Using statistical learning methods for impact evaluation: the effect of European Union membership on economic growth

Constantin Colonescu

MacEwan University

FIRST DRAFT

Abstract. Studies known as impact evaluation or treatment effect evaluation are traditionally based on regression models that include categorical covariates. Another method, known as "synthetic counterfactual analysis" tries to determine how some variables would have evolved in the absence of the event of interest. This paper proposes "statistical learning" as an alternative method of impact evaluation. Statistical learning fits an econometric model to a subset of the data (the training set) and tests its predictions on another subset of the data (the test set). The parameters of the econometric model are determined by evaluating the model's performance in predicting the variable of interest in the test data subset. Using prediction methods in impact evaluation problems is a novelty. The method is exemplified on the effect of EU membership on a member country's GDP. The results, however, are not yet satisfactory, probably since the current statistical learning methods are not suitable for panel data.

1. Introduction

How would the world look like if event *X* would not have happened? Of course, nobody knows with certainty. The question, though, is more than a matter of curiosity. An answer would cast light on the consequences of event *X*. Event *X* could be a persistent change, such as a change in taxes, a new vaccination program, a change in legislation banning smoking in public places, increasing or abolishing tuition fees for tertiary education, lowering trade tariffs, a country's becoming a member of some international organization, or the election of a new president or prime minister. Event *X* could also be a one-time happening, such as a natural or human-caused disaster. Answering such questions is the concern of impact evaluation, or treatment effect methods.

Given adequate information about today's state of the world, what would be the state of the world of tomorrow? Or, alternatively, given adequate information about a number of entities (Hastie, Tibshirani and Friedman 2009) states, provinces, companies, and individuals), how can one describe another, similar entity? Statistical learning tries to predict the values of some variable *Y* by finding the best-predicting model, based on existing information about bout *Y* and its covariates *X*. While assumptions or theories about how some factors influence a dependent variable are important for standard regression analysis, statistical learning focuses more on prediction rather than interpretation. While standard regression models emphasize inference, interpretation and hypothesis testing, statistical learning focuses on prediction, with less concern about the interpretability of a model.

How can statistical learning help solving an impact evaluation problem? Impact evaluation can be related to the problem of determining a *counterfactual*, which is how the world would evolve if some event wouldn't have happened. Statistical learning methods could allow one to "predict" the fictitious, counterfactual state of the world. Then, by comparing the counterfactual with the real state of the world, one can make some statement of the impact of the event under scrutiny. This study focuses on contrasting the a few alternative methods of impact evaluation, of which statistical learning is new, using European Union membership as an example.

Recent studies on the effect of EU membership on economics growth are (Cuaresma, Ritzberger-Grunwald and Silgoner 2008), (Böwer and Turrini 2009) and (Campos, Coricelli and Moretti 2014). The former two articles use simple difference estimators ((Hill, Griffiths and Lim 2011), while the latter uses a synthetic counterfactual method (Abadie, Diamond and Hainmueller 2010). The authors find some evidence of positive effects, with large heterogeneity across countries.

2. Data

The data come from the World Bank (WB 1990), under the form of a panel data over the period of 1995 to 2010, including up to 101 countries. Ten countries have become members in 2004 (Cyprus, the Czech Republic, Estonia, Hungary, Poland, the Slovak Republic, Slovenia, Lithuania, Latvia, and Malta), Romania and Bulgaria in 2007 and Croatia in 2013. The European Union has currently 28 member states. (Croatia, Cyprus, Lithuania, and Latvia are excluded from the dataset due to incomplete data.)

