of foreign pressures, the risk of cross-border conflict, the risk of terrorism and the risk of ethnic tensions. To analyze political risks, a variable evaluating the democratic accountability was chosen. Actual variables contained in the database are further observed for analytical purposes on the scale shown in Table 1. Therefore, higher values of these variables are interpreted as higher economic, security or political risks.

Data describing military expenditures as a share in GDP was obtained from the Stockholm International Peace Research Institute (SIPRI) database (SIPRI, 2018). Data characterizing selected determinants of military expenditures in the form of quantified socio-economic, security and political risks are from the PRS database (PRS, 2018). The detailed description of data from PRS database is in Table 1.

Variables	Description	Measurement
Budget Balance as a Percentage of GDP	The estimated central government budget balance as a percentage of the estimated GDP.	e.g. 4.0 plus % of GDP, 0 points; 3 to 3.9, 0.5 points; and e.g. 8 to 8.9, 6.5 points; -30.0 bellow, 10 points
Foreign Debt as a Percentage of GDP	The estimated gross foreign debt in a given year is expressed as a percentage of GDP.	e.g. 0 to 4.9, 0 points; 5 to 9.9, 0.5 points and 200 plus, 10 points
GDP per Capita	The estimated GDP is expressed as a percentage of the average of the estimated total GDP of all countries.	e. g. 250 plus (% of average), 0 points; 200 to 249.99, 1 point; and e.g. up to 9.9, 10 points
Actual GDP Growth	The annual change in the estimated GDP, at constant 1990 prices, of a given country is expressed as a percentage increase or decrease.	e.g. 6 % change plus, 0 points; change 5 to 5.9, 0.5 points; and e.g. 5.0 to 5.9, 9.5 points; 6.0 below, 10 points
Inflation	The estimated annual inflation rate (the unweighted average of the Consumer Price Index) is calculated as a percentage change.	e.g. 130% change plus, 0 points; 2.0 below, 10 points
Foreign Pressures	A score of 0 points equates to Very Low Risk and a score of 10 points to Very High Risk.	a minimum score of 0 points, a maximum score of 10 points
Cross-Border Conflict	A score of 0 points equates to Very Low Risk and a score of 10 points to Very High Risk.	a minimum score of 0 points, a maximum score of 10 points
Terrorism	A score of 0 points equates to Very Low Risk and a score of 10 points to Very High Risk.	a minimum score of 0 points, a maximum score of 10 points
Ethnic Tensions	Higher ratings are given to countries where racial and ethnic tensions are high. Lower ratings are given to countries where tensions are minimal.	a minimum score of 0 points, a maximum score of 10 points
Democratic Accountability	This is a measure of how responsive the government is to its people	a minimum score of 0 points, a maximum score of 10 points

Table 1: Description of the data from the PRS database used in panel models

4 Empirical results

In order to model the development of NATO military expenditures, we decided to use the panel data models described in the previous section, namely the fixed and random effect models, and GMM and GMM-SYS models.

At first, we applied a fixed and random model with lagged response variable MILEX_{t-1}. Table 5 contains estimates and standard errors of the full model with all explanatory variables and the final models with statistically significant estimates (up to the significance level of 0.10). The final models were determined by the strategy of backward selection (elimination) starting from the model containing all explanatory variables. According to Hausman test (Wooldridge, 2002), it can be claimed that the fixed and random effect models are not equivalent (p-value is $2.036 \cdot 10^{-9}$). In this case, the results of the fixed effect model are recommended. It should be noted that OLS parameter estimates of the fixed model with a lagged value of the response variable as a regressor are biased. That is the reason to use another methods or models. In addition, the residuals of this model show significant autocorrelation, see Table 2.

Models based on the general methods of moments are often used to estimate dynamic panel models for "short" panels (T is small compared to n). In our case, we have $T = 17$ and $n = 27$. We decided to employ two-step GMM and GMM-SYS methods to estimate the parameters. The results are summarized in Table 5. Tables 3 and 4 contain several tests on estimated models. According to the results of the Sargan test, the instrument variables are valid in all models (Hsiao, 2014; Wooldridge, 2002). When the idiosyncratic errors in the panel are independently and identically distributed, the first-differenced errors will become first-order autocorrelated. It means that the results of the Arellano-Bond test (order 1) should indicate the presence of autocorrelation. This phenomenon can be observed in all estimated models. Nevertheless, for the higher order of autocorrelation (we tested order 2), the residuals are not correlated.