The following variables serve as predictors in the next models:

- *LGDP* = log of real GDP per capita, in 2005 U.S. dollars
- *LGoEx* = log of government expenditure, % of GDP
- *LGoEx_1* is the one-period lag of LGoEx
- *Capital* = gross capital formation, % of GDP
- *Saving* = gross domestic savings, % of GDP
- *IndShr* = industry share, % of GDP
- *EUmember* = 1 if a country is an EU member in a given year;
- *EUmember1* = EUmember (t+1), to model pre-accession effects
- *crisis* =1 if Year = 2008 or Year =2009

3. Empirical considerations and results using existing methods

To put the results of a statistical learning exercise in perspective, some alternative models are investigated. A first model estimates a difference-in-differences (DID) parameter (*EUmember*) in a dynamic panel data (Arellano and Bond 1991). The dynamic panel model works with first differences in variables; first differences eliminate the time-invariant, country-specific factors and reduce autocorrelation. The model also uses higher lags as instrumental variables to address endogeneity issues. Table 1 shows the results. The *EUmember* indicator variable comes out significant and important, suggesting an average effect of about 6% advantage in GDP per capita for member countries as compared to similar non-members. Table 2 uses the same dynamic model as Table 1, but includes a *crisis* indicator for the years of 2008 and 2009. The inclusion of these extra terms has not changed the previous result in what the variable *EUmember* is concerned. To check the robustness of this result, Table 2 includes the variables *EUmember1*, which accounts for the fact that a new member begins the accession process before the actual data of accession. Table 3 proposes another robustness exercise: Estimating the same model, but under the form of a fixed-effect panel data with time dummies. The shortcoming of such DID models is that they only give average effects, which do not account for country specific circumstances.

Another method of impact evaluation is the synthetic counterfactual analysis (Abadie, Diamond and Hainmueller 2010). Instead of trying to find a model that provides reliable inference and hypothesis testing, the synthetic counterfactual method tries to find a subset in the data that can best predict past values of the dependent variable. When this subset is found, it is used to forecast the "counterfactual," i.e., the values of the dependent variable that would have occurred in the absence of the "treatment." A similar study, using fewer control countries has been performed by (Campos, Coricelli and Moretti 2014) to estimate the effect of EU membership on GDP per capita.

Figure 1 shows the actual and the calculated counterfactual GDP per capita in a few countries of the 2004 and 2007 waves of EU enlargement. These graphs have been constructed using the "Synth" package in R (Diamond and Hainmueller 2011). They extend (Campos, Coricelli and Moretti 2014) to three new member states, Bulgaria, Cyprus, and Romania, and a few extra years after the date the integration took place. While for many countries the results are similar to (Campos, Coricelli and Moretti 2014), results do not match for Poland and the Slovak Republic. The extension of the period of analysis from the end-of-period 2009 of (Campos, Coricelli and Moretti 2014) to the end-of-period of 2012 in the present study reveals some interesting cases. One is Hungary, which appeared to enjoy an advantage from EU membership only after 2010; On the other hand, Bulgaria and Latvia appear to be losing from membership after 2010.

4. Statistical learning as a tool for constructing a counterfactual

Unlike standard econometrics models, which focus on explanation and inference, both synthetic counterfactual and statistical learning focus on prediction (Shmueli 2010). Explanation and inference require keeping models reasonably simple, a condition that is not necessary when prediction is pursued. However, one may notice that an explanatory model that does a poor job in predicting even intra-sample observations is unlikely to provide reliable inference. On the other hand, by disregarding standard econometrics problems, such as serial correlation and endogeneity, predictive models are unlikely to serve their purpose of providing accurate predictions. The idea of using statistical learning for the construction of a counterfactual works in two steps: (i) find and optimize a model for accurate in-sample predictions (the training and testing phase), and (ii) use the model to predict the counterfactual. The training and testing phase consists in dividing the dataset in a "training" and a "test" subsamples and tuning a specified model to best predict the test samples. Figure 2 illustrates some results obtained by the use of the "caret" package in R (Kuhn 2008). Two countries and two regression methods serve as examples of using a statistical learning methodology. The method marked "lm" is based on a standard linear regression model, while the other, marked "brnn" uses a "Bayesian regularized neural network" model (Rodriguez and Gianola 2015). The results are similar between the two methods but are inconsistent, in some cases, with the results discussed in the previous sections (see, for instance, the cases of Poland and the Czech Republic.)