The quality of the regression fit can be measured by the standard error

$$\widehat{\mathrm{SE}} = \frac{1}{n(T-t_0)-k} \sum_{i=1}^{n} \sum_{t=t_0}^{T} (y_{it} - \hat{y}_{it})^2, \tag{10}$$

where $i = 1, 2, \ldots, n$, $t = 1, 2, \ldots, T$ and $t_0 = 2$ for the fixed model and $t_0 = 3$ for the GMM and GMM-SYS model, k is a number of estimated parameters. These errors are displayed in Table 5. The fixed effect models offer, according to $\widehat{\mathrm{SE}}$, better fit than GMM and GMM-SYS models. The models based on GMM method are all comparable.

If we compare the final estimated models, we can see that the fixed effect model and the GMM model contains the same regressors, the lagged value of military expenditures (MILEX_{t-1}), the risk for budget balance ($\mathrm{BALANCE}_{t-1}$) and the risk for GDP (GPD_{t-1}). The parameter estimates are almost the same. The GMM-SYS estimates differ significantly. The final model is formed, except for the the lagged value of military expenditures (MILEX_{t-1}), by the risk for foreign debt (DEBT_{t-1}), the risk for GDP (GPD_{t-1}), the risk for inflation ($\mathrm{INFLATION}_{t-1}$), the risk for terrorism ($\mathrm{TERRORISM}_{t-1}$) and the risk for ethnic tension (ETHNIC_{t-1}). The estimated parameter corresponding to the regressor MILEX_{t-1} is very close to 1, which can cause problems with model stability (non-stationarity). Given the above, we would prefer the regression model estimated by GMM.

	Full model		*Final model*	
Test	**Statistics**	**p-value**	**Statistics**	**p-value**
Breusch-Godfrey/Wooldridge test (order 1)	15.846	$6.872 \cdot 10^{-5}$	15.706	$7.399 \cdot 10^{-5}$
Breusch-Godfrey/Wooldridge test (order 2)	16.504	0.00026	16.173	0.00031
Durbin-Watson test	1.663	$7.769 \cdot 10^{-5}$	1.657	0.00015
Wooldridge's test	5.502	0.019	4.853	0.0276

Table 2: Tests of correlation in residuals – fixed effect models

	Full model		*Final model*	
Test	**Statistics**	**p-value**	**Statistics**	**p-value**
Sargan test	11.982	1	23.221	1
Arellano-Bond test (order 1)	−2.670	0.00759	−3.188	0.00143
Arellano-Bond test (order 2)	−1.367	0.17159	−1.386	0.16567

Table 3: Tests of correlation in residuals – GMM models

	Full model		*Final model*	
Test	**Statistics**	**p-value**	**Statistics**	**p-value**
Sargan test	19.193	1	22.635	1
Arellano-Bond test (order 1)	−3.714	0.00020	−3.576	0.00035
Arellano-Bond test (order 2)	−1.386	0.16587	−1.407	0.15957

Table 4: Tests of correlation in residuals – GMM-SYS models

	FE full	FE final	GMM full	GMM final	GMM-SYS full	GMM-SYS final
MILEX_{t-1}	0.787^{***}	0.791^{***}	0.809^{***}	0.814^{***}	0.986^{***}	0.978^{***}
	(0.030)	(0.024)	(0.152)	(0.073)	(0.062)	(0.020)
BUDGET_{t-1}	-0.015^{**}	-0.017^{***}	-0.018^{*}	-0.031^{**}	-0.006	
	(0.007)	(0.007)	(0.009)	(0.013)	(0.015)	
DEBT_{t-1}	0.012		0.007^{*}		0.006^{*}	0.006^{*}
	(0.005)		(0.011)		(0.004)	(0.003)
GDP_{t-1}	-0.013^{***}	-0.014^{***}	-0.018^{***}	-0.013^{**}	-0.013^{**}	-0.012^{***}
	(0.004)	(0.004)	(0.004)	(0.007)	(0.005)	(0.004)
GDPPC_{t-1}	-0.012		-0.019		0.0002	
	(0.012)		(0.018)		(0.005)	
INFLATION_{t-1}	0.005		-0.018		-0.038^{**}	-0.035^{***}
	(0.012)		(0.028)		(0.017)	(0.009)
FOREIGN_{t-1}	-0.009		-0.002		0.004	
	(0.009)		(0.009)		(0.009)	
CONFLICT_{t-1}	0.010		0.013		0.0003	
	(0.008)		(0.015)		(0.013)	
TERRORISM_{t-1}	0.004		-0.006		0.008	0.008^{*}
	(0.008)		(0.011)		(0.009)	(0.005)
ETHNIC_{t-1}	-0.010		0.011		0.009	0.009^{**}
	(0.016)		(0.025)		(0.007)	(0.004)
DEMOCRATIC_{t-1}	0.017		0.041		0.005	
	(0.013)		(0.030)		(0.009)	
$\widehat{\text{SE}}$	0.12763	0.12741	0.15760	0.15645	0.15442	0.15429
Note:						$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 5: Panel data models for military expenditures (MILEX_t); standard errors are in parenthesis