After several attempts of finding the best predictive model, it became clear that the larger the “training” subsample, the poorer the prediction, which is a counterintuitive result. The best prediction was found when only the time series for the country under study was used, which is an important loss of information. Apparently, models that would take advantage of the panel structure of the data are not yet available in statistical learning packages. To somehow use the panel structure of the data, time and country indicator variables were included in the models, which did not help much in stabilizing the results. Perhaps one way to incorporate panel features in the available packages would be to include in the dataset, as separate variables, lags or first differences of the existing independent variables. While doing so may address the autocorrelation problem, it does not address the endogeneity problem, other than including a “control” group of countries.

5. Conclusion

This first attempt of using statistical learning (or predictive modeling) methods in an impact evaluation problem is rather disappointing. The existing statistical learning methods seem to perform poorly when the data have a time dimension. They also seem to be very sensitive to the selection of the countries in the control group, as well as to slight errors in the imputation of missing values

Bibliography

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*, 2010: 493-505.

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

Böwer, Uwe, and Alessandro Turrini. *EU accession: A road to fast-track convergence? Economic Papers 393.* Economic Papers, Directorate General for Economic and Financial Affaires, European Commission, Brussels: European Commission, 2009.

Campos, Nauro, Fabrizio Coricelli, and Luigi Moretti. *Economic Growth and Political Integration:Estimating the Benefits from Membership in the European Union Using the Synthetic Counterfactuals Method.* Discussion Paper Series No. 8162, Bonn: Institute for the Study of Labour (IZA), 2014.

Cuaresma, Jesus Crespo, Doris Ritzberger-Grunwald, and Maria Antoinette Silgoner. "Growth, convergence and EU Membership." *Applied Economics*, 2008: 643–656.

Diamond, Alexis, and Jens Hainmueller. " Synth: An R Package for Synthetic Control Methods in Comparative Case Studies." *Journal of Statistical Software*, 2011: 1-17.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: Data mining, inference, and prediction.* Stanford, California: Springer, 2009.

Hill, Carter, William Griffiths, and Guay Lim. *Principles of Econometrics.* Hoboken: John Wiley & Sons, Inc., 2011.

Kuhn, Max. "Building Predictive Models in R Using the caret Package." *Journal of Statistical Software* 28, no. 5 (November 2008).

Rodriguez, Paulino Perez, and Daniel Gianola. *Package 'brnn'.* cran-r.project, 2015.

Shmueli, Galit. "To explain or to predict." *Statistical Science*, 2010: 289-310.

WB. *World Development Indicators.* Online database, Washington, DC: World Bank, 1990.

Appendix

Table 1: 1-step dynamic panel, using 1515 observations
Included 101 cross-sectional units
Including equations in levels
H-matrix as per Ox/DPD
Dependent variable: LGDP

	coefficient	std. error	z	p-value	
LGDP(-1)	0.965987	0.00921085	104.9	0.0000	***
const	0.103063	0.0365248	2.822	0.0048	***
LGoEx	0.0379830	0.0152658	2.488	0.0128	**
Capital	0.00283375	0.000587965	4.820	1.44e-06	***
Capital_1	-0.00174599	0.000609948	-2.863	0.0042	***
Savings	0.00208080	0.000460772	4.516	6.31e-06	***
Savings_1	-0.000689950	0.000338864	-2.036	0.0417	**
IndShr	0.00109667	0.000454282	2.414	0.0158	**
EUmember	0.0571678	0.0189016	3.024	0.0025	***

Sum squared resid 3.927293 S.E. of regression 0.051066

Number of instruments = 127
Test for AR(1) errors: z = -5.09481 [0.0000]
Test for AR(2) errors: z = -1.89142 [0.0586]
Sargan over-identification test: Chi-square(118) = 602.382 [0.0000]
Wald (joint) test: Chi-square(8) = 326029 [0.0000]

Table 2: 1-step dynamic panel, using 1414 observations
Included 101 cross-sectional units
Including equations in levels
H-matrix as per Ox/DPD
Dependent variable: LGDP
Asymptotic standard errors

	coefficient	std. error	z	p-value	
LGDP(-1)	0.952304	0.00549156	173.4	0.0000	***
const	0.145123	0.0194646	7.456	8.94e-014	***
LGoEx	0.0535078	0.00794947	6.731	1.69e-011	***
Capital	0.00274672	0.000265189	10.36	3.86e-025	***
Capital_1	-0.00124902	0.000270566	-4.616	3.91e-06	***
Savings	0.00223448	0.000257874	8.665	4.52e-018	***
Savings_1	-0.000194038	0.000255877	-0.7583	0.4483	
IndShr	0.00117924	0.000163142	7.228	4.89e-013	***
crisis	-0.0251879	0.00245873	-10.24	1.26e-024	***
EUmember	0.0378071	0.0137653	2.747	0.0060	***
EUmember1	0.0485567	0.0134808	3.602	0.0003	***

Sum squared resid 5.343033 S.E. of regression 0.061711

Number of instruments = 114
Test for AR(1) errors: z = -7.41179 [0.0000]
Test for AR(2) errors: z = -0.78265 [0.4338]
Sargan over-identification test: Chi-square(103) = 291.358 [0.0000]
Wald (joint) test: Chi-square(10) = 1.33745e+006 [0.0000]

Table 3: Fixed-effects, using 1515 observations
Included 101 cross-sectional units
Time-series length = 15
Dependent variable: LGDP

	coefficient	std. error	t-ratio	p-value	
const	7.89250	0.0535663	147.3	0.0000	***
LGoEx	0.0380811	0.0166706	2.284	0.0225	**
Capital	0.000646942	0.000730625	0.8855	0.3761	
Capital_1	0.00199073	0.000749876	2.655	0.0080	***
Savings	0.00415222	0.000642106	6.467	1.38e-010	***
Savings_1	0.00354197	0.000610960	5.797	8.32e-09	***
IndShr	0.00194444	0.000825465	2.356	0.0186	**
EUmember	0.0954430	0.0180153	5.298	1.36e-07	***
dt_2	-0.344427	0.0141922	-24.27	8.44e-109	***
dt_3	-0.315454	0.0142015	-22.21	8.00e-094	***
dt_4	-0.291054	0.0141805	-20.53	5.26e-082	***
dt_5	-0.277976	0.0142531	-19.50	4.34e-075	***
dt_6	-0.260625	0.0142159	-18.33	2.01e-067	***
dt_7	-0.246700	0.0141216	-17.47	6.00e-062	***
dt_8	-0.223284	0.0141570	-15.77	1.09e-051	***
dt_9	-0.196715	0.0140909	-13.96	1.49e-041	***
dt_10	-0.163879	0.0139618	-11.74	2.10e-030	***
dt_11	-0.133636	0.0139356	-9.590	3.91e-021	***
dt_12	-0.0964917	0.0139628	-6.911	7.32e-012	***
dt_13	-0.0547706	0.0139744	-3.919	9.31e-05	***
dt_14	-0.0302399	0.0140746	-2.149	0.0318	**
dt_15	-0.0362623	0.0141839	-2.557	0.0107	**

Mean dependent var	8.103585	S.D. dependent var	1.733330
Sum squared resid	13.53449	S.E. of regression	0.098570
LSDV R-squared	0.997025	Within R-squared	0.672597
LSDV F(121, 1393)	3857.609	P-value(F)	0.000000
Log-likelihood	1424.140	Akaike criterion	-2604.280
Schwarz criterion	-1954.853	Hannan-Quinn	-2362.465
rho	0.911875	Durbin-Watson	0.167191

Figure 1: The results from applying a synthetic counterfactual method

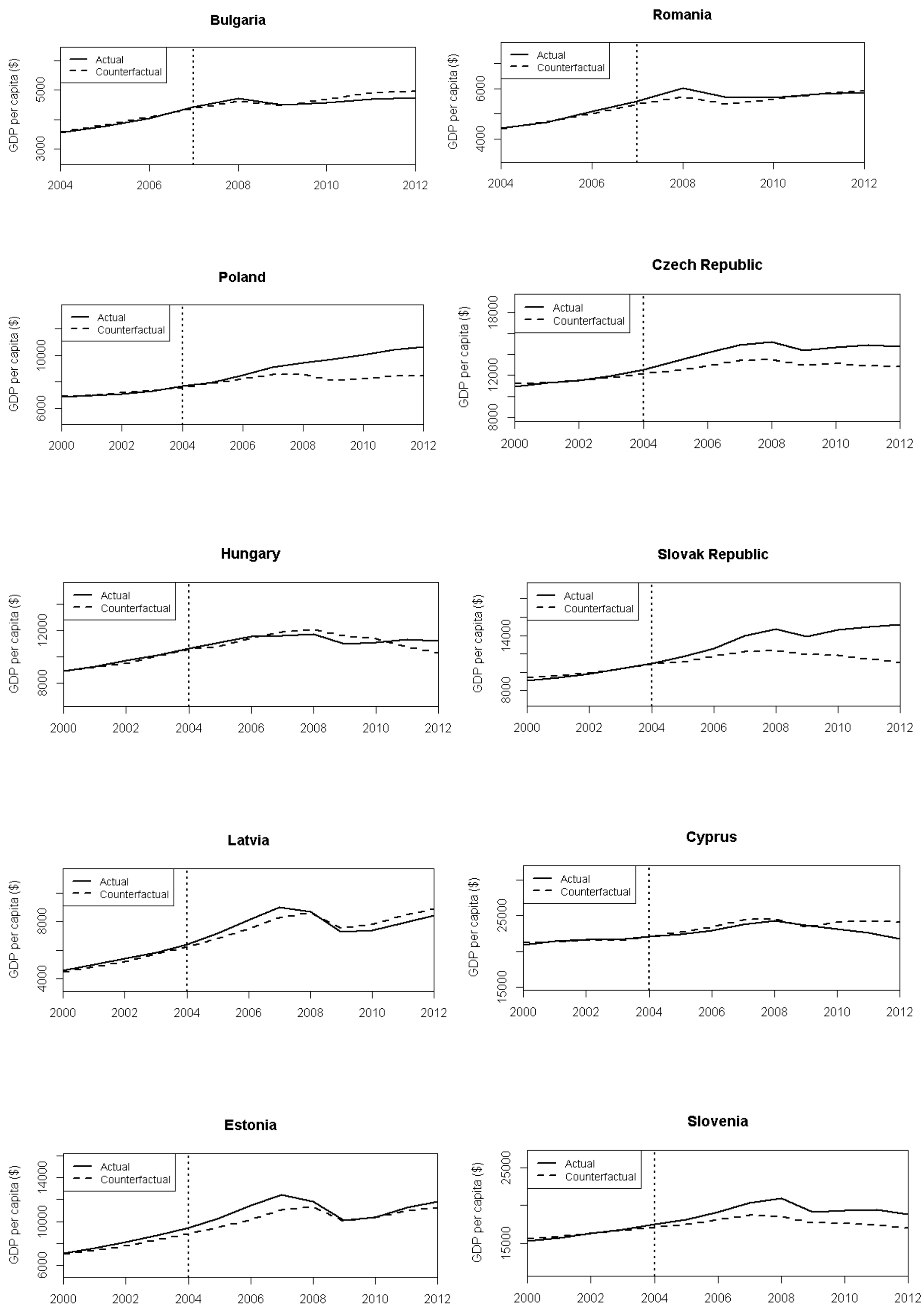

Figure 2: The results from applying statistical learning methods