5 Conclusion

We applied three models to describe military expenditures in NATO member countries from 2001 to 2017. We have found that despite the theoretical flaws, the fixed effect model with the lagged response variable as the regressor gives the best fit to data. The parameter estimates of this model and model based on GMM are rather comparable. The estimates obtained by the GMM-SYS method give different results for given data. We can conclude that, according to our findings, military expenditures in the NATO members countries are mainly determined by the economic situation.

Acknowledgements

The paper was supported by FML Development Projects AERO.

References

Ali, H. E., 2012. Military Expenditure and Inequality in the Middle East and North Africa: a Panel Analysis. *Defence and Peace Economics*, 23(6), 575–589.

Ambler, T. and J. Neubauer, 2017. Defense expenditure and economic growth in Visegrad group countries: a panel data analysis. In *35th International Conference Mathematical Methods in Economics, Hradec Králové*, 2017, 6–11.

Anderson T. and C. Hsiao, 1981. Estimation of Dynamic Models With Error Components. *Journal of the American Statistical Association*, 76(375), 598–606.

Arellano, M. and S. Bond, 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations, *The Review of Economic Studies*, 58, 277–297.

Blundell R. and S. Bond, 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*, 87(1), 115–143.

Croissant,Y. and G. Millo, 2008. Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software*, 27(2). Accessible from: http://www.jstatsoft.org/v27/i02/.

Dunne, J. P. and S. Perlo-Freeman, 2003. The demand for military spending in developing countries: a dynamic panel analysis. *Defense and Peace Economics*, 14, 461–474.

Dunne, J. P., Perlo-Freeman, S. and A. Soydan, 2010. Military expenditure and debt in small industrialised economies: A panel analysis, *Defence and Peace Economics* 15(2), 125–132.

Hsiao, C., 2014. *Analysis of Panel Data*. 3rd edition. Cambridge University Press.

Odehnal, J. and J. Neubauer, 2015. Ekonomické determinanty vojenských výdajů – kauzální analýza. *Ekonomický časopis*, 63(10), 1019–1032.

Political Risk Service Group (ICRG Methodology). 2018. Accessed June 15 2018. http://www.prsgroup.com/ICRG_Methodology.aspx

Wooldridge, J. M., 2002. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press: Cambridge, Massachusetts.

Stockholm International Peace Research Institute. 2018. Accessed July 10 2018. https://www.sipri.org/databases.

Yildirim, J, Sezgin, S. and N. Öcal, 2005. Military Expenditure and Economic Growth in Middle Eastern Countries: a Dynamic Panel Data Analysis, *Defence and Peace Economics* 16(4), 283–295.

Jiří Neubauer, Assoc. prof., Ph.D.
Department of Quantitative Methods, Faculty of Military Leadership, University of Defence
Kounicova 65, 662 10, Brno, Czech Republic
E-mail: jiri.neubauer@unob.cz
Telephone: +420 973 442 029

Jakub Odehnal, Assoc. prof., Ph.D.
Department of Economics, Faculty of Military Leadership, University of Defence
Kounicova 65, 662 10, Brno, Czech Republic
E-mail: jakub.odehnal@unob.cz
Telephone: +420 973 442 197

Todor Ivanov, Dr.
Department of Computer Science, West Connecticut State University
181 White Street, CT 06810, Danbury, USA
E-mail: ivanovt@wcsu.edu
Telephone: +1203 837 9